Scheduling of Physicians with Time-Varying Productivity Levels in Emergency Departments

Farzad Zaerpour
Faculty of Business and Economics, The University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada
f.zaerpour@uwinnipeg.ca

Marco Bijvank
Haskayne School of Business, University of Calgary, Calgary, Alberta T2N 1N4, Canada
marco.bivank@haskayne.ucalgary.ca

Huiyin Ouyang
Faculty of Business and Economics, The University of Hong Kong, Pok Fu Lam Road, Hong Kong
oyhy@hku.hk

Zhankun Sun
Department of Management Sciences, College of Business, City University of Hong Kong, Kowloon, Hong Kong
zhankun.sun@cityu.edu.hk

Emergency department (ED) overcrowding and long patient wait times have become a worldwide problem. We propose a novel approach to assigning physicians to shifts such that ED wait times are reduced without adding new physicians. In particular, we extend the physician rostering problem by including heterogeneity among emergency physicians in terms of their productivity (measured by the number of new patients seen in 1 hour) and by considering the stochastic nature of patient arrivals and physician productivity. We formulate the physician rostering problem as a two-stage stochastic program and solve it with a sample average approximation and the L-shaped method. To formulate the problem, we investigate the major drivers of physician productivity using patient visit data from our partner ED, and find that the individual physician, shift hour, and shift type (e.g., day or night) are the determining factors of ED productivity. A simulation study calibrated using real data shows that the new scheduling method can reduce patient wait times by as much as 13% compared to the current scheduling system at our study ED. We also demonstrate how to incorporate physician preference in scheduling through physician clustering based on productivity. Our simulation results show that EDs can receive almost the full benefit of the new scheduling method even when the number of clusters is small.

Key words: Emergency department; Time-varying productivity; Physician Scheduling; Stochastic optimization

History: Received: November 2019; accepted: September 2021 by Sergei Savin after two revisions.

1. Introduction

The goal of emergency departments (EDs) is to provide immediate medical treatments to patients who require urgent care. In recent years, these hospital departments have become more crowded in many countries (including the US and Canada), for various reasons: the aging of the population, higher patient volumes, and the growing complexity of patient conditions and treatments. Consequently, patients who seek care in EDs may experience long wait times. This phenomenon is known as *ED overcrowding*, which results in prolonged pain and suffering, patient dissatisfaction, and dangers to public safety (Hoot and Aronsky 2008, Bernstein et al. 2009, Pines et al. 2011). As it is not always feasible to increase ED treatment capacity, mainly due to budget constraints, it is crucial to allocate existing resources efficiently to alleviate ED overcrowding. To improve patient flow in EDs, personnel schedules need to match the time-varying demand for emergency care. According to Defraeye and Van Nieuwenhuyse (2016), the personnel scheduling process is often decomposed into four subproblems: (i) forecasting: predicting demand for each time period during the scheduling horizon; (ii) staffing: determining the required number of workers for each scheduling period to meet specific performance targets at minimal cost; (iii) shift scheduling: creating shift schedules and determining how many workers are needed for each shift type to cover the staffing requirements; and (iv) rostering (or employee scheduling): assigning employees to shifts. We focus on the fourth subproblem (i.e., the rostering problem for ED personnel). This research is motivated by a case study in an ED in Calgary, Alberta, Canada.

There are two major personnel rostering problems in EDs, namely the nurse rostering problem (NRP) and the physician rostering problem (PRP). The PRP is fundamentally different from the NRP (Carter and Lapierre 2001), especially in Canada. In most Canadian EDs, nurses work under a collective agreement, whereas physicians are not employed by hospitals and do not have collective labor contracts. Consequently, physician schedules are predominantly driven by the need to satisfy as much as possible a large number of (often conflicting) rules and personal preferences. Physicians are the most expensive ED resource and are frequently considered the bottleneck in the delivery of emergency care (Bucheli and Martina 2004). As a result, the scheduling of emergency physicians is crucial for the delivery of high-quality, timely care. In this paper, we study a PRP in which we include characteristics of real-life EDs, such as the stochastic nature of time-varying arrivals and physician-specific productivity levels. The *productivity* of an emergency physician is defined as the number of new patients seen by the physician in 1 hour, which is commonly referred to as the *patient-per-hour rate* (or *PPH rate*) in the emergency medicine literature (Bukata et al. 2018, Joseph et al. 2018).

The combinatorial nature of the PRP makes it difficult to solve. Large sets of contractual and individual constraints contribute to the complexity of the problem. According to Gendreau et al. (2007), the constraints of the PRP can be classified into four categories: (i) *supply and demand constraints* define how many physicians are available and how many should work during different time periods of a day in the planning horizon; (ii) *workload constraints* regulate the number of shifts assigned to physicians during a certain time period; (iii) *ergonomic constraints* cover hospital rules regarding rest periods after a certain (set of) shift(s); and (iv) *fairness constraints* aim at distributing the assignment of particular types of shifts among the physicians during the scheduling horizon. In addition to the hospital rules, individual physicians may be allowed to express preferences concerning their schedules. It is usually not possible to respect all constraints and preferences. Therefore, a distinction is made between compulsory rules that have to be satisfied (known as *hard constraints*) and flexible rules that can be violated (known as *soft constraints*). This paper aims to develop a formulation of the PRP to create a schedule that satisfies the hard constraints. Our proposed formulation also takes into consideration the uncertainties regarding patient volume, arrival times, service times, and physician heterogeneity in their non-stationary productivity levels. Including these practically relevant characteristics requires an understanding of how productivity may change during shifts and the development of approaches to solve the new PRP. This has been rarely studied in the literature, although effective physician scheduling that matches capacity with demand for emergency care is crucial for reducing ED overcrowding and prolonged patient wait times.

We make three contributions in this paper. First, through data analysis we find that it is sufficient to consider shift hour, shift type (daytime or night shift), and the individual physician to predict the productivity of an emergency physician. Second, we study a PRP in which shifts are assigned to physicians based on their non-stationary productivity levels to minimize the mismatch between the available ED capacity and the non-stationary demand for emergency care. We propose a two-stage stochastic programming formulation in which the productivity levels and the time-varying patient demand are modeled as stochastic variables. This has not been studied in the literature, as most rostering problems ignore the random components and aim to satisfy hospital scheduling rules and physician preferences (Van den Bergh et al. 2013, Erhard et al. 2018). We use sample average approximation to express the extensive form of the stochastic programming formulation and subsequently use the L-shaped method to solve the problem. Third, we derive managerial insights based on a case study at an ED in Calgary, Alberta, Canada. In practice, ED physicians are allowed to exchange shifts among themselves even after the schedule is created. To mitigate the negative impact of such exchanges, we group physicians into different clusters based on their productivity so that physicians within the same cluster have similar PPH rates. Our results suggest that the performance of the schedule with clustering is near optimal even when the number of clusters is fairly small.

The outline of the paper is as follows. In Section 2, we review the relevant literature on personnel scheduling and worker productivity. In Section 3, we empirically investigate the determinants of physician productivity. In Section 4, we formulate the physician rostering problem as a two-stage stochastic program and propose a solution method. We investigate the impact of the optimal schedule through a simulation study in Section 5 and study how to account for physician preferences through physician clustering in Section 6. We conclude our paper in Section 7.

2. Literature Review

Our research is relevant to multiple streams of the literature. In Section 2.1, we discuss various personnel rostering problems that have been studied in the literature with a particular focus on ED physician scheduling. Studies on physician productivity and service times as a function of workload are discussed in Section 2.2. The notion of productivity levels in personnel scheduling problems is presented in Section 2.3.

2.1. Personnel Rostering Problem in EDs

The problem of shift scheduling for physicians has been studied extensively in the literature (see Erhard et al. 2018 for an overview). We only focus on papers that study physician rostering in EDs.

Beaulieu et al. (2000) are among the first to present a mixed 0-1 mathematical programming formulation for the assignment of emergency physicians to three distinct 8-hours shifts during a day. They propose a decomposition strategy to solve the problem. Carter and Lapierre (2001) extract the characteristics of a generic emergency physician rostering problem from six hospitals in greater Montreal, Canada. In contrast to Beaulieu et al. (2000) and Carter and Lapierre (2001), Rousseau et al. (2002) construct non-cyclic rosters and develop a solution approach that is a hybrid of constraint programming, local search and genetic algorithms. Gendreau et al. (2007) apply four solution techniques to solve the PRP: mathematical programming, column generation, tabu search and constraint programming. A genetic algorithm is proposed by Puente et al. (2009). The simultaneous creation and assignment of shifts (including on-call services) is studied by Brunner et al. (2009), who use a heuristic decomposition strategy. This work is extended by Brunner et al. (2011) to incorporate part-time workers and solve the problem by a branch-and-price algorithm, and by Stolletz and Brunner (2012) to include different measures of fairness. Brunner and Edenharter (2011) extend the problem by including the experience levels of physicians and specifying that each patient requires treatment based on a minimum experience level. Gunawan and Lau (2013) and Bruni and Detti (2014) schedule physicians to cover the demand for certain duties and/or departments in a hospital. The only work that includes physician productivity levels to the PRP is Camiat et al. (2021). All formulations of the PRP in these papers are similar in that they are multi-objective optimization problems that incorporate a large number of constraints and physician preferences, where the demand for physicians or shifts is deterministic.

The stochastic components in patient arrivals and service times are mostly included in the second and third steps of personnel scheduling problems (Defraeye and Van Nieuwenhuyse 2016). Common approaches to solving these problems include queueing theory and simulation; see Defraeye and Van Nieuwenhuyse (2013) and EL-Rifai et al. (2015) for recent developments. Another approach to accounting for stochastic demand and service times in personnel rostering problems is stochastic programming. Bagheri et al. (2016) study a stochastic nurse scheduling problem where they include ergonomic constraints (nurses cannot work the day after a night shift) and distribution constraints (each nurse should work a minimum number of shifts). Campbell (2011) and Gnanlet and Gilland (2014) investigate the workforce planning over multiple departments with the consideration of cross-training.

With regard to the PRP, we make two main contributions. First, we include heterogeneity in physician productivity levels (similar to Camiat et al. 2021) and show that clustering physicians in three different categories is sufficient to capture the benefits to account for differences between physician productivity. Second, we formulate the PRP as a stochastic program to capture the uncertainty in patient arrivals and physician productivity in an ED.

2.2. Productivity of Physicians in EDs

Shift work is very common in healthcare. A number of studies have analyzed the impact of shift duration on the productivity of emergency physicians. Hart and Drall (2007) conclude that shifts with a duration of 8 to 9 hours result in higher average PPH rates than 12-hour shifts, and Foster et al. (2015) conclude that 7-hour shifts are the best among 6-, 7-, and 8-hour shifts. In contrast, in a pilot study, Yang et al. (2008) suggest no difference in average PPH rate. Extended shift duration has also been associated with adverse effects on patient outcomes and with increased accidents and medical errors (e.g., diagnostic errors and medication errors). Fatigue (both physical and mental) is one of the most common concerns associated with shift work and extended shift duration. Other factors also influence the productivity of emergency physicians during their shifts. Chan (2018) study the end-of-shift phenomenon and conclude that physicians accept fewer patients in the last 2 to 4 hours of 9-hour shifts to avoid handing more patients over to other physicians.

Finally, the impact of workload on patient service times and physician productivity has attracted attention from the operations management (OM) community. KC and Terwiesch (2009) are among the first to conclude that a high load on the healthcare system leads to decreased service times in a cardiothoracic surgery setting. Kuntz and Sülz (2013) study the service times of individual emergency physicians under different workload levels. They conclude that the service time of physicians with more professional experience is shorter when the system load is high and longer when the system load is low. Other empirical work on the impact of workload on service times include Armony et al. (2015), Berry Jaeker and Tucker (2017), and Delasay et al. (2016). These studies provide insights into the impact of various factors on physician productivity; however, these factors cannot be used directly to determine staffing levels or physician rosters. More interestingly, we illustrate in our data analysis that shift hour, the individual physician, and the type of shift contribute more to the predictability of physician productivity than workload measures such as the number of patients in the ED or the waiting area, or who are waiting to be admitted to the hospital.

2.3. Productivity and Employee Heterogeneity in Personnel Scheduling

The final stream of the literature combines workforce planning and employee productivity (as discussed in the two previous sections), but not necessarily in a healthcare setting. One of the earliest works in the area of personnel scheduling with consideration of productivity levels is by Li et al. (1991), who study a shift scheduling problem in which a relative productivity measure is included to distinguish between full-time and part-time employees. A similar approach is used in multi-skill workforce planning, where employees can be cross-trained; for example, see Brusco et al. (1998) and Brusco and Johns (1998) on such staffing problems, where employees have a lower productivity in secondary skills.

A productivity parameter for individual employees is rarely included in rostering problems when workers are assigned to shifts. Goodale and Thompson (2004) take individual productivity levels from a normal distribution and the labor cost of each employee is approximately proportional to their productivity levels. Employees are grouped by their productivity level in Thompson and Goodale (2006). In Akbari et al. (2013), employees' productivity level is exogenous and dependent on the shift they are assigned. Furthermore, when they are scheduled to work consecutive shifts on the same day, their productivity decreases by a constant factor. Camiat et al. (2021) use a constant productivity level for each employee (or physician), which is only dependent on the type of shift and not on the hour of the shift. In these papers, the authors formulate a deterministic optimization model and use heuristic procedures to assign workers to shifts. Campbell (2011) and Gnanlet and Gilland (2014) are the exceptions as they use stochastic programming.

