Data-driven optimization methodology for admission control in critical care units

Amirhossein Meisami1 · Jivan Deglise-Hawkinson2 · Mark E. Cowen3 · Mark P. Van Oyen1

Received: 8 December 2017 / Accepted: 15 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
The decision of whether to admit a patient to a critical care unit is a crucial operational problem that has significant influence on both hospital performance and patient outcomes. Hospitals currently lack a methodology to selectively admit patients to these units in a way that patient health risk metrics can be incorporated while considering the congestion that will occur. The hospital is modeled as a complex loss queueing network with a stochastic model of how long risk-stratified patients spend time in particular units and how they transition between units. A Mixed Integer Programming model approximates an optimal admission control policy for the network of units. While enforcing low levels of patient blocking, we optimize a monotonic dual-threshold admission policy. A hospital network including Intermediate Care Units (IMCs) and Intensive Care Units (ICUs) was considered for validation. The optimized model indicated a reduction in the risk levels required for admission, and weekly average admissions to ICUs and IMCs increased by 37% and 12%, respectively, with minimal blocking. Our methodology captures utilization and accessibility in a network model of care pathways while supporting the personalized allocation of scarce care resources to the neediest patients. The interesting benefits of admission thresholds that vary by day of week are studied.

Keywords Critical care units · Data-driven optimization · Dual-threshold admission policy · Mortality risk · Patient flow · Capacitated network

1 Introduction
Healthcare costs in the US are extremely high, with $759.1 billion annually in hospital costs alone [32]. Among critical care units/wards, Intensive Care Units (ICUs) are particularly expensive, with over $55 billion in estimated annual costs as of 2005 [14]. Although less expensive than ICUs, Intermediate Care units (IMCs) are also more costly than general/regular medical units. IMCs provide a level of care between intensive care and general units, and their inclusion in hospitals’ network flow has shown benefits for both patients and system utilization [2]. These units have less equipment and a higher ratio of patients to nurses; therefore, IMCs are less costly to operate than ICUs. Considering the steep financial burden imposed by critical care units and the importance of patient outcomes, it is essential to investigate methodologies to improve admission decisions in the higher level of care units/wards such as ICUs and IMCs. This research generates a powerful method and a novel perspective to optimize a personalized risk-based admission policy. We leverage a recently developed estimate of mortality risk metric as a surrogate measure of each patient’s true bed placement requirements. We integrate this metric with the hospital’s bed capacity using a network model of patient flow through units/wards and a data-informed methodology.
One important component of this problem is the management of a hospital units’ bed capacity in the face of congestion. Hospital census (the number of patients in the hospital at a given time) and unit/ward level census are both variable, with a weekly pattern of variation (see [16] as one example in addition to others cited therein). Under-utilization of capacity reduces potential hospital revenue, whereas over-utilization leads to patient blocking, decreased quality of care, and job dissatisfaction among nurses [1, 16]. By blocking, we mean either that a patient is diverted to another hospital because the desired hospital is not accepting patients due to limited resources, or that a patient is placed in a different unit because the desired unit is full. Both types of blocking are undesirable. Placing a patient in a suboptimal unit may compromise quality of care and can lead to worse patient outcomes (i.e., increased length of stay, mortality rate, and readmission rate [6]).

The number of ICU beds in the U.S. is growing, but this growth is slower than the growth of demand for ICUs [15]. Some predict a shortage of ICU beds [33]. As the population in the U.S. continues to age rapidly, more stress will be placed on our healthcare systems because older patients require more frequent and more intensive healthcare services [31, 33]. Compounding this effect, hospitals across the country are reducing the number of available beds. This is especially troubling when we consider that the effects of ICU congestion are felt throughout the hospital, as patients are frequently competing for ICU beds [20, 23]. This paper sheds light on how to efficiently allocate limited ICU and IMC resources.

ICU care is beneficial to patients [3, 29] and may decrease patient mortality rates during the first 24 hours of care, especially for those patients with high risk [29, 30]. ICU care is especially beneficial for critically ill patients; it has been shown to markedly improve outcomes for these critically ill patients [11]. Conversely, ICU care may not be beneficial for patients who are sufficiently well [11, 35].

