
Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# A Robust Optimization Model for Managing Elective Admission in a Public Hospital

Fanwen Meng, Jin Qi, Meilin Zhang, James Ang, Singfat Chu, Melvyn Sim

OPERATIONS RESEARCH
Vol. 63, No. 6, November–December 2015, pp. 1452–1467
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2015.1423


# A Robust Optimization Model for Managing Elective Admission in a Public Hospital

**Fanwen Meng**
Department of Health Services and Outcomes Research, National Healthcare Group, Singapore 138543,
fanwen_meng@nhg.com.sg

**Jin Qi**
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong,
jinqi@ust.hk

**Meilin Zhang, James Ang, Singfat Chu, Melvyn Sim**
NUS Business School, National University of Singapore, Singapore 119077
{214101@gmail.com, bizangsk@nus.edu.sg, bizchucl@nus.edu.sg, dscsimm@nus.edu.sg}

The admission of emergency patients in a hospital is unscheduled, urgent, and takes priority over elective patients, who are usually scheduled several days in advance. Hospital beds are a critical resource, and the management of elective admissions by enforcing quotas could reduce incidents of shortfall. We propose a distributionally robust optimization approach for managing elective admissions to determine these quotas. Based on an ambiguous set of probability distributions, we propose an optimized *budget of variation* approach that maximizes the level of uncertainty the admission system can withstand without violating the expected bed shortfall constraint. We solve the robust optimization model by deriving a second order conic problem (SOCP) equivalent of the model. The proposed model is tested in simulations based on real hospital admission data, and we report favorable results for adopting the robust optimization models.

*Keywords*: robust optimization; elective admission; healthcare; hospital operations.
*Subject classifications*: health care: hospital; programming: stochastic.
*Area of review*: Optimization.
*History*: Received May 2012; revisions received October 2013, November 2014, June 2015; accepted August 2015. Published online in *Articles in Advance* October 26, 2015.

## 1. Introduction

Overcrowding in a hospital emergency department (ED) causes long waiting times associated with patient satisfaction (e.g., Thompson and Yarnold 1995, Conner-Spady et al. 2011, Sun et al. 2012) and more importantly, morbidity and mortality (e.g., Librero et al. 2004, Guttmann et al. 2011, Plunkett et al. 2011). ED overcrowding is often caused by availability (or rather the shortage) of hospital beds (e.g., Wardrope and Driscoll 2003, Hwang and Concato 2004, Hoot and Aronsky 2008). However, beds are a critical resource in hospital operations, which need highly trained personnel to man these beds. Work has been done in the area of the acquisition and utilization of bed resources (e.g., Harper and Shahani 2002, Kao and Tung 1981, Cochran and Roche 2008, Teow and Tan 2008). Harper and Shahani (2002) acknowledged the complexity of the internal dynamics of a hospital (especially bed management), and used a simulation model for patient flows and bed matching over time.

Typically, day-of-week (DoW) patterns of a hospital exhibit a wide range of variations: emergency admissions are unpredictable, whereas elective admissions are scheduled by the hospital. Nevertheless, often the relative variation is largest in elective admissions, and larger in discharges than emergency admissions (Proudlove et al. 2007). On days with high bed occupancy, long wait time is encountered. On days with low bed occupancy, beds are underutilized. We have tightness of usage on one hand and looseness on the other. It is not the desired state.

Elective surgeries account for the majority of elective admissions, though medical electives (nonsurgical cases) do make up some of these admissions. Elective surgeries are procedures planned in advance and can be divided into day surgery (DS), same-day surgery admission (SDA) and inpatient admission (IP). DS cases do not “consume” beds, whereas SDA cases require beds to accommodate patient day after surgery. IP cases require beds one day before the surgery.

In general, hospitals will admit all patients with emergent needs. In Singapore, there are four levels of patient acuity category (PAC) for ED attendances, with PAC1 being the most serious and the PAC4 the least. All PAC1 patients and a large portion of PAC2 patients usually need to be admitted to the hospital. As such, in a tight bed situation, the trade-off is to reduce the number of beds designated for elective admissions. A more prudent and sensible approach would be to make adjustments on a dynamic basis. When emergency cases are fewer, then more beds could be assigned to elective cases and vice versa. This leads to an optimal control policy, which is to maximize bed utilization on a daily basis by controlling the number of elective admissions. This requires a more prudent scheduling of operating theatre sessions. However, a higher level of complexity in planning ensues because of the high degree of uncertainty involved in bed availability and its effect on admission rates.

Various models for managing patient admissions have been proposed in the literature. Esogbue and Singh (1976) developed a method for determining optimal distribution of beds in a ward using cut-off level via shortage and holding costs. They assumed Poisson patient arrival distribution and negative exponential distribution for length of stay. Kao and Tung (1981) proposed an approach for periodically reallocating beds to services to minimize the expected overflows, using queueing models to approximate the population dynamics. In fact, queueing theory and stochastic simulation are the main methodological tools in studies of bed allocation and bed capacity (Vassilacopoulos 1985, Gorunescu et al. 2004, Cochran and Roche 2008, Lamiri et al. 2008). The underlying rationale for researchers relying on these methodological tools is the uncertain nature of the hospital unit vis-à-vis the number of patients as a result of random arrivals and random lengths of stay. A thorough review on OR applications in healthcare services in the United Kingdom can be found in Proudlove et al. (2007).

The admission of emergency inpatients is unscheduled, and they are usually warded within hours. In contrast, admission of elective patients is less pressing, and they can be warded on the day of admission or even several weeks later. The flexibility vis-à-vis elective patients allows hospitals to manage the flow of elective patients in such a way as to "smooth out" the daily bed occupancy, a modus operandus known as "elective smoothing." This will ensure that on days with spikes in emergency cases, the admission rate for elective patients can be reduced. The converse applies. Some hospitals in Singapore have already incorporated this mechanism into their decision support systems, and it has led to improvements when elective patient flow is high (Teow et al. 2007). In these hospitals, the admission quotas for elective patients are obtained by solving a deterministic linear optimization problem without taking into account the variability of patient arrivals and stay durations. Whereas this achieves smoothing in expectation, it is conceivable that efficacy would diminish when variability is high.

Because of difficulties in obtaining true probability distributions and solving stochastic optimization problems, it is common in real world deployment of optimization technology to ignore uncertainty. A fine level of analysis would be required to obtain the distributions of patient admission and departure profile as a function of admission quotas, which may not necessarily lead to a computational tractable optimization problem. In recent years, robust optimization offers an attractive alternative for addressing uncertainty in optimization modeling without having to specify exact probability distributions. In many interesting cases, the approach leads to computationally tractable optimization problems; see, for instance, Ben-Tal and Nemirovski (1998), Bertsimas and Sim (2004), El Ghaoui et al. (1998). In classical robust optimization, uncertainty is represented by an *uncertainty set*, which is usually a simple geometric convex set such as a $l$-norm ball intersected with the *support set*, the minimal convex set that contains the uncertainty. The modeler requires to articulate her *ambiguity attitude*[1] by specifying the *budget of uncertainty* parameter, which relates to the size of the uncertainty set.

Although there are several proposed uncertainty sets and heuristics for specifying budgets of uncertainty, these approaches may not naturally characterize the uncertainty relating to patient movements within the hospital. In this paper, we adopt the distributionally robust optimization approach for managing elective admission in hospital, where uncertainty is characterized by the support set and a restricted ambiguous set of probability distributions (or *ambiguity set* for short); see, for instance, Chen et al. (2007, 2010), Delage and Ye (2010) and Goh and Sim (2010, 2011). Similar to the uncertainty set in classical robust optimization, the proposed ambiguity set is adjustable via a so-called *budget of variation* parameter, which is the bound on the coefficient of variation of the uncertainty parameters. The ambiguity set is enlarged by increasing the budget of variation, which leads to greater uncertainty in the patient movements.

Quite apart from the usual paradigm of robust optimization, we propose an approach to optimize the budget of variation while ensuring that the worst-case expected maximum bed requirement over the planning horizon falls below the bed capacity of the hospital. This approach is inspired by the actual problem for which we have access to the data to attest the performance. The key challenge we face is to model uncertainty in a way while keeping the computations tractable so that we can obtain consistent improvement over the static approach for which uncertainty is ignored. Interestingly, this could be achieved by solving a small collection of computationally tractable optimization problems. We also study the performance of this approach in a case study using real data.

The rest of this paper is organized as follows. In §2, we establish a distributionally robust optimization model for managing elective admission in hospital with incomplete information of uncertainties. We then investigate deterministic formulation to this model by deriving a second order conic optimization problem (SOCP) in §3. Numerical experiments using real data are carried out in §4.

*Notation.* We denote a random variable with a tilde, such as $\tilde{z}$. Matrices and vectors are represented as upper and lower case boldface characters, respectively. If $\mathbf{x}$ is a vector, we use the notation $x_i$ to denote the $i$th component of the vector. We represent uncertainty by a state-space $\Omega$ and a set ($\sigma$-algebra) $\mathcal{F}$ of events. We use the notation $\tilde{x} \geqslant \tilde{y}$ to denote state-wise dominance over all attributes, i.e., $\tilde{x}(\omega) \geqslant \tilde{y}(\omega)$ for all $\omega \in \Omega$. We use $\mathbb{P}$ to denote a probability measure on $\Omega$, and $\mathbb{E}_{\mathbb{P}}(\tilde{x})$, $\sigma_{\mathbb{P}}(\tilde{x})$, and $\text{cv}_{\mathbb{P}}(\tilde{x})$ denote, respectively, the expectation, standard deviation, and coefficient of variation of $\tilde{x}$ under $\mathbb{P}$.

## 2. Model Formulation

In this section, we develop a robust optimization model for managing elective admissions in hospitals. First, we list some notations such as the parameters and variables used in the model development.

*Parameters*

- $T$: total number of days in the planning horizon starting from day 0, indexed by $t = 0, 1, \ldots, T-1$;
- $L$: maximum length of stay of patients in hospitals (in day);
- $\mathcal{T}$: index set of days describing patients' stay status in the planning horizon, where $\mathcal{T} := \mathcal{T}^{--} \cup \mathcal{T}^{+}$ with $\mathcal{T}^{--} = \{-L+1, \ldots, -1\}$ and $\mathcal{T}^{+} = \{0, \ldots, T-1\}$. For a negative $t \in \mathcal{T}$, the index $t$ is referred to as the $|t|$th day ahead of the starting point (i.e., day 0) in the planning horizon;
- $\tilde{p}_{t,l}$: number of emergency inpatients who might arrive or have arrived on the $t$th day and will likely be warded for at least $l$ days, $t \in \mathcal{T}$ and $l \in \{1, \ldots, L\}$;
- $\tilde{a}_{t,l}$: number of elective inpatients who might be admitted or have been admitted on the $t$th day according to the schedule and will likely be warded for at least $l$ days, $t \in \mathcal{T}$ and $l \in \{1, \ldots, L\}$;
- $X$: the feasible region of admissible quotas defined by their daily bounds and weekly capacity requirements.

*Decision variables*

- $\eta_t$: quota for elective admissions for the $t$th day within the planning horizon, where $t = 0, 1, \ldots, T-1$.

We consider a planning horizon of $T$ days indexed by $t = 0, 1, \ldots, T-1$. For simplicity of model presentation, we assume that all inpatients are of the same type. We can easily refine the model to consider quotas for different types of inpatients that may be characterized by gender, discipline, and so forth. We detail how this is implemented in a public hospital. At the beginning of day $t = 0$, say at 8 AM when the hospital bed management unit will make bed allocation plan for the next few days, the quotas $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_{T-1})'$ will be determined and integrated within hospital decision support system for assignment of admissions. In general, when the requests arrive, hospital administrators will work with the patients or their caregivers and assign patient admission dates, which would largely depend on the availability of bed quotas and patients' schedule preferences. The booking system is similar to the airline booking system. When an elective bed request is submitted, the patient could only be assigned to the days where the quota is strictly greater than the number of patients that have been assigned, which is represented as $\underline{\boldsymbol{\eta}}$. Suppose this new elective patient is scheduled to be admitted on day $l$; the corresponding $\underline{\eta}_l$ will be updated as $\underline{\eta}_l = \underline{\eta}_l + 1$. The patient may commit on a schedule or, if the waiting time is too long, decide not to proceed with the treatment in this hospital. As we proceed to the next day, the process is repeated and a new set of quotas will be computed using the latest information on patient admission and discharge status. In this study, we do not investigate specific bed assignment rules for elective requests. Instead, our focus is on the daily bed capacity planning of elective admissions within a given planning period.