Among the papers reviewed, the one closest to ours is Camiat et al. (2021), which also study the PRP by considering individual physician productivity in a deterministic setting and assuming that each physician has a shift-hour-independent PPH rate. Next, we explicitly discuss the differences between our work and Camiat et al. (2021). First, we include the stochastic nature of the problem in our optimization to better match reality. Our simulation results show that by considering stochastic factors, the average patient wait time can be reduced by 8.4% compared to solving the PRP in a deterministic setting (as in Camiat et al. 2021). Second, we perform an empirical analysis to investigate the determining factors of PPH and to predict PPH, whereas Camiat et al. (2021) use the average of the PPH observed in historical data as the predictor, which depends on the day of the week and the shift type. Consequently, sufficient observations where each physician is assigned to all combinations of shift types and days of the week are required for accurate prediction. Third, we incorporate shift-hour-dependent PPH in our model, as suggested by our empirical analysis, whereas Camiat et al. (2021) use a constant PPH rate for each physician. Our simulation results show that in a deterministic setting, as in Camiat et al. (2021), when the shift-hour-dependent PPH is considered in the PRP, the average wait time can be reduced by 4.3%. When we further consider the stochastic factors in the PRP, the average wait time is reduced by 12.3% compared to the schedule from Camiat et al. (2021). Finally, we demonstrate that a schedule in which physicians with similar productivity levels are clustered can achieve near optimal performance, which makes our results more implementable compared to Camiat et al. (2021).

In conclusion, we are the first authors to include heterogeneous employee productivity levels in personnel scheduling problems that are dependent on the hour of the shift, in addition to considering the stochastic nature of the physician rostering problem.

3. Empirical Study of Physician Productivity

Our study is based on data from an ED in the city of Calgary, Alberta, Canada. During our study period (from August 2013 to July 2015), there were approximately 75,000 patient visits to the ED each year. Like other EDs across North America, this ED was dealing with physician shortages and increased patient volumes, resulting in prolonged wait times. Our data set contains patient visit records from this ED over a 2-year period from August 2013 to July 2015. In this section, we describe the patient flow and physician scheduling rules in our study ED. We then investigate factors that impact the PPH rates of ED physicians.

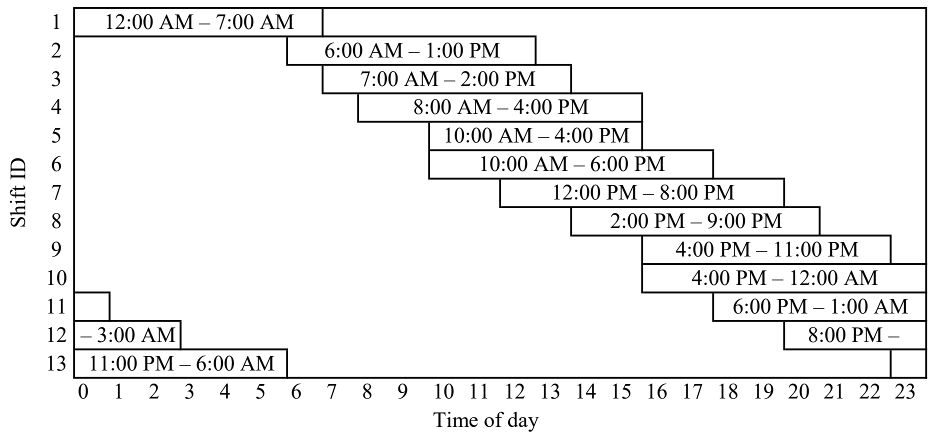
3.1. Patient Flow

The study hospital operates in a manner similar to many hospitals in North America. Upon arrival, a patient is seen by a triage nurse. The nurse performs a quick assessment of the patient (e.g., measuring vital signs and recording the chief complaint), assigns a triage score indicating the acuity of the patient's condition, and then creates an electronic record for the patient. Our study hospital follows a triage protocol called the Canadian Triage and Acuity Scale (CTAS), a five-level scoring system with 1 being most urgent and 5 being least urgent. After triage, all patients wait in a common waiting room and the patient's electronic record is placed in a virtual queue to be processed by one of the emergency physicians. This marks the beginning of the *waiting phase*. Physicians use a computer terminal to access the information of patients waiting to be seen, including their CTAS score, waiting time so far, other clinical information (e.g., chief complaint codes), and diagnostic test results. Based on the patient information, a physician decides which patient to see next (Ding et al. 2019, Li et al. 2021). A physician's selection of a new patient for an initial assessment marks the end of the waiting phase and the beginning of the *treatment phase*. During treatment, the physician may order diagnostic tests and re-assess the patient after the test results become available. This process may be repeated. The treatment phase ends when the physician makes a decision to either discharge the patient from the ED or admit the patient to the hospital for further treatment. An admitted patient has to wait in an ED bed before being transferred to an inpatient bed. This is referred to as *ED boarding*, during which the patient may require some attention from the nursing staff, but the physician is effectively done with the patient.

3.2. Current Practice of Physician Scheduling

During the study period, there were 15 shifts (thus 15 physicians) scheduled every day, two of which were dedicated to the fast track where patients with minor conditions were treated. The remaining 13 physicians worked in the main ED area during their scheduled shifts. The shift lengths in the main area vary between 6 to 8 hours. See Figure 1 for the start and end times for each of the 13 shifts in the main ED area. In this paper, we only focus on the main ED area, as it is the most congested part of the ED and most physicians are scheduled to work in this area.

Figure 1 The start and end times of the thirteen shifts in the main treatment area of our study ED.

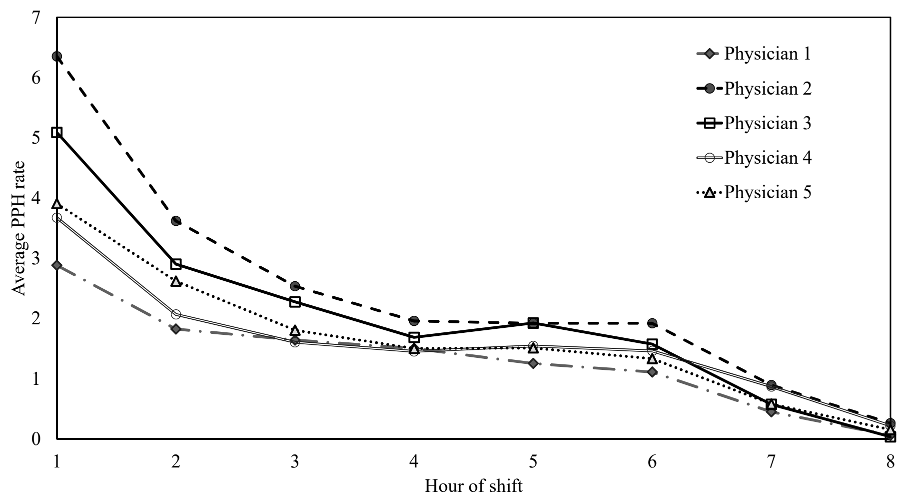


Physicians were assigned to these shifts based on a number of scheduling rules. The most common rules can be grouped into three categories: balance rules, pattern rules, and weekend rules. Balance rules ensure that physicians work an equal number of shifts in total (e.g., at least 12 and at most 14 shifts) and by type (e.g., day or night), proportional to FTE and the amount of time off requested. Pattern rules specify that physicians need to be scheduled at least 2 and at most 4 days in a row. Additionally, after a physician is scheduled for a consecutive number of days, she should have at least 2 and at most 4 days off. Night shifts should be scheduled on consecutive days, because no day shift is allowed for 3 days after a night shift and there need to be at least 21 days after a group of consecutive night shifts. In general, weekend rules are considered soft constraints that favor scheduling physicians for an entire weekend (defined as the first shift on Friday until the last shift starting on Sunday) and avoiding scheduling physicians to work on two consecutive weekends. However, in our study ED, there is a hard constraint on the maximum number of weekends that a physician is scheduled (again proportional to FTE).

The hospital produces the shift schedule several months in advance. Hence, physicians are not encouraged to express preferences (e.g., applying for time off during their scheduled shift days). Instead, they can exchange shifts with other physicians, as observed in our study ED.

Our data analysis shows that the productivity of a physician, measured by the number of new patients seen per hour (i.e., the PPH rate) changes significantly during the course of her shift and the pattern depends on individual physicians. See Figure 2 for the average hourly productivity of the five physicians who worked the most shifts in the main ED area over the 2-year study period. In particular, we observe that the PPH rate (i) decreases with the hour of the shift and (ii) depends on the physician. Next, we examine the factors that impact a physician's PPH rate and develop a model to predict a physician's PPH rate, with the ultimate goal of incorporating the prediction into the physician rostering problem so as to alleviate the mismatch between the supply and demand of emergency care.

Figure 2 **The average productivity of five physicians during each hour of their shifts.**



3.3. Empirical Investigation of the PPH Rate

In this section, we empirically examine factors that impact physician productivity. Our data include 6-hour, 7-hour, and 8-hour shifts. The 6-hour shifts are considered flexible shifts in the study ED and the data for this particular type of shift are less reliable. Hence, the 6-hour shifts are not included in our study. A total of 7,911 shifts remain in the data, 5,031 of which are 7-hour shifts (63.6%) and 2,880 of which are 8-hour shifts (36.4%). More than 100 physicians worked at least one shift during the study period. However, a significant number worked there on exchange from other hospitals in the same healthcare region and do not normally work in our study ED. To simplify the problem, we select the 52 physicians who have worked the most 7-hour shifts during the study period in the main area. We present the corresponding results for 7-hour shifts in the paper. We also select the 52 physicians who worked the most 8-hour shifts and perform the same analysis for 8-hour shifts. The results are qualitatively similar to that of the 7-hour shifts and thus are deferred to Appendix A.

3.3.1. Choice of Variables Our objective is to investigate the factors that drive a physician's productivity. Hence, the outcome variable is the number of new patients seen by a physician during a specific hour of her shift (i.e., the PPH). We denote the productivity of physician i during the m-th hour of her shift j by PPH_{ijm}.

Our data include timestamps of the patient flow and treatment process for each patient and the start and end times of each physician's shift, which allows us to calculate PPH_{ijm}.

As for the independent variables, a physician's characteristics, such as age, experiences, level of training, and risk attitude, might affect her productivity. However, such information is difficult to collect for each physician and is not available in our data. Hence, we define *Physician* as a unique identifier of an individual physician. As shown in Figure 2, the shift hour also affects a physician's productivity. Hence, we define a categorical variable *ShiftHour* to indicate the hour of a shift and include it in the model. We define a dichotomous variable *NightShift* to indicate whether a shift starts between 5 PM and midnight. We include *NightShift* in the model to verify whether it has an impact on a physician's productivity.

The workload and congestion level of the ED during the hour when PPH is measured have been reported to affect physician productivity; for example, see KC and Terwiesch (2009), Kuntz and Sülz (2013), Armony et al. (2015), Berry Jaeker and Tucker (2017), and Batt and Terwiesch (2017) for extensive evidence that the load of healthcare systems and the workload of emergency physicians have a statistically significant impact on their service times and thus their PPH rates. Hence, we define three variables, *EDCensus*, *WaitRoomCensus*, and *BoarderCensus*, which measure the time-averaged number of patients in the ED, in the waiting room, and waiting for inpatient beds, respectively, during the hour when the PPH rate is measured. We control for ED congestion levels by including them in our model. In a recent study, Duan et al. (2020) show that task switching has a significant impact on physicians' productivity and patient outcomes. However, because task switching is a decision at the operational level and is difficult to consider in the physician rostering problem, which is a tactical-level decision, we do not consider physician task switching in our model.

The study ED is in a teaching hospital, and physicians oftentimes perform teaching (or supervising) duties during their shifts. There are two types of learners: emergency medicine residents and medical students. It is known that the type of learners has an impact on physician productivity (Bhat et al. 2014). Hence, we define a categorical variable *Learner* with three levels, indicating respectively whether there is a resident learner, a student learner, or no learner in the shift. Physicians often need to take patients whose treatments at the ED are not yet complete from another physician whose shift is ending. These patients are called handover patients. Similarly, a physician who will be off duty soon needs to hand over her patients to other physicians on duty. This patient care handover takes time and effort from the physicians on both sides, and thus may have an impact on physician productivity. We use *HandoverTaken* to denote the number of handover patients a physician takes over from other physicians, and let *Handover* denote the number of handover patients given by this physician to others. We control for both variables in our model and check whether they improve the model fit. The summary statistics of all variables are provided in Table 1.

Table 1 Summary statistics for variables of interest.