Hence, some studies in the ICU admission literature include severity measures in the decision making processes [18, 27]. Such information may not be accessible until the patient is placed in the ICU, which limits its potential benefits. However, Cowen et al. have developed a mortality risk metric (MRM) [9] that is available prior to bed assignment. The MRM system has been operational in our partner hospital for a few years and we use this measure as one ICU admission criterion in our model, given the potential benefits of ICU care may be easier to demonstrate in patients with an elevated risk of dying [8]. Moreover, the MRM metric is not only associated with the risk of death, but also correlates significantly with the risk of other adverse health events [9]. The concept that drives implementation in our partner hospital is that MRM is an effective summary measure for a patient’s comorbidities and her overall health state that complements the severity of the patient’s chief diagnosis and acuity level for nursing care [10]. So, we are using MRM as a surrogate for a patient’s need for an ICU/IMC bed to determine an optimal admission policy. But our method is general and could incorporate other metrics such as the Rothman index [25].

This paper introduces a more complex model of patient placement into a network of ICU and IMC units. Many patients are placed in the higher level of care units for reasons of nursing acuity, that is, the increased nursing hours per patient required relative to the average patient, because of the complexities of the treatments administered and daily care needs. Because of the broader range of treatments available in ICU and IMC units, such patients require specific treatments available only in ICU or possibly IMC unit. We call patients who are either high acuity, or simply require a treatment only found in an ICU or IMC, “mandatory” patients (and we differentiate ICU mandatory from IMC mandatory to designate the unit that they need) [22]. In contrast to mandatory placements in ICUs and IMCs, we consider a third placement criterion, risk of dying, that may qualify a patient for a higher level of care. We call patients who are not mandatory, i.e., who do not require treatment specific to an ICU or IMC unit, “discretionary”.

This distinction allows us to investigate the opportunities to admit critical discretionary patients to ICUs or IMCs. This is expected to result in improved health outcomes and reduced mortality. In essence, we study the trade-off between blocking future mandatory patients and admitting high risk discretionary patients.

Let us consider some of the issues that motivate optimizing the admission control to ICUs and IMCs. Currently, patients are regularly placed in the ICU who could receive equally beneficial care in a non-ICU unit. For instance, most low risk patients never use specific ICU services when in the ICU (low risk patients are here defined as those with mortality risk < 0.06%) [24, 35]. This seems to be present in the data from our partner hospital. During an eighteen-month period the number of ICU admissions for low risk patients were 20% higher than those for high risk patients while a considerable number of high risk patients were still admitted to the non-ICU units. Some hospitals use Intermediate Care Units in addition to ICUs to provide an intermediate level of care. Although less than for ICUs, there is significant competition for IMC beds, which are much scarcer than regular/general beds. Given the limited resources available in the critical care units, this motivates the necessity of a methodological framework to admit patients to higher levels of care in a selective manner.

There is a need in the medical community for a good predictive model identifying which patients would and would not benefit from ICU or IMC care. Lacking good information, doctors ration (i.e., deny an ICU bed to some patients
to manage capacity) less accurately than they would given perfect information. The current practice of assigning ICU beds is often ‘subconscious rationing’ rather than a data driven policy [14]. Doctors make resource allocation decisions based on clinical judgment at the moment, which may or may not be cost effective or beneficial in the long run [31]. Additionally, some studies suggest that doctors are overconfident in their ability to judge ICU need [31].

Given this state of practice, an approach to ICU admissions that uses scientific tools and methods seems beneficial. It is worth noting that although our partner hospital has a “triage” coordinator overseeing admission to bed units, the hospital has still pursued the use of a personalized mortality risk metric as a key component of its admission decision process. With the advent of electronic medical records (EMRs), new possibilities are emerging for data driven, analytical models that use past data to inform future behavior.

The main thesis of our study is that we can approximately optimize the ICU and IMC patient admission policy to get the neediest patients into the ICU and IMC units. This is expected to improve patient outcomes by reducing patient complications and mortality as well as reducing unplanned transfers to the ICU. Reducing transfers is important because of the longer length of stays and higher mortality rate observed in transferred patients [12]. Unplanned transfers from a non-ICU bed to the ICU are especially harmful because in this case the patient needs immediate transfer or her condition is likely to deteriorate rapidly, and an appropriate bed cannot always be found sufficiently quickly [26]. In addition, considering stochastic arrivals of specific patient types and their historical paths in the hospital, we will be able to adjust the admission decision such that it delivers patient probabilistic needs while tracking units’ utilization to support hospital profitability.

Unlike the prior literature, we consider an integrated model of several ICUs and several IMCs, and we aggregate all the other “regular” bed units of the hospital in a unit we call General Medical Beds (GMB). This allows us to focus on the flow details of the higher level of care units while accounting for total bed capacity of the hospital and the fact that patients in the GMB units may at any time during their stay require elevation of care to an IMC or ICU bed.