For the public hospital of this study, the average request for elective admissions would exceed the assigned capacity because of the hospital's policy of catering more beds for emergency admissions. Besides, when a request for elective admission is made, the database does not capture the negotiating processes to report the unconstrained demand on each day. Hence, in our model, the uncertain parameter $\tilde{a}_{t,l}$ corresponds to the elective patients who have scheduled their appointment and arrived or might arrive on the $t$th day and would be warded for at least $l$ days. To avoid misunderstanding, we will call it uncertain admissions instead of uncertain arrivals.

The feasible space, $X \subseteq \mathbb{Z}^T$, of elective admissible quotas is usually defined based on practical considerations such as the bounds for daily admissions, the minimum requirement on the weekly accumulated admissions. The feasible set $X$ should be specified accordingly to exclude trivial results such as zero assigned quotas for elective patients. For instance, since the hospital sets aside a portion of her capacity to serve elective patients, we enforce by constraining the total quotas during the planning horizon to match the desired average number of elective patients. In the rolling horizon implementation, it is also imperative to ensure that the new set of quotas is able to accommodate previously assigned elective admissions. For example, if 15 elective admissions have already been assigned on day $t = 6$, we would impose a constraint $\eta_6 \geqslant 15$.

We next describe the dynamics of patient flow. Note that by definition, inpatients are patients who are warded for at least one day. To account for the total number of inpatients on the $t$th day, we need to keep track of the admission status up to $L-1$ days before the planning horizon. Recall that uncertain factor $\tilde{p}_{t,l}$ (or $\tilde{a}_{t,l}$) represents the number of emergency (or elective) inpatients being admitted on the $t$th day, $t \in \mathcal{T}$ and would be warded for at least $l$ days, $l \in \{1, \ldots, L\}$. For instance, $\tilde{p}_{1,1}$ refers to the total number of emergency inpatients on day $t = 1$, and its value is uncertain. If $\tilde{d}$ of these patients are discharged on day $t = 2$, then $\tilde{p}_{1,2} = \tilde{p}_{1,1} - \tilde{d}$. Likewise, $\tilde{a}_{-1,2}$ refers to the number of elective inpatients that were admitted on the previous day ($t = -1$) and would be warded for at least two days. At the beginning of day $t = 0$, doctors may not have reviewed the cases for discharge. Hence, the parameter $\tilde{a}_{-1,2}$ is generally uncertain. For our purpose, we need to account for the number of inpatients during the planning horizon, i.e., on the days in $\mathcal{T}^{+}$. For inpatients who were admitted on day $t \in \mathcal{T}^{--}$, only the inpatients with the length of stay of at least $l$ days, $l \geqslant 1-t$, may remain warded in the hospital during the planning horizon. On the other hand, for patients being admitted on day $t \in \mathcal{T}^{+}$, only the information associated with inpatients with length of stay at least $l$ days, $l \leqslant \min\{L, T-t\}$ will be needed to compute the quotas. Hence, for notational convenience, we define $\mathcal{L}_t = \{\max\{1, 1-t\}, \max\{1, 1-t\}+1, \ldots, \min\{L, T-t\}\}$, $t \in \mathcal{T}$.

We now account for the total number of inpatients on the $t$th day during the planning horizon, $t \in \mathcal{T}^{+}$. For example, the total number of inpatients on day $t=0$ can be computed as follows:

$$
\begin{aligned}
&\tilde{a}_{0,1} + \tilde{p}_{0,1} + \quad &&\text{(admissions on } t=0)\\
&\tilde{a}_{-1,2} + \tilde{p}_{-1,2} + \quad &&\text{(admissions on } t=-1 \text{ and warded for at least two days)}\\
&\cdots + \tilde{a}_{-L+1,L} + \tilde{p}_{-L+1,L} \quad &&\text{(admissions on } t=-L+1 \text{ and warded for up to } L \text{ days).}
\end{aligned}
$$

In general, it follows that the total inpatients on day $t \in \mathcal{T}^{+}$ can be computed as

$$\sum_{(\tau, l) \in \mathcal{U}_t} (\tilde{a}_{\tau, l} + \tilde{p}_{\tau, l}),$$

where the index set $\mathcal{U}_t$ is given by

$$\mathcal{U}_t = \big\{(\tau, l)\colon \tau \in \mathcal{T},\ l \in \mathcal{L}_\tau,\ l + \tau = t+1\big\}.$$

A bed shortfall occurs whenever the total number of inpatients exceeds the bed capacity, which we denote by $c_t$, $t \in \mathcal{T}^{+}$. Note that for generality, we assume that bed capacity, which encompasses the physical beds and manpower availability, is time dependent. Before we can specify an optimization problem, we first need to account for the uncertainty concerning patient admission and departure.

## Characterizing Patient Admissions and Departures Uncertainty

We describe a nonparametric approach for characterizing the uncertainty on patient admissions and departures using information obtained from patient movement records. Our aim is to introduce a model of uncertainty without imposing excessive burden on the information requirement, which may otherwise deter practical implementation. Instead of ignoring variability and assuming deterministic parameters taking values at their empirical averages, which is usually done in practice, we assume that the parameters are random variables with known means. However, their precise distributions are unavailable but belong to a restricted ambiguity set. To avoid being overly conservative, we control the "size" of the ambiguity set by specifying the budget of variation, $\mu$, which is the upper bound of the coefficients of variations of all the uncertain parameters.

We next show how the uncertain parameters $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ are interrelated, which is the basis for characterizing the support of the uncertainty. Observe that by definition, $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ are nonincreasing in $l$. For inpatients being admitted before $t=0$, their total admissions are known, but their durations of stay may be uncertain. Let $p_t^0$ and $a_t^0$, $t \in \mathcal{T}^{--}$, be, respectively, the number of remaining emergency and elective inpatients who have been admitted on day $t$ and are still being warded up to the beginning of day 0. The support of the uncertain parameters $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$ is given by

$$
\begin{aligned}
&p_t^0 \geqslant \tilde{p}_{t,l} \geqslant \tilde{p}_{t,l'} \geqslant 0,\\
&a_t^0 \geqslant \tilde{a}_{t,l} \geqslant \tilde{a}_{t,l'} \geqslant 0,
\end{aligned}
$$

for all $t \in \mathcal{T}^{--}$, $l, l' \in \mathcal{L}_t, l' > l$. Similarly, for inpatients arriving during the planning horizon $t \in \mathcal{T}^{+}$, the support of the associated uncertain parameters $\tilde{p}_{t,l}$, $\tilde{a}_{t,l}$ is given by

$$
\begin{aligned}
&p_t^0 \geqslant \tilde{p}_{t,l} \geqslant \tilde{p}_{t,l'} \geqslant 0,\\
&\eta_t \geqslant \tilde{a}_{t,l} \geqslant \tilde{a}_{t,l'} \geqslant 0,
\end{aligned}
$$

for all $t \in \mathcal{T}^{+}$, $l, l' \in \mathcal{L}_t, l' > l$. For the emergency patients, the input parameter $p_t^0$ is a prescribed upper bound of $\tilde{p}_{t,l}$. This setting is consistent with the practice, since the numbers of daily arrivals could not be arbitrarily large, which are actually constrained by the population of the catchment area of the hospital under consideration. In contrast to emergency admissions, elective patients usually adhere to their scheduled appointments for surgery treatment on the same day or the following day if presurgical preparations are required. Hence, according to the admission process we have described, the number of patients being admitted at the $t$th day and would be warded at least $l$ days, $\tilde{a}_{t,l}$, is an endogenous random variable that depends on the quota, $\eta_t$. If $\eta_t = 0$, then it is clear that $\tilde{a}_{t,l} = 0$ for all $l \in \mathcal{L}_t$. We provide an example to illustrate this dependency. Suppose at $t=1$, $\eta_1 = 10$, $\eta_2 = 1$, $\eta_3 = 15$, and the number of assigned electives, $\underset{\smile}{\eta}_1 = 10$, $\underset{\smile}{\eta}_2 = 0$, $\underset{\smile}{\eta}_3 = 10$, the hospital would be able to schedule new elective patients at $t=2$ or $t=3$ but not at $t=1$. Since the quotas are fully assigned at $t=1$, if every elective patient turns up, then $\tilde{a}_{1,1} = 10$. Hence, $\tilde{a}_{t,l}$ is highly dependent on $\eta_t$.

Instead of assuming a probability distribution, we specify the ambiguity set such that for each distribution, $\mathbb{P}$ in the set, the uncertain parameters are random variables with known mean values and their coefficients of variations are bounded above by $\mu$. Specifically, for inpatients being admitted before $t=0$, i.e., $t \in \mathcal{T}^{--}$, we assume that

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l},\\
&\mathbb{E}_{\mathbb{P}}(\tilde{a}_{t,l}) = \bar{a}_{t,l},
\end{aligned}
$$

for all $l \in \mathcal{L}_t$, where $\bar{p}_{t,l}$ and $\bar{a}_{t,l}$ are, respectively, the empirical averages of $\tilde{p}_{t,l}$ and $\tilde{a}_{t,l}$. Since these patients are already admitted, in principle, the parameters $\bar{p}_{t,l}$, $\bar{a}_{t,l}$ may be inferred from the patients' likely duration of stay assessed by their doctors. If such information is unavailable, then one may also use values that are empirically estimated from historical records.

Observe that during the planning horizon, $t \in \mathcal{T}^+$, the uncertain parameters $\tilde{p}_{t,l}$, $t \in \mathcal{T}^+$ are associated with inpatients who have yet to be admitted at the hospital. Hence, we are able to obtain the empirical averages from patient movement records as follows:

$$\mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l},$$

for all $t \in \mathcal{T}^+, l \in \mathcal{L}_t$. Unlike the previous case, the elective patients $\tilde{a}_{t,l}, t \in \mathcal{T}^+, l \in \mathcal{L}_t$ are associated with the quotas $\eta_t$. Clearly, the dependency on $\eta_t$ would impact how the parameters should be estimated and how the model can be solved efficiently. Accordingly, we define new random variables $\tilde{\alpha}_{t,l}$ as

$$\tilde{\alpha}_{t,l} = \frac{\tilde{a}_{t,l}}{\eta_t},$$

for all $t \in \mathcal{T}^+, l \in \mathcal{L}_t$ to represent the proportion of patients who will be warded at least $l$ days with respect to the quota $\eta_t$. From the historical data on $\tilde{a}_{t,l}$ and $\eta_t$, we could determine its empirical average as

$$\mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) = \bar{\alpha}_{t,l},$$

for all $t \in \mathcal{T}^+, l \in \mathcal{L}_t$.

To formulate a tractable and scalable model that could be solved by commercial solvers, we impose the following assumption.

Assumption 1. *The descriptive statistics of $\tilde{\alpha}_{t,l}$ are independent on $\eta_t$.*

Though Assumption 1 may be restrictive, it has important ramifications on the computational tractability of the model, which we will explain in §3. It leads to simpler estimation of the descriptive statistics from data.

For notational simplicity, since we have the complete information for the quotas assigned before $t = 0$, we define

$$\tilde{\alpha}_{t,l} = \frac{\tilde{a}_{t,l}}{\eta_t}, \quad \bar{\alpha}_{t,l} = \frac{\bar{a}_{t,l}}{\eta_t}, \quad \alpha_t^0 = \frac{a_t^0}{\eta_t}, \quad \forall\, t \in \mathcal{T}^{--},\ l \in \mathcal{L}_t,$$
$$\alpha_t^0 = 1, \quad \forall\, t \in \mathcal{T}^+.$$

Hence, we can empirically determine the values for $\{\bar{\alpha}_{t,l}\colon t \in \mathcal{T}, l \in \mathcal{L}_t\}$ and $\{\alpha_t^0\colon t \in \mathcal{T}^{--}\}$ from the data.