Variables	7-hour shifts			8-hour shifts		
	All shifts	DayShift	NightShift	All shifts	DayShift	NightShift
PPH						
Mean (SD)	1.87 (1.51)	1.70 (1.50)	1.99 (1.51)	1.63 (1.40)	1.61 (1.40)	1.68 (1.41)
Range	[0, 14]	[0, 10]	[0, 14]	[0, 11]	[0, 11]	[0, 11]
EDCensus						
Mean (SD)	46.94 (11.06)	45.39 (11.84)	48.04 (10.33)	51.17 (10.73)	49.95 (10.90)	54.74 (9.35)
Range	[1.81, 91.75]	[6.75, 87.83]	[1.81, 91.75]	[6.75, 87.83]	[6.75, 87.83]	[12.45, 87.10]
WaitRoomCensus						
Mean (SD)	12.55 (6.39)	10.32 (5.86)	14.14 (6.28)	12.63 (6.13)	11.63 (5.80)	15.58 (6.14)
Range	[0, 37.32]	[0, 37.32]	[0, 36.87]	[0.06, 37.14]	[0.06, 37.14]	[0.76, 37.14]
HandoverTaken						
Mean (SD)	0.73 (1.74)	0.71 (1.88)	0.74 (1.63)	0.39 (1.25)	0.35 (1.19)	0.51 (1.41)
Range	[0, 18]	[0, 14]	[0, 18]	[0, 12]	[0, 12]	[0, 11]
Handover						
Mean (SD)	0.06 (0.41)	0.08 (0.48)	0.05 (0.36)	0.07 (0.46)	0.07 (0.47)	0.06 (0.43)
Range	[0, 10]	[0, 9]	[0, 10]	[0, 9]	[0, 9]	[0, 7]
BoarderCensus						
Mean (SD)	9.26 (4.62)	10.08 (5.09)	8.67 (4.15)	9.54 (4.80)	10.02 (4.99)	8.11 (3.83)
Range	[0, 33.37]	[0, 33.37]	[0, 29.30]	[0, 33.37]	[0, 33.37]	[0, 29.53]
Learner						
Resident $n(\%)$	602 (2.5%)	469 (4.6%)	133 (0.9%)	392 (2.4%)	320 (2.7%)	72 (1.8%)
Student $n(\%)$	14 (<0.1%)	14 (0.1%)	0 (0%)	40 (0.2%)	24 (0.2%)	16 (0.4%)
No learner $n(\%)$	23,933 (97%)	9,730 (95%)	14,203 (99%)	15,584 (97%)	11,608 (97%)	3,976 (98%)
Observations	24,549	10,213	14,336	16,016	11,952	4,064
Shift count	3,507	1,459	2,048	2,002	1,494	508

3.3.2. Model Development and Results The dependent variable is a count variable. Hence, a generalized linear model from the Poisson family (or negative binomial family) is likely to be a better choice than a linear regression model. Let $\mathbb{E}(PPH|\boldsymbol{X})$ denote the mean of the predicted Poisson distribution that fits a physician's productivity level. Then, the model is specified as follows:

$$\log \mathbb{E}(PPH|\boldsymbol{X}) = \beta_0 + \beta' \boldsymbol{X}, \tag{1}$$

where X represents the independent variables (or a subset of them) as discussed in Section 3.3.1. We estimate the model specification in Eq. (1) using Poisson regression and negative binomial regression, which is a generalization of Poisson regression that relaxes the highly restrictive assumption that the variance is equal to the mean. We start by including all of the variables and gradually remove variables to check the goodness of fit with and without the variables of interest. We estimate six models and find no significant improvement from using the negative binomial model over the Poisson model by comparing the corresponding values of the log-likelihood (or AIC, BIC). Moreover, a Chi-square goodness-of-fit test fails to reject the assumption of a Poisson model (with a p-value > 0.999). Hence, we only present the results from the Poisson model in Table 2.

Table 2 Estimation results for the effect of various factors on PPH rates for 7-hour shifts. Model 1 includes all variables discussed in Section 3.3.1. Models 2-6 gradually remove variables from Model 1.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1.662***	1.663***	1.668***	1.63***	1.546***	1.393***
	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.03)
Physician (base = MD003)						
MD004	−0.22***	−0.22***	−0.226***	−0.218***	−0.212***	−0.203***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
MD005	0.01	0.017	0.015	0.017	0.018	0.025
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
MD006	0.072	0.081·	0.078	0.079·	0.08·	0.082·
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
⋮						
MD349	−0.427***	−0.424***	−0.423***	−0.414***	−0.398***	−0.395***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
ShiftHour (base=Hour1)						
Hour2	−0.491***	−0.491***	−0.491***	−0.497***	−0.502***	−0.499***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hour3	−0.753***	−0.753***	−0.752***	−0.754***	−0.763***	−0.758***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Hour4	−0.881***	−0.881***	−0.878***	−0.874***	−0.885***	−0.877***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Hour5	−0.954***	−0.954***	−0.949***	−0.938***	−0.947***	−0.937***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Hour6	−1.111***	−1.111***	−1.105***	−1.092***	−1.103***	−1.092***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Hour7	−1.952***	−1.952***	−1.945***	−1.945***	−1.956***	−1.944***
	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)	(0.026)
NightShift (base=DayShift)	0.165***	0.16***	0.15***	0.153***	0.182***	0.163***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
EDCensus	−0.011***	−0.01***	−0.008***	−0.008***	−0.003***	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	
WaitRoomCensus	0.015***	0.015***	0.012***	0.012***		
	(0.001)	(0.001)	(0.001)	(0.001)		
HandoverTaken	−0.017***	−0.017***	−0.016***			
	(0.002)	(0.002)	(0.002)			
Handover	−0.114***	−0.113***	−0.113***			
	(0.02)	(0.02)	(0.02)			
BoarderCensus	0.008***	0.007***				
	(0.001)	(0.001)				
Learner (base=No Learner)						
Resident	0.15***					
	(0.026)					
Student	0.021					
	(0.106)					
Log Likelihood	-35,985	-35,997	-36,014	-36,053	-36,116	-36,157
AIC	72,102	72,122	72,154	72,227	72,353	72,432
BIC	72,637	72,641	72,665	72,722	72,839	72,910
Observations	24,549	24,549	24,549	24,549	24,549	24,549

Notes. This table reports the estimation results from the Poisson regression. Robust standard errors are shown in the parentheses. ***p<0.001; **p<0.01; *p<0.05

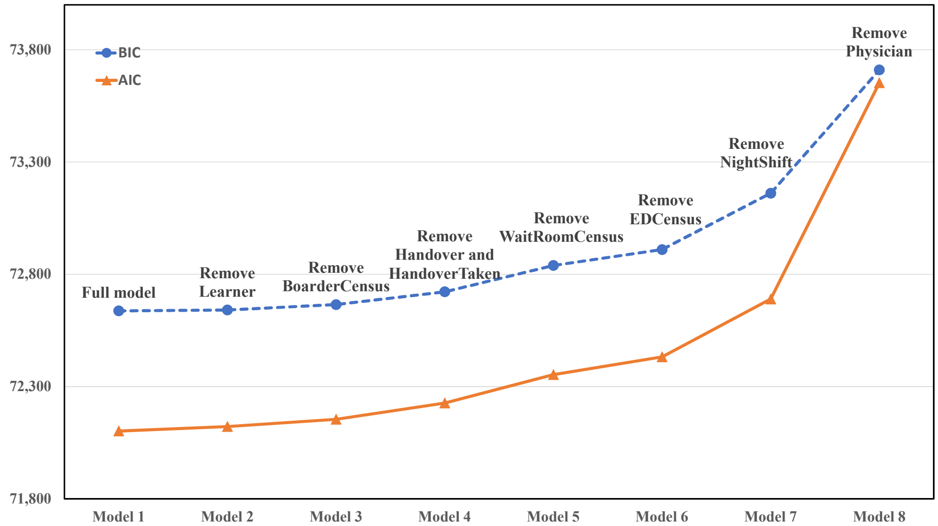
From Table 2, we observe that across all six models, a physician's productivity decreases with the hour of the shift (*ShiftHour*) and is higher during night shifts (*NightShift* = 1). Fatigue (both physical and mental) may explain why the PPH rate decreases with the hour of the shift. Another possible explanation is that a physician needs to spend time and effort on existing patients under her care. The number of existing patients accumulates through the course of the shift until the end of the shift approaches (Chan 2018). As for the effect of an individual physician, some levels of the variable *Physician* are significant (e.g., MD004 and MD349 in Table 2) and some are not (e.g., MD005 and MD006), which shows that there is heterogeneity in productivity among physicians. However, some physicians are more homogeneous than others. This in fact motivates the clustering of physicians based on their productivity, which we discuss in detail in Section 6.

We plot the AIC/BIC of the six models presented in Table 2 in Figure 3. While Table 2 shows that all the variables included in the model are statistically significant, we want to identify a subset of the variables that are more important in explaining the variations in a physician's productivity and that are easier to incorporate into our physician rostering model. To make this comparison more explicit, we further estimate two other models, namely, Model 7 which removes *NightShift* from Model 6, and Model 8 which removes both *Physician* and *NightShift* from Model 6. Through Figure 3, we observe that removing *Learner, BoarderCensus*, *Handover, HandoverTaken*, *WaitRoomCensus* and/or *EDCensus* only marginally increases AIC/BIC; on the other hand, removing either (or both) *Physician* and *NightShift* significantly increases AIC or BIC, indicating a much worse model fit. If we further remove the variable *ShiftHour* from Model 8, the AIC (BIC) increases dramatically from 73,653 (73,710) to 85,434 (85,426). We conclude that the three variables *Physician*, *ShiftHour*, and *NightShift* are the most important driving factors of the PPH rate. Hence, we do not include the other variables in our model due to their negligible impact and the difficulty of including them in our optimization model. In other words, we will use Model 6 in Table 2 to predict the average PPH rate of a physician during a particular hour of her shift in the stochastic optimization model. Note that in the interest of space, we include the coefficients for part of the physician variable in Table 2, and the complete estimation results are provided in Table A.2 of Appendix A.

4. An Optimization Model of the Physician Rostering Problem

In this section, we formulate the physician rostering problem. The objective function is to minimize the total hourly mismatch between the patient demand for emergency care (measured by newly arriving patients per hour) and the service provided by the ED (measured by the total PPH rate over all physicians working in the ED during a particular hour). We introduce a two-stage stochastic programming formulation of the rostering problem, where the uncertainty associated with time-varying ED arrivals and hourly physician productivity is included.

Figure 3 **The AIC and BIC of the Poisson models. A lower values of AIC/BIC indicates a better model fit.**



Most measures of ED productivity in the OM literature focus on throughput, such as length of stay (LOS) and service (or processing) times (Song et al. 2015, Batt and Terwiesch 2017, Song et al. 2018). In the healthcare literature, physician productivity is usually measured as the number of patients treated by a physician (or PPH rate). The previous section shows that the number of patients newly treated by a physician in an hour can be modeled with a Poisson distribution, where the average PPH rate for individual physicians can be estimated with a regression model. Aside from using the PPH rate as a productivity measure, we have also explored the relationships between PPH rates and the more established productivity measures from the OM literature (see Appendix B). We conclude that physicians with a higher PPH rate have a shorter average patient LOS and service times. Hence, it is plausible that matching physician productivity in terms of PPH rates with uncertain demands in the PRP will result in similar conclusions to a study focusing on ED throughput in our optimization formulation. However, including heterogeneity in PPH rates is less complex and more acceptable from a practical perspective, as it is widely accepted by ED physicians that they see different numbers of patients during their shifts.

4.1. Formulation of the Physician Rostering Problem

Let I be the set of physicians, J be the set of days of the planning period (where $|J| = n$), and K be the set of shifts per day. Without loss of generality, day 1 is assumed to be a Monday and day n is a Sunday. Each day is divided into hourly intervals $t \in T = \{0, 1, \ldots, 23\}$. Furthermore, we define the subset K_D as the set of daytime shifts and K_N as the set of night shifts. Hence, K_D and K_N are subsets of K.

Let Λ_{jt} denote the number of patient arrivals in hour $t \in T$ on day $j \in J$, and Ψ_{im}^{jk} denote the productivity of physician $i \in I$ in the m-th hour of the shift if the physician is scheduled to work shift $k \in K$ on day $j \in J$. To relate the t-th hour of a day to the m-th hour of shift k, let $f_{kt} = m$ if the m-th hour of shift k equals hour t, and $f_{kt} = 0$ otherwise. Consequently, we define $\Psi_{i0}^{jk} = 0$ for all $i \in I, j \in J, k \in K$. Note that for any candidate schedule, the productivity of a physician, i.e., Ψ_{im}^{jk}, can be estimated by Model 6 in Table 2, where $Physician = i$, $ShiftHour = m$, and $NightShift$ can be determined by the hour of day t.

Note that we divide the planning horizon into n days since the patient arrival pattern and the staffing levels are recurring on a daily basis in our case study. If this was not the case, and the arrival rates and number of shifts (or their timing) would also depend on the day of the week, then the planning horizon could be divided in specific days of the week instead of (generic) days. This would have no impact on the number of decision variables for the PRP since K and T become seven times larger whereas J becomes one seventh of its original value. In our model, we choose to present the results with days as the recurring time interval, but one can generalize it into models with weeks (even months) as the recurring time interval.