Built upon approximations of the stochastic arrival and flow processes, this paper contributes a data-driven Mixed Integer Programming (MIP) optimization methodology in equilibrium that can generate optimal key system metrics. Different policies can be applied to enforce desired managerial perspectives. We will consider two admission strategies. The first strategy requires the model to generate constant admission thresholds for every day of the week, while the second strategy allows daily variation across the days of the week. We show that the latter will not only increase the number of admitted patients to critical units, but it will also reduce the blocking rates in the long run. We validate our results via extensive numerical simulations on synthetic and real data sets.

2 Literature review

Green showed that as many as 90% of hospitals lack capacity to accommodate high risk patients arriving into their ICUs [13]. To address this issue, there has been a rise in ICU management literature in recent years. These studies can be classified into two broad categories: 1) ICU capacity management and 2) ICU admission control. Chalfin et al. found that in case of full capacity, new ICU admission requests will be delayed and patients with critical conditions may be boarded in the emergency department [5]. This can endanger the patient’s life as delays in ICU admission can significantly raise the mortality rate [4, 5].

On the admission side, many researchers employed queueing theory and Markov decision process theory to study the admission control and elective admission policies for critical care units [28, 34]. These studies focus on a single patient type and/or a single resource, and disregard hospital network effect. Kim et. al. examined the ICU admission problem via econometrics [21]. They showed the importance of ICU admission policies and derive a threshold policy for admitting patients to the ICU. While Kim et al. studied a single unit, this paper models multiple patient classes and a network of units in the problem of ICU admission policies. In this paper we clarify that many patients require one or more specific treatments that are available only in an ICU of a specific type or in some cases a set of IMC units. Kim et al. assume that patients can go to any ICU, regardless of diagnosis, so they only consider one combined ICU. Our collaboration with hospital staff, indicated that, in our partner hospital, patients are rarely placed off-service (i.e., in the wrong type of ICU), and if they are, it is only for a few hours until a bed opens up in the desired ICU. In this paper, our optimization based model will examine the ICU and IMC admission problem when considering multiple ICUs and IMCs, where different units treat different types of patients.

Mathews and Long also study the ICU admission problem [22]. Their contribution builds on the work of KC and Terwiesch [20] by considering the impact of census on patient non-service wait time in the ICU, namely how long patients wait to be transferred after they are done with treatment, as a function of census. Although they consider the emergency department (ED) versus non-ED distinction, acute and non-acute conditions for each patient may not be easily interpretable and measurable in practice. We model the class of mandatory ICU or IMC placed patients who cannot obtain a required treatment in a GMB type bed. This
has a large impact on the accuracy of the results of the models. For instance, mandatory patients comprise almost 20% of the overall body of patients in our partner hospital.

Step Down Units (SDUs) are considered to be reliable alternatives for ICUs, and there has been a surge in studying the influence of these units on patient outcome [7] and hospital performance [2]. Long and Mathews consider both an ICU and an IMC for a single type platform. In contrast, we will examine the ICU admission problem with multiple ICUs and IMCs. This is important for two main reasons. First, a patient’s diagnosis often uniquely determines which ICU or IMC the patient should go to. It can be a patient safety issue if a patient is placed in the wrong ICU or IMC. Second, there are complex network flows between the various ICUs and IMCs, which affect overall system performance. We should clarify that IMCs are very similar to Progressive Care Units (PCUs) or Step Down Units. IMCs are nearly identical in character to SDUs, and the main difference between IMCs and PCUs is that PCUs have more flexible staffing, allowing a patient’s nursing ratio to shift as her care needs change. Whereas a patient would be transferred out of an IMC when her care needs lessen, in a similar situation she would remain in a PCU, just with a lower nurse:patient ratio. This distinction is not essential to this paper, and though we will talk about IMCs for the remainder of the paper, one can apply our results to PCUs or SDUs as staffing decisions are not our focus.

