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
HKU Business School, The University of 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 assign physicians to shifts such that ED wait times are reduced without adding new physicians. In particular, we extend the physician rostering problem by including the heterogeneity between emergency physicians with regard to their productivity levels (measured as patient-per-hour rate) and by including 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 perform a data analysis to investigate the major drivers of physician productivity levels using patient visit data from our partner ED, and find that individual physicians, shift hour, and shift type (e.g., day or night) are the dominating factors of ED productivity. A simulation study calibrated using real data shows that the new scheduling from our formulation can reduce patient wait times by as much as 16%, compared to the current scheduling at our study ED. We also demonstrate how to incorporate physician preference in scheduling through physician clustering based on their productivity levels. Our simulation results show that EDs can receive almost full benefit even when the number of clusters is fairly small.

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

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 are becoming more crowded, partially due to an increasingly aging society, continuously rising patient volumes, and growing complexity of patient conditions and treatments. Consequently, patients seeking care in EDs may experience long wait times. This phenomenon is known as *ED overcrowding*, whose immediate results are prolonged pain and suffering, dissatisfied patients, and public safety being at risk (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. In order to improve patient flow in EDs, personnel schedules need to match with 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; (iv) *Rostering*: assigning employees to shifts. Our paper focuses on the fourth subproblem, i.e., the rostering problem of ED personnel. This research is motivated by a case study in an emergency department in Calgary, Canada.

There exist two major personnel rostering problems at EDs, namely, the nurse rostering problem (NRP) and the physician rostering problem (PRP). PRPs are fundamentally different from NRPs (Carter and Lapierre 2001), especially in Canada. In most Canadian EDs, nurses work under a collective agreement, while physicians are not employed by hospitals and do not have collective labor contracts. Consequently, physician schedules are predominantly driven by a fair satisfaction of 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 plays a crucial role in the delivery of high-quality, timely care. In this paper, we study a physician rostering problem where 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 one hour, which is commonly referred to as the *patient-per-hour rate* (referred as *the PPH rate* thereafter) 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), 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 at different 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); (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 might 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 (refer to as *hard constraints*) and flexible rules that can be violated (refer to 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 and service times, and physician heterogeneity in their non-stationary productivity levels. Including these practically relevant characteristics requires an understanding how productivity levels may change during shifts and the development of solution approaches to solve the new physician rostering problem. This has been rarely studied in the literature, despite that matching capacity with demand for emergency care using effective physician scheduling is very important to reduce ED overcrowding and prolonged patient wait times. The paper closest to our work is by Camiat et al. (2019), who also study the PRP with individual physician productivity levels. However, we include the stochastic nature of the problem to improve the solution to the PRP. Furthermore, we perform an empirical analysis to specify the productivity levels. Based on our study of the data from practice, we define a non-stationary PPH rate, whereas Camiat et al. (2019) use a constant PPH rate for each physician. Finally, we demonstrate that the schedule when physicians with similar productivity levels are clustered can achieve near optimal performance which makes our results more implementable.

Our contribution in this paper is threefold. First, through data analysis we find that it is sufficient to consider shift hour, shift type (daytime or night shift) and individual physician to predict the productivity of an emergency physician. In our second contribution, 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 where 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 at satisfying 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. In our third contribution, we derive managerial insights based on a case study at an emergency department in Calgary, Canada. In practice, ED physicians are allowed to exchange shifts among themselves even after the schedule is created. To mitigate the negative impact of exchanging shifts, 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. Section 7 concludes our paper.

2. Literature Review

Our research is relevant to multiple streams of literature. In Section 2.1, we discuss the different personnel rostering problems that have been studied in the literature with a particular focus on ED physician scheduling. The notion of productivity levels in personnel scheduling problems is presented in Section 2.2. Studies on physician productivity and service times as function of workload are discussed 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 the 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) with part-time workers and a branch-and-price algorithm to solve the problem, and by Stolletz and Brunner (2012) to include different measures of fairness. Brunner and Edenharter (2011) extend the problem by including experience levels of physicians and 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. (2019). All formulations of the PRP in these papers have in common 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 step of personnel scheduling problems (Defraeye and Van Nieuwenhuyse 2016). Common approaches to solve 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 account 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) as well as distribution constraints (each nurse should work a minimum number of shifts). Campbell (2011) and Gnanlet and Gilland (2014) investigate the planning of workforce over multiple departments with the consideration of cross-training.

2.2. Productivity and Employee Heterogeneity in Personnel Scheduling

One of the earliest works in personnel scheduling with productivity levels is by Li et al. (1991), who study a shift scheduling problem where a relative productivity 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; e.g., 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 hardly 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). The productivity level in Akbari et al. (2013) is exogenous and dependent on the shift they are assigned. Furthermore, when employees are scheduled to work consecutive shifts on the same day, their productivity decreases by a constant factor. Camiat et al. (2019) 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.

2.3. Productivity of Physicians in EDs

Shift work is very common in health care. There are a number of studies that analyze the impact of the shift length duration on the productivity of emergency physicians. Hart and Drall (2007) conclude that shifts with a duration of 8 to 9 hours result in a higher average PPH rates compared to 12-hour shifts while Foster et al. (2015) conclude 7-hour shifts are the best among 6, 7, and 8-hour shifts. In contrast, the pilot study of Yang et al. (2008) suggests no difference in average PPH rate. Extended shift duration also has been associated with adverse effects on patient outcomes, and increased accidents and health care errors (such as diagnostic errors and medication errors). Fatigue (both physically and mentally) 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) studies the end-of-shift phenomenon and concludes that physicians accept fewer patients in the last 2 to 4 hours prior to the end of 9-hour shifts to avoid more patients to be handed over to other physicians. The decline in PPH rate begins earlier when there is more shift overlap at the end of shift.

Finally, the impact of workload on patient service times and physician productivity has attracted interests from operations management community. KC and Terwiesch (2009) are among the first to conclude that a high load on healthcare systems 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 physicians with more professional experience have a shorter service time when the system load is high and a longer service time when the system load is low. Other empirical works on the impact of workload on service times include Armony et al. (2015), Berry Jaeker and Tucker (2017), and Delasay et al. (2016). There studies provide insights on the impact of different factors on physician productivity, however, these factors cannot be used directly to determine staffing levels or physician rostering.