We formulate the PRP as a two-stage stochastic binary problem. The stochastic program formulation is denoted by PRP_0 thereafter. In the first stage, before observing the uncertain arrival volumes Λ_{jt} and physician productivity $\Psi_{i f_{kt}}^{jk}$, we decide on the assignment of physicians to each shift of each day during the planning horizon. The decision variables x_{ijk} represent whether or not physician $i \in I$ is assigned to shift $k \in K$ on day $j \in J$. More specifically,

$$x_{ijk} = \begin{cases} 1, & \text{if physician } i \text{ is assigned to shift } k \text{ on day } j, \\ 0, & \text{otherwise.} \end{cases}$$

After the patient arrivals and physician productivity during each hour are observed, the second-stage problem calculates any mismatch between the demand for emergency care and the supply provided in terms of physician productivity for a given assignment. The PRP_0 problem can be formulated as follows:

$$\min_{x} \quad \mathbb{E}_{\Lambda,\Psi}\left[Q(\Lambda,\Psi,x)\right] \tag{2}$$

The objective (2) minimizes the mismatch between the demand for emergency care and the supply measured by physician productivity, denoted by $Q(\Lambda,\Psi,x)$. See the detailed definition of $Q(\Lambda,\Psi,x)$ in Eq. (16).

$$\sum_{i\in I} x_{ijk} = 1, \qquad j \in J, k \in K \tag{3}$$

Constraints (3) ensure that exactly one physician is assigned to each shift of the day.

$$\sum_{k\in K} x_{ijk} \leq 1, \qquad i \in I, j \in J \tag{4}$$

Constraints (4) enforce that a physician is not assigned to more than one shift per day.

$$\sum_{j\in \tilde{J}} \sum_{k\in \tilde{K}} x_{ijk} \geq L(\tilde{J},\tilde{K}), \qquad i \in I, \tilde{J} \in J', \tilde{K} \in K' \tag{5}$$

$$\sum_{j \in \tilde{J}} \sum_{k \in \tilde{K}} x_{ijk} \leq U(\tilde{J}, \tilde{K}), \qquad i \in I, \tilde{J} \in J', \tilde{K} \in K' \tag{6}$$

Constraints (5) and (6) specify the minimum and maximum number of shifts of certain types that can be performed by a physician over the planning horizon, where $J' \subset P(J)$ is a set of subsets of J taken from the power set $P(J)$ (i.e., $\tilde{J} \subseteq J$) and similarly $K' \subset P(K)$ (such that $\tilde{K} \subseteq K$). These subsets capture combinations of days and types of shifts (for example, weekend and night shifts). Furthermore, $L(\tilde{J}, \tilde{K})$ and $U(\tilde{J}, \tilde{K})$ are the minimum and maximum number of shifts for the combination of subsets $\tilde{J}$ and $\tilde{K}$.

Constraints (3) through (6) correspond to the *balance rules* in our problem formulation. The next eight constraints correspond to the *pattern rules*.

$$\sum_{k \in K_N} x_{i(j-1)k} + \sum_{k \in K_D} x_{ijk} \leq 1, \qquad i \in I, j \in J \setminus \{1\} \tag{7}$$

Constraints (7) guarantee that a physician who is assigned to a night shift must not be assigned to a day shift on the next day.

$$\sum_{k \in K_N} x_{i(j-3)k} + \sum_{j'=0}^{2} \sum_{k \notin K_N} x_{i(j-j')k} \leq 1, \qquad i \in I, j \in J \setminus \{1, 2, 3\} \tag{8}$$

Constraints (8) specify that a physician who is assigned to a night shift cannot be assigned to a shift of another type on the next three days.

$$3 \sum_{k \in K_N} x_{i(j-20)k} + \sum_{j'=0}^{17} \sum_{k \in K_N} x_{i(j-j')k} \leq 3, \qquad i \in I, j \in J \setminus \{1, 2, \ldots, 20\} \tag{9}$$

Constraints (9) specify that a physician who is assigned to a group of (at most three) consecutive night shifts cannot be assigned to a night shift for 20 days.

$$\sum_{k \in K} x_{i2k} \geq \sum_{k \in K} x_{i1k}, \qquad i \in I \tag{10}$$

$$\sum_{k \in K} x_{ijk} \geq \sum_{k \in K} x_{i(j-1)k} - \sum_{k \in K} x_{i(j-2)k}, \qquad i \in I, j \in J \setminus \{1, 2\} \tag{11}$$

$$\sum_{j'=0}^{4} \sum_{k \in K} x_{i(j-j')k} \leq 4, \qquad i \in I, j \in J \setminus \{1, 2, 3, 4\} \tag{12}$$

Constraints (10), (11), and (12) correspond to at least 2 and at most 4 consecutive working days for a physician, respectively.

$$\left(1 - \sum_{k \in K} x_{ijk}\right) \geq \sum_{k \in K} x_{i(j-2)k} - \sum_{k \in K} x_{i(j-1)k}, \qquad i \in I, j \in J \setminus \{1, 2\} \tag{13}$$

$$\sum_{j'=0}^{4} \left(1 - \sum_{k \in K} x_{i(j-j')k}\right) \leq 4, \qquad i \in I, j \in J \setminus \{1, 2, 3, 4\} \tag{14}$$

Constraints (13) and (14) express the minimum and maximum number of days off after a consecutive number of days that a physician is scheduled.

$$x_{ijk} \in \{0,1\} \qquad i \in I, j \in J, k \in K \qquad (15)$$

Constraints (15) require the decision variables of the first stage to take on binary values.

Note that we do not include the *weekend rules* in our problem formulation since they are soft constraints. However, they can easily be formulated. The following constraints assign the same physician to all three consecutive weekend days:

$$\sum_{j'=0}^{2} \sum_{k \in K} x_{i(j+j')k} = 3, \quad i \in I, \; j \in \{5, 12, \ldots, n-2\}.$$

A physician should not work the next weekend after working a weekend:

$$\sum_{j'=0}^{2} \sum_{k \in K} x_{i(j+j')k} + \sum_{j'=7}^{9} \sum_{k \in K} x_{i(j+j')k} \leq 3, \quad i \in I, \; j \in \{5, 12, \ldots, n-9\}.$$

Similarly, we can easily include physician preferences to the problem formulation. For instance, the constraints $\sum_{k \in K} x_{ijk} \geq Z_{ij}, i \in I, j \in J$, require that physician i works on day j if she has a preference to work on that day (i.e., if $Z_{ij} = 1$). If physician i has a preference not to work on day j (i.e., if $Z'_{ij} = 1$), we can add the following constraints to our model to ensure that no shift is assigned to that physician on that day: $\sum_{k \in K} x_{ijk} \leq 1 - Z'_{ij}, i \in I, j \in J$. These constraints can also be added as soft constraints such that any deviations from either the weekend rules or physician preferences are penalized as a weighted sum in objective function (2).

The function $Q(A, P, x)$ is then used to evaluate the mismatch in the assignment x between the arrivals A and the productivity P, where A and P represent a matrix of realizations from arrivals Λ and productivities Ψ, respectively. Specifically, the mismatch $Q(A, P, x)$ over the planning horizon is defined as

$$\begin{aligned}
Q(A,P,x) = \min_{M} \quad & \sum_{j \in J} \sum_{t \in T} M_{jt} && (16)\\
\text{s.t.} \quad & A_{jt} + M_{j(t-1)} - \sum_{i \in I} \sum_{k \in K} x_{ijk} P^{jk}_{i f_{kt}} \leq M_{jt}, \quad \forall j \in J, t \in T \setminus \{0\}, && (17)\\
& A_{10} - \sum_{i \in I} \sum_{k \in K} x_{i1k} P^{1k}_{i f_{k0}} \leq M_{10}, && (18)\\
& A_{j0} + M_{(j-1),23} - \sum_{i \in I} \sum_{k \in K} x_{ijk} P^{jk}_{i f_{k0}} \leq M_{j0}, \quad j \in J \setminus \{1\}, && (19)\\
& M_{jt} \geq 0, \quad j \in J, t \in T. && (20)
\end{aligned}$$

In the second-stage model, the variable M_{jt} represents the mismatch between the patient demand and the

ED productivity in hour $t \in T$ on day $j \in J$. The objective function (16) sets these variables to the lowest possible values. Constraints (17) to (19) specify the hourly mismatch between patient demand and the total physician productivity. The left-hand side of the constraints (17) to (19) represents the effective patient demand minus the number of new patients treated during hour t on day j, where the effective patient demand consists of new patients arriving to the ED and patients still waiting in the ED from the previous hour.

The PRP_0 is a two-stage stochastic integer program (SIP) with binary variables in the first stage, and continuous variables in the second stage. The second-stage problem is always feasible and it is a bounded LP. As a result, the function $Q(\cdot)$ is piece-wise linear and convex (Birge 1997). We refer to the decision variables in the first-stage model as *assignment variables* and the decision variables in the second-stage model as *mismatch variables*.

4.2. Solving the Physician Rostering Problem

The objective function of the PRP_0 includes expected values over the random matrices Λ and Ψ. We use *sample average approximation* (SAA) to overcome the difficulty in evaluating this expected value. Using Monte Carlo sampling, we first obtain a set of independent and identically distributed samples from the random variables in the matrices. These realizations are called *scenarios*. Let $A_{jt}(s)$ be the realization of Λ_{jt} in scenario $s \in S$, i.e., this represents the number of new patient arrivals in hour $j \in J$ on day $t \in T$ under scenario $s \in S$. The expected value is then replaced by the sample average over these scenarios, where we assign equal weight $1/|S|$ to each scenario:

$$\mathbb{E}_{\Lambda,\Psi}\left[Q(\Lambda,\Psi,x)\right] \approx \frac{1}{|S|}\sum_{s\in S} Q(A(s),P(s),x). \tag{21}$$

With the scenarios fixed, we can obtain the deterministic equivalent of the SAA problem, which is included in Appendix C. In this formulation, the first-stage decision variables x are the same as defined before, whereas the second-stage decision variables are specified for each scenario. So, $M_{jt}(s)$ denotes the surplus demand for emergency care in hour j on day t under scenario s.

To solve the SAA problem, we use a decomposition strategy where we iteratively add constraints since most of the constraints are not active in an optimal solution. In particular, we use an iterative algorithm known as the *L-shaped method*, first introduced by Van Slyke and Wets (1969) for two-stage SIP models. (See Laporte and Louveaux (1993) and Angulo et al. (2016) for more developments when binary variables are included.) We decompose our problem into a master problem (MP), where shift assignment decisions are made, and a set of subproblems (SP) where the mismatch variables are set for each scenario $s \in S$. Observe that for fixed values of x, the deterministic equivalent of the SAA problem (as presented in Appendix C) decomposes into $|S|$ independent subproblems. Each subproblem is the same as the LP formulation in (16)-(20) (i.e., subproblem s corresponds to solving $Q(A(s), P(s), x)$, whereas the master problem for iteration v is as follows:

$$
\begin{aligned}
(MP) \quad \min_{x} \quad & \theta \\
\text{s.t.} \quad & Ax = b, && (22)\\
& E^{\ell} x + \theta \geq e^{\ell}, \ \ell = 1, \ldots, v-1, && (23)\\
& x \in \{0,1\}^{|I|\times|J|\times|K|}, \ \theta \in \mathbb{R}, && (24)
\end{aligned}
$$

where (x^v, θ^v) is the optimal solution in iteration v. The constraints (22) correspond to the constraints (3)-(14) from the PRP_0 and the constraints (23) are the optimality cuts included from the previous iterations. Note that if optimality constraint (23) is not present, we set θ to be $-\infty$ and do not consider it in the computation of x^v. To derive these cuts, we consider the dual of $Q(A(s), P(s), x^v)$, where $\pi^{\ell}(s)$, $\rho^{\ell}(s)$ and $\gamma^{\ell}(s)$ are the dual variables associated with constraints (17), (18) and (19), respectively, for each scenario $s \in S$ in iteration ℓ. We then define $E^{\ell} x$ and e^{ℓ}, $\forall \ell = 1, \ldots, v$, as follows:

$$
E^{\ell} x = \frac{1}{|S|} \sum_{s \in S} \left(\sum_{j \in J} \sum_{t \in T \setminus \{0\}} \pi^{\ell}_{jt}(s) \sum_{k \in K} \sum_{i \in I} x_{ijk} P^{ik}_{if_{kt}}(s) + \rho^{\ell}(s) \sum_{k \in K} \sum_{i \in I} x_{i1k} P^{1k}_{if_{k0}} + \sum_{j \in J \setminus \{1\}} \gamma^{\ell}_{j}(s) \sum_{k \in K} \sum_{i \in I} x_{ijk} P^{jk}_{if_{kt}}(s) \right), \tag{25}
$$

and

$$
e^{\ell} = \frac{1}{|S|} \sum_{s \in S} \left(\sum_{j \in J} \sum_{t \in T \setminus \{0\}} \pi^{\ell}_{jt}(s) \, A_{jt}(s) + \rho^{\ell}(s) \, A_{1,0}(s) + \sum_{j \in J \setminus \{1\}} \gamma^{\ell}_{j}(s) \, A_{j,0}(s) \right). \tag{26}
$$

Let $\omega^v = e^{\ell} - E^{\ell} x^v$. If $\theta^v \geq \omega^v$, then we stop the algorithm with the optimal solution given by x^v. Otherwise, we add the optimality cut and perform the next iteration.