Recently, Helm, Van Oyen, and coauthors developed methodologies for optimal elective hospital admissions [16, 17]. Of these, [16] is by far the closest to this paper. We extend these methods to the critical care environment by introducing a novel individualized/personalized framework. The primary focus of elective admission models exploits the ability of elective patients to be delayed until a later day that would improve the patient flow. This can be understood as a change in the weekly scheduling of surgeries to achieve a weekly block-scheduling pattern that facilitates patient flow. Our paper takes patients as they arrive and selectively promotes the highest risk patients to a higher level of care unit or demotes patients of low risk to a lower level of care. This purpose requires a new definition of patient types, the approach to the modeling of patient flow pathways and the complex math programming formulation. We also develop an exact approach for computing the standard deviation of the census in our model. Interestingly, this paper develops a methodology to compute an optimal dual-threshold admission policy for IMCs and ICUs.

3 System design and modeling assumptions

We generate optimal admission policies through a binary matrix, \([\theta^k_{udi}]\), whose elements, \(\theta^k_{udi}\), are the risk threshold decision variables of our MIP model indicating whether we admit a patient of class \(k\) with risk level \(i\) to unit \(u\) on day \(d\). Once the binary matrix is constructed, we can then create a daily dual-threshold admission policy. Individual ICU and IMC admission policies will be easily accessible by tracking decision variables in the matrix. We check the monotonic admission-risk patterns and define the threshold for ICUs, IMCs and general beds as a global rule. To establish such well-defined thresholds, however, we need to discretize the mortality risk metric into several bins such that we have roughly equally frequent observations in each risk interval. For instance, if we consider \(r\) different risk levels, then the admission rule to unit \(u\) for patients of class \(k\) on day \(d\) is fully defined by the array \([\theta^k_{udi1}, \theta^k_{udi2}, ..., \theta^k_{udi(r-1)}, \theta^k_{uird}].\)

To further differentiate mandatory patients, we will call the group of patients who require treatment that is only available in the ICU, “ICU-mandatory”. These ICU-mandatory patients should receive priority for ICU admission whenever there is a conflict with a discretionary patient. Similarly, the “IMC-mandatory” patients should receive priority for IMC admission. Although we usually treat them separately and attach a desired unit for each patient, when convenient we will sometimes lump ICU-mandatory and IMC-mandatory patients as a single group of mandatory patients.

In our optimal admission policy for critical care units, we consider an infinite horizon and cyclo-stationary model with a seven-day week as the period. Thus, in equilibrium, the same policy is applied every week, but it may vary by day of week to accommodate day of week effects. This model matches the weekly schedule of clinics, operating rooms, etc.

We work on a dataset that spans two years, with more than 200,000 data points representing over 70,000 distinct patients at our partner hospital. The dataset contains the information regarding patients’ mortality risk, specialty, arrival and discharge, and the units they visited during their stay by day and time. The hospital network consists of three different ICUs as well as five distinct IMCs.

For the sake of intuition, before developing the more realistic model, we employ a simple queueing model with a single patient class to provide insight into how the admission to ICU’s based on a mortality risk threshold reveals the average sum of MRM’s as a function of the threshold and the number of beds. This simple single-class \(M/M/s/s\) model with rejection when an arrival finds the ICU full with \(s\) patients assumes Poisson arrivals and exponentially distributed length of stay (LOS) in the ICU. We will revisit these assumptions in Section 3.1.

Our scoring metric is the time-average sum of MRMs over patients in an ICU bed. We will discuss this objective function in more detail below and extend it to consider IMC beds also in the network model. The idea is that
the patient mortality outcomes are proportional to the mortality risk, which justifies maximizing the average mortality in these beds that best care for the sickest patients within the bed capacity. The relationship between mortality risk and poor health outcomes is not known precisely; however, the mortality risk is associated with the risk of unplanned transfers, cardiopulmonary arrests, intra-hospital complications and 30 day readmissions [9]. Therefore, we believe this objective is appropriate within the current limits of knowledge. Figure 1 indicates that for a given demand model, as the ICU capacity increases the optimal MRM threshold decreases. It also shows that as we reduce the MRM threshold below the optimal level, the average score declines, which means that initial admission of lower risk patients impeded the entrance of more critical patients. Similarly, raising the MRM threshold above the optimal level diminishes the overall score of the system, which reflects the underutilization of ICUs.

### 3.1 From queueing to optimization

The above queueing model gives good intuition about the dynamics of the system, and it effectively communicates the concept to the hospital staff. On the other hand, the model is limited by the over-simplification of the problem. Therefore, in order to gain a more accurate and general framework, as well as to get richer results, we create a Mixed Integer Programming (MIP) model that allows us to build a data driven model and to capture network effects, including the interplay between the ICUs, IMCs, and the rest of the hospital. For instance, if a patient is admitted to a specific ICU on a given day, there is a chance (which depends on the patient’s characteristics) that she will be in a specific IMC after a given number of days. See Fig. 2 for an illustration of possible patients’ movements in the hospital network, which allows for multiple visits to a station (feedback), and extends from the time of admission until discharge from the hospital. We will discuss in the next section how this stochastic patient stay pattern is realized in our modeling framework.