3. Empirical Study of Physician Productivity

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

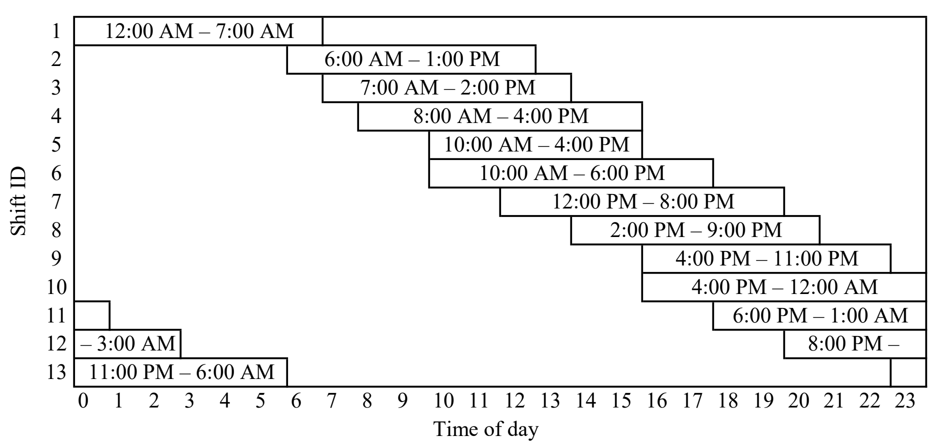
3.1. Patient Flow

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 will perform a quick assessment of the patient (e.g., measuring vital signs and recording the chief complaint), assign a triage score indicating the patient's acuity, and then create an electronic record for the patient. Our study hospital adopts a triage protocol called Canadian Triage and Acuity Scale (CTAS), which is 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 CTAS score, arrival time, and 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. 2020). The selection of a new patient by a physician to start the 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 are available. This process may repeat itself. The treatment phase ends when the physician makes a decision that either discharges the patient from the ED or admits 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 are 15 shifts (thus 15 physicians) in our study ED every day, two of which are dedicated to *fast track* where patients with minor conditions are treated. The remaining 13 physicians work at the *main ED area* during their scheduled shifts. The shift lengths at the main area vary between 6 to 8 hours. See Figure 1 for the start and end times for each of the 13 shifts at 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.

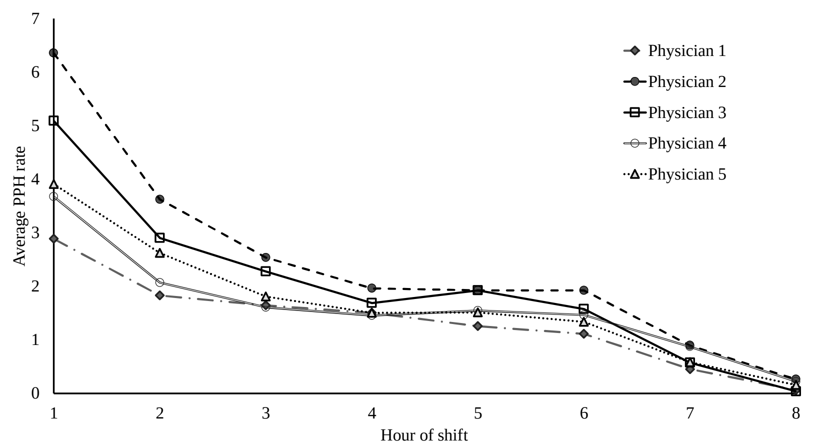
Figure 1 The start and end times of the thirteen shifts in the main treatment area of our study ED.



Physicians are assigned to these shifts based on a number of scheduling rules. The most common rules are 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. After a consecutive number of days that a physician is scheduled, the physician should have at least 2 and at most 4 days off. Night shifts should be scheduled on consecutive days, since no day shift is allowed for three days after a night shift and there need to be at least 21 days after a group of consecutive night shifts. *Weekend rules* are soft constraints that favor physicians to be scheduled for an entire weekend (defined as the first shift on Friday until the last shift started on Sunday) and to avoid that a physician is scheduled to work on two consecutive weekends. There is also a 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 (for instance, apply for time off during their scheduled shift days). Instead, physicians can exchange shifts with other physicians, which has been observed in our study ED.

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



Our data analysis shows that the productivity of a physician, measured by the number of new patients seen by a physician per hour (i.e., the PPH) changes significantly during her shift and the pattern depends heavily on physicians. See Figure 2 for the average hourly productivity of five physicians who have worked the most shifts in the main ED area over the two-year study period. In particular, we observe that the PPH rate (i) decreases with the hour of 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 supply and demand of emergency care.

3.3. Empirical Investigation of the PPH Rate

In this section, we empirically examine factors that drive physician productivity. In our data, there are more than 100 physicians that have worked at least one shift during our study period. However, a significant number of them are exchanging from other hospitals in the same health region, and do not reside in our study ED. To simplify the problem, we choose 52 physicians, and each of them has worked at least 100 shifts during the study period in the main ED area. There are in total 39,163 shifts, 23,667 of which are 7-hour shifts (60.4%) and 15,496 are 8-hour shifts (39.6%). The shift from 10am until 4pm is considered a flexible shift, and the data for this particular shift is less reliable. Therefore, this 6-hour shift is not included in our study. We present the results on 7-hour shifts in the main paper, and defer the results on 8-hour shifts, which are qualitatively similar to that of the 7-hour shifts, to Appendix A.

3.3.1. Choice of Variables Our objective is to investigate factors that drive a physician's productivity. Hence, the dependent variable is the number of new patients seen by a physician during a specific hour of her shift, i.e., PPH. We denote the productivity of physician i during the m-th hour of her shift j by PPH_{ijm}. Our data includes various timestamps of the patient flow and treatment process for each patient, the shift starting and ending time of each physician, etc., 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 have an impact on physician productivity. We define *Physician* as a unique identifier of an individual physician. Secondly, the hour of the shift also affects the productivity (as shown in Figure 2). Hence, we define *Hour* to indicate the hour of shift. Lastly, the workload and congestion level of the ED during the hour when PPH is measured might also affect a physician productivity; see, e.g., KC and Terwiesch (2009), Kuntz and Sülz (2013), Armony et al. (2015), Berry Jaeker and Tucker (2017) for extensive evidence that the load of healthcare systems and the workload of emergency physician have a significant impact on their service times and thus their PPH rates. However, these factors can not be estimated accurately at the time of assigning physicians to shifts. Hence, we include a binary variable *Night*, which is true if the shift starting time is from 5 PM to 12 AM. We believe this variable can serve as a proxy of the level of system load and physician workload. Note that all three independent variables are categorical.

3.3.2. Model Development and Results The dependent variable is a count variable. Hence, a generalized linear model with Poisson family (or negative binomial family) would potentially be a better choice than a linear regression model. Let $\mathbb{E}(\mathit{PPH}|\mathit{Physician, Hour, Night})$ denote the mean of the predicted Poisson distribution that fits physicians' productivity level. Then, the model is specified as follows:

$$\log \mathbb{E}(\mathit{PPH}|\mathit{Physician, Hour, Night}) = \beta_0 + \beta_P \mathit{Physician} + \beta_H \mathit{Hour} + \beta_N \mathit{Night}. \tag{1}$$

We consider an alternative model with both the main effects and the interactions of the independent variables in (1), specified as follows:

$$\begin{aligned}\log \mathbb{E}(\mathit{PPH}|\mathit{Physician}, \mathit{Hour}, \mathit{Night}) = {} & \beta_0 + \beta_P \mathit{Physician} + \beta_H \mathit{Hour} + \beta_N \mathit{Night} \\ & + \beta_{PN} \mathit{Physician} \times \mathit{Night} + \beta_{PH} \mathit{Physician} \times \mathit{Hour} + \beta_{NH} \mathit{Night} \times \mathit{Hour}.\end{aligned} \tag{2}$$

We fit the two models in (1) and (2) using Poisson regression as well as negative binomial regression, which is a generalization of Poisson regression because it loosens the highly restrictive assumption that the variance is equal to the mean. We also estimate two linear regression models with the same specifications

Table 1 Regression results for Poisson model, negative binomial model, and linear model with model specifications in (1) (no interaction) and (2) (with interaction).