4.3. Upper and Lower Bounds on the Optimal Solution

Since the solution from the L-shaped method is only an approximation of the true optimal solution, it is important to evaluate the deviation in optimality. Next, we apply a Monte Carlo bounding technique (Mak et al. 1999) to obtain the upper bound and lower bound on the optimal objective value, which can be used to evaluate the quality of any solutions. Let z^* denote the optimal value of the objective of the PRP_0 problem, which can be approximated by a sample average when solving the extensive form (PRP_{0m}) for a given set of scenarios S_m (see Appendix C). Hence, we have

$$
z^*_m = \frac{1}{|S_m|} \sum_{s \in S_m} Q(A(s), P(s), x^*_m),
$$

where z_m^* and x_m^* are respectively the optimal objective function value and optimal solution for the set of scenarios S_m. By Mak et al. (1999), we have

$$\mathbb{E}\left[z_m^*\right] = \mathbb{E}\left[\min_{x_m}\left\{\frac{1}{|S_m|}\sum_{s\in S_m} Q(A(s),P(s),x_m)\right\}\right] \le z^*.$$

Hence, the lower bound on z^* can be obtained by solving the extensive form of the PRP_0 problem (see Appendix C) for multiple sets of scenarios, where each set of scenarios is independently generated. More specifically, a lower bound, denoted by $\overline{L}(n_l)$, is

$$\overline{L}(n_l) = \frac{1}{n_l}\sum_{j=1}^{n_l} z_j^* \le z^*, \tag{27}$$

where n_l is the number of sets of scenarios. Note that the value of the objective function for an individual set of scenarios (among all n_l sets) can exceed the optimal value z^*. However, $\overline{L}(n_l)$ provides a lower bound when there are sufficiently many sets of scenarios.

To obtain an upper bound, consider a feasible solution $\hat{x}$. It can be the solution obtained by the L-shaped method after a few iterations for a given set of scenarios S. The corresponding value of the objective function can be estimated as follows:

$$z(\hat{x},S') = \frac{1}{|S'|}\sum_{s\in S'} Q(A(s),P(s),\hat{x}),$$

where S' is the set of scenarios that are used to evaluate the objective value of $\hat{x}$. Note that by the L-shaped method, S' should be independent of S which is used to construct $\hat{x}$. Due to the suboptimality of $\hat{x}$, a straightforward estimate of the upper bound on z^*, denoted by $\overline{U}(n_u)$, is the average objective function value over n_u sets of scenarios. Hence, we have

$$\overline{U}(n_u) = \frac{1}{n_u}\sum_{j=1}^{n_u} z(\hat{x},S_j) \ge z^*. \tag{28}$$

Since both the lower and upper bounds are estimated from samples of scenarios, the sample variances of the lower and upper bounds can be estimated by $s_l^2(n_l)$ and $s_u^2(n_u)$, respectively, where

$$s_l^2(n_l) = \frac{1}{n_l-1}\sum_{m=1}^{n_l}\left(z_m^* - \overline{L}(n_l)\right)^2 \text{ and } s_u^2(n_u) = \frac{1}{n_u-1}\sum_{m=1}^{n_u}\left(z(\hat{x},S_m) - \overline{U}(n_u)\right)^2.$$

As a result, the confidence interval (with significance level α) for the optimality gap at $\hat{x}$ is given by

$$\left[0,\ [\overline{U}(n_u) - \overline{L}(n_l)]^+ + \hat{\epsilon}_u + \hat{\epsilon}_l\right], \tag{29}$$

where $[A]^+ = \max\{A,0\}$, $\hat{\epsilon}_u = s_u(n_u)t_{n_u-1,\alpha/2}/n_u$, $\hat{\epsilon}_l = s_l(n_l)t_{n_l-1,\alpha/2}/n_l$, and $t_{n,\alpha}$ equals the $1-\alpha$ percentile of a t-distribution with n degrees of freedom. If the confidence interval is tight, then the optimality gap at $\hat{x}$ is small.

5. Comparison of Different Physician Schedules

In the previous section, we propose a stochastic optimization model to solve for the optimal physician schedule. To evaluate the impact of a physician's schedule, taking stochastic arrivals and heterogeneous physician productivity into consideration, we compare the performance of such a schedule with alternative schedules through a simulation study.

5.1. Study Setting

We use the simulation software Simio (version 12) to simulate the ED process in our study hospital. The system parameters are estimated from our 2-year dataset. In our setting, a physician schedule is the assignment of the 52 physicians into the 13 shifts (see Figure 1) per day over a 4-week planning horizon. In practice, the schedule is planned every 6 months. However, the schedules largely repeat themselves. Hence, we focus on 4 weeks to demonstrate the benefits of our proposed schedule. We also assume that all 13 shifts are 7-hour shifts. The shift start times are shown in Figure 1. We modify the shift end times to make the shift duration 7 hours.

The ED operations is modeled as a single station queue with a time-dependent number of servers (physicians). The patient arrival process follows a non-homogeneous Poisson process with hourly rates estimated from the average number of patient arrivals in each hour of the day over the 2-year study period (see, e.g., Kim and Whitt 2014). Upon arrival, patients enter a single queue and wait for treatment if all physicians are busy. For any time of day under a given physician schedule, we know exactly how many physicians are working, who they are, and which shift a physician is working on. When a physician completes the treatment of a patient, the physician signs up the next patient waiting in the queue. Note that we do not consider any patient prioritization in the simulation, since it is complex and does not impact our performance measure, i.e., the average time that patients have to wait before being seen by a physician. As we established in Section 3.3, the number of new patients a physician could treat per hour, i.e., the PPH, follows a Poisson distribution with rate μ_{ijh} for physician i in the h-th hour of shift j. Hence, in our simulation model, the service time of a patient being treated by physician i in the h-th hour of shift j follows an exponential distribution with rate μ_{ijh}, where μ_{ijh} is predicted from the Poisson model specified in Eq. (1) in Section 3.3.

5.2. Alternatives of Physician Schedules

In this section, we describe the physician schedules that we compare with our simulation model. Each schedule assigns the 52 physicians to the 13 shifts per day over a 28-day planning horizon. To make it fair, each physician has to be assigned to exactly 7 shifts. No weekend rules or physician preferences are included to allow more flexibility in the assignment decisions. We will relax this assumption in Section 5.4

Table 3 The computational results of three candidate solutions to the physician rostering problem based on 20, 40 and 60 scenarios.

Candidate solution	$\hat{x}^{20}$	$\hat{x}^{40}$	$\hat{x}^{60}$
Number of scenarios	20	40	60
Lower bound: $\overline{L}(10)$	19,085	20,771	21,851
Upper bound: $\overline{U}(500)$	22,539	22,539	22,539
95% confidence interval	[0, 3701]	[0, 1905]	[0, 811]
CPU time (in minutes)	17	43	118

by allowing physician preferences (both for weekend versus weekday shifts and versus night shifts) and in Section 6 through physician clustering.

Optimal Schedule: We solve the PRP_0 problem using the solution method discussed in Section 4.2. The quality of the solution improves as the number of scenarios increases, and the confidence interval on the optimality gap becomes tighter. However, the computational effort increases exponentially at the same time. We solve the PRP_0 problem based on 20, 40 and 60 scenarios, and the corresponding solutions are denoted by $\hat{x}^{20}$, $\hat{x}^{40}$, and $\hat{x}^{60}$, respectively. We also calculate the lower and upper bounds for each solution with $n_l = 10$ and $n_u = 500$. Table 3 presents the results of the three candidate solutions.

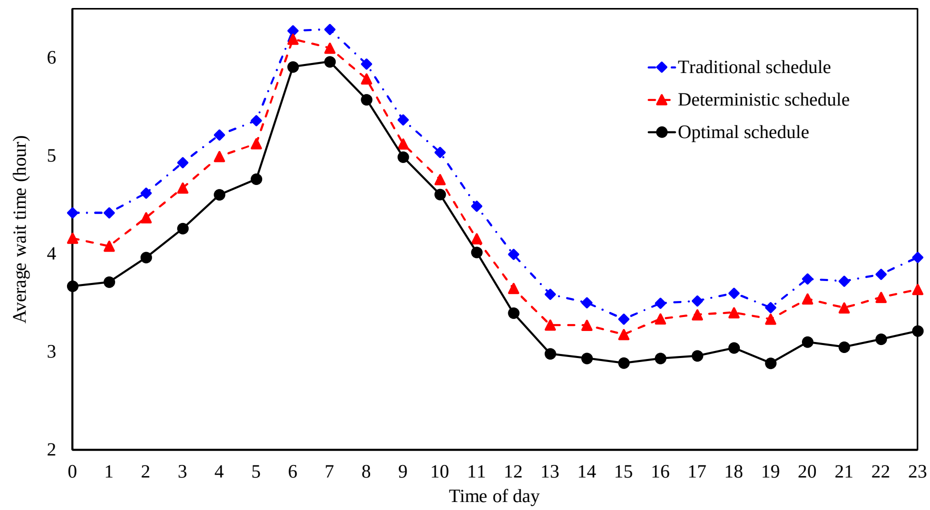
The deviation in objective values of the solution with 60 scenarios (i.e., $\hat{x}^{60}$) from that of the optimal solution has a tight confidence interval. More specifically, the length of the confidence interval is 3.7% of the lower bound and 3.6% of the upper bound. In other words, the L-shaped method with 60 scenarios provides an reasonably accurate solution for the PRP_0 within 2 hours of computation time. Hence, $\hat{x}^{60}$ is a good approximation of the optimal schedule, and it can serve as a benchmark for other alternative policies in the simulation study.

Deterministic Schedule: The deterministic schedule is obtained by solving the deterministic version of the PRP_0 problem (see the formulation in Section 4.1) with all random variables replaced by their corresponding average values. Hence, the deterministic schedule takes the heterogeneity in physician productivity into consideration but not the randomness in patient arrivals and physician productivity. The computational complexity of the deterministic schedule is significantly lower than its stochastic counterpart.

Traditional Schedule: We also consider a schedule generated by the scheduler in the hospital using a traditional approach, where neither the heterogeneity in physician productivity nor the stochastic nature of patient arrivals and physician productivity are considered. Hence, the traditional schedule is based on the status quo practice of our study ED where all physicians are regarded as homogeneous.

In addition to the schedules discussed above, we also consider a schedule derived from Camiat et al. (2021) in our simulation experiments. As discussed in Section 2.3, Camiat et al. (2021) is the only other study

Figure 4 **Comparison of the average ED wait times under three different schedules using simulation.**



in which heterogeneity in physician productivity is considered when solving the PRP. In their approach, the stochastic nature of the physician rostering problem is not included (similar to the deterministic schedule) and the PPH rate for each physician is assumed to be constant during a physician's shift. In contrast, we use time-varying PPH rates when solving the PRP.

5.3. Simulation Results

We simulate the ED operations over 28 days and evaluate the average time that a patient waits before seen by a physician over 200 replications. The average wait times under the traditional schedule, the schedule based on Camiat et al. (2021), the deterministic schedule, and the optimal schedule are 4.28 hours, 4.23 hours, 4.05 hours, and 3.71 hours, respectively. In our discussion, we use the traditional schedule as our baseline for comparison. Next, we compare it against the scheduling approach proposed by Camiat et al. (2021). When the heterogeneity in physician productivity is considered when assigning physicians to shifts, the average wait times can be reduced by 5.4% (from 4.28 hours to 4.05 hours). When the stochastic nature of the ED environment is included, the average ED wait times can be further reduced by 8.4% (from 4.05 hours to 3.71 hours). In other words, our formulation of the physician rostering problem can potentially reduce the overall average ED wait times by 13.3%. Figure 4 shows the average time-of-day patient wait times over 200 simulation replications under each of these three schedules. Note that the performance of the schedule based on Camiat et al. (2021) is excluded, as it is similar to the traditional schedule. We observe that the optimal schedule performs significantly better than the other two schedules for each of the 24 hours.

When considering the schedule based on Camiat et al. (2021), we conclude that including the non-stationarity of individual physicians' PPH rates reduces the average wait time by 4.3% (from 4.23 hours to

4.05 hours). Overall, the optimal schedule reduces the average wait time by 12.3% compared to the schedule based on the approach proposed by Camiat et al. (2021) (from 4.23 hours to 3.71 hours).

5.4. Physician Preferences

To generate the physician rosters for our numerical results in the previous section, we included constraints such that all physicians have to perform an equal number of shifts by type (i.e., day versus night shifts). To create more flexibility such that physician preferences can be included and to better understand the impact of physician preferences on average wait times, we study the following four additional settings:

- Setting 1: Each physician must work seven shifts.
- Setting 2: Each physician must work a minimum of four shifts and a maximum of eight shifts.
- Setting 3: Setting 2 plus 25% of the physicians (randomly chosen) are not willing to work night shifts and 50% of the physicians (randomly chosen) are not willing to work weekend shifts.
- Setting 4: Setting 2 plus 25% of the physicians (randomly chosen) are not willing to work day shifts and 50% of the physicians (randomly chosen) are not willing to work weekend shifts.

Recall that we have 52 physicians that need to be assigned to 13 shifts per day and we use a planning horizon of 28 days. This means that on average each physician has to be assigned to seven shifts during the planning horizon. In that regard, Setting 1 is comparable to the setting of the results obtained in Section 5.3, except that Setting 1 has more flexibility in assigning particular types of shifts to physicians. Setting 2 is needed to create more flexibility when physician preferences are included in Settings 3 and 4. When this flexibility is not provided in Settings 3 and 4 (i.e., imposing physician preferences on Setting 1), there can be instances where it is not possible to find any feasible solution to the PRP. However, imposing the constraint that physicians can be assigned at most eight shifts instead of exactly seven shifts, can result in physician rosters where a physician is assigned only one extra shift at most. In other words, we provide minimal additional flexibility to allow for physician preferences.