Employing an MIP model, we can find an optimal dual-threshold-type policy. The dual-threshold policy is reasonable and probably optimal if one accepts the preferences of the clinical leadership. They specifically desire to announce the MRM risk threshold for ICU admission and the lower one for IMC admission. A dynamic, state-dependent policy should perform better; however, a costly and complex decision support system would be required. A patient with an MRM score at or above the ICU-threshold, is approved for ICU admission, and will then be considered for IMC admission if there is no room in the ICU. If patient does not meet this threshold but meets the IMC-threshold, the patient is approved for IMC admission. A patient who does not meet either of these thresholds is sent to a non-ICU/IMC unit, which is modeled using the GMB unit to avoid unnecessary complications in the model. Our partner hospital is already using MRM thresholds to guide ICU and IMC admission decisions, but the thresholds were not optimized prior to this work.

We note that an MIP framework enables us to vary the thresholds by day of week and by unit. Another advantage of the MIP modeling is that we can introduce multiple classes of patients, which is crucial for our problem as we are dealing with mandatory and discretionary patients. This methodology, therefore, lets us use a threshold policy for discretionary patients while automatically approving all mandatory patients for ICU or IMC admission. We can also approximate the queueing dynamics.

---

**Fig. 1** Left: Score as a function of the MRM threshold used, when the ICU has 25 beds. The arrow indicates the optimal threshold. Right: Score as a function of MRM threshold, for 5 (bottom curve), 10, 15, 20, 25, 30, 35, 40, and 46 (top curve) beds in the ICU
3.2 Mixed integer program formulation

Our goal is to compute an admission policy that will admit the neediest discretionary patients into the ICUs and IMCs. To do this, we propose the following: mandatory patients (i.e., those who require treatment specific to the ICU or IMC) are always approved for admission to the relevant unit. To be specific, ICU-mandatory patients are always approved for admission to the appropriate ICU and similarly for IMC-mandatory patients. All other patients, who we are calling discretionary, may be approved for admission to an ICU or IMC based on their respective MRM scores. We will apply the dual threshold-type policy on these discretionary patients with separate thresholds for ICU admission and for IMC admission. The intuition here is that high risk patients would benefit most from ICU care, but would also benefit from IMC care, and that less risky patients would benefit from IMC care without requiring the extra level of care given by ICU admittance.

We explained briefly in the introduction why our partner hospital believes the degree of a patient’s neediness can be approximately quantified by her MRM risk score and our objective is to maximize the sum of risk scores over patients admitted to the ICUs and IMCs. Based on discussions with the medical staff at our partner hospital, we found that ICU and IMC admissions are roughly equally beneficial for discretionary patients. Thus, our objective function is to maximize

$$\sum_{d \in D} \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I} \setminus (\mathcal{M} \cup \mathcal{M}')} \sum_{u \in \mathcal{U} \setminus \{GMB\}} \theta^k_{udi} R_i,$$

where $\mathcal{I}$ is the set of risk groups (obtained by partitioning patients by MRM) and $\mathcal{M}$ and $\mathcal{M}'$ define the IMC and ICU mandatory patients, respectively. Also, $D$ is the set of days of the week (Sunday, Monday, etc.), $\mathcal{U}$ is the set of units/wards, and $\mathcal{K}$ is the set of patient classes. In this context, a patient’s class is determined by her treatment needs and diagnosis; we would use this to determine to which units a patient may be admitted to. For instance, a cardiac class patient would likely go to the cardiac ICU instead of the surgical ICU (SICU) if she required ICU services. Our decision variable, $\theta^k_{udi}$, is a binary variable that is equal to 1 if we allow patients of class $k$ and risk group $i$ to be admitted to unit $u$ on day $d$, and equal to 0 otherwise. We let $\Theta$ refer to the matrix $[\theta^k_{udi}]$. Though we do not have an explicit threshold variable, the thresholds can be inferred from $\Theta$. $R_i$ is the expected risk score for a patient in risk group $i$. Our objective is simple and fair, admitting as many of the patients as possible while ensuring that the overall risk is maximized.
stratified high-risk groups of patients as capacity will allow and admitting higher risk patients over lower risk ones.