	Poisson model		Negative binomial model		Linear model	
	Interaction	No interaction	Interaction	No interaction	Interaction	No interaction
Log likelihood	-29,020	-29,627	-33,922	-34,247	-32,647	-33,423
AIC	58,884	59,372	68,688	68,612	66,140	66,966
BIC	62,291	59,848	72,094	69,089	69,554	67,450

Notes. Bigger log likelihood and smaller AIC (or BIC) both mean a better model fit.

in (1) and (2) with *PPH* as the dependent variable. Hence, we estimate six models in total. We examine the log likelihood (also AIC and BIC) from the model estimation, with results shown in Table 1.

We observe that the Poisson models perform better than negative binomial and linear models in terms of achieving significantly bigger likelihood (smaller AIC and BIC). Moreover, the comparison between estimation results of model specifications in (1) and (2) shows that removing the interaction terms from the model specification only decreases the likelihood by 2.1% and increases the AIC by 0.8%. On the other hand, it decreases the BIC by 3.9% due to the smaller number of independent variables. Hence, the Poisson model with specification in (1) is preferable (referred to as *the base model* thereafter). A Chi-square goodness-of-fit test strongly supports the assumption of Poisson model (with p-value > 0.999).

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, or 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 and might have an impact on physician productivity as well. We let the independent variable *Handover* denote the number of handover patients a physician takes over from the previous shift. We next estimate models with the variables *Learner* and/or *Handover* added to the base model, and check whether they improve the model fit.

The estimation results of all models (including the base model) from Poisson regression are shown in Table 2. In the interest of space, we include partial coefficients for the physician variable, and the complete results of the base model are provided in Table A.2 in Appendix A. We observe that across all four models, a physician's productivity decreases with the hour of shift (*Hour*) and is higher during *Night* shifts. Fatigue (both physically and mentally) may explain why PPH decreases with the hour of shift. Another possible explanation is that physicians need to spend time and efforts on existing patients under her care. The number of existing patients accumulates with the hour of shift until approaching the end of a shift due to the end of

Table 2 Regression results for the effect of individual physician and shift information on PPH rates from 7-hour shifts.

	Base model	Base model with *Learner*	Base model with *Handover*	Base model with *Learner&Handover*
(Intercept)	$0.508^{***}(0.031)$	$0.496^{***}(0.031)$	0.553^{***} (0.033)	0.543^{***} (0.033)
Physician2	$0.482^{***}(0.038)$	$0.470^{***}(0.038)$	0.476^{***} (0.038)	0.465^{***} (0.039)
Physician3	$0.352^{***}(0.043)$	$0.341^{***}(0.043)$	0.358^{***} (0.043)	0.349^{***} (0.043)
Physician4	$0.195^{***}(0.045)$	$0.186^{***}(0.045)$	0.195^{***} (0.045)	0.186^{***} (0.045)
⋮				
Physician52	$0.231^{***}(0.052)$	$0.216^{***}(0.052)$	0.235^{***} (0.052)	0.221^{***} (0.052)
Hour2	$-0.245^{***}(0.017)$	$-0.245^{***}(0.017)$	-0.245^{***} (0.017)	-0.245^{***} (0.017)
Hour3	$-0.367^{***}(0.018)$	$-0.367^{***}(0.018)$	-0.367^{***} (0.018)	-0.367^{***} (0.018)
Hour4	$-0.437^{***}(0.018)$	$-0.437^{***}(0.018)$	-0.437^{***} (0.018)	-0.437^{***} (0.018)
Hour5	$-0.564^{***}(0.019)$	$-0.564^{***}(0.019)$	-0.564^{***} (0.019)	-0.564^{***} (0.019)
Hour6	$-1.420^{***}(0.025)$	$-1.420^{***}(0.025)$	-1.420^{***} (0.025)	-1.420^{***} (0.025)
Hour7	$-3.599^{***}(0.069)$	$-3.599^{***}(0.069)$	-3.599^{***} (0.069)	-3.599^{***} (0.069)
Night	$0.236^{***}(0.012)$	$0.242^{***}(0.013)$	0.238^{***} (0.012)	0.243^{***} (0.013)
Learner (Student)		-0.009 (0.020)		-0.015 (0.020)
Learner (Resident)		$0.066^{***}(0.014)$		0.064^{***} (0.014)
Handover			-0.009^{***} (0.002)	-0.009^{***} (0.002)
Log likelihood	-29,627	-29,615	-29,619	-29,606
AIC	59372	59,351	59,358	59,336
BIC	59,848	59,843	59,842	59,837

Notes. Standard error in parenthesis; $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
The complete results of the base model are provided in Table A.2 in Appendix A.

shift effect (Chan 2018). As for the effect of an individual physician, some levels of the variable *Physician* are significant (e.g., Physician 2,3,4 in Table 2), and some are not (see more details in Table A.2 in Appendix A), 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 physician based on their productivity, which we discuss in detail in Section 6.

It is interesting to note that the effect of a student learner on the productivity is not significant, while having a resident learner could increase the productivity. This is consistent with findings in the literature (Bhat et al. 2014). The intuition is that residents specialized in emergency medicine can perform assessments of new patients independently under the supervision of ED physicians, which increases the overall productivity. On the other hand, the handover patients taken from a previous shift decrease a physician's productivity as she needs to spend efforts on the existing patients.

Introducing new variables *Handover* and/or *Learner* into the base model can improve the model fit (bigger log likelihood and smaller AIC/BIC) but only marginally. Moreover, it has little impact on the coefficients of other variables. Hence, we decide not to include the two variables in our model due to their negligible impacts as well as the difficulty of including them in our optimization model. The coefficients of the base model in Table 1 are used to predict the PPH of a physician during a particular hour of her shift in the stochastic optimization model.

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 the 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 uncertainty associated with time-varying ED arrivals and hourly physician productivity is included.

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 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 mismatch between the demand for emergency care and the supply provided in terms of physician productivity for a given assignment. The formulation and a detailed discussion of the constraints in PRP_0 are provided in Appendix B.