Finally, the coefficients of variation of these parameters are bounded above by $\mu$, which we call the *budget of variation*, as follows:

$$\mathrm{cv}_{\mathbb{P}}(\tilde{p}_{t,l}) \leqslant \mu,$$
$$\mathrm{cv}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) \leqslant \mu,$$

for all $t \in \mathcal{T}, l \in \mathcal{L}_t$. Hence, $\mu = 0$, implies that the parameters are almost surely certain and take values at their means. On the other extreme with $\mu = \infty$, then essentially the variabilities of these parameters are not constrained by $\mu$, but could otherwise be limited by the support. We present the ambiguity set as a function of the budget of variation, $\mu$, as follows:

$$\mathbb{F}(\mu) = \left\{ \mathbb{P}\colon \begin{array}{l} \mathbb{P}\left( \left\{ \begin{array}{l} p_t^0 \geqslant \tilde{p}_{t,l} \geqslant \tilde{p}_{t,l'} \geqslant 0, \\ \quad \forall (t,l'), (t,l) \in \mathcal{I}, l' > l \\ \alpha_t^0 \geqslant \tilde{\alpha}_{t,l} \geqslant \tilde{\alpha}_{t,l'} \geqslant 0, \\ \quad \forall (t,l'), (t,l) \in \mathcal{I}, l' > l \end{array} \right\} \right) = 1 \\ \mathbb{E}_{\mathbb{P}}(\tilde{p}_{t,l}) = \bar{p}_{t,l}, \quad \forall (t,l) \in \mathcal{I}, \\ \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{t,l}) = \bar{\alpha}_{t,l}, \quad \forall (t,l) \in \mathcal{I}, \\ \sigma_{\mathbb{P}}(\tilde{p}_{t,l}) \leqslant \bar{p}_{t,l}\mu, \quad \forall (t,l) \in \mathcal{I}, \\ \sigma_{\mathbb{P}}(\tilde{\alpha}_{t,l}) \leqslant \bar{\alpha}_{t,l}\mu, \quad \forall (t,l) \in \mathcal{I}, \end{array} \right\},$$

where

$$\mathcal{I} := \{(t,l)\colon t \in \mathcal{T},\ l \in \mathcal{L}_t\}.$$

and the parameters for characterizing the ambiguous set of distributions, $\{p_t^0, \alpha_t^0\colon t \in \mathcal{T}\}$, $\{\bar{p}_{t,l}\colon (t,l) \in \mathcal{I}\}$, $\{\bar{\alpha}_{t,l}\colon (t,l) \in \mathcal{I}\}$ are values obtained from the patient movement records. Observe that the set $\mathbb{F}(\mu)$ is nondecreasing in $\mu$, i.e.,

$$\mathbb{F}(\mu) \subseteq \mathbb{F}(\mu'), \quad \forall\, \mu' \geqslant \mu.$$

Remark 1. For tractability purpose, the discrete nature of the uncertain admissions of emergency and elective patients are not characterized in the ambiguity set. The relaxation of integer random variables to continuous ones is a common technique used in robust optimization to obtain tractable formulations. If we confine to integer random variables, we would have to enumerate exponentially many scenarios to obtain an exact formation, which would lead to intractability. This integrality gap can be significant. For instance, consider a univariate random variable, $\tilde{p}$ taking values in $\{0, 1\}$ and two distributionally ambiguity sets

$$\mathbb{F}_1 = \left\{\mathbb{P}\colon \mathbb{E}_{\mathbb{P}}(\tilde{p}) = 0.5,\ \mathbb{E}_{\mathbb{P}}(\tilde{p}^2) \leqslant 0.1^2,\ \mathbb{P}(\tilde{p} \in [0,1]) = 1\right\}$$

and

$$\mathbb{F}_2 = \left\{\mathbb{P}\colon \mathbb{E}_{\mathbb{P}}(\tilde{p}) = 0.5,\ \mathbb{E}_{\mathbb{P}}(\tilde{p}^2) \leqslant 0.1^2,\ \mathbb{P}(\tilde{p} \in \{0,1\}) = 1\right\}.$$

Observe that the set, $\mathbb{F}_1$ is a conservative approximation of $\mathbb{F}_2$, which is an empty set. Hence, the "integrality gap" can be arbitrarily bad. The ambiguity set we consider in the bed management problem is far more complex, and we do not see how the "integrality gap" can be eliminated in a computationally efficient manner. Likewise, there are other types of distributional ambiguity information that do not lead to computationally tractable formations. Among others, we are unable to obtain tight and tractable formulations for ambiguity information such as higher order moments with support, independence of random variables, and so forth. As in the spirit of robust optimization models, the goal here is to model uncertainty in its entire generality while keeping the model within the framework for which current state-of-the-art commercial solvers can deliver. In our computational study, we observe the importance of adjusting the conservativeness of the ambiguity set through the parameter $\mu$. Hence, instead of fixing $\mu$, we will propose an approach of maximizing the size of the ambiguity set (hence, the level of conservativeness) subject to a threshold constraint. In our computational studies, we observe this approach provides significant improvement over an approach with fixed ambiguity set. As will become clearer, the ambiguity set, as we have defined, enables us to obtain the solution by solving a sequence of tractable optimization problems.

Remark 2. It would also be possible to extend our model to allow for multiple layers of uncertainty, for example, by providing confidence intervals of mean estimates in the distributional ambiguity set. The key issue we face is how we can calibrate the model of uncertainty so that we can have a reliable performance from our data. After experimenting with several distributionally robust optimization models such as incorporating variance estimates, confidence intervals of mean estimates, we observe from our available data that the current model provides consistency in performance improvement.

### Distributionally Robust Optimization Models

To circumvent the difficulties of obtaining probability distributions and solving the complex stochastic model, the elective smoothing approach ignores uncertainty and solves the following deterministic optimization problem:

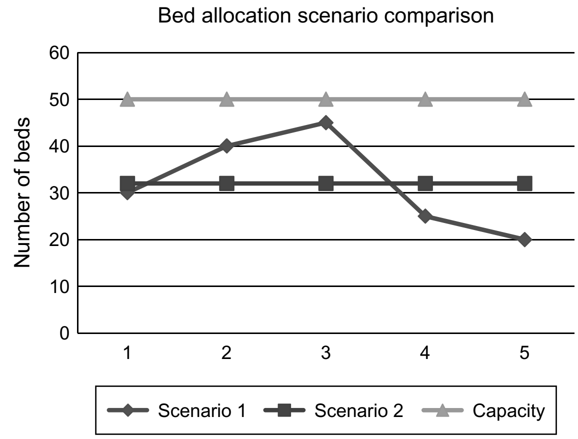
$$Z_{\mathrm{D}} = \min_{\boldsymbol{\eta}\in X}\left(\max_{t\in\mathcal{T}^+}\left\{\sum_{(\tau,l)\in\mathcal{U}_t}(\bar{\alpha}_{\tau,l}\eta_\tau+\bar{p}_{\tau,l})-c_t\right\}\right). \tag{1}$$

The decision variables here are the quotas $\boldsymbol{\eta}=(\eta_\tau)_{\tau\in\mathcal{T}^+}$. In the hospital of our study, bed capacity planning is performed for a period in multiples of weeks. Here we assume that $T=7k$, where $k$ is a positive integer and the feasible set is represented as follows:

$$X=\left\{\boldsymbol{\eta}\in\mathbb{Z}^T\colon \underline{\omega}_t\leqslant\eta_t\leqslant\bar{\omega}_t,\ t=0,\ldots,T-1;\right.$$
$$\left.\sum_{t=0}^{6}\eta_{(t+7i)}=\omega_i,\ i=0,\ldots,k-1\right\}, \tag{2}$$

where $\underline{\omega}_t$, $\bar{\omega}_t$ represent the lower and upper bounds of daily quota for elective admissions, respectively, and $\omega_i$ represents weekly quota capacity for elective admissions in the planning horizon. The aim of the objective function is not merely to minimize bed shortages that might occur during the planning horizon, but rather to "smooth out" the

**Figure 1.** An illustrative example of bed allocation policy.




daily bed occupancy by minimizing the maximum occupancy over the horizon. This is a service-inspired criterion to better accommodate for fluctuations in bed demands, which is known in the literature as elective smoothing. As illustrated in Figure 1, even in the absence of bed shortages, the minmax criterion would favor the bed allocation in Scenario 2, which is uniformly distributed, over that of Scenario 1.

The model ignores the potential impact of uncertainty and could lead to severe shortfalls in hospital beds whenever bad scenarios arise. A natural extension of the elective smoothing approach to incorporate uncertainty is to minimize the worst-case expected maximum bed excess over the planning horizon as follows:

$$Z_{\mathrm{R}}(\mu) = \min_{\boldsymbol{\eta}\in X}\sup_{\mathbb{P}\in\mathbb{F}(\mu)}\mathbb{E}_{\mathbb{P}}\left(\max_{t\in\mathcal{T}^+}\left\{\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau+\tilde{p}_{\tau,l})-c_t\right\}\right). \tag{3}$$

In the absence of uncertainty, i.e., $\mu=0$, it is clear that Model (1) is the same as Model (3); hence, $Z_{\mathrm{D}}=Z_{\mathrm{R}}(0)$. As we increase the budget of variation $\mu$, the model takes into consideration more potential variations in the admission process. In this approach, the onus is on the modeler to set the budget of variations, $\mu$. We refer to this model as the robust model with fixed budget.

**Optimized Robust Model.** The main challenge of Model (3) is how to specify the value of $\mu$ that would yield the desired level of performance in controlling bed shortfalls. Intuitively, a meagerly or overly specified budget of variation, $\mu$ may not adequately protect against potential bed shortfalls when the actual uncertainty is realized. In practice, the parameter $\mu$ has to be tuned accordingly so that it gives the best overall performance on real data.

We note that Model (3) is only a means to cope with the issue of bed shortfalls. In a well-managed hospital, it is imperative that bed capacity should exceed average demands, which implies $Z_{\mathrm{D}}=Z_{\mathrm{R}}(0)\leqslant 0$. Extending this notion to incorporate uncertainty, if $Z_{\mathrm{R}}(\mu) \leqslant 0$, for $\mu > 0$, then we are guaranteed a solution that ensures that for all $\mathbb{P} \in \mathbb{F}(\mu)$, the expected maximum bed excess across the time periods is less than zero, i.e.,

$$\mathbb{E}_{\mathbb{P}}\Big(\max_{t\in\mathcal{T}^+}\Big\{\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau+\tilde{p}_{\tau,l})-c_t\Big\}\Big)\leqslant 0,\quad \forall\,\mathbb{P}\in\mathbb{F}(\mu).$$

In light of the above discussion, we propose another robust optimization approach, i.e., to find the most reliable solution that would protect against the worst uncertainty that might lead to bed shortfalls. In other words, we hope to maximize the level of uncertainty that the system can absorb without going into bed shortages. Hence, we push the boundary of uncertainty by maximizing the budget of variation, $\mu$ subject to $Z_{\mathrm{R}}(\mu)\leqslant 0$ as follows:

$$\begin{aligned}\mu^* = \max\ & \mu\\ \text{s.t.}\ & Z_{\mathrm{R}}(\mu)\leqslant 0,\\ & \mu\in[0,\infty).\end{aligned} \tag{4}$$

Since the set $\mathbb{F}(\mu)$ is nondecreasing in $\mu$, the function $Z_{\mathrm{R}}(\mu)$ is also nondecreasing in $\mu$. As a result, Model (4) is feasible and finite if and only if $Z_{\mathrm{R}}(0)\leqslant 0$ and $Z_{\mathrm{R}}(\infty)\geqslant 0$. Moreover, it is reasonable to assume that the inequalities are strict so that the bed capacity is sufficient to meet average demands but also not overly excessive. As opposed to the robust model, we refer to this model as the optimized robust model.

We note that the target value zero on the right-hand side of the constraint can further be adjusted accordingly to match the service level desired by the hospital. For simplicity, we leave it at zero.

The optimal solution of Model (4) can easily be obtained by binary search and solving a sequence of subproblems in the form of Model (3) so that $Z_{\mathrm{R}}(\mu^*)=0$.