To link the operational constraints of the problem to our objective function, we have to characterize the stochastic paths of different patient classes who are admitted to ICUs, IMCs, and the regular beds of GMB. Incorporating binary decision variables, we can define a new version of the stochastic location process introduced by Helm et al. [16]. In that work the same patient population was admitted to the same unit, but the day of week of admission could be changed to improve the patient flow. Emergency patients were not changed, but elective patients could be delayed until a day of admission that would improved the flow of the hospital. That paper also considered a version that increased the hospital throughput by selectively admitting more patient of any type desired subject to the capacity limits of the hospital.

In our setting, the patient of a given type cannot wait to be admitted. Moreover, we intentionally selectively change the first unit of admission based on his/her personalized risk metric. Some discretionary patients will be elevated to a higher level unit (and a distribution on the initial unit of admission must be used to capture detailed patient type information that is not explicitly modeled). Other discretionary patients with a low risk will be demoted to a lower level of care. Including a patient’s mortality risk conveys a broad range of information related to a patient’s medical history, lab results, and demographics [9]. This will improve the model’s performance in replicating patient flow in the most appropriate units of the network.

We must construct a model of how long patients will spend in any given unit and their likelihood of moving to another unit. We view this as a stochastic location process over time, which is indexed to the weekday of admission and the patient’s class. This can be defined as a stochastic model that captures length of stay (in days) in each unit and the units that patients visit during their hospital stay. Additionally, we assume that the stochastic itineraries of different patient classes are independent throughout the hospital, which fits with our offered load approach to the census distribution. Also, to derive the stochastic location of patients, we build the model based on the less congested periods in our dataset to preserve the independent itinerary assumption (Fig. 3).

Without any constraints, it is easy to see that the optimal admission policy would prescribe $d_{kui} = 1$ for all $k, u, d$ and $i$, meaning that everyone is approved for ICU and IMC admission. This corresponds to the intuition that, given infinite resources, it would be beneficial to make the entire hospital into an ICU. No hospital can afford that expense; however, so we need to constrain the system by the available capacity while maintaining a certain service level. Of course, one can use this approach to quantify the higher level of care capacity needed to send all patients above a given risk level to an ICU or similar objective.

To formally define our stochastic location process, we will employ a modified version of the stochastic location process modeling approach developed in [16]. Let $P_{d'kui}(r)$ be the probability that a patient of class $k$ and risk group $i$ who arrived on day $d'$ is in unit $u$, $r$ days after admission, given that the patient was initially admitted to unit $u'$. This measure is evaluated on a weekly basis for $n$ weeks ($r = d - d' + 7n$). Also, let $\tilde{F}_{d'k}(a)$ be the probability that there are more than $a$ arrivals of class $k$ and risk group $i$ on day $d'$ and consider $n$ as the counter of the weeks.

To compute the congestion under any policy, given $\Theta$, we seek estimates of the mean and variance of the census for each unit on each day. As will be described below, using the mean and variance of a stochastic location process, we can approximate the units’ capacity and overall patient flow in the hospital. By keeping a sufficiently low probability that the census deviations above the mean go beyond the capacity of the unit, we can limit the blocking and thereby ensure that the new system will work within the hospital’s unit network capacity constraints. Small amounts of blocking are not a practical problem, because the hospital is very experienced at reacting to blocking. The following theorems show that the mean and the variance of the units’ daily census can be calculated linearly in $\Theta$.

**Theorem 1** The mean census in unit $u$ on day $d$ is given by

$$\mu_{du} = \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a=0}^{\infty} P_{d'kui}(r) \tilde{F}_{d'k}(a) 6^{u_{du}}.$$  

**Theorem 2** The variance of census in unit $u$ on day $d$ is given by

$$\nu_{du} = \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a=0}^{\infty} \left( P_{d'kui}(r) \tilde{F}_{d'k}(a) \right)^2 6^{u_{du}}.$$  

Theorems 1 and 2 state that the mean and variance of daily unit census can be calculated linearly in $\Theta$ for any arrival process. Clearly, these are data-driven estimations based on historical records for each risk stratum/group, which were available by limiting the number of risk strata in the model without assuming a specific arrival distribution. The proofs for these theorems are given in Appendix B.