The simulation results suggest that the average patient wait times for the optimal physician schedules under Settings 1, 2, 3, and 4 are 3.64 hours, 3.01 hours, 3.54 hours, and 3.72 hours, respectively. When we compare the results of the new settings to the results obtained in Section 5.3, when there are constraints on the number of certain types of shifts for each physician instead of physician preferences (i.e., an average wait time of 3.71 hours for the optimal schedule), we draw the following conclusions: First, not specifying the type of shifts reduces the average wait time by only 1.9% (from 3.71 hours to 3.64 hours). Conversely, imposing constraints how many shifts of each type to assign to physicians is not very restrictive. Second, providing some flexibility in the number of shifts to perform by a physician reduces the average wait time by 17.3% (from 3.64 hours to 3.01 hours). Consequently, physicians with a higher PPH rate will be assigned an additional shift, whereas physicians with a lower PPH rate will be assigned fewer shifts. We understand that this creates unfairness, but it allows us to study the impact of physician preferences. Third, the additional constraints imposed in Settings 3 and 4 increase the average wait times by 17.6% and 23.6%, respectively. This means that when most physicians have specific preferences regarding which type of shifts they want to be assigned to, the average wait time will increase by roughly 20% (depending on the type of preferences).

Note that the wait time reductions obtained by including the stochastic nature of patient arrivals and heterogeneity in physician productivity levels (or PPH rates) in the PRP (i.e., the results in Section 5.3) remain intact when physician preferences are included, because the deterministic and traditional schedules also require the inclusion of constraints regarding physician preferences to make that comparison. One can argue that the wait time reductions will be offset or less significant because there will be less flexibility in assigning physicians to shifts. Instead of including physician preferences, we propose a different approach that is observed at our study hospital. In particular, the traditional schedule allows more flexibility in practice since physicians can exchange shifts among themselves based on their preferences. This could undermine the superior performance of the optimal schedule over the traditional schedule if it is taken into consideration in our problem formulation. To alleviate this issue, we group physicians with similar productivity levels into clusters and allow shift exchanges among physicians within the same cluster. This new setting is explored in more detail in the next section.

6. Optimal Schedules with Physician Clustering

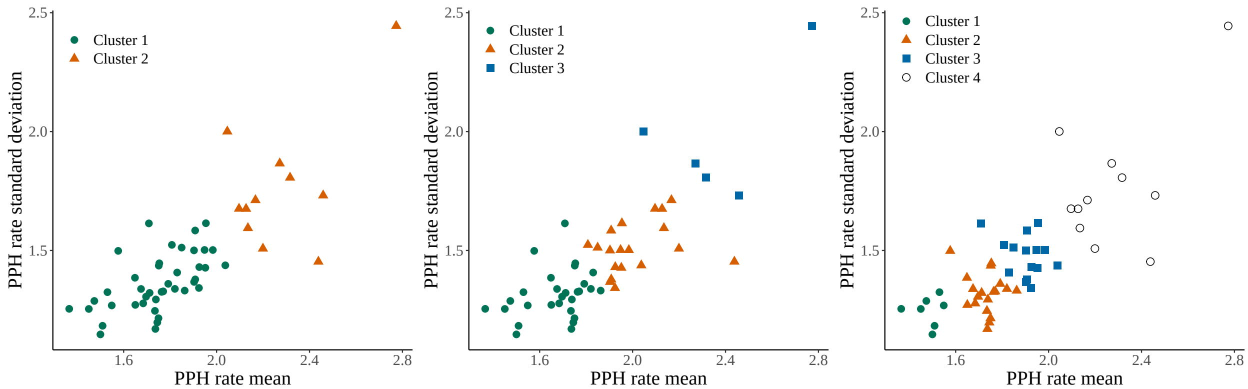
Our simulation results show that the optimal schedule considering the heterogeneity in physician productivity performs significantly better than the traditional schedule in terms of lower average patient wait times. However, as mentioned in Section 5.4, physicians may exchange shifts in practice due to their preferences, which may diminish the benefit of the optimal schedule.

6.1. Physician Clustering

It is plausible that some physicians share similarities in terms of their productivity levels, and hence, shift exchanges among them have little impact on system performance. We apply the k-means clustering method to group physicians into different clusters, using the mean and standard deviation of the PPH rates as the clustering factors. We apply the Elbow method and find that it is best to group physicians into four clusters. We also obtain the clustering results with two and three clusters. Figure 5 shows the clustering results for the 52 physicians. We solve the PRP_0 to get the optimal schedules when physicians are grouped into two, three, and four clusters, respectively, and evaluate their performances through simulation.

We assume that physicians within the same cluster have the same PPH rates. Hence, in the prediction model for PPH, instead of using the variable *Physician*, we define a categorical variable *Cluster* to indicate

Figure 5 **Cluster 52 physicians by their PPH rate mean and standard deviation.**



whether a physician belongs to a particular cluster. The Poisson regression model with clustering is specified as follows:

$$\log \mathbb{E}(PPH|Cluster, Hour, NightShift) = \beta_0 + \beta_C Cluster + \beta_H Hour + \beta_N NightShift. \tag{30}$$

We estimate the model in Eq. (30) for four clustering models that group the physicians into one cluster, two clusters, three clusters, and four clusters, respectively. The results of the four clustering models are summarized in Table 4. We observe that the cluster variables are significant in explaining the variation of physician productivity, and the model with a larger number of clusters fits the data better (bigger log likelihood and smaller AIC/BIC). However, the values of the AIC (and the log likelihood and BIC) for models with four clusters are close to those of the model with no clustering (see Model 6 in Table 2).

6.2. Simulation Results with Physician Clustering

In this section, we solve the PRP_0 problems with physicians grouped into one, two, three, and four clusters and obtain their corresponding optimal schedules. The PPH rates used in the optimization model are estimated by Eq. (30) with coefficients given in Table 4. Note that in the model where physicians are grouped into one cluster, the productivity of all physicians is considered to be homogeneous. Hence, the optimal schedule under this model corresponds to the traditional schedule (see Section 5.2). We compare their corresponding average wait times with that under the optimal schedule, which was fitted when each individual physician is considered as a cluster (see Section 5.2), through our simulation model in Section 5. In the simulation model, a physician's PPH rate under any schedule is estimated by Eq. (1) with coefficients given in Table 2.

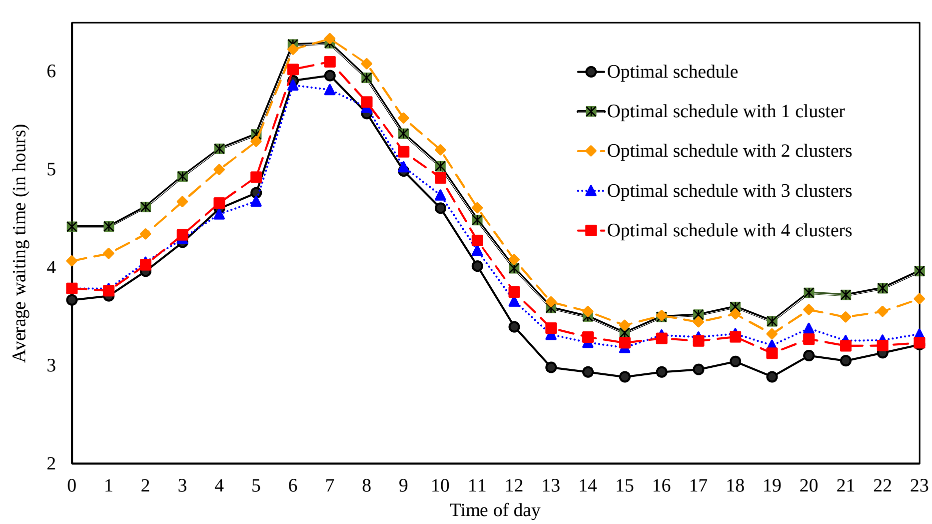
Our simulation results show that the average wait time under the optimal schedule is 3.71 hours, and that under the schedules with one, two, three, and four clusters are 4.28 hours, 4.21 hours, 3.88 hours and 3.92 hours, respectively. Hence, schedules with one or two clusters are not recommended due to their poor

Table 4 **Regression results from the Poisson model for various ways of clustering physicians.**

	1 Cluster	2 Clusters	3 Clusters	4 Clusters
(Intercept)	1.266***	1.476***	1.339***	1.197***
	(0.008)	(0.011)	(0.009)	(0.009)
ShiftHour (base=Hour1)				
Hour2	−0.499***	−0.499***	−0.499***	−0.499***
	(0.011)	(0.01)	(0.01)	(0.01)
Hour3	−0.758***	−0.758***	−0.758***	−0.758***
	(0.012)	(0.012)	(0.012)	(0.012)
Hour4	−0.877***	−0.877***	−0.877***	−0.877***
	(0.013)	(0.013)	(0.013)	(0.013)
Hour5	−0.937***	−0.937***	−0.937***	−0.937***
	(0.014)	(0.013)	(0.013)	(0.013)
Hour6	−1.092***	−1.092***	−1.092***	−1.092***
	(0.016)	(0.016)	(0.015)	(0.015)
Hour7	−1.944***	−1.944***	−1.944***	−1.944***
	(0.026)	(0.026)	(0.026)	(0.026)
NightShift (base = DayShift)	0.157***	0.148***	0.15***	0.155***
	(0.008)	(0.008)	(0.008)	(0.008)
Cluster (base=Cluster1)				
Cluster2		0.27***	0.181***	0.162***
		(0.009)	(0.008)	(0.013)
Cluster3			0.363***	0.25***
			(0.012)	(0.013)
Cluster4				0.436***
				(0.014)
LogLikelihood	-36,684	-36,365	-36,311	-36,241
AIC	73,385	72,748	72,641	72,504
BIC	73,450	72,821	72,722	72,593
Observations	24,549	24,549	24,549	24,549

Notes. Robust standard errors are shown in the parentheses. ***p<0.001; **p<0.01; *p<0.05

Figure 6 **Average ED wait times when physicians are grouped in clusters with similar productivity levels.**



performances. Nonetheless, the optimal schedule with three clusters performs significantly better than the schedule that does not consider heterogeneity in physician productivity (the schedule with one cluster) in that the average wait time decreases by close to 10% (from 4.28 hours to 3.88 hours). Moreover, the performance of the optimal schedule with three clusters does not deviate greatly from the optimal schedule with an optimality gap of 4.6% (from 3.71 hours to 3.88 hours), which is the price of allowing shift exchanges among physicians. The performance of the schedule with four clusters is statistically indifferent compared to that of three clusters.

If we take a closer look at the average hourly wait times (shown in Figure 6), the optimal schedule and the schedule with three clusters are very close during periods when the wait times are high (from 5 AM until 11 AM). Hence, the schedule with three clusters can achieve similar performance to the optimal schedule when the ED congestion level is high. The schedule with four clusters performs slightly worse than that with three clusters. This suggests that hospital managements do not have to consider the productivity levels of each individual physician when assigning physicians to shifts. The benefit of considering heterogeneity in physician productivity is achieved by three clusters of productivity levels, where physicians within each cluster are considered homogeneous. Physician clustering can help address physician preferences in shift scheduling by allowing shift exchanges among physicians of the same group.

7. Conclusions and Future Research

In this paper, we study the PRP in EDs. Our formulation captures the stochastic nature of ED operations and physician-specific shift-hour-dependent productivity, which is measured by the number of new patients seen by the physician during each hour of her shift (the PPH rate). This is in contrast to the literature where random components are not modeled in most rostering problems and the productivity of physicians is mostly treated as a constant. Our analysis using data from an ED in Calgary, Alberta, Canada, indicates that the individual physician, hour of the shift, and shift type are the dominant factors in explaining the variations in ED productivity. We incorporate these findings into our stochastic programming formulation of the PRP and propose a solution method to solve it. A simulation study shows that the new rostering solution can reduce average ED wait times as much as 13% over the current scheduling method implemented in our study ED. Furthermore, ED physicians are allowed to exchange shifts among themselves in practice even after the schedule is created. To mitigate the negative impact of exchanging shifts on the near-optimal assignment, we group physicians into different clusters based on their productivity so that physicians within the same cluster have similar PPH rates. Our results show that EDs can receive significant benefit in terms of reduced patient wait times when the number of clusters is fairly small.

Our study opens a number of directions for future research in scheduling processes. First, including the stochastic environment in scheduling problems makes these problems more realistic, and they can improve operational performance measures and service levels. This is an underexplored area in healthcare settings although they suffer seriously from uncertainties. Second, taking physician heterogeneity (or employee heterogeneity in more general settings) into consideration can significantly improve employee scheduling and thus system performance. Beyond the physician rostering problem explored in this paper, both aspects can be included in staffing problems as well. It would also be interesting to study dynamic schedule adjustment based on physician workload and ED occupancy levels, for instance through surge calls for additional staff members in case of high patient demand for emergency care. Another direction for future research is an empirical study of whether physician-specific characteristics (e.g., age, experience, training, and education) can be used to estimate their productivity. A better understanding of the factors that determine a physician's productivity in the individual level can help derive best practices so as to improve operational performance in the system level.