## 3. Tractable Formulation

In this section, we first study the inner maximization problem of Model (3) and formulate it as a second order cone programming problem. Subsequently, we develop a tractable formulation of Model (3) in the form of a deterministic SOCP. Since the problem is easy to solve when $\mu=0$, we will focus on the case for which $\mu>0$. We first focus on the inner maximization problem in Model (3), i.e.,

$$\begin{aligned}\sup\ & \mathbb{E}_{\mathbb{P}}\Big(\max_{t\in\mathcal{T}^+}\Big\{\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau+\tilde{p}_{\tau,l})-c_t\Big\}\Big)\\ \text{s.t.}\ & \mathbb{E}_{\mathbb{P}}(\tilde{p}_{\tau,l})=\bar{p}_{\tau,l},\quad \forall(\tau,l)\in\mathcal{J},\\ & \mathbb{E}_{\mathbb{P}}(\tilde{p}_{\tau,l}^2)\leqslant\bar{p}_{\tau,l}^2(1+\mu^2),\quad\forall(\tau,l)\in\mathcal{J},\\ & \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{\tau,l})=\bar{\alpha}_{\tau,l},\quad\forall(\tau,l)\in\mathcal{J},\\ & \mathbb{E}_{\mathbb{P}}(\tilde{\alpha}_{\tau,l}^2)\leqslant\bar{\alpha}_{\tau,l}^2(1+\mu^2),\quad\forall(\tau,l)\in\mathcal{J},\\ & \mathbb{P}\big\{(\tilde{p}_{\tau,l},\tilde{\alpha}_{\tau,l})_{(\tau,l)\in\mathcal{J}}\in W_p\times W_a\big\}=1,\end{aligned} \tag{5}$$

where

$$W_p:=\big\{(p_{\tau,l})_{(\tau,l)\in\mathcal{J}}:\ p_\tau^0\geqslant p_{\tau,l}\geqslant p_{\tau,l'}\geqslant 0,\ \forall(\tau,l),\,(\tau,l')\in\mathcal{J},\ l'>l\big\},$$

$$W_a:=\big\{(\alpha_{\tau,l})_{(\tau,l)\in\mathcal{J}}:\ \alpha_\tau^0\geqslant \alpha_{\tau,l}\geqslant \alpha_{\tau,l'}\geqslant 0,\ \forall(\tau,l),\,(\tau,l')\in\mathcal{J},\ l'>l\big\}.$$

Note also that $W_p$ and $W_a$ are actually the cross products of a number of sets with respect to parameters $\tau\in\mathcal{T}$. In other words, $W_p$ and $W_a$ can be rewritten as

$$W_p=\Pi_{\tau\in\mathcal{T}}W_p^\tau,\qquad W_a=\Pi_{\tau\in\mathcal{T}}W_a^\tau,$$

where

$$W_p^\tau:=\big\{(p_{\tau,l})_{l\in\mathcal{L}_\tau}:\ p_\tau^0\geqslant p_{\tau,l}\geqslant p_{\tau,l'}\geqslant 0,\forall l,l'\in\mathcal{L}_\tau,\ l'>l\big\},\quad \tau\in\mathcal{T},$$

$$W_a^\tau:=\big\{(\alpha_{\tau,l})_{l\in\mathcal{L}_\tau}:\ \alpha_\tau^0\geqslant \alpha_{\tau,l}\geqslant \alpha_{\tau,l'}\geqslant 0,\forall l,l'\in\mathcal{L}_\tau,\ l'>l\big\},\quad \tau\in\mathcal{T}.$$

Problem (5) is a maximization problem over a probability distribution function, which is generally an intractable optimization problem; see, for instance, Murty and Kabadi (1987). However, under our model of uncertainty, we will show an equivalent formulation of Problem (5), namely its dual problem, is a minimization problem in the form of SOCP. As a result, Problem (5) can be readily solved by existing commercialized SOCP solvers, such as CPLEX and MOSEK.

For convenience in description, for each $t\in\mathcal{T}^+$, let $z_{\tau,l}^t$ denote the indicator function defined by

$$z_{\tau,l}^t=\begin{cases}1, & \text{if } \tau+l=t+1,\\ 0, & \text{otherwise},\end{cases}$$

for any $(\tau,l)\in\mathcal{J}$. Noticing that $\mathcal{U}_t=\{(\tau,l)\in\mathcal{J}:\ \tau+l=t+1\}$, item $\sum_{(\tau,l)\in\mathcal{U}_t}(\tilde{\alpha}_{\tau,l}\eta_\tau+\tilde{p}_{\tau,l})-c_t$ in the objective of (5) can then be expressed as below:

$$\sum_{(\tau,l)\in\mathcal{J}}(\tilde{\alpha}_{\tau,l}\eta_\tau+\tilde{p}_{\tau,l})z_{\tau,l}^t-c_t,\quad\forall\,t\in\mathcal{T}^+.$$

By applying duality theory, we derive an equivalent formulation of Problem (5) as follows:

Theorem 1. *Problem* (5) *has the same optimal objective value as the following optimization problem*:

$$\begin{aligned}\inf_{\rho,(s_{\tau,l},u_{\tau,l},v_{\tau,l},w_{\tau,l})_{(\tau,l)\in\mathcal{J}}}\ & \Big\{\rho+\sum_{(\tau,l)\in\mathcal{J}}\bar{p}_{\tau,l}s_{\tau,l}+\sum_{(\tau,l)\in\mathcal{J}}\bar{p}_{\tau,l}^2(1+\mu^2)u_{\tau,l}+\sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}_{\tau,l}v_{\tau,l}\\ &\quad+\sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}_{\tau,l}^2(1+\mu^2)w_{\tau,l}\Big\}\\ \text{s.t.}\ & \sum_{\tau\in\mathcal{T}}\pi_1^t(\mathbf{s}_\tau-\mathbf{z}_\tau^t,\mathbf{u}_\tau)+\sum_{\tau\in\mathcal{T}}\pi_2^t(\mathbf{v}_\tau-\eta_\tau\mathbf{z}_\tau^t,\mathbf{w}_\tau)\\ &\quad+\rho+c_t\geqslant 0,\quad\forall\,t\in\mathcal{T}^+,\\ & u_{\tau,l},\ w_{\tau,l}\geqslant 0,\quad\forall(\tau,l)\in\mathcal{J},\end{aligned} \tag{6}$$

*where for* $t \in \mathcal{T}^+$, $\tau \in \mathcal{T}$,

$$\pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau) := \min\left\{ \sum_{l \in \mathcal{L}_\tau} ((s_{\tau,l} - z_{\tau,l}^t) p_{\tau,l} + u_{\tau,l} p_{\tau,l}^2) \,\Big|\, (p_{\tau,l})_{l \in \mathcal{L}_\tau} \in W_p^\tau \right\},$$

$$\pi_2^t(\mathbf{v}_\tau - \eta_\tau \mathbf{z}_\tau^t, \mathbf{w}_\tau) := \min\left\{ \sum_{l \in \mathcal{L}_\tau} ((v_{\tau,l} - \eta_\tau z_{\tau,l}^t) \alpha_{\tau,l} + w_{\tau,l} \alpha_{\tau,l}^2) \,\Big|\, (\alpha_{\tau,l})_{l \in \mathcal{L}_\tau} \in W_a^\tau \right\},$$

*and* $\rho \in \Re$, $\mathbf{s}_\tau = (s_{\tau,l})_{l \in \mathcal{L}_\tau}$, $\mathbf{u}_\tau = (u_{\tau,l})_{l \in \mathcal{L}_\tau}$, $\mathbf{v}_\tau = (v_{\tau,l})_{l \in \mathcal{L}_\tau}$, $\mathbf{w}_\tau = (w_{\tau,l})_{l \in \mathcal{L}_\tau}$, $\mathbf{z}_\tau^t = (z_{\tau,l}^t)_{l \in \mathcal{L}_\tau}$.

Proof. The dual problem of Problem (5) can be written as

$$\begin{aligned}
\inf_{\rho, (s_{\tau,l}, u_{\tau,l}, v_{\tau,l}, w_{\tau,l})_{(\tau,l) \in \mathcal{J}}} & \left\{ \rho + \sum_{(\tau,l) \in \mathcal{J}} \bar{p}_{\tau,l} s_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} \bar{p}_{\tau,l}^2 (1+\mu^2) u_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} \bar{\alpha}_{\tau,l} v_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} \bar{\alpha}_{\tau,l}^2 (1+\mu^2) w_{\tau,l} \right\} \\
\text{s.t. } & \rho + \sum_{(\tau,l) \in \mathcal{J}} s_{\tau,l} p_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} u_{\tau,l} p_{\tau,l}^2 + \sum_{(\tau,l) \in \mathcal{J}} v_{\tau,l} \alpha_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} w_{\tau,l} \alpha_{\tau,l}^2 \\
& \geqslant \sum_{(\tau,l) \in \mathcal{U}_t} (\alpha_{\tau,l} \eta_\tau + p_{\tau,l}) - c_t, \quad \forall\, t \in \mathcal{T}^+, \\
& (p_{\tau,l}, \alpha_{\tau,l})_{(\tau,l) \in \mathcal{J}} \in W_p \times W_a, \\
& u_{\tau,l}, w_{\tau,l} \geqslant 0, \quad \forall\, (\tau,l) \in \mathcal{J},
\end{aligned} \tag{7}$$

where $s_{\tau,l}, u_{\tau,l}, v_{\tau,l}, w_{\tau,l}$ and $\rho$ are the Lagrange multipliers corresponding to the equality/inequality constraints concerning the first and second moments of $\tilde{p}_{\tau,l}$ and $\tilde{\alpha}_{\tau,l}$, $(\tau,l) \in \mathcal{J}$, together with the implicit constraint that $\mathbb{E}_{\mathbb{P}}[1] = 1$. Evidently, the multipliers, $u_{\tau,l}$ and $w_{\tau,l}$, $(\tau,l) \in \mathcal{J}$, corresponding to the inequality constraints are all nonnegative. This property, as we shall see, is very important in the subsequent analysis. Note that since the parameters of the ambiguity set are obtained empirically, there exist probability distributions that are feasible; hence, strong duality holds according to Shapiro (2001). Moreover, since $\mu > 0$, a distribution, $\mathbb{P}$ for which

$$\mathbb{P}\{(\tilde{p}_{\tau,l}, \tilde{\alpha}_{\tau,l})_{(\tau,l) \in \mathcal{J}} = (\bar{p}_{\tau,l}, \bar{\alpha}_{\tau,l})_{(\tau,l) \in \mathcal{J}}\} = 1$$

would lead to expectation constraints that are strictly feasible.

Note that the system of inequality constraints in Problem (7) consists of infinitely many constraints. Using the notation of vector $z_{\tau,l}^t$, we can express the inequality system in the dual problem (7) as

$$\begin{aligned}
& \sum_{(\tau,l) \in \mathcal{J}} ((s_{\tau,l} - z_{\tau,l}^t) p_{\tau,l} + u_{\tau,l} p_{\tau,l}^2) + \sum_{(\tau,l) \in \mathcal{J}} ((v_{\tau,l} - \eta_\tau z_{\tau,l}^t) \alpha_{\tau,l} + w_{\tau,l} \alpha_{\tau,l}^2) \geqslant -\rho - c_t, \\
& \qquad \forall\, t \in \mathcal{T}^+, \ (p_{\tau,l}, \alpha_{\tau,l})_{(\tau,l) \in \mathcal{J}} \in W_p \times W_a,
\end{aligned}$$

or equivalently,

$$\begin{aligned}
& \min\left\{ \sum_{(\tau,l) \in \mathcal{J}} (s_{\tau,l} - z_{\tau,l}^t) p_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} u_{\tau,l} p_{\tau,l}^2 \,\Big|\, (p_{\tau,l})_{(\tau,l) \in \mathcal{J}} \in W_p \right\} \\
& + \min\left\{ \sum_{(\tau,l) \in \mathcal{J}} (v_{\tau,l} - \eta_\tau z_{\tau,l}^t) \alpha_{\tau,l} + \sum_{(\tau,l) \in \mathcal{J}} w_{\tau,l} \alpha_{\tau,l}^2 \,\Big|\, (\alpha_{\tau,l})_{(\tau,l) \in \mathcal{J}} \in W_a \right\} \geqslant -\rho - c_t, \\
& \qquad \forall\, t \in \mathcal{T}^+.
\end{aligned}$$

Note that the objectives in the above system are separable in $(p_{\tau,l})_{l \in \mathcal{L}_\tau}$, $(\alpha_{\tau,l})_{l \in \mathcal{L}_\tau}$ for $\tau \in \mathcal{T}$, respectively. Noticing that $W_p$ and $W_a$ can be written as the cross products of some sets with respect to the parameter $\tau \in \mathcal{T}$, thereby the "min" and "sum" operators on the left-hand side are exchangeable. By recalling the definitions of $W_p^\tau, W_a^\tau$, we have

$$\sum_{\tau \in \mathcal{T}} \pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau) + \sum_{\tau \in \mathcal{T}} \pi_2^t(\mathbf{v}_\tau - \eta_\tau \mathbf{z}_\tau^t, \mathbf{w}_\tau) \geqslant -\rho - c_t, \quad \forall\, t \in \mathcal{T}^+.$$

Thus, the desired result follows immediately. This completes the proof. □

Note that the equivalent formulation (6) in Theorem 1 is a deterministic counterpart of the objective function, $Z_{\mathrm{R}}(\mu)$, of robust optimization Model (3). To derive a tractable reformulation, in what follows, we investigate the underlying minimization problems in the constraints of (6), i.e., $\pi_i^t$, $i = 1, 2$, $t \in \mathcal{T}^+$. First, for any $\tau \in \mathcal{T}$, define an index set $\mathcal{L}_\tau^+ := \mathcal{L}_\tau \cup \{1 + \min\{L, T - \tau\}\}$. We state the results as follows.