The function $Q(A,P,x)$ is 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

$$Q(A,P,x) = \min_{M} \quad \sum_{j\in J}\sum_{t\in T} M_{jt} \tag{3}$$

$$\text{s.t.} \quad A_{jt} + M_{j(t-1)} - \sum_{i\in I}\sum_{k\in K} x_{ijk} P^{jk}_{i f_{kt}} \leq M_{jt}, \qquad \forall j\in J, t\in T\setminus\{0\}, \tag{4}$$

$$A_{10} - \sum_{i\in I}\sum_{k\in K} x_{i1k} P^{1k}_{i f_{k0}} \leq M_{10}, \tag{5}$$

$$A_{j0} + M_{(j-1),23} - \sum_{i\in I}\sum_{k\in K} x_{ijk} P^{jk}_{i f_{k0}} \leq M_{j0}, \qquad j\in J\setminus\{1\}, \tag{6}$$

$$M_{jt} \geq 0, \qquad j\in J, t\in T. \tag{7}$$

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 (3) sets these variables to the lowest possible values. Constraints (4) to (6) specify the hourly mismatch between patient demand and the total physician productivity. The left-hand side of the constraints (4) to (6) 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:

$$\mathbb{E}_{\Lambda,\Psi}\left[Q(\Lambda,\Psi,x)\right] \approx \frac{1}{|S|}\sum_{s\in S} Q(A(s),P(s),x). \tag{8}$$

With the scenarios fixed, we can obtain the deterministic equivalent of the SAA problem, which is included in Appendix B. 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 B) decomposes into $|S|$ independent subproblems. Each subproblem is the same as the LP formulation in (3)-(7) (i.e., subproblem s corresponds to solving $Q(A(s), P(s), x)$), whereas the master problem for iteration v is as follows:

$$
\begin{aligned}
(MP) \quad & \min_{x} \quad \theta \\
\text{s.t.} \quad & Ax = b, && (9)\\
& E^{\ell} x + \theta \geq e^{\ell}, \ \ell = 1, \ldots, v-1, && (10)\\
& x \in \{0,1\}^{|I| \times |J| \times |K|}, \ \theta \in \mathbb{R}, && (11)
\end{aligned}
$$

where (x^v, θ^v) is the optimal solution in iteration v. The constraints (9) correspond to the constraints (B.2)-(B.13) from the PRP_0 and the constraints (10) are the optimality cuts included from the previous iterations. Note that if optimality constraint (10) 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 (4), (5) and (6), 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:

$$
E^{\ell} x = \frac{1}{|S|} \sum_{s \in S} \left(\sum_{j \in J} \sum_{t \in T \setminus \{0\}} \pi^{\ell}_{jt}(s) \sum_{k \in K} \sum_{i \in I} x_{ijk} P^{ik}_{if_{kt}}(s) + \rho^{\ell}(s) \sum_{k \in K} \sum_{i \in I} x_{i1k} P^{1k}_{if_{k0}} + \sum_{j \in J \setminus \{1\}} \gamma^{\ell}_{j}(s) \sum_{k \in K} \sum_{i \in I} x_{ijk} P^{jk}_{i f_{kt}}(s) \right), \tag{12}
$$

and

$$
e^{\ell} = \frac{1}{|S|} \sum_{s \in S} \left(\sum_{j \in J} \sum_{t \in T \setminus \{0\}} \pi^{\ell}_{jt}(s)\, A_{jt}(s) + \rho^{\ell}(s)\, A_{1,0}(s) + \sum_{j \in J \setminus \{1\}} \gamma^{\ell}_{j}(s)\, A_{j,0}(s) \right). \tag{13}
$$

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 of 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 of the optimal objective value, which can evaluate the quality of our solution. 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 B). 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 and optimal solution for the set of scenarios S_m. By Mak et al. (1999), we have

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

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

$$\overline{L}(n_l) = \frac{1}{n_l}\sum_{j=1}^{n_l} z_j^* \leq z^*, \tag{14}$$

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

$$\overline{U}(n_u) = \frac{1}{n_u}\sum_{j=1}^{n_u} z(\hat{x}, S_j) \geq z^*. \tag{15}$$

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

$$\left[0,\ [\overline{U}(n_u) - \overline{L}(n_l)]^+ + \hat{\epsilon}_u + \hat{\epsilon}_l \right], \tag{16}$$

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 previous sections, we proposed an stochastic optimization model to solve for the optimal physician schedule. To evaluate the impact of a physician schedule taking stochastic arrivals and heterogeneous physician productivity into consideration, we compare the performance of such a schedule with other alternatives through a simulation study.

5.1. Study Setting

We use a simulation software Simio (version 12) to simulate the ED process in our study hospital. The system parameters are estimated from our two-year data. In our setting, a physician schedule is the assignment of the 52 physicians into the 13 shifts (see Figure 1) per day over a four week planning horizon. In practice, the schedule is planned every 6 months. However, the schedules largely repeat themselves. Hence, we focus on four 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 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 waiting 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 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 hth hour of shift j. Hence, in our simulation model, the service time of a patient being treated by physician i in the hth hour of shift j follows an exponential distribution with rate μ_{ijh}, where μ_{ijh} is predicted from the Poisson model specified in (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 by allowing physician preferences through physician clustering.

Optimal Schedule We solve the PRP_0 problem using the solution method 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 efforts 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.

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

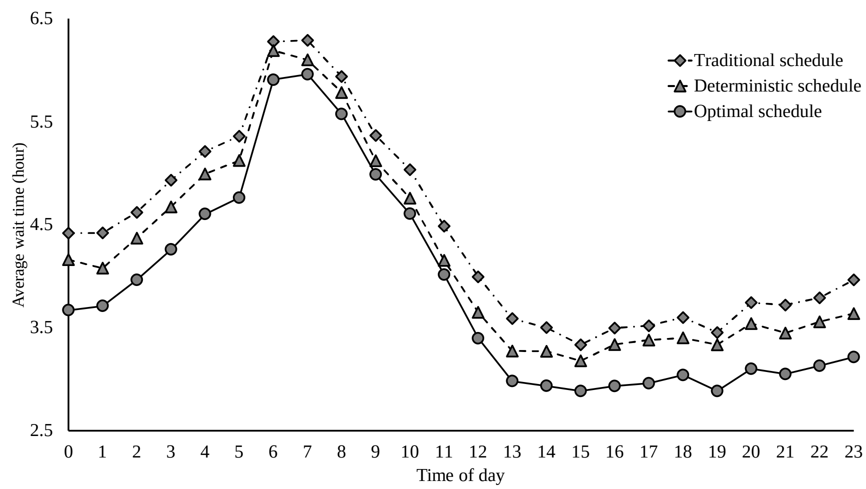
The deviation in objective values of the solution with 60 scenarios (i.e., $\hat{x}^{60}$) from that of the optimal solution has a tight confidence interval. More specifically, the length of the confidence interval is 3.7% of the lower bound and 3.6% of the upper bound. In other words, the L-shaped method with 60 scenarios provides an reasonably accurate solution for the PRP_0 within two 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 Appendix B) 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 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.

5.3. Simulation Results

Figure 3 **Comparison of the average ED wait times under three different schedules using simulation.**



We simulate the ED operations for 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 deterministic schedule, and the optimal schedule are 4.28 hours, 4.05 hours, and 3.71 hours, respectively. Hence, when the heterogeneity in physician productivity is considered when assigning physicians to shifts, the average wait times can be reduced by 5.6% (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 9.2% (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 close to 16%. Figure 3 shows the average patient wait times over the 200 simulation replications over the time of day under each of the three schedules. We observe that the optimal schedule performs significantly better than the other two schedules for each of the 24 hours. However, we should mention that the traditional schedule allows more flexibility in that physicians can exchange shifts among themselves based on their preferences, which could potentially undermine the superior performance of the optimal schedule over the traditional schedule if it was taken into consideration in our formulation. Next, we alleviate this issue by physician clustering.