We incorporate finite capacity by limiting blocking. By blocking, we mean that a patient cannot be admitted to the preferred unit because it is full. If we do not constrain blocking, our model would lead to a first-come-first-served policy, which would be unacceptable due to the blocking of many of the incoming mandatory patients. To enforce
a required service level for our system, we model the right tail of the census of each unit on each day of the week using its own normal/Gaussian distribution. Note that the mean and variance of the census are calculated by Theorems 1 and 2, respectively. The normality of workloads is common in healthcare workload analysis literature, and it has been shown to be a good model in similar settings [19]. Admittedly, allowing for non-normal distributions considerably complicates the model and often leads to nonlinearity of the optimization. Because stability requires that the capacity must exceed average demand for service, the key is to estimate the overload in each unit by examining the right tail of the census distribution, given a desired service level for our system, we model the right tail of the census of each unit on each day of the week is respected. The optimization will constrain the blocking probability \( (1 - \alpha) \) so that the desired expected number of blockings by unit and day of week is kept to a small fraction of arrival volume, but this can be modified and applied for different hospitals in accordance with their policies and requirements.

In the derivation of \( v_{du} \), the variance of the census, we find that it is linear in \( \Theta \) (Online Appendix B). This would make the standard deviation non-linear in \( \Theta \), as \( \sigma_{du} \) is the square root of \( v_{du} \), which in turn would make (2) nonlinear. To avoid this issue, we achieve linearity by first rearranging (2) as

\[
\sigma_{du} \leq C_{du} - \mu_{du} .
\]  

We then raise both sides of Eq. 3 to the second power to replace \( \sigma_{du} \) with \( v_{du} \):

\[
z_{\alpha}^{2}v_{du} \leq C_{du}^{2} + \mu_{du}^{2} - 2C_{du}\mu_{du} .
\]  

The equation above is equivalent to Eq. 3 only when both sides of Eq. 3 are non-negative. Hence, we enforce \( C_{du} - \mu_{du} \geq 0 \) as a new set of constraints in our model. To linearize this equation, we take advantage of the binary characteristics of our main decision variables, \( \theta_{udi} \). As defined in theorem 1, \( \mu \) is constructed via a linear combination of \( \theta_{udi} \). Therefore, given \( (\theta_{udi})^{2} = \theta_{udi} \), the square of mean census can be written as:

\[
C_{du}^{2} = \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k' \in K} \sum_{a = 0}^{n} \sum_{r = 0}^{\infty} \left( P_{d'k'i}(r)(\hat{P}_{d'ki}(a))\right)^{2} \theta_{udi}^{k} \]

\[
\quad + \sum_{i \in I} \sum_{d' \in D} \sum_{u'' \in U} \sum_{k'' \in K} \sum_{i' \in I} \sum_{d'' \in D} \sum_{u'' \in U} \sum_{k'' \in K} \theta_{udi}^{k} \theta_{udu}^{k'} \theta_{d'd'i'}^{k''} .
\]  

It can be seen that (5) includes multiplications of two binary decision variable, which results in nonlinearity. However, we can avoid this problem by introducing new binary variables and a new set of constraints. In fact, we can replace each of the nonlinear terms generated in Eq. 5 by one new binary variable and three new constraints. Let’s call our new constraints:\n
\[
\theta_{udi}^{k} \leq \theta_{udi}^{k} \leq \theta_{udi}^{k} \leq \theta_{udi}^{k} .
\]
binary variables $\omega_i$. Now, the linear equivalent of Eq. 5 can be defined by the following equations:

$$
\mu_{du}^2 = \sum_{i \in \mathcal{I}} \sum_{d' \in \mathcal{D}, u' \in \mathcal{U}, k \in \mathcal{K}_{n \neq 0}, d'' = 0} \sum_{i \in \mathcal{I}} \sum_{d' \in \mathcal{D}, u' \in \mathcal{U}, k \in \mathcal{K}_{n \neq 0}, d'' = 0} \left( P_{du}^{\prime\prime}(r) \cdot \bar{P}_{d'(d''(a))} \right)^2 \delta_{a di}^k
$$

$$
= \sum_{i \in \mathcal{I}} \sum_{d' \in \mathcal{D}, u' \in \mathcal{U}, k \in \mathcal{K}_{n \neq 0}, d'' = 0} \sum_{i \in \mathcal{I}} \sum_{d' \in \mathcal{D}, u' \in \mathcal{U}, k \in \mathcal{K}_{n \neq 0}, d'' = 0} \omega_{a u d i} \leq \theta_{a d i}^k
$$