Proposition 1. *Given* $\gamma \in \Re$. *For* $t \in \mathcal{T}^+$, *the following statements hold true.*

(i) *For any* $\tau \in \mathcal{T}^{--}$, *the system of inequality*

$$\pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau) \geqslant \gamma \tag{8}$$

*is second order cone representable in the sense that there exist* $\lambda_{\tau,l}^t \geqslant 0$, $l \in \mathcal{L}_\tau^+$, *such that* (8) *is equivalent to*

$$\begin{aligned}
& \sum_{l \in \mathcal{L}_\tau} y_{\tau,l}^t + p_\tau^0 \lambda_{\tau,1-\tau}^t + \gamma \leqslant 0, \\
& 4 u_{\tau,l} y_{\tau,l}^t \geqslant (s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t)^2, \quad \forall\, l \in \mathcal{L}_\tau, \\
& y_{\tau,l}^t \geqslant 0, \quad \forall\, l \in \mathcal{L}_\tau.
\end{aligned} \tag{9}$$

(ii) *For any* $\tau \in \mathcal{T}^+$, *the system of inequality* (8) *is second order cone representable in the sense that there exist* $\lambda_{\tau,l}^t \geqslant 0$, $l \in \mathcal{L}_\tau^+$, *such that* (8) *is equivalent to*

$$\begin{aligned}
& \sum_{l \in \mathcal{L}_\tau} y_{\tau,l}^t + p_\tau^0 \lambda_{\tau,1}^t + \gamma \leqslant 0, \\
& 4 u_{\tau,l} y_{\tau,l}^t \geqslant (s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t)^2, \quad \forall\, l \in \mathcal{L}_\tau, \\
& y_{\tau,l}^t \geqslant 0, \quad \forall\, l \in \mathcal{L}_\tau.
\end{aligned} \tag{10}$$

(iii) *For any* $\tau \in \mathcal{T}^{--}$, *the system of inequality*

$$\pi_2^t(\mathbf{v}_\tau - \eta_\tau \mathbf{z}_\tau^t, \mathbf{w}_\tau) \geqslant \gamma \tag{11}$$

*is second order cone representable in the sense that there exist* $\lambda_{\tau,l}^t \geqslant 0$, $l \in \mathcal{L}_\tau^+$, *such that*

$$\begin{aligned} &\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + \alpha_\tau^0 \lambda_{\tau,1-\tau}^t + \gamma \leqslant 0,\\ &4w_{\tau,l} y_{\tau,l}^t \geqslant (v_{\tau,l} - \eta_\tau z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t)^2, \quad \forall l \in \mathcal{L}_\tau,\\ &y_{\tau,l}^t \geqslant 0, \quad \forall l \in \mathcal{L}_\tau. \end{aligned} \tag{12}$$

(iv) *For any* $\tau \in \mathcal{T}^+$, *the system of inequality* (11) *is second order cone representable in the sense that there exist* $\lambda_{\tau,l}^t \geqslant 0$, $l \in \mathcal{L}_\tau^+$, *such that*

$$\begin{aligned} &\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + \alpha_\tau^0 \lambda_{\tau,1}^t + \gamma \leqslant 0,\\ &4w_{\tau,l} y_{\tau,l}^t \geqslant (v_{\tau,l} - \eta_\tau z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t)^2, \quad \forall l \in \mathcal{L}_\tau,\\ &y_{\tau,l}^t \geqslant 0, \quad \forall l \in \mathcal{L}_\tau. \end{aligned} \tag{13}$$

Proof. (i). By definition, problem $\pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau)$ can be written as

$$\begin{aligned} &\pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau)\\ &= \min \sum_{l\in\mathcal{L}_\tau} (u_{\tau,l} p_{\tau,l}^2 + (s_{\tau,l} - z_{\tau,l}^t) p_{\tau,l})\\ &\quad \text{s.t.} \quad p_\tau^0 \geqslant p_{\tau,l} \geqslant p_{\tau,l'} \geqslant 0, \quad \forall l, l' \in \mathcal{L}_\tau,\ l' > l. \end{aligned} \tag{14}$$

Note that the above problem is a quadratic programming problem in which the coefficients concerning the second degree are nonnegative as $u_{\tau,l} \geqslant 0$ for $l \in \mathcal{L}_\tau$ by Theorem 1. To solve this problem, we consider its dual as given below:

$$\max_{\boldsymbol{\lambda}_\tau^t \geqslant 0} \zeta(\boldsymbol{\lambda}_\tau^t), \tag{15}$$

where $\zeta(\boldsymbol{\lambda}_\tau^t)$ is the associated Lagrange dual function, $\boldsymbol{\lambda}_\tau^t \in \Re^{|\mathcal{L}_\tau|+1}$ denotes the vector of the corresponding Lagrange multipliers, and for any given set $S$, $|S|$ denotes the cardinality of $S$.

Let $\mathbf{p}_\tau = (p_{\tau,l})_{l\in\mathcal{L}_\tau}$. For convenience in description and without loss of generality, we assume the indices of the entries in vector $\boldsymbol{\lambda}_\tau^t$ are consistent with those of $\mathbf{p}_\tau$, i.e., $\boldsymbol{\lambda}_\tau^t = (\lambda_{\tau,l}^t)_{l\in\mathcal{L}_\tau^+}$. Applying some basic operations, it gives the Lagrange dual function as follows:

$$\begin{aligned} &\zeta(\boldsymbol{\lambda}_\tau^t)\\ &:= \min_{(p_{\tau,l})_{l\in\mathcal{L}_\tau}} \Bigg\{ \sum_{l\in\mathcal{L}_\tau} \big(u_{\tau,l} p_{\tau,l}^2 + (s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t) p_{\tau,l}\big)\\ &\qquad\qquad - p_\tau^0 \lambda_{\tau,1-\tau}^t \Bigg\}. \end{aligned}$$

Note that the Slater's condition holds true since the interior of the feasible region of Problem (14) is nonempty. By the strong duality theorem, the system of inequality (8) can then be written as what follows. There exist $\lambda_{\tau,l}^t \geqslant 0$, $l \in \mathcal{L}_\tau^+$, such that

$$\begin{aligned} &\min_{(p_{\tau,l})_{l\in\mathcal{L}_\tau}} \Bigg\{ \sum_{l\in\mathcal{L}_\tau} \big(u_{\tau,l} p_{\tau,l}^2 + (s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t) p_{\tau,l}\big) \Bigg\}\\ &\qquad - p_\tau^0 \lambda_{\tau,1-\tau}^t \geqslant \gamma, \end{aligned}$$

which, by virtue of the separability of the above minimization problem in $p_{\tau,l}$, can be further reformulated as

$$\begin{aligned} &\sum_{l\in\mathcal{L}_\tau} \Big( \min_{p_{\tau,l}} \big\{ u_{\tau,l} p_{\tau,l}^2 + (s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t) p_{\tau,l} \big\} \Big)\\ &\qquad - p_\tau^0 \lambda_{\tau,1-\tau}^t \geqslant \gamma. \end{aligned} \tag{16}$$

To investigate the quadratic programming problems on the left-hand side of (16), we consider the following two cases: (a) $u_{\tau,l} > 0$ for all $l \in \mathcal{L}_\tau$; (b) $u_{\tau,l} = 0$ for some $l \in \mathcal{L}_\tau$, respectively.

For case (a), solving the optimality condition of each minimization problem involved, i.e., $2u_{\tau,l} p_{\tau,l} + s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t = 0$, $l \in \mathcal{L}_\tau$, we immediately derive the optimal solution and the corresponding optimal value, which are denoted by $p_{\tau,l}^*$ and $f_{\tau,l}^*$ as follows:

$$\begin{aligned} &p_{\tau,l}^* = \frac{1}{2u_{\tau,l}} \big(z_{\tau,l}^t - s_{\tau,l} + \lambda_{\tau,l+1}^t - \lambda_{\tau,l}^t\big), \quad l \in \mathcal{L}_\tau,\\ &f_{\tau,l}^* = -\frac{1}{4u_{\tau,l}} \big(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\big)^2, \quad l \in \mathcal{L}_\tau. \end{aligned}$$

Substituting the optimal value $f_{\tau,l}^*$ to the inequality (16), it yields that

$$\sum_{l\in\mathcal{L}_\tau} \frac{1}{4u_{\tau,l}} \big(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\big)^2 + p_\tau^0 \lambda_{\tau,1-\tau}^t \leqslant -\gamma. \tag{17}$$

To derive a second-order cone representation, we introduce the additional variables $y_{\tau,l}^t$, $l \in \mathcal{L}_\tau$, $t \in \mathcal{T}^+$ such that

$$\frac{1}{4u_{\tau,l}} \big(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\big)^2 \leqslant y_{\tau,l}^t, \quad l \in \mathcal{L}_\tau.$$

Thereby, system (17) is equivalent to

$$\begin{aligned} &\sum_{l\in\mathcal{L}_\tau} y_{\tau,l}^t + p_\tau^0 \lambda_{\tau,1-\tau}^t \leqslant -\gamma,\\ &4u_{\tau,l} y_{\tau,l}^t \geqslant \big(s_{\tau,l} - z_{\tau,l}^t + \lambda_{\tau,l}^t - \lambda_{\tau,l+1}^t\big)^2, \quad \forall l \in \mathcal{L}_\tau,\\ &y_{\tau,l}^t \geqslant 0, \quad \forall l \in \mathcal{L}_\tau, \end{aligned} \tag{18}$$

which is a second-order cone representation as desired.

For case (b), the analysis is similar to case (a) but becomes much simpler, as the underlying problem reduces to a linear programming problem in this case. Noticing that $\pi_1^t(\mathbf{s}_\tau - \mathbf{z}_\tau^t, \mathbf{u}_\tau)$ is lower bounded by a constant $\gamma$, we then have $s_{\tau,l} - z^t_{\tau,l} + \lambda^t_{\tau,l} - \lambda^t_{\tau,l+1} = 0$ and $p^*_{\tau,l} = 0$. Thereby, system (18) is valid as well.

(ii)–(vi). The arguments for these cases are similar to case (i). For brevity, here we omit the details. This completes the proof. □

Using Theorem 1 and Proposition 1, we are ready to derive the following result concerning the tractability of robust optimization Model (3), which is a main result of this paper.