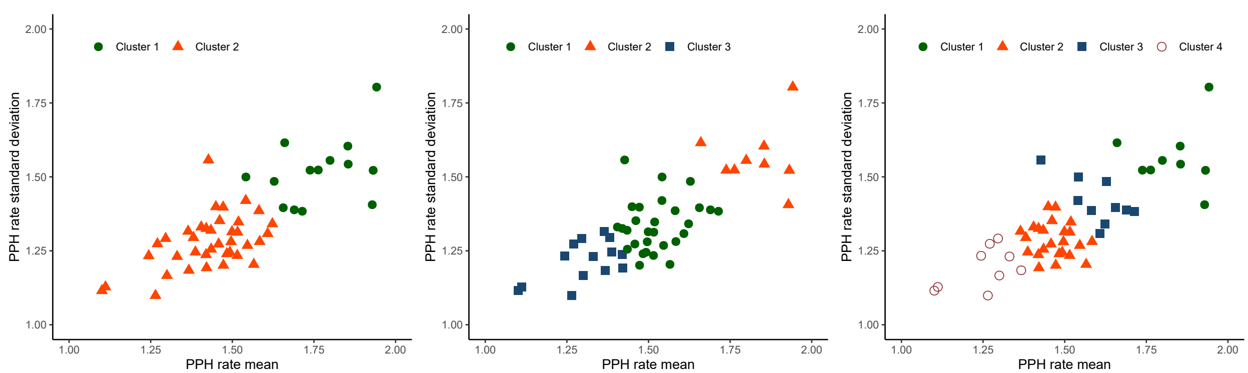
6. Optimal Schedules with Physician Clustering

Our simulation results show that the optimal schedule considering the heterogeneity in physician productivity performs significantly better the traditional schedule in terms of lower average patient wait times. However, as we mentioned in Section 3.2 that in practice, physicians may exchange shifts due to their preferences, which may diminish the benefit of the optimal schedule. In this section, we divide physicians into different clusters, and allow shift exchange for physicians within the same cluster.

6.1. Physician Clustering

It is plausible that some physicians share some similarity in terms of their productivity, and hence shift exchange among them has little impact on the system performance. We apply the k-means clustering method to divide 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 divide physicians into 4 clusters. We also obtain the clustering results with 2 and 3 clusters. Figure 4 shows the clustering results for the 52 physicians. We solve the PRP_0 to get the optimal schedules when physicians are grouped into 2, 3, and 4 clusters, respectively, and evaluate their performances through simulation.

Figure 4 Cluster 52 physicians by their PPH rate mean and standard deviation.



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

$$\log PPH = \beta_0 + \beta_C Cluster + \beta_H Hour + \beta_N Night. \tag{17}$$

We estimate the model in (17) for four clustering models that divide the physicians into 1 cluster, 2 clusters, 3 clusters, and 4 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 larger number of clusters fits the data better (bigger log likelihood and smaller AIC/BIC). On the other hand, the values of the AIC (and the log likelihood and BIC) for models with 3 clusters and 4 clusters are both very close to that of the model with no clustering.

Table 4 Regression coefficients for cluster models.

Variables	1 Cluster	2 Clusters	3 Clusters	4 Clusters
(Intercept)	0.709^{***} (0.014)	0.879^{***} (0.016)	0.723^{***} (0.015)	0.906^{***} (0.018)
Hour2	-0.245^{***} (0.017)	-0.245^{***} (0.017)	-0.245^{***} (0.017)	-0.245^{***} (0.017)
Hour3	-0.367^{***} (0.018)	-0.367^{***} (0.018)	-0.367^{***} (0.018)	-0.367^{***} (0.018)
Hour4	-0.437^{***} (0.018)	-0.437^{***} (0.018)	-0.437^{***} (0.018)	-0.437^{***} (0.018)
Hour5	-0.564^{***} (0.019)	-0.564^{***} (0.019)	-0.564^{***} (0.019)	-0.564^{***} (0.019)
Hour6	-1.420^{***} (0.025)	-1.420^{***} (0.025)	-1.420^{***} (0.025)	-1.420^{***} (0.025)
Hour7	-3.599^{***} (0.069)	-3.599^{***} (0.069)	-3.599^{***} (0.069)	-3.599^{***} (0.069)
Night	0.238^{***} (0.012)	0.233^{***} (0.012)	0.235^{***} (0.012)	0.231^{***} (0.012)
Cluster2		-0.234^{***} (0.012)	0.181^{***} (0.014)	-0.228^{***} (0.015)
Cluster3			-0.169^{***} (0.014)	-0.135^{***} (0.018)
Cluster4				-0.384^{***} (0.019)
Log likelihood	-29,965	-29,784	-29,742	-29,733
AIC	59,946	59,587	59,503	59,489
BIC	60,011	59,659	59,584	59,577

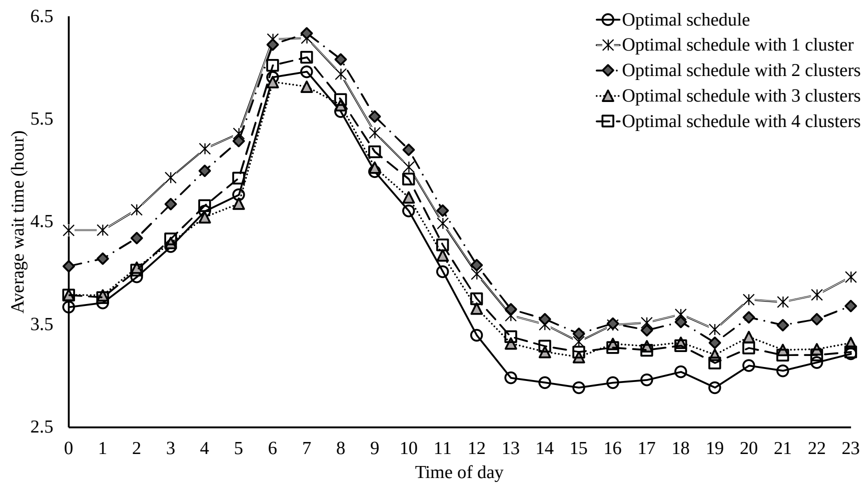
Notes. Standard error in parenthesis; $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

6.2. Simulation Results with Physician Clustering

In this section, we solve the PRP_0 problems with physicians grouped into 1, 2, 3, and 4 clusters and obtain their corresponding optimal schedules. The PPH rates used in the optimization model are estimated by (17) with coefficients given in Table 4. Note that in the model where physicians are grouped into one cluster, the productivity of all physicians are considered 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 (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 1, 2, 3, 4 clusters are 4.28 hours, 4.21 hours, 3.88 hours and 3.92 hours, respectively. Hence, the schedules with 1 or 2 clusters are not recommended due to their poor performances. On the other hand, the optimal schedule with 3 clusters performs significantly better than the schedule that does not consider the heterogeneity in physician productivity (the schedule with 1 cluster) in that the average

Figure 5 Average ED wait times when physicians are grouped in clusters with similar productivity levels.



wait time decreases by close to 10% (from 4.28 hours to 3.88 hours). Moreover, the performance of the optimal schedule with 3 clusters does not deviate too much 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 exchange among physician. The performance of the schedule with 4 clusters is statistically indifferent compared with that of 3 clusters.