Acknowledgments

The authors gratefully acknowledge the assistance of Dr. Mozart Menezes (Professor at NEOMA Business School) and Dr. Eddy Lang (Department Head and Professor, University of Calgary) for helpful discussions. The authors also acknowledge the assistance of Dongmei Wang for the data support of this research. The work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. ECS 21500517, ECS 27501020, GRF 11504620, and CRF C7162-20GF).

References

Akbari, M., M. Zandieh, B. Dorri. 2013. Scheduling part-time and mixed-skilled workers to maximize employee satisfaction. *International Journal of Advanced Manufacturing Technology*, 64 1017-1027.

Angulo, G., S. Ahmed, S.S. Dey. 2016. Improving the integer L-shaped method. *INFORMS Journal on Computing*, 28 (3), 483-499.

Armony, M., S. Israelit, A. Mandelbaum, Y.N. Marmor, Y. Tseytlin, G.B. Yom-Tov. 2015. On patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems*, 5 146-194.

Bagheri, M., A.G. Devin, A. Izanloo. 2016. An application of stochastic programming method for nurse scheduling problem in real world hospital. *Computers & Industrial Engineering*, 96 192-200.

Batt, R.J., C. Terwiesch. 2017. Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science*, 63 3531-3551.

Beaulieu, H., J.A. Ferland, B. Gendron, P. Michelon. 2000. A mathematical programming approach for scheduling physicians in the emergency room. *Health Care Management Science*, 3 193-200.

Bernstein, S.L., D. Aronsky, R. Duseja, S. Epstein, D. Handel, U. Hwang, M. McCarthy, K. John McConnell, J.M. Pines, N. Rathlev, R. Schafermeyer, F. Zwemer, M. Schull, B.R. Asplin. 2009. The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*, 16 1-10.

Berry Jaeker, J.A., A.L. Tucker. 2017. Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science*, 63 1042-1062.

Bhat, Rahul, Jeffrey Dubin, Kevin Maloy. 2014. Impact of learners on emergency medicine attending physician productivity. *Western Journal of Emergency Medicine*, 15 (1), 41.

Birge, John R. 1997. State-of-the-art-survey—stochastic programming: Computation and applications. *INFORMS Journal on Computing*, 9 (2), 111-133.

Bruni, R., P. Detti. 2014. A flexible discrete optimization approach to the physician scheduling problem. *Operations Research for Health Care*, 3 191-199.

Brunner, J.O., J.F. Bard, R. Kolisch. 2009. Flexible shift scheduling of physicians. *Health Care Management Science*, 12 285-305.

Brunner, J.O., J.F. Bard, R. Kolisch. 2011. Midterm scheduling of physicians with flexible shifts using branch-and-price. *IIE Transactions*, 43 84-109.

Brunner, J.O., G.M. Edenharter. 2011. Long term staff scheduling of physicians with different experience levels in hospitals using column generation. *Health Care Management Science*, 14 189-202.

Brusco, M.J., T.R. Johns. 1998. Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences*, 29 499-515.

Brusco, M.J., T.R. Johns, J.H. Reed. 1998. Cross-utilization of a two-skilled workforce. *International Journal of Operations & Production Management*, 18 (6), 555-564.

Bucheli, B., B. Martina. 2004. Reduced length of stay in medical emrgency department patients: A prospective controlled study on emergency physician staffing. *European Journal of Emergency Medicine*, 11 29-34.

Bukata, W.R., H. Murray, P. Atkinson. 2018. Measuring and understanding causes of variability in emergency physician performance are essential to improve emergency department efficiency. *Canadian Journal of Emergency Medicine*, 20 (6), 821-825.

Camiat, Fanny, Marìa I Restrepo, Jean-Marc Chauny, Nadia Lahrichi, Louis-Martin Rousseau. 2021. Productivity-driven physician scheduling in emergency departments. *Health Systems*, 10 (2), 104-117.

Campbell, G.M. 2011. A two-stage stochastic program for scheduling and allocating cross-trained workers. *Journal of the Operational Research Society*, 62 1038-1047.

Carter, M.W., S.D. Lapierre. 2001. Scheduling emergency room physicians. *Health Care Management Science*, 4 347-360.

Chan, D.C. 2018. The efficiency of slacking off: Evidence from the emergency department. *Econometrica*, 86 997-1030.

Defraeye, M., I. Van Nieuwenhuyse. 2013. Controlling excessive waiting times in small service systems with time-varying demand: An extension of the ISA algorithm. *Decision Support Systems*, 54 1558-1567.

Defraeye, M., I. Van Nieuwenhuyse. 2016. Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, 58 4-25.

Delasay, M., A. Ingolfsson, B. Kolfal. 2016. Modeling load and overwork effets in queueing systems with adaptive service rates. *Operations Research*, 64 867-885.

Ding, Yichuan, Eric Park, Mahesh Nagarajan, Eric Grafstein. 2019. Patient prioritization in emergency department triage systems: An empirical study of the canadian triage and acuity scale (CTAS). *Manufacturing & Service Operations Management*, 21 (4), 723-741.

Duan, Yige, Yiwen Jin, Yichuan Ding, Mahesh Nagarajan, Garth Hunte. 2020. The cost of task switching: Evidence from the emergency department. *Available at SSRN 3756677*, .

EL-Rifai, O., T. Garaix, V. Augusto, X Xie. 2015. A stochastic optimization model for shift scheduling in emergency departments. *Health Care Management Science*, 18 289-302.

Erhard, M., J. Schoenfelder, A. Fügener, J.O. Brunner. 2018. State of the art in physician scheduling. *European Journal of Operational Research*, 265 1-18.

Foster, Michele, Zhankun Sun, Dongmei Wang, Grant Innes, Laurie-Ann Baker, Andrew McRae, Eddy Lang. 2015. Optimal shift duration for emergency physician efficiency, effectiveness and safety: A comparison of 6, 7, and 8-hour shifts. *Canadian Association of Emergency Physicians Annual Conference*, .

Gendreau, M., J. Ferland, B. Gendron, N. Hail, B. Jaumard, S. Lapierre, G. Pesant, P. Soriano. 2007. Physician scheduling in emergency rooms. E.K. Burke, H. Rudová, eds., *The Practice and Theory of Automated Timetabling VI (PATAT'06, Selected papers). Lecture Notes in Computer Science, vol 3867*. Springer Berlin Heidelberg, 53-66.

Gnanlet, A., W.G. Gilland. 2014. Impact of productivity on cross-training configurations and optimal staffing decisions in hospitals. *European Journal of Operational Research*, 238 254-269.

Goodale, J.C., G.M. Thompson. 2004. A comparison of heuristics for assigning individual employees to labor tour schedules. *Annals of Operations Research*, 128 47-63.

Gunawan, A., H.C. Lau. 2013. Master physician scheduling problem. *Journal of the Operational Research Society*, 64 410-425.

Hart, A., S. Drall. 2007. Productivity: Do 8-9 hour shifts make a difference? *Annals of Emergency Medicine*, 50 69-70.

Hoot, N.R., D. Aronsky. 2008. Systematic review of emergency department crowding:causes, effects, and solutions. *Annals of Emergency Medicine*, 52 126-136.

Joseph, J.W., S. Davis, E.H. Wilker, M.L. Wong, O. Litvak, S.J. Traub, L.A. Nathanson, L.D. Sanchez. 2018. Modelling attending physician productivity in the emergency department: a multicentre study. *Emergency Medicine Journal*, 35 (5), 317-322.

KC, Diwas., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, 55 1486-1498.

Kim, Song-Hee, Ward Whitt. 2014. Choosing arrival process models for service systems: Tests of a nonhomogeneous poisson process. *Naval Research Logistics (NRL)*, 61 (1), 66-90.

Kuntz, L., S. Sülz. 2013. Treatment speed and high load in the emergency department - does staff quality matter? *Health Care Management Science*, 16 366-376.

Laporte, G., F.V. Louveaux. 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letter*, 13 (3), 133-142.

Li, C., E.P. Robinson Jr., V.A. Mabert. 1991. An evaluation of tour scheduling heuristics with differences in employee productivity and cost. *Decision Sciences*, 22 700-718.

Li, Wenhao, Zhankun Sun, L Jeff Hong. 2021. Who is next: Patient prioritization under emergency department blocking. *Operations Research, Forthcoming*, .

Mak, W.K., D.P. Morton, R.K. Wood. 1999. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24 47-56.

Pines, J.M., J.A. Hilton, E.J. Weber, A.J. Alkemade, H. Al Shabanah, P.D. Anderson, M. Bernhard, A. Bertini, A. Gries, S. Ferrandiz, V.A. Kumar, V.P. Harjola, B. Hogan, B. Madsen, S. Mason, G. Ohlén, T. Rainer, N. Rathlev,

E. Revue, D. Richardson, M. Sattarian, M.J. Schull. 2011. International perspectives on emergency department crowding. *Academic Emergency Medicine*, 18 1358-1370.

Puente, J., A. Gómez, I. Fernández, P. Priore. 2009. Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms. *Computers & Industrial Engineering*, 56 1232-1242.

Rousseau, L.M., M. Gendreau, G. Pesant. 2002. A general approach to the physician rostering problem. *Annals of Operations Research*, 115 193-205.

Song, H., A. L. Tucker, K. L. Murrell. 2015. The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science*, 61 3032-3053.

Song, H., A. L. Tucker, K. L. Murrell, D. R. Vinson. 2018. Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science*, 64 2628-2649.

Stolletz, R., J.O. Brunner. 2012. Fair optimization of fortnightly physician schedules with flexible shifts. *European Journal of Operational Research*, 219 622-629.

Thompson, G.M., J.C. Goodale. 2006. Variable employee productivity in workforce scheduling. *European Journal of Operational Research*, 170 376-390.

Van den Bergh, J., J. Beliën, P. De Bruecker, E. Demeulemeester. 2013. Personnel scheduling: A literature review. *European Journal of Operational Research*, 226 367-385.

Van Slyke, R. M., R. Wets. 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17 (4), 638-663.

Yang, A., P. Pryor, M.S. Radeos, C. Stanescu. 2008. Emergency physician productivity: A comparison of 8-hour versus 12-hour shifts. *Annals of Emergency Medicine*, 52 110.

Appendix A. Supplements of Regression Results

The estimation results from Poisson regression for 8-hour shifts are provided in Table A.1. In the interest of space, the coefficients of individual physicians for 7-hour shifts and 8-hour shifts are shown in Tables A.2 and A.3, respectively. The estimation results when physicians are grouped in two, three, and four clusters are provided in Table 4.

Appendix B. Relationship PPH Rate and ED Productivity Measures

To investigate the relationship between physician productivity measured as PPH rate and other ED productivity performance measures, such as length of stay (LOS) and service (or processing) times of patients in the ED, we use the following log-linear regression model

$$\log LOS_{ijk} = \alpha + \boldsymbol{\beta}\, Cluster_j + \boldsymbol{\gamma} \boldsymbol{Y} + \boldsymbol{\delta} \boldsymbol{Z}, \tag{B.1}$$

where LOS_{ijk} is the LOS of patient i who is treated by physician j in shift type k and $Cluster_j$ indicates the cluster assigned to physician j. The vector $\boldsymbol{Y}$ consists of all control variables related to the patient (age, gender, CTAS, ambulance arrival) and the vector $\boldsymbol{Z}$ consists of all shift-characteristic variables of the physician (daytime, weekend). The age of a patient is categorized in four age groups: younger than 25 years, between 25 and 45 years, between 45 and 65 years, and at least 65 years of age. We denote them by *Age1*, *Age2*, *Age3* and *Age4*, respectively. The CTAS level of a patient in captured by categorical variables. Besides LOS, we use the same log-linear regression model for patient service times.

The results of the regression models with two clusters and three clusters are presented in Table B.1. Based on these results, we conclude that patients who are treated by a physician with a high PPH rate have a lower average LOS and lower average service time.