Theorem 2. *Robust optimization Model* (3) *is equivalent to the following SOCP*

$$\begin{aligned}
\inf_{\rho,\boldsymbol{\eta},(\mathbf{S}_\tau,\mathbf{U}_\tau,\mathbf{V}_\tau,\mathbf{W}_\tau)_{\tau\in\mathcal{T}},(\boldsymbol{\lambda}^p_\tau,\boldsymbol{\lambda}^a_\tau,\mathbf{y}^p_\tau,\mathbf{y}^a_\tau)_{\tau\in\mathcal{T}}} &\Bigg\{\rho+\sum_{(\tau,l)\in\mathcal{J}}\bar{p}_{\tau,l}s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}}\bar{p}^2_{\tau,l}(1+\mu^2)u_{\tau,l}+\sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}_{\tau,l}v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}^2_{\tau,l}(1+\mu^2)w_{\tau,l}\Bigg\}\\
\text{s.t. } &\sum_{(\tau,l)\in\mathcal{J}} y^{t,p}_{\tau,l}+\sum_{\tau\in\mathcal{T}^{--}} p^0_\tau\lambda^{t,p}_{\tau,1-\tau}+\sum_{\tau\in\mathcal{T}^+} p^0_\tau\lambda^{t,p}_{\tau,1}+\sum_{(\tau,l)\in\mathcal{J}} y^{t,a}_{\tau,l}\\
&\quad+\sum_{\tau\in\mathcal{T}^+}\alpha^0_\tau\lambda^{t,a}_{\tau,1}+\sum_{\tau\in\mathcal{T}^{--}}\alpha^0_\tau\lambda^{t,a}_{\tau,1-\tau}\leqslant\rho+c_t,\quad\forall t\in\mathcal{T}^+,\\
&4u_{\tau,l}y^{t,p}_{\tau,l}\geqslant(s_{\tau,l}-z^t_{\tau,l}+\lambda^{t,p}_{\tau,l}-\lambda^{t,p}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&4w_{\tau,l}y^{t,a}_{\tau,l}\geqslant(v_{\tau,l}-\eta_\tau z^t_{\tau,l}+\lambda^{t,a}_{\tau,l}-\lambda^{t,a}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&\lambda^{t,p}_{\tau,l},\lambda^{t,a}_{\tau,l}\geqslant0,\quad\forall t\in\mathcal{T}^+,\tau\in\mathcal{T},l\in\mathcal{L}^+_\tau,\\
&y^{t,p}_{\tau,l},y^{t,a}_{\tau,l}\geqslant0,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&u_{\tau,l},w_{\tau,l}\geqslant0,\quad\forall(\tau,l)\in\mathcal{J},\\
&\boldsymbol{\eta}\in X.
\end{aligned}\tag{19}$$

Proof. First, we rewrite the system of inequality constraints of Problem (6) as

$$\sum_{\tau\in\mathcal{T}}\pi^t_1(\mathbf{s}_\tau-\ \mathbf{z}^t_\tau,\mathbf{u}_\tau)+\sum_{\tau\in\mathcal{T}}\pi^t_2(\mathbf{v}_\tau-\eta_\tau\mathbf{z}^t_\tau,\mathbf{w}_\tau)\geqslant-\rho-c_t,\quad\forall t\in\mathcal{T}^+.\tag{20}$$

Then according to Proposition 1 and applying some necessary operations, for each $t\in\mathcal{T}^+$, there exist some Lagrange multipliers $\lambda^{t,p}_{\tau,l}$ and $\lambda^{t,a}_{\tau,l}$, $l\in\mathcal{L}^+_\tau$, $\tau\in\mathcal{T}$, such that (20) is equivalent to

$$\begin{aligned}
&\sum_{\tau\in\mathcal{T}^+}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,p}_{\tau,l}+p^0_\tau\lambda^{t,p}_{\tau,1}\Bigg)+\sum_{\tau\in\mathcal{T}^{--}}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,p}_{\tau,l}+p^0_\tau\lambda^{t,p}_{\tau,1-\tau}\Bigg)\\
&\quad+\sum_{\tau\in\mathcal{T}^+}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,a}_{\tau,l}+\alpha^0_\tau\lambda^{t,a}_{\tau,1}\Bigg)\\
&\quad+\sum_{\tau\in\mathcal{T}^{--}}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,a}_{\tau,l}+\alpha^0_\tau\lambda^{t,a}_{\tau,1-\tau}\Bigg)\leqslant\rho+c_t,\quad\forall t\in\mathcal{T}^+,\\
&4u_{\tau,l}y^{t,p}_{\tau,l}\geqslant(s_{\tau,l}-z^t_{\tau,l}+\lambda^{t,p}_{\tau,l}-\lambda^{t,p}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,\ (\tau,l)\in\mathcal{J},\\
&4w_{\tau,l}y^{t,a}_{\tau,l}\geqslant(v_{\tau,l}-\eta_\tau z^t_{\tau,l}+\lambda^{t,a}_{\tau,l}-\lambda^{t,a}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,\ (\tau,l)\in\mathcal{J},\\
&y^{t,p}_{\tau,l},y^{t,a}_{\tau,l}\geqslant0,\quad\forall t\in\mathcal{T}^+,\ (\tau,l)\in\mathcal{J}.
\end{aligned}$$

On the other hand, according to Theorem 1, Model (3) is actually a "min-min" two-stage problem. Thus, Model (3) is equivalent to the following problem:

$$\begin{aligned}
\inf_{\substack{\rho,\boldsymbol{\eta},(\mathbf{S}_\tau,\mathbf{U}_\tau,\mathbf{V}_\tau,\mathbf{W}_\tau)_{\tau\in\mathcal{T}},\\(\boldsymbol{\lambda}^p_\tau,\boldsymbol{\lambda}^a_\tau,\mathbf{y}^p_\tau,\mathbf{y}^a_\tau)_{\tau\in\mathcal{T}}}} &\Bigg\{\rho+\sum_{(\tau,l)\in\mathcal{J}}\bar{p}_{\tau,l}s_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}}\bar{p}^2_{\tau,l}(1+\mu^2)u_{\tau,l}+\sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}_{\tau,l}v_{\tau,l} + \sum_{(\tau,l)\in\mathcal{J}}\bar{\alpha}^2_{\tau,l}(1+\mu^2)w_{\tau,l}\Bigg\}\\
\text{s.t. } &\sum_{\tau\in\mathcal{T}^+}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,p}_{\tau,l}+p^0_\tau\lambda^{t,p}_{\tau,1}\Bigg)+\sum_{\tau\in\mathcal{T}^{--}}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,p}_{\tau,l}+p^0_\tau\lambda^{t,p}_{\tau,1-\tau}\Bigg)\\
&\quad+\sum_{\tau\in\mathcal{T}^+}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,a}_{\tau,l}+\alpha^0_\tau\lambda^{t,a}_{\tau,1}\Bigg)+\sum_{\tau\in\mathcal{T}^{--}}\Bigg(\sum_{l\in\mathcal{L}_\tau}y^{t,a}_{\tau,l}+\alpha^0_\tau\lambda^{t,a}_{\tau,1-\tau}\Bigg)\\
&\quad\leqslant\rho+c_t,\quad\forall t\in\mathcal{T}^+,\\
&4u_{\tau,l}y^{t,p}_{\tau,l}\geqslant(s_{\tau,l}-z^t_{\tau,l}+\lambda^{t,p}_{\tau,l}-\lambda^{t,p}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&4w_{\tau,l}y^{t,a}_{\tau,l}\geqslant(v_{\tau,l}-\eta_\tau z^t_{\tau,l}+\lambda^{t,a}_{\tau,l}-\lambda^{t,a}_{\tau,l+1})^2,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&\lambda^{t,p}_{\tau,l},\lambda^{t,a}_{\tau,l}\geqslant0,\quad\forall t\in\mathcal{T}^+,\tau\in\mathcal{T},l\in\mathcal{L}^+_\tau,\\
&y^{t,p}_{\tau,l},y^{t,a}_{\tau,l}\geqslant0,\quad\forall t\in\mathcal{T}^+,(\tau,l)\in\mathcal{J},\\
&u_{\tau,l},w_{\tau,l}\geqslant0,\quad\forall(\tau,l)\in\mathcal{J},\\
&\boldsymbol{\eta}\in X.
\end{aligned}\tag{21}$$

This completes the proof. □

Our ability to solve the model and deploy the solution in practice critically depends on the model's computational tractability and efficiency. According to Theorem 2, we obtain a SOCP reformulation of Model (3). Since the feasible set $X$ is integral, the problem becomes a SOCP problem with integrality constraints, which can be solved by state-of-the-art commercial solvers such as CPLEX. Note that Assumption 1 allows us to obtain a tractable formulation of the problem. However, if we have more elaborate models, such as the standard deviation of $\tilde{a}_{t,l}$ being a function of $\sqrt{\eta_t}$, then it would lead to a nonlinear, nonconvex optimization problem, which we do not know how to solve to optimality.

## 4. Numerical Studies

We have access to one year of data (366 days) for the purpose of evaluating and comparing the performance of various models. Our data set consists of daily admission and length of stay of both emergency and elective patients throughout the year of 2008. Emergency patients, averaging about 119 daily, account for about 82% of daily admissions. Their mean length of stay at 3.57 days exceeds that of elective patients by about one day.

Our investigation of seasonality in daily emergency admissions reveals volatility across the days rather than across the months. The admission pattern of emergency patients over a typical week is known with the highest admission rates on Mondays and Tuesdays and a relatively less busy period from Thursdays to Sundays. In addition, this pattern seems quite consistent over a year. Further, the patterns of elective admissions more or less mirror those appearing in the emergency admissions. Our descriptive data analysis demonstrated DoW pattern of the hospital appeared to be in accordance with the overall trend of public health systems (Teow et al. 2007, Teow and Tan 2008, Sun et al. 2012).

### Simulation Setup

To extend the numerical study beyond a year, we impute our current data to provide a longer period for performance analysis. Because of the weekly pattern of the emergency admissions, the sample admissions for a given day in a week is randomly chosen from the same day of the week within the data. In this way, we can obtain a sequence of stationary samples of the emergency and elective inpatients admissions processes.

The hospital is initialized with zero admitted patients and the numerical study commences on Day $T_0$, ($T_0 = 2003$, the day after 286 weeks) to allow some time for the state of the hospital to achieve some level of stationarity. We also obtain the empirical averages of parameters based on the samples from Day 1 through Day $T_0 - 1$. Specifically, given the weekly periodicity of the data, we obtain the empirical averages of the parameters $(\tilde{p}_{t,l}, \tilde{\alpha}_{t,l})$ based on the day of the week that $t$ falls on. To be consistent with previous time index, Day $T_0$ would be referenced as $t = 0$.

We adopt a rolling horizon approach on a weekly basis in our simulation study. Specifically, we solve the elective admission problem over a planning horizon of $T = 7k$ days, $k \in \{1, 2, 3\}$ and implement only the quotas obtained for the first *seven* days. The parameter, $L$ is fixed at 14. In each problem we solve, we restrict the daily quota within the range $[5, 80]$ and weekly quota to $\omega$ admissions. Explicitly, the feasible set of admissible quotas is as follows:

$$X = \left\{ \boldsymbol{\eta} \in \mathbb{Z}^T \colon \eta_t \in [5, 80],\ t = 0, \ldots, T-1; \right.$$
$$\left. \sum_{t=0}^{6} \eta_{(t+7i)} = \omega,\ i = 0, \ldots, T/7 - 1 \right\}.$$

**Table 1.** Configuration settings for simulation study.

| Configurations | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $c$ | 600 | 600 | 600 | 620 | 620 | 620 | 650 | 650 | 650 |
| $\omega$ | 203 | 203 | 203 | 245 | 245 | 245 | 301 | 301 | 301 |
| $T$ | 7 | 14 | 21 | 7 | 14 | 21 | 7 | 14 | 21 |

Subsequently, we evaluate the performance for the week based on the samples of the emergency and elective inpatients admissions processes we have generated. We repeat the computations and evaluations by moving to the next week, until the evaluation period is reached.

In our numerical study, we compare the solutions of four different strategies as follows:

1. *Uniform quota model*: Quota is distributed evenly to maintain a weekly total quota of $\omega$. Hence, $\omega$ is chosen to be a multiple of 7.
2. *Deterministic model*: Quota is obtained by solving Model (1) with $\mu = 0$.
3. *Robust model*: Quota is obtained by solving Model (3) with different budget of variations $\mu$.
4. *Optimized robust model*: Quota is obtained by solving Model (4).

In Table 1, we present different configurations of our numerical study, with each configuration differing in the parameters $c$ (i.e., $c_t = c$, $\forall\, t \in \mathcal{T}$), $\omega$, and $T$. Note that as we increase the capacity, $c$, we also increase the average weekly quota, $\omega$ so that the incidents of bed shortages would be apparent in the numerical study. This would better elucidate the performance of various approaches.[2] The evaluation period is fixed to 100,1 days (i.e., 143 weeks). Under these settings, we can obtain the solutions of our models within reasonable time. For the mixed integer model, it requires about 10 to 20 seconds to obtain the optimal solution on a 12-core 2.4 GHz Mac Pro computer using the CPLEX solver. By relaxing the integrality constraints, we are able to solve the problems within two to three seconds. Moreover, in most cases, the solutions after rounding to their nearest integers are identical to the optimal integer solutions.