If we take a closer look at the average hourly wait times (shown in Figure 5), they are very close between the optimal schedule and the schedule with 3 clusters during periods when the wait times are high (from 5 AM to 11 AM). Hence, the schedule with 3 clusters can achieve similar performance to the optimal schedule when the ED congestion level is high. The schedule with 4 clusters performs slightly worse than that with 3 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 be achieved by 3 clusters of productivity levels, where physicians within each cluster are considered homogeneous. Physician clustering can help address physician preference in shift scheduling by allowing shift exchange among physicians of the same group.

7. Conclusions and Future Research

In this paper, we study the physician rostering problem (PRP) in emergency departments. Our formulation captures the stochastic nature of ED operations as well as the physician-specific shift-hour-dependent productivity which is measured by the number of new patients seen by the physician during each hour (the PPH rate). This is in contrast to existing 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 emergency department in Calgary, Canada, supports that individual physicians, hour of the shift, and shift type are the dominating factors of ED productivity. We incorporate these findings into our two-stage stochastic programming formulation of the PRP_0 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 16% over the current existing scheduling. 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 (3 clusters).

Our study opens a number of directions for future research on scheduling processes (defined in Section 1). 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 unexplored area in healthcare settings even though they suffer most from uncertainties. Second, taking physician heterogeneity into consideration (or employees in more general scheduling settings) can significantly improve employee schedules and thus system performance. Besides the physician rostering problem explored in this paper, both aspects can be included in staffing problems as well. One can also study correcting schedules dynamically based on the workload and occupancy levels, for instance by including surge calls for additional staff members in case of high patient demand for emergency care. Another direction for future research is an empirical study whether physician-specific characteristics (such as age, years of experience, training and education, etc.) can be used to estimate their productivity. When there is a better understanding which factors determine the productivity of a physician, best practices might be derived from such a study. Furthermore, nudging mechanisms can be studied to influence the physician productivity whenever it is needed most.

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 a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. 9048104).

References

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

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

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

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

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

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

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

R. Bhat, J. Dubin, and K. Maloy. Impact of learners on emergency medicine attending physician productivity. *Western Journal of Emergency Medicine*, 15(1):41, 2014.

J. R. Birge. State-of-the-art-survey—stochastic programming: Computation and applications. *INFORMS journal on computing*, 9(2):111–133, 1997.

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

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

J. Brunner, J. Bard, and R. Kolisch. Flexible shift scheduling of physicians. *Health Care Management Science*, 12: 285–305, 2009.

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

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

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

B. Bucheli and B. Martina. 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, 2004.

W. Bukata, H. Murray, and P. Atkinson. 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, 2018.

F. Camiat, M. Restrepo, J. Chauny, N. Lahrichi, and L. Rousseau. Productivity-driven physician scheduling in emergency departments. *Health Systems*, 2019.

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

M. Carter and S. Lapierre. Scheduling emergency room physicians. *Health Care Management Science*, 4:347–360, 2001.

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

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

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

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

Y. Ding, E. Park, M. Nagarajan, and E. Grafstein. 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, 2019.

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

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

M. Foster, Z. Sun, D. Wang, G. Innes, L.-A. Baker, A. McRae, and E. Lang. 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*, 2015.

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

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

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

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

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

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

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

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

S.-H. Kim and W. Whitt. Choosing arrival process models for service systems: Tests of a nonhomogeneous poisson process. *Naval Research Logistics (NRL)*, 61(1):66–90, 2014.

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

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

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

W. Li, Z. Sun, and L. J. Hong. Who is next: Patient prioritization under emergency department blocking. *Available at SSRN 3590979*, 2020.

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

J. Pines, J. Hilton, E. Weber, A. Alkemade, H. Al Shabanah, P. Anderson, M. Bernhard, A. Bertini, A. Gries, S. Ferrandiz, V. Kumar, V. Harjola, B. Hogan, B. Madsen, S. Mason, G. Ohlén, T. Rainer, N. Rathlev, E. Revue, D. Richardson, M. Sattarian, and M. Schull. International perspectives on emergency department crowding. *Academic Emergency Medicine*, 18:1358–1370, 2011.

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

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

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

G. Thompson and J. Goodale. Variable employee productivity in workforce scheduling. *European Journal of Operational Research*, 170:376–390, 2006.

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

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

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

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 to 2, 3 and 4 clusters are provided in Table 4.

Table A.1 Regression results for the effect of individual physician and shift information on PPH rates from 8-hour shifts. Coefficients for all physicians of the base model are provided in Table A.3

	Base model		Base model with *Learner*		Base model with *Handover*		Base model with *Learner&Handover*	
(Intercept)	0.668^{***}	(0.049)	0.644^{***}	(0.050)	0.707^{***}	(0.050)	0.683^{***}	(0.050)
Physician2	0.651^{***}	(0.063)	0.627^{***}	(0.063)	0.651^{***}	(0.063)	0.627^{***}	(0.063)
⋮								
Physician52	0.253^{***}	(0.077)	0.228^{**}	(0.077)	0.263^{***}	(0.077)	0.239^{**}	(0.077)
Hour2	-0.355^{***}	(0.022)	-0.355^{***}	(0.022)	-0.355^{***}	(0.022)	-0.355^{***}	(0.022)
Hour3	-0.553^{***}	(0.024)	-0.553^{***}	(0.024)	-0.553^{***}	(0.024)	-0.553^{***}	(0.024)
Hour4	-0.685^{***}	(0.025)	-0.685^{***}	(0.025)	-0.685^{***}	(0.025)	-0.685^{***}	(0.025)
Hour5	-0.647^{***}	(0.025)	-0.647^{***}	(0.025)	-0.647^{***}	(0.025)	-0.647^{***}	(0.025)
Hour6	-0.961^{***}	(0.027)	-0.961^{***}	(0.027)	-0.961^{***}	(0.027)	-0.961^{***}	(0.027)
Hour7	-2.098^{***}	(0.043)	-2.098^{***}	(0.043)	-2.098^{***}	(0.043)	-2.098^{***}	(0.043)
Hour8	-4.281^{***}	(0.123)	-4.281^{***}	(0.123)	-4.281^{***}	(0.123)	-4.281^{***}	(0.123)
Night	0.045^{**}	(0.017)	0.047^{**}	(0.017)	0.064^{***}	(0.018)	0.067^{***}	(0.018)
Student Learner			-0.018	(0.022)			-0.013	(0.022)
Resident Learner			0.103^{***}	(0.017)			0.106^{***}	(0.017)
Handover					-0.015^{***}	(0.003)	-0.015^{***}	(0.003)
Log likelihood	-18,037		-18,013		-18,026		-18,001	
AIC	36,193		36,150		36,174		36,129	
BIC	36,652		36,624		36,640		36,611	

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

Appendix B. Mathematical Formulation of the Rostering Problem

The PRP_0 problem can be formulated as follows:

$$\min_{x} \quad \mathbb{E}_{\Lambda,\Psi}\left[Q(\Lambda,\Psi,x)\right] \tag{B.1}$$

The objective (B.1) minimizes the expected mismatch between the demand for emergency care and the supply provided in terms of physician productivity.

$$\sum_{i \in I} x_{ijk} = 1, \qquad j \in J, k \in K \tag{B.2}$$

Constraints (B.2) ensure that exactly one physician is assigned to each shift of the day.