Table A.1 Estimation results for the effect of various factors on PPH rates for 8-hour shifts.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1.654***	1.656***	1.661***	1.675***	1.551***	1.438***
	(0.036)	(0.036)	(0.036)	(0.036)	(0.034)	(0.027)
Physician (base = MD003)						
MD004	−0.201**	−0.198**	−0.197**	−0.194**	−0.194**	−0.195**
	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)	(0.043)
MD005	−0.013	−0.015	−0.014	−0.016	−0.01	−0.014
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
MD006	0.025	0.029	0.03	0.029	0.035	0.026
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
⋮						
MD349	−0.465***	−0.453***	−0.452***	−0.449***	−0.455***	−0.461***
	(0.058)	(0.058)	(0.058)	(0.058)	(0.059)	(0.059)
ShiftHour (base=Hour1)						
Hour2	−0.356***	−0.356***	−0.354***	−0.358***	−0.367***	−0.374***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Hour3	−0.688***	−0.688***	−0.686***	−0.711***	−0.716***	−0.728***
	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
Hour4	−0.905***	−0.905***	−0.901***	−0.929***	−0.935***	−0.953***
	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.017)
Hour5	−1.039***	−1.039***	−1.034***	−1.043***	−1.039***	−1.061***
	(0.02)	(0.02)	(0.019)	(0.019)	(0.02)	(0.019)
Hour6	−1.009***	−1.009***	−1.003***	−1.002***	−0.994***	−1.018***
	(0.02)	(0.02)	(0.019)	(0.019)	(0.019)	(0.019)
Hour7	−1.342***	−1.342***	−1.336***	−1.335***	−1.317***	−1.342***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Hour8	−2.428***	−2.428***	−2.421***	−2.436***	−2.408***	−2.433***
	(0.046)	(0.046)	(0.045)	(0.045)	(0.045)	(0.045)
NightShift (base = DayShift)	0.034*	0.033*	0.038*	0.034*	0.066***	0.044**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)
EDCensus	−0.008***	−0.008***	−0.008***	−0.009***	−0.003***	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
WaitRoomCensus	0.016***	0.016***	0.017***	0.017***		
	(0.001)	(0.001)	(0.001)	(0.001)		
HandoverTaken	−0.03***	−0.03***	−0.03***			
	(0.005)	(0.005)	(0.005)			
Handover	−0.08***	−0.079***	−0.08***			
	(0.021)	(0.021)	(0.021)			
BoarderCensus	−0.002	−0.002				
	(0.001)	(0.001)				
Learner (base=No Learner)						
Resident	0.137***					
	(0.032)					
Student	0.025					
	(0.089)					
Observations	16,016	16,016	16,016	16,016	16,016	16,016
Log Likelihood	-21,755	-21,762	-21,762	-21,784	-21,845	-21,854
AIC	43,645	43,653	43,652	43,693	43,812	43,828
BIC	44,159	44,152	44,144	44,169	44,281	44,289

Notes. This table reports the estimation results from the Poisson regression. Robust standard errors are shown in the parentheses. ***p<0.001; **p<0.01; *p<0.05

Table A.2 Regression coefficients for all physicians in Model 7 for 7-hour shifts (Base = *MD003*).

Physician	Coefficient		Physician	Coefficient		Physician	Coefficient	
MD004	-0.203^{***}	(0.041)	*MD028*	-0.188^{***}	(0.042)	*MD065*	-0.351^{***}	(0.037)
MD005	0.025	(0.04)	*MD030*	−0.011	(0.042)	*MD066*	-0.187^{***}	(0.044)
MD006	$0.082^{\cdot}$	(0.036)	*MD032*	-0.183^{***}	(0.04)	*MD096*	-0.254^{***}	(0.037)
MD007	-0.101^{*}	(0.037)	*MD033*	-0.183^{**}	(0.047)	*MD099*	-0.324^{***}	(0.04)
MD008	$-0.099^{\cdot}$	(0.039)	*MD034*	−0.065	(0.039)	*MD100*	-0.135^{*}	(0.042)
MD009	-0.171^{***}	(0.042)	*MD037*	−0.068	(0.041)	*MD111*	−0.089	(0.043)
MD011	-0.171^{**}	(0.044)	*MD040*	-0.359^{***}	(0.044)	*MD128*	-0.249^{***}	(0.041)
MD012	-0.458^{***}	(0.051)	*MD041*	-0.209^{***}	(0.042)	*MD136*	-0.136^{*}	(0.041)
MD015	0.036	(0.038)	*MD048*	-0.272^{***}	(0.042)	*MD177*	−0.042	(0.04)
MD016	-0.153^{**}	(0.039)	*MD049*	-0.123^{*}	(0.039)	*MD183*	-0.15^{**}	(0.042)
MD018	-0.131^{**}	(0.038)	*MD050*	-0.351^{***}	(0.044)	*MD234*	-0.355^{***}	(0.044)
MD019	-0.148^{**}	(0.039)	*MD054*	-0.158^{**}	(0.04)	*MD238*	0.131^{*}	(0.039)
MD021	-0.239^{***}	(0.044)	*MD056*	-0.135^{**}	(0.04)	*MD247*	-0.147^{**}	(0.042)
MD022	−0.034	(0.041)	*MD058*	−0.01	(0.041)	*MD251*	-0.108^{*}	(0.039)
MD023	-0.293^{***}	(0.045)	*MD059*	-0.242^{***}	(0.04)	*MD256*	0.005	(0.043)
MD025	0.244^{***}	(0.037)	*MD062*	0.054	(0.04)	*MD272*	-0.253^{***}	(0.043)
MD027	-0.194^{***}	(0.039)	*MD063*	$0.096^{\cdot}$	(0.039)	*MD349*	-0.395^{***}	(0.047)

Notes. Standard error in parenthesis; $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.3 Regression coefficients for all physicians in Model 7 for 8-hour shifts (Base = *MD003*).

Physician	Coefficient		Physician	Coefficient		Physician	Coefficient	
MD004	-0.195^{**}	(0.043)	*MD033*	-0.233^{***}	(0.041)	*MD100*	-0.147^{*}	(0.046)
MD005	−0.014	(0.038)	*MD034*	−0.029	(0.037)	*MD111*	−0.086	(0.041)
MD006	0.026	(0.038)	*MD036*	$-0.107^{\cdot}$	(0.047)	*MD128*	-0.219^{***}	(0.041)
MD007	−0.067	(0.043)	*MD037*	$-0.101^{\cdot}$	(0.047)	*MD173*	-0.28^{***}	(0.048)
MD008	$-0.092^{\cdot}$	(0.041)	*MD040*	-0.386^{***}	(0.048)	*MD177*	−0.034	(0.04)
MD009	-0.173^{**}	(0.041)	*MD041*	-0.245^{***}	(0.047)	*MD181*	-0.299^{***}	(0.047)
MD010	$-0.097^{\cdot}$	(0.039)	*MD048*	-0.376^{***}	(0.045)	*MD183*	-0.153^{**}	(0.047)
MD013	-0.389^{***}	(0.038)	*MD049*	-0.138^{**}	(0.044)	*MD206*	-0.258^{***}	(0.041)
MD015	−0.027	(0.045)	*MD050*	-0.376^{***}	(0.043)	*MD215*	-0.245^{***}	(0.042)
MD016	-0.165^{**}	(0.042)	*MD056*	-0.148^{**}	(0.039)	*MD223*	-0.303^{***}	(0.046)
MD018	-0.146^{*}	(0.047)	*MD058*	−0.026	(0.049)	*MD238*	0.056	(0.04)
MD022	0.063	(0.047)	*MD059*	-0.187^{**}	(0.048)	*MD247*	-0.18^{**}	(0.045)
MD023	-0.281^{***}	(0.044)	*MD062*	−0.01	(0.044)	*MD251*	-0.168^{**}	(0.041)
MD025	0.219^{***}	(0.044)	*MD065*	-0.399^{***}	(0.042)	*MD252*	-0.532^{***}	(0.055)
MD027	-0.143^{*}	(0.042)	*MD066*	-0.254^{***}	(0.044)	*MD254*	-0.248^{***}	(0.049)
MD030	−0.077	(0.044)	*MD095*	$-0.113^{\cdot}$	(0.042)	*MD271*	-0.416^{***}	(0.054)
MD032	-0.328^{***}	(0.046)	*MD099*	-0.38^{***}	(0.045)	*MD349*	-0.461^{***}	(0.059)

Notes. Standard error in parenthesis; $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table B.1 Results of the log-linear regression models for the effect of physician PPH clusterings on length of stay and service (or processing) times.

	Length of Stay (LOS)		Service Time	
	2 Clusters	3 Clusters	2 Clusters	3 Clusters
Cluster (base = Low PPH)				
Medium PPH		0.000 (0.006)		0.007 (0.008)
High PPH	-0.078*** (0.006)	-0.084*** (0.008)	-0.063*** (0.008)	-0.073*** (0.010)
CTAS (base = CTAS1)				
CTAS2	0.312*** (0.014)	0.312*** (0.014)	0.208*** (0.018)	0.208*** (0.018)
CTAS3	0.229*** (0.014)	0.228*** (0.014)	-0.018 (0.019)	-0.018 (0.019)
CTAS4	-0.009 (0.015)	-0.010 (0.015)	-0.383*** (0.02)	-0.384*** (0.020)
CTAS5	-0.175*** (0.018)	-0.177*** (0.018)	-0.636*** (0.024)	-0.638*** (0.024)
Age (base = Age1)				
Age2	0.055*** (0.009)	0.055*** (0.009)	0.107*** (0.012)	0.107*** (0.012)
Age3	0.202*** (0.009)	0.201*** (0.009)	0.273*** (0.012)	0.273*** (0.012)
Age4	0.396*** (0.010)	0.396*** (0.010)	0.458*** (0.013)	0.458*** (0.013)
Gender (base = Female)	-0.006 (0.005)	-0.006 (0.005)	-0.024*** (0.007)	-0.024*** (0.007)
Arrival Mode (base = No Ambulance)	0.219*** (0.006)	0.218*** (0.006)	0.335*** (0.008)	0.334*** (0.008)
NightShift (base = DayShift)	0.122*** (0.006)	0.120*** (0.006)	-0.140*** (0.007)	-0.142*** (0.007)
WeekendShift (base =WeekdayShift)	-0.107*** (0.006)	-0.107*** (0.006)	-0.042*** (0.008)	-0.042*** (0.008)
(Intercept)	1.165*** (0.016)	1.158*** (0.016)	0.645*** (0.021)	0.637*** (0.022)
Observations	76,044	76,044	75,367	75,367
Adjusted R^2	0.1149	0.1143	0.1327	0.1326

Notes. Standard error in parenthesis; $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Appendix C. The Extensive Form of the Rostering Problem

For a given set of scenarios S_m, the extensive form of the original problem formulation of the rostering problem (PRP_m) is given as follows. Note that the constraints have similar explanations as those in the original PRP problem (see Section 4.1). Hence, we do not repeat them here.

$$
\begin{aligned}
(PRP_{0m}) \qquad \min_x \quad & \frac{1}{|S_m|} \sum_{s \in S_m} \sum_{j \in J} \sum_{t \in T} M_{jt}(s) \\
\text{s.t.} \quad & A_{jt}(s) + M_{j(t-1)}(s) - \sum_{i \in I} \sum_{k \in K} x_{ijk} P^{jk}_{i f_{kt}}(s) \le M_{jt}(s), && \forall j \in J, t \in T \setminus \{0\}, s \in S_m, \\
& A_{1,0}(s) - \sum_{i \in I} \sum_{k \in K} x_{i1k} P^{1,k}_{i f_{k0}}(s) \le M_{1,0}, && s \in S_m, \\
& A_{j,0}(s) + M_{(j-1),23}(s) - \sum_{i \in I} \sum_{k \in K} x_{ijk} P^{jk}_{i f_{k0}}(s) \le M_{j,0}(s), && j \in J \setminus \{1\}, s \in S_m, \\
& \sum_{i \in I} x_{ijk} = 1, && j \in J, k \in K, \\
& \sum_{k \in K} x_{ijk} \le 1, && i \in I, j \in J, \\
& \sum_{j \in \tilde{J}} \sum_{k \in \tilde{K}} x_{ijk} \ge L(\tilde{J}, \tilde{K}), && i \in I, \tilde{J} \in J', \tilde{K} \in K', \\
& \sum_{j \in \tilde{J}} \sum_{k \in \tilde{K}} x_{ijk} \le U(\tilde{J}, \tilde{K}), && i \in I, , \tilde{J} \in J', \tilde{K} \in K',
\end{aligned}
$$

$$\sum_{k\in K_N} x_{i(j-1)k} + \sum_{k\in K_D} x_{ijk} \le 1, \qquad i\in I, j\in J\setminus\{1\},$$

$$\sum_{k\in K_N} x_{i(j-3)k} + \sum_{j'=0}^{2}\sum_{k\notin K_N} x_{i(j-j')k} \le 1, \qquad i\in I, j\in J\setminus\{1,2,3\},$$

$$3\sum_{k\in K_N} x_{i(j-20)k} + \sum_{j'=0}^{17}\sum_{k\in K_N} x_{i(j-j')k} \le 3, \qquad i\in I, j\in J\setminus\{1,2,\dots,20\},$$

$$\sum_{k\in K} x_{i2k} \ge \sum_{k\in K} x_{i1k}, \qquad i\in I,$$

$$\sum_{k\in K} x_{ijk} \ge \sum_{k\in K} x_{i(j-1)k} - \sum_{k\in K} x_{i(j-2)k}, \qquad i\in I, j\in J\setminus\{1,2\},$$

$$\sum_{j'=0}^{4}\sum_{k\in K} x_{i(j-j')k} \le 4, \qquad i\in I, j\in J\setminus\{1,2,3,4\},$$

$$1-\sum_{k\in K} x_{ijk} \ge \sum_{k\in K} x_{i(j-2)k} - \sum_{k\in K} x_{i(j-1)k}, \qquad i\in I, j\in J\setminus\{1,2\},$$

$$\sum_{j'=0}^{4}\left(1-\sum_{k\in K} x_{i(j-j')k}\right) \le 4, \qquad i\in I, j\in J\setminus\{1,2,3,4\},$$

$$x_{ijk}\in\{0,1\}, \qquad i\in I, j\in J, k\in K,$$

$$M_{jt}(s)\ge 0, \qquad j\in J, t\in T, s\in S_m.$$