### Numerical Results

In Table 2, we report the total bed shortages of different models under different configurations. We use bold text to highlight the model with the best performance. We note that the optimized robust model performs relatively well against other models. In particular, it has significant performance improvements over the uniform quota and deterministic models. Notably, as evident from Figure 2, which is based on the study of Configuration 1, the performance of the optimized robust model almost dominates that of the deterministic and uniform quota models in terms of the management of bed shortages at all periods. Although there are instances where the robust model with fixed budget may perform marginally better than the optimized robust model,

**Table 2.** Total bed shortages for different models under given configurations.

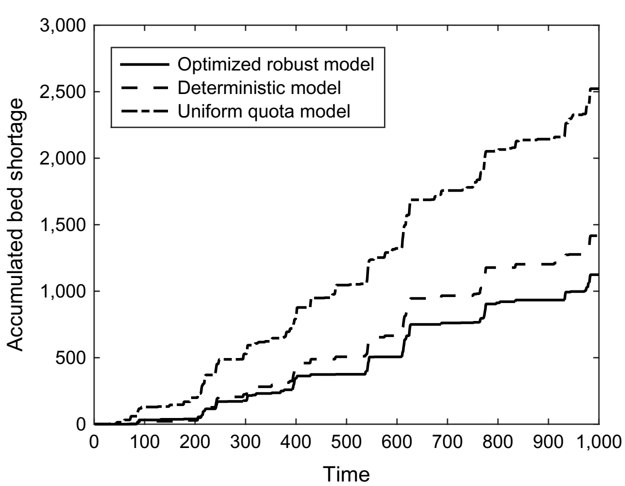
| Config. | Uni. quota model | Deterministic model | Opt. robust model | Robust model ($\mu =$) 0.01 | 0.02 | 0.05 | 0.1 |
|---|---|---|---|---|---|---|---|
| 1 | 2,438 | 1,418 | 1,148 | 1,072 | **1,068** | 1,395 | 2,212 |
| 2 | 2,438 | 1,429 | 1,115 | 1,175 | **1,088** | 2,091 | 2,583 |
| 3 | 2,438 | 1,562 | **1,085** | 1,182 | 1,132 | 2,389 | 2,438 |
| 4 | 2,969 | 1,778 | **1,441** | 1,469 | 1,509 | 1,799 | 1,975 |
| 5 | 2,969 | 1,796 | 1,470 | 1,536 | **1,435** | 2,335 | 2,679 |
| 6 | 2,969 | 2,019 | **1,426** | 1,454 | 1,489 | 2,473 | 2,464 |
| 7 | 2,012 | 1,235 | 982 | 1,025 | **911** | 1,108 | 2,109 |
| 8 | 2,012 | 1,230 | 942 | 994 | **917** | 1,178 | 1,522 |
| 9 | 2,012 | 1,379 | 962 | **924** | 987 | 1,326 | 1,377 |

we find that the performance of the former could be rather sensitive to the choice of $\mu$.

In Tables 3 and 4, we compare the performance of the optimized robust model against the deterministic and uniform quota models with higher bed capacities. We use bold text to emphasize the total bed shortage values that are closest to those obtained via the optimized robust model. We observe that the uniform quota model would require about 11 additional beds, whereas the deterministic model would require about four additional beds to achieve similar performance as the optimized robust model. Although the savings of beds may be a small fraction of the total bed capacity, it is a still noteworthy improvement since increasing the number of beds is usually not an imminent option.

Although there are enough beds in the hospital to meet average demands, because of the uncertain arrivals of patients and their durations of stay, bed shortfalls are inevitable in any hospital that is operating near its capacity. Whenever the bed shortage occurs, it could severely impact the safety of patients at the ED, because they would have to wait for an extended period of time to be admitted. Consequently, apart from reporting the average bed shortages, we also evaluate other performance metrics such as the fractions of the evaluation period with bed shortfalls and maximum number of shortages in a day. Tables 5 and 6 present the performance of the different models evaluated based on the maximum bed shortages in a day and the fraction of the evaluation period with bed shortfalls. The optimized robust model yields relatively good results in terms of having lower maximum bed shortages and fewer days with bed shortfalls. As in the previous study, we also observe that the performance of the robust model with fixed budget is sensitive to the choice of $\mu$. We summarize the performance improvement of the optimized robust model over the uniform quota and deterministic models in Table 7. In particular, in terms of overall bed shortages, the optimized robust model can offer over 24% improvement over the deterministic model and 53% improvement over the uniform quota model.

**Figure 2.** Accumulated bed shortages for the uniform quota, the deterministic, and the optimized robust models.




We observe in our computational study that as we increase the budget of variation $\mu$, the performance level of robust Model (3) initially improves but then deteriorates as $\mu$ increases further. Hence, this underscores the importance of specifying the appropriate value for the parameter $\mu$. We observe that the robust model for fixed budget $\mu = 0.02$ performs relatively well, and incidentally the optimal budgets in the optimized robust models are close to that value. Noting that poor choice of $\mu$ can lead to inferior solutions, the optimized robust model has an advantage of providing relatively good performance without the need of specifying the parameter $\mu$.

**Solution Patterns.** We show in Figures 3 and 4 the box plots for the optimal daily elective quota derived from both the deterministic and the optimized robust models. The solution patterns are similar under different configuration settings. We observe that the optimal elective quotas may not be higher on those days with lower emergency arrivals, which may be counterintuitive, yet nevertheless reasonable because of the complexity of the optimization models. It is interesting to note that the variations of the optimized robust model solutions are relatively modest compared to the solutions obtained by the deterministic model.

**Table 3.** Total bed shortage comparison between the optimized robust model and the uniform quota model with additional beds.

| | | Uniform quota model with additional beds | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Configuration | Optimized model | (+1) | (+2) | (+4) | (+6) | (+8) | (+10) | (+12) | (+14) |
| 1 | 1,148 | 2,291 | 2,149 | 1,881 | 1,643 | 1,439 | **1,260** | 1,107 | 972 |
| 2 | 1,115 | 2,291 | 2,149 | 1,881 | 1,643 | 1,439 | 1,260 | **1,107** | 972 |
| 3 | 1,085 | 2,291 | 2,149 | 1,881 | 1,643 | 1,439 | 1,260 | **1,107** | 972 |
| 4 | 1,441 | 2,798 | 2,635 | 2,334 | 2,059 | 1,812 | 1,595 | **1,408** | 1,243 |
| 5 | 1,470 | 2,798 | 2,635 | 2,334 | 2,059 | 1,812 | 1,595 | **1,408** | 1,243 |
| 6 | 1,426 | 2,798 | 2,635 | 2,334 | 2,059 | 1,812 | 1,595 | **1,408** | 1,243 |
| 7 | 982 | 1,890 | 1,788 | 1,584 | 1,470 | 1,245 | 1,098 | **965** | 846 |
| 8 | 942 | 1,890 | 1,788 | 1,584 | 1,470 | 1,245 | 1,098 | **965** | 846 |
| 9 | 962 | 1,890 | 1,788 | 1,584 | 1,470 | 1,245 | 1,098 | **965** | 846 |

**Table 4.** Total bed shortage comparison between the optimized robust model and the deterministic model with additional beds.

| | | Deterministic model with additional beds | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Configuration | Optimized model | (+1) | (+2) | (+3) | (+4) | (+5) | (+6) | (+7) | (+8) |
| 1 | 1,148 | 1,331 | 1,246 | **1,164** | 1,088 | 1,016 | 947 | 883 | 822 |
| 2 | 1,115 | 1,338 | 1,251 | 1,168 | **1,093** | 1,022 | 953 | 888 | 824 |
| 3 | 1,085 | 1,458 | 1,361 | 1,269 | 1,185 | **1,107** | 1,032 | 959 | 888 |
| 4 | 1,441 | 1,674 | 1,579 | 1,487 | **1,398** | 1,313 | 1,231 | 1,153 | 1,080 |
| 5 | 1,470 | 1,689 | 1,588 | **1,490** | 1,398 | 1,310 | 1,227 | 1,149 | 1,075 |
| 6 | 1,426 | 1,896 | 1,782 | 1,672 | 1,569 | **1,470** | 1,378 | 1,287 | 1,204 |
| 7 | 982 | 1,159 | 1,086 | 1,015 | **948** | 886 | 826 | 769 | 716 |
| 8 | 942 | 1,150 | 1,075 | 1,004 | **937** | 874 | 815 | 758 | 704 |
| 9 | 962 | 1,291 | 1,208 | 1,130 | 1,058 | **988** | 923 | 862 | 804 |

**Table 5.** Maximum bed shortages in a day for different models.

| | | | | Robust model ($\mu =$) | | | |
|---|---|---|---|---|---|---|---|
| Configuration | Uniform quota model | Deterministic model | Optimized robust model | 0.01 | 0.02 | 0.05 | 0.1 |
| 1 | 56 | 58 | **46** | 53 | 49 | 52 | 66 |
| 2 | 56 | 59 | **47** | 53 | 50 | 64 | 66 |
| 3 | 56 | 60 | 51 | 53 | **49** | 66 | 66 |
| 4 | 62 | 65 | **52** | 58 | 55 | 61 | 66 |
| 5 | 62 | 65 | **53** | 59 | 55 | 66 | 66 |
| 6 | 62 | 65 | 57 | 58 | **55** | 66 | 66 |
| 7 | 57 | 61 | **49** | 54 | 51 | 56 | 57 |
| 8 | 57 | 61 | **49** | 55 | 52 | 56 | 56 |
| 9 | 57 | 61 | 52 | 55 | **51** | 56 | 56 |

**Table 6.** Fraction of the evaluation period with bed shortfalls (in %) for different models.

| | | | | Robust model ($\mu =$) | | | |
|---|---|---|---|---|---|---|---|
| Configuration | Uniform quota model | Deterministic model | Optimized robust model | 0.01 | 0.02 | 0.05 | 0.1 |
| 1 | 15.1 | 8.9 | 8.3 | **7.7** | 8.0 | 9.6 | 12.8 |
| 2 | 15.1 | 9.5 | 8.4 | **7.8** | 8.3 | 12.9 | 15.1 |
| 3 | 15.1 | 10.8 | 8.1 | **7.9** | 8.8 | 13.6 | 14.1 |
| 4 | 17.4 | 10.9 | 9.9 | 9.5 | **9.0** | 10.9 | 16.8 |
| 5 | 17.4 | 11.1 | **9.6** | 9.8 | 9.6 | 14.5 | 15.9 |
| 6 | 17.4 | 12.9 | **9.8** | 9.8 | 9.8 | 15.0 | 15.0 |
| 7 | 12.0 | 7.8 | 7.1 | **6.4** | 6.5 | 7.3 | 11.8 |
| 8 | 12.0 | 8.1 | 6.6 | **6.4** | 6.7 | 8.1 | 9.9 |
| 9 | 12.0 | 9.0 | 7.1 | **6.3** | 6.8 | 8.7 | 9.0 |

**Table 7.** Performance improvement of the optimized robust model over the uniform quota and the deterministic models.

| | | Configurations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Performance (%) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Uniform quota model | Total shortage | 52.9 | 54.3 | 55.5 | 51.5 | 50.5 | 52.0 | 51.2 | 53.2 | 52.2 |
| | Max shortage | 17.9 | 16.1 | 8.9 | 16.1 | 14.5 | 8.1 | 14.0 | 14.0 | 8.8 |
| | Shortage days | 45.0 | 44.4 | 46.4 | 43.1 | 44.8 | 43.7 | 40.8 | 45.0 | 40.8 |
| Deterministic model | Total shortage | 19.0 | 22.0 | 30.5 | 19.0 | 18.2 | 29.4 | 20.5 | 23.4 | 30.2 |
| | Max shortage | 20.7 | 20.3 | 15.0 | 20.0 | 18.5 | 12.3 | 19.7 | 19.7 | 14.8 |
| | Shortage days | 6.7 | 11.6 | 25.0 | 9.2 | 13.5 | 24.0 | 9.0 | 18.5 | 21.1 |