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

Table A.2 Regression coefficients for all physicians in the base model for 7-hour shifts.

Physician	Coefficient		Physician	Coefficient		Physician	Coefficient	
Physician2	0.484^{***}	(0.054)	Physician19	0.287^{***}	(0.067)	Physician36	0.275^{***}	(0.077)
Physician3	0.387^{***}	(0.060)	Physician20	0.378^{***}	(0.067)	Physician37	0.269^{***}	(0.066)
Physician4	0.224^{***}	(0.063)	Physician21	$0.122^{\cdot}$	(0.071)	Physician38	0.232^{***}	(0.071)
Physician5	-0.032	(0.078)	Physician22	0.397^{***}	(0.070)	Physician39	0.402^{***}	(0.073)
Physician6	0.246^{***}	(0.062)	Physician23	0.335^{***}	(0.063)	Physician40	-0.220^{*}	(0.098)
Physician7	-0.001	(0.075)	Physician24	0.236^{***}	(0.071)	Physician41	0.315^{***}	(0.070)
Physician8	0.217^{***}	(0.064)	Physician25	0.386^{***}	(0.076)	Physician42	0.248^{**}	(0.078)
Physician9	0.226^{***}	(0.059)	Physician26	0.014	(0.120)	Physician43	0.216^{**}	(0.082)
Physician10	0.227^{**}	(0.070)	Physician27	0.165^{*}	(0.069)	Physician44	0.262^{**}	(0.080)
Physician11	0.076	(0.060)	Physician28	0.111	(0.073)	Physician45	0.019	(0.079)
Physician12	0.177^{**}	(0.067)	Physician29	0.077	(0.073)	Physician46	0.123	(0.088)
Physician13	0.207^{**}	(0.067)	Physician30	0.102	(0.077)	Physician47	0.450^{***}	(0.060)
Physician14	-0.042	(0.082)	Physician31	0.172^{*}	(0.072)	Physician48	0.231^{**}	(0.085)
Physician15	0.234^{***}	(0.064)	Physician32	0.051	(0.068)	Physician49	0.282^{***}	(0.073)
Physician16	0.208^{**}	(0.070)	Physician33	0.204^{**}	(0.066)	Physician50	-0.062	(0.095)
Physician17	0.101	(0.070)	Physician34	0.104	(0.076)	Physician51	0.198^{**}	(0.073)
Physician18	0.017	(0.075)	Physician35	0.324^{***}	(0.071)	Physician52	0.228^{**}	(0.075)

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 the base model for 8-hour shifts.

Physician	Coefficient		Physician	Coefficient		Physician	Coefficient	
Physician2	0.651^{***}	(0.063)	Physician19	0.378^{***}	(0.066)	Physician36	0.361^{***}	(0.067)
Physician3	0.401^{***}	(0.063)	Physician20	0.375^{***}	(0.064)	Physician37	0.457^{***}	(0.072)
Physician4	0.217^{**}	(0.067)	Physician21	0.272^{***}	(0.070)	Physician38	0.177^{**}	(0.068)
Physician5	-0.011	(0.072)	Physician22	0.460^{***}	(0.065)	Physician39	0.351^{***}	(0.080)
Physician6	0.338^{***}	(0.071)	Physician23	0.333^{***}	(0.073)	Physician40	-0.104	(0.082)
Physician7	-0.061	(0.076)	Physician24	0.208^{**}	(0.070)	Physician41	0.314^{***}	(0.074)
Physician8	0.344^{***}	(0.065)	Physician25	0.386^{***}	(0.061)	Physician42	0.193^{**}	(0.071)
Physician9	0.201^{**}	(0.077)	Physician26	0.059	(0.064)	Physician43	0.338^{***}	(0.072)
Physician10	0.213^{**}	(0.066)	Physician27	0.070	(0.076)	Physician44	0.309^{***}	(0.073)
Physician11	0.175	(0.110)	Physician28	0.193^{**}	(0.069)	Physician45	0.128	(0.087)
Physician12	0.241^{***}	(0.067)	Physician29	0.072	(0.072)	Physician46	0.110	(0.076)
Physician13	0.226^{***}	(0.068)	Physician30	0.206^{**}	(0.077)	Physician47	0.420^{***}	(0.086)
Physician14	0.073	(0.073)	Physician31	0.198^{**}	(0.076)	Physician48	0.338^{***}	(0.065)
Physician15	0.267^{***}	(0.068)	Physician32	0.048	(0.073)	Physician49	0.339^{***}	(0.075)
Physician16	0.266^{***}	(0.069)	Physician33	0.282^{***}	(0.073)	Physician50	0.019	(0.078)
Physician17	0.228^{**}	(0.076)	Physician34	$0.128^{\cdot}$	(0.071)	Physician51	0.209^{*}	(0.092)
Physician18	-0.088	(0.082)	Physician35	0.358^{***}	(0.073)	Physician52	0.253^{***}	(0.077)

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

Constraints (B.3) enforce that a physician is not assigned to more than one shift per day.

$$\sum_{j\in\tilde{J}}\sum_{k\in\tilde{K}} x_{ijk} \geq L(\tilde{J},\tilde{K}), \qquad i\in I, \tilde{J}\in J', \tilde{K}\in K' \tag{B.4}$$

$$\sum_{j\in\tilde{J}}\sum_{k\in\tilde{K}} x_{ijk} \leq U(\tilde{J},\tilde{K}), \qquad i\in I, \tilde{J}\in J', \tilde{K}\in K' \tag{B.5}$$

Table A.4 **Regression coefficients for models with physician clustering for 8-hour shifts.**

Variables	1 Cluster	2 Clusters	3 Clusters	4 Clusters	
(Intercept)	0.905^{***} (0.015)	1.069^{***} (0.019)	0.921^{***} (0.016)	1.086^{***}	(0.022)
Hour2	-0.355^{***} (0.022)	-0.355^{***} (0.022)	-0.355^{***} (0.022)	-0.355^{***}	(0.022)
Hour3	-0.553^{***} (0.024)	-0.553^{***} (0.024)	-0.553^{***} (0.024)	-0.553^{***}	(0.024)
Hour4	-0.685^{***} (0.025)	-0.685^{***} (0.025)	-0.685^{***} (0.025)	-0.685^{***}	(0.025)
Hour5	-0.647^{***} (0.025)	-0.647^{***} (0.025)	-0.647^{***} (0.025)	-0.647^{***}	(0.025)
Hour6	-0.961^{***} (0.027)	-0.961^{***} (0.027)	-0.961^{***} (0.027)	-0.961^{***}	(0.027)
Hour7	-2.098^{***} (0.043)	-2.098^{***} (0.043)	-2.098^{***} (0.043)	-2.098^{***}	(0.043)
Hour8	-4.281^{***} (0.123)	-4.281^{***} (0.123)	-4.281^{***} (0.123)	-4.281^{***}	(0.123)
Night	0.046^{**} (0.017)	0.046^{**} (0.017)	0.048^{**} (0.017)	0.045^{**}	(0.017)
Cluster2		-0.223^{***} (0.016)	0.164^{***} (0.020)	-0.207^{***}	(0.020)
Cluster3			-0.176^{***} (0.018)	-0.110^{***}	(0.024)
Cluster4				-0.370^{***}	(0.025)
Log likelihood	-18,256	-18,165	-18,144	-18,137	
AIC	36,530	36,350	36,309	36,299	
BIC	36,599	36,427	36,393	36,391	

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

Constraints (B.4) and (B.5) 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 (B.2) through (B.5) correspond to the *balance rules* in our case study. The next eight constraints correspond to the *pattern rules*.

$$\sum_{k \in K_N} x_{i(j-1)k} + \sum_{k \in K_D} x_{ijk} \leq 1, \qquad i \in I, j \in J \setminus \{1\} \tag{B.6}$$

Constraints (B.6) guarantee that a physician who is assigned to a night shift must not be assigned to a day shift on the next day.

$$\sum_{k \in K_N} x_{i(j-3)k} + \sum_{j'=0}^{2} \sum_{k \notin K_N} x_{i(j-j')k} \leq 1, \qquad i \in I, j \in J \setminus \{1,2,3\} \tag{B.7}$$

Constraints (B.7) 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.

$$3 \sum_{k \in K_N} x_{i(j-20)k} + \sum_{j'=0}^{17} \sum_{k \in K_N} x_{i(j-j')k} \leq 3, \qquad i \in I, j \in J \setminus \{1,2,\ldots,20\} \tag{B.8}$$

Constraints (B.8) 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.

$$\sum_{k \in K} x_{i2k} \geq \sum_{k \in K} x_{i1k}, \qquad i \in I \tag{B.9}$$

$$\sum_{k \in K} x_{ijk} \geq \sum_{k \in K} x_{i(j-1)k} - \sum_{k \in K} x_{i(j-2)k}, \qquad i \in I, j \in J \setminus \{1,2\} \tag{B.10}$$

$$\sum_{j'=0}^{4}\sum_{k\in K} x_{i(j-j')k} \leq 4, \qquad i\in I, j\in J\setminus\{1,2,3,4\} \qquad \text{(B.11)}$$

Constraints (B.9), (B.10), and (B.11) correspond to at least 2 and at most 4 consecutive working days for a physician, respectively.

$$\Big(1-\sum_{k\in K} x_{ijk}\Big) \geq \sum_{k\in K} x_{i(j-2)k} - \sum_{k\in K} x_{i(j-1)k}, \qquad i\in I, j\in J\setminus\{1,2\} \qquad \text{(B.12)}$$

$$\sum_{j'=0}^{4}\Big(1-\sum_{k\in K} x_{i(j-j')k}\Big) \leq 4, \qquad i\in I, j\in J\setminus\{1,2,3,4\} \qquad \text{(B.13)}$$

Constraints (B.12) and (B.13) express the minimum and maximum number of days off after a consecutive number of days that a physician is scheduled.

$$x_{ijk}\in\{0,1\} \qquad i\in I, j\in J, k\in K \qquad \text{(B.14)}$$

Constraints (B.14) 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 (B.1).

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 Appendix B). Hence, we do not repeat them here.

$$(PRP_{0m}) \qquad \min_{x} \quad \frac{1}{|S_m|}\sum_{s\in S_m}\sum_{j\in J}\sum_{t\in T} M_{jt}(s)$$

$$
\begin{aligned}
\text{s.t.}\quad & A_{jt}(s) + M_{j(t-1)}(s) - \sum_{i\in I}\sum_{k\in K} x_{ijk} P^{jk}_{i f_{kt}}(s) \le M_{jt}(s), && \forall j\in J, t\in T\setminus\{0\}, s\in S_m,\\
& A_{1,0}(s) - \sum_{i\in I}\sum_{k\in K} x_{i1k} P^{1,k}_{i f_{k0}}(s) \le M_{1,0}, && s\in S_m,\\
& A_{j,0}(s) + M_{(j-1),23}(s) - \sum_{i\in I}\sum_{k\in K} x_{ijk} P^{jk}_{i f_{k0}}(s) \le M_{j,0}(s), && j\in J\setminus\{1\}, s\in S_m,\\
& \sum_{i\in I} x_{ijk} = 1, && j\in J, k\in K,\\
& \sum_{k\in K} x_{ijk} \le 1, && i\in I, j\in J,\\
& \sum_{j\in\tilde{J}}\sum_{k\in\tilde{K}} x_{ijk} \ge L(\tilde{J},\tilde{K}), && i\in I, \tilde{J}\in J', \tilde{K}\in K',\\
& \sum_{j\in\tilde{J}}\sum_{k\in\tilde{K}} x_{ijk} \le U(\tilde{J},\tilde{K}), && i\in I,, \tilde{J}\in J', \tilde{K}\in K',\\
& \sum_{k\in K_N} x_{i(j-1)k} + \sum_{k\in K_D} x_{ijk} \le 1, && i\in I, j\in J\setminus\{1\},\\
& \sum_{k\in K_N} x_{i(j-3)k} + \sum_{j'=0}^{2}\sum_{k\notin K_N} x_{i(j-j')k} \le 1, && i\in I, j\in J\setminus\{1,2,3\},\\
& 3\sum_{k\in K_N} x_{i(j-20)k} + \sum_{j'=0}^{17}\sum_{k\in K_N} x_{i(j-j')k} \le 3, && i\in I, j\in J\setminus\{1,2,\ldots,20\},\\
& \sum_{k\in K} x_{i2k} \ge \sum_{k\in K} x_{i1k}, && i\in I,\\
& \sum_{k\in K} x_{ijk} \ge \sum_{k\in K} x_{i(j-1)k} - \sum_{k\in K} x_{i(j-2)k}, && i\in I, j\in J\setminus\{1,2\},\\
& \sum_{j'=0}^{4}\sum_{k\in K} x_{i(j-j')k} \le 4, && i\in I, j\in J\setminus\{1,2,3,4\},\\
& 1 - \sum_{k\in K} x_{ijk} \ge \sum_{k\in K} x_{i(j-2)k} - \sum_{k\in K} x_{i(j-1)k}, && i\in I, j\in J\setminus\{1,2\},\\
& \sum_{j'=0}^{4}\left(1 - \sum_{k\in K} x_{i(j-j')k}\right) \le 4, && i\in I, j\in J\setminus\{1,2,3,4\},\\
& x_{ijk}\in\{0,1\}, && i\in I, j\in J, k\in K,\\
& M_{jt}(s) \ge 0, && j\in J, t\in T, s\in S_m.
\end{aligned}
$$