**Figure 3.** Boxplot for daily quotas of the deterministic model under Configuration 1.

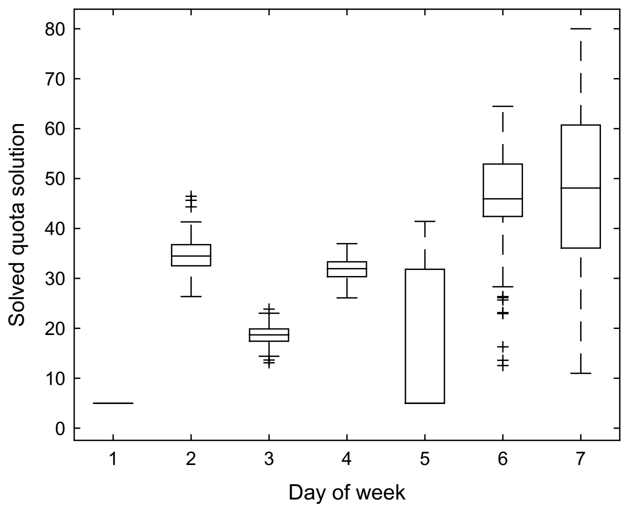
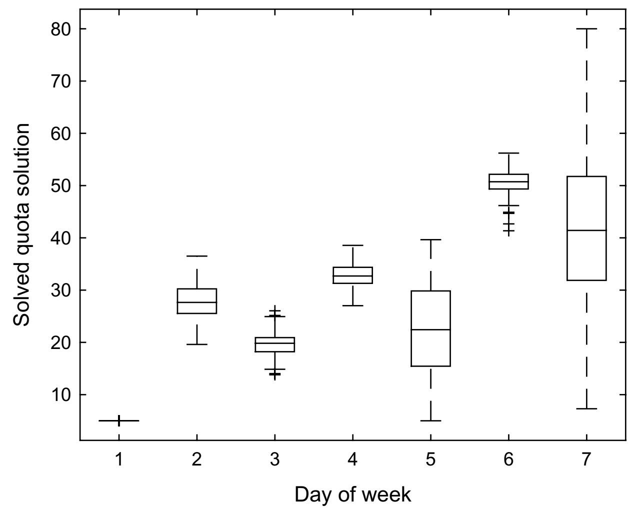



**Figure 4.** Boxplot for daily quotas of the optimized budget model under Configuration 1.




We observe that for fixed values of $(c, \omega)$, the duration of the planning horizon $T$ may affect the performance of the models. Interestingly, its impact on the performance of the robust and the optimized robust models is marginal. In contrast, extending the planning horizon may deteriorate the performance of the deterministic model. Similarly, in Figure 5, we observe that the solutions of the optimized robust model are not significantly influenced by the dura-

**Figure 5.** Daily elective quota of the optimized robust model under Configurations 1, 2, 3.

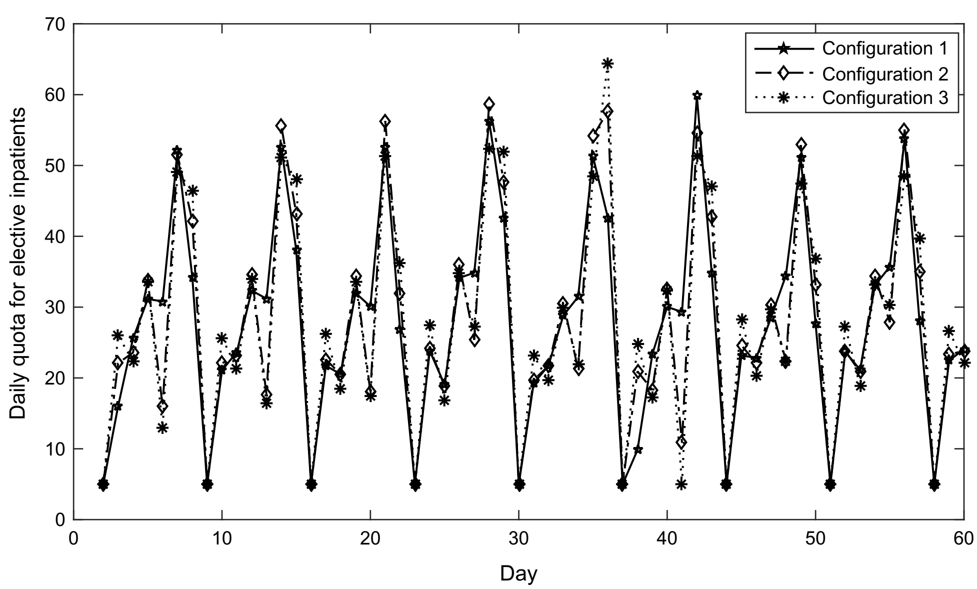
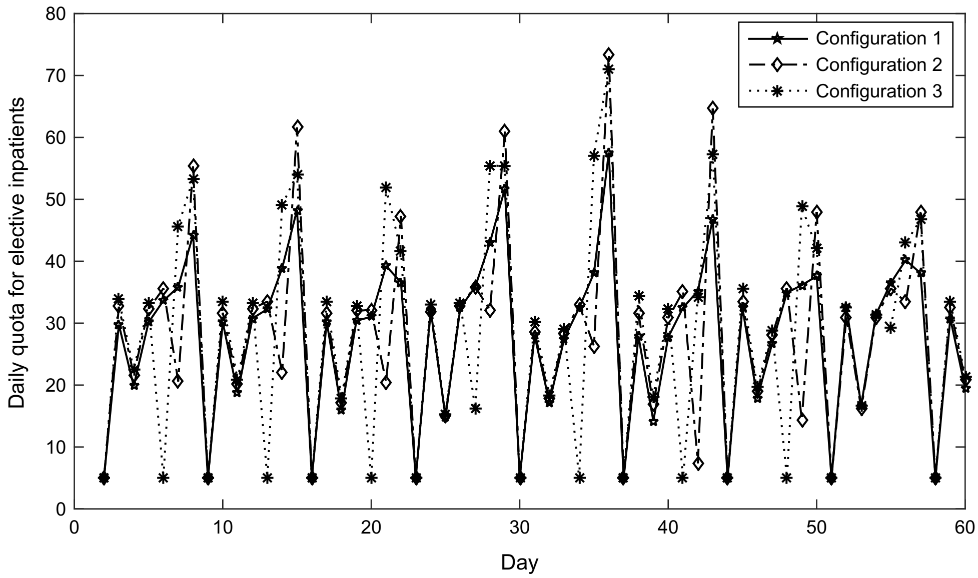


**Figure 6.** Daily elective quota of the deterministic model under Configurations 1, 2, 3.




tions of the planning horizon, $T \in \{7, 14, 21\}$ as compared to the case for the deterministic model reflected in Figure 6.

Though it is not reported in this study, we have also experimented with several other distributionally robust optimization models, such as incorporating variance estimates, confidence intervals of mean estimates. From the simulation study, the current model that we introduced provides consistently good performance without the need to have parameters' estimation beyond their first moments.

## Acknowledgments

The authors are grateful to Daizhuo Chen (Columbia University) and Palvannan R. K. (National Healthcare Group, Singapore) for their constructive inputs and discussions during the earlier stage of manuscript preparation. The authors thank Joe Sim (CEO, National University Hospital, Singapore) for the connections and support of this project and Kiok-Liang Teow (National Healthcare Group, Singapore) for his feedback and suggestions for this paper. The authors are also grateful to the associate editor and the anonymous referees for their careful readings and suggestions to improve the paper. This research work was supported by National University of Singapore—Global Asia Institute, Data Analytics Program. The work of the second author is also supported by Hong Kong RGC Project (T32-102/14N).

## Endnotes

1. We distinguish between risk and ambiguity. Risk deals with uncertainty with known distributions whereas ambiguity does not.
2. As a disclaimer, the parameters we use in the numerical study do not correspond to any hospital in Singapore.

## References

Ben-Tal A, Nemirovski A (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.

Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1): 35–53.

Chen W, Sim M, Sun J, Teo CP (2010) From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Oper. Res.* 58(2):470–485.

Chen X, Sim M, Sun P (2007) A robust optimization perspective on stochastic programming. *Oper. Res.* 55(6):1058–1107.

Cochran JK, Roche K (2008) A queuing-based decision support methodology to estimate hospital inpatient bed demand. *J. Oper. Res. Soc.* 59:1471–1482.

Conner-Spady BL, Sanmartin C, Johnston GH, McGurran JJ, Kehler M, Noseworthy TW (2011) The importance of patient expectations as a determinant of satisfaction with waiting times for hip and knee replacement surgery. *Health Policy* 101:245–252.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):596–612.

El Ghaoui L, Oustry F, Lebret H (1998) Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* 9:33–52.

Esogbue AO, Singh AJ (1976) A stochastic model for an optimal priority bed distribution problem in a hospital ward. *Oper. Res.* 24(5):884–898.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Oper. Res.* 58(4):902–917.

Goh J, Sim M (2011) Robust optimization made easy with ROME. *Oper. Res.* 59(4):973–985.

Gorunescu F, McClean SI, Millard PH (2004) Using a queueing model to help plan bed allocation in a department of geriatric medicine. *Health Care Management Sci.* 5:307–312.

Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA (2011) Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada. *British Medical J.* 342:d2983, http://dx.doi.org/10.1136/bmj.d2983.

Harper PR, Shahani AK (2002) Modelling for the planning and management of bed capacities in hospitals. *J. Oper. Res. Soc.* 53:11–18.

Hoot NR, Aronsky D (2008) Systematic review of emergency department crowding: Causes, effects, and solutions. *Ann. Emergency Medicine* 52:126–136.

Hwang U, Concato J (2004) Care in the emergency department: How crowded is overcrowded? *Acad. Emergency Medicine* 11:1097–1101.

Kao EPC, Tung GG (1981) Bed allocation in a public health care delivery system. *Management Sci.* 27(5):507–521.

Lamiri M, Xie X, Dolgui A, Grimaud F (2008) A stochastic model for operating room planning with elective and emergency demand for surgery. *Eur. J. Oper. Res.* 185:1026–1037.

Librero J, Marin M, Peiro S, Munujos AV (2004) Exploring the impact of complications on length of stay in major surgery diagnosis-related groups. *Internat. J. Quality in Health Care* 16:51–57.

Murty KG, Kabadi SN (1987) Some NP-complete problems in quadratic and nonlinear programming. *Math. Programming* 39:117–129.

Plunkett PR, Byrne DG, Breslin T, Bennett K, Silke B (2011) Increasing wait times predict increasing mortality for emergency medical admissions. *Eur. J. Emergency Medicine* 18:192–196.

Proudlove NC, Black S, Fletcher A (2007) OR and the challenge to improve the NHS: Modelling for insight and improvement in inpatients flow. *J. Oper. Res. Soc.* 58:145–158.

Shapiro A (2001) On duality theory of conic linear problems. *Semi-Infinite Programming: Recent Advances* (Kluwer Academic Publishers, Dordrecht, Netherlands), 135–165.

Sun Y, Teow KL, Heng BH, Ooi CK, Tay SY (2012) Real time prediction of waiting time in the emergency department, using quantile regression. *Ann. Emergency Medicine* 60:299–308.

Teow KL, Tan WS (2008) Allocation of hospital beds in an existing hospital. *J. Oper. Logist.* 2:2.

Teow KL, Heng BH, Chong Y, Sim J (2007) Using mathematical programming to optimize bed occupancy by smoothing elective against emergency admissions. Working paper, Department of Health Services and Outcomes Research, National Healthcare Group, Singapore.

Thompson DA, Yarnold PR (1995) Relating patient satisfaction to waiting time perceptions and expectations: The disconfirmation paradigm. *Acad. Emergency Medicine* 2:1057–1062.

Vassilacopoulos G (1985) A simulation model for bed allocation to hospital inpatient departments. *Simulation* 45:233–241.

Wardrope J, Driscoll P (2003) Turbulent times. *Emergency Medicine J.* 20:116.

---

**Fanwen Meng** is an operations research specialist at the Department of Health Services and Outcomes Research of National Healthcare Group, Singapore. His research interests include capacity planning, resource allocation/scheduling in hospital, medical decision making, robust optimization, logistics, and supply chain management.

**Jin Qi** is an assistant professor at the Department of Industrial Engineering and Logistic Management at Hong Kong University of Science and Technology. She is broadly interested in the analytics of services systems, especially, the healthcare system, with the goals of eliciting operational insights and providing solutions for supporting decision making in practice.

**Meilin Zhang** is a postdoctoral fellow in the Global Asia Institute at National University of Singapore. Her research interests include robust optimization, healthcare operations management, and big data analytics.

**James SK Ang** is an associate professor in the Department of Decision Sciences and is research director, of the Logistics Institute Asia Pacific, National University of Singapore. His current research interests are supply chain management, healthcare, and operations management.

**Singfat Chu** is an associate professor of business analytics at the NUS Business School. His current research focus is on the efficiency and equity aspects of the certificate of entitlement scheme, which manages vehicle ownership in Singapore.

**Melvyn Sim** is a professor in the Department of Decisions Sciences, NUS Business School. His research interests fall broadly under the categories of decision making and optimization under uncertainty with applications ranging from finance, supply chain management, and healthcare to engineered systems.