$$
\omega_{a u d i} \leq \theta_{a d i}^k
$$

$$
\omega_{a u d i} \geq \theta_{a d i}^k + \theta_{a d i}^k - 1
$$

Hence, we are able to overcome the nonlinearity of the workload control constraint by using Eq. 6 in Eq. 4. It is worth mentioning that the nonlinearity issue imposed by the standard deviation of unit census in (3) could be alternatively resolved by applying numerical approximation. One can use Newton’s method to approximate $\sigma_{du}$ by

$$
\frac{1}{2} \left( \frac{\delta_{du}}{\sigma_{du}} + \sigma_{du} \right),
$$

where $\delta_{du}$ is the historical standard deviation of the census in unit $u$ on day $d$. Plugging it into our initial Eq. 3 gives:

$$
\mu_{du} + \frac{z}{2} \left( \frac{\delta_{du}}{\sigma_{du}} + \sigma_{du} \right) \leq C_{du}
$$

This approximation can avoid the use of extra binary decision variables and constraints to mitigate the complexity of the algorithm. We introduce the above exact methodology; however, the approximate formulation of the problem can also be found in the Appendix A.

In addition to the operational constraints discussed above there are other technical and problem-specific considerations that we need to observe. As mentioned earlier, we have to ensure that all ICU-mandatory and IMC-mandatory patients are approved for ICU and IMC admission, respectively. We do so by requiring:

$$
\sum_{u \in \mathcal{U}_{ICU}} \theta_{u d i}^k = 1,
$$

$$
\sum_{u \in \mathcal{U}_{IMC}} \theta_{u d i}^k = 1,
$$

where superscript $i = M$ denotes that a patient is ICU-mandatory, $i = M'$ denotes that a patient is IMC-mandatory, $\mathcal{U}_{ICU}$ is the set of ICUs ($\mathcal{U}_{ICU} \subset \mathcal{U}$), and $\mathcal{U}_{IMC}$ is the set of IMCs ($\mathcal{U}_{IMC} \subset \mathcal{U}$).

We also ensure that each patient stream is initially sent to only one unit. We can partition the patient streams in such a way that splitting patient streams between units would give no advantage. It is worth noting that we are aggregating the main hospital (i.e., all GMB units) into one unit because, for our purposes, there is little value in modeling those flows in detail.

$$
\sum_{u \in \mathcal{U}} \theta_{u d i}^k = 1
$$

Our model must also respect limitations beyond units' capacities and the mandatory decision rules. Therefore, we constrain the number of admitted patients of a certain class to each unit on each day to be less than the maximum number allowed by the hospital:

$$
\lambda_{di} \leq \beta_{ud i}^k
$$

where $\lambda_{di}$ is the mean arrival rate of patients of class $k$ and risk group $i$ on day $d$, and $\beta_{ud i}^k$ is the maximum allowed number of admissions for patients of a certain class with a certain risk level to each unit on each day. $\beta_{ud i}^k$ can be easily set to the maximum historical value from the data. This constraint also ensures that the initial admission unit is conforming to the patient requirements and the current hospital practice.

It is also important to note that we want to maintain threshold level consistency for all days and in all units such that on a given day no higher risk patient goes to a lower level of care while a lower risk patient is admitted to a higher level of care. To ensure this pattern, we have to include monotonicity constraints that enforce the required structure:

$$
\theta_{u d i}^k \leq \theta_{u d (i+1)}^k \quad \forall u \in \mathcal{U}_{ICU},
$$

$$
\theta_{u d i}^k \leq \theta_{u d (i+1)}^k \quad \forall u \in \mathcal{U}_{IMC}, u' \in \mathcal{U}_{ICU},
$$

$$
\theta_{u d i}^k \leq \theta_{u d (i+1)}^k \quad \forall u \in \mathcal{U}_{GMB}, u' \in \mathcal{U}_{IMC},
$$

where $\mathcal{U}_{GMB}$ refers to general medical beds. Please also note that the risk groups are set to be ordinal. Additional constraints can be added, if desired. For instance, one could enforce a constant threshold across units and/or days of the week.

4 Numerical results

In this section we validate and compare our results using almost two years of data from our partner hospital corresponding to over 70,000 distinct patients. We first validate the mean census process estimations for different units on different days of the week. This hospital has three ICUs and five IMCs, and we labeled all other units as general medical beds. We confirm that our estimation of the unit workload nearly equals the actual unit census throughout the hospital by day of week.

Table 1 compares the weekly mean census for the estimated model versus the historical observation. The results indicate that the errors of our estimates derived from Theorems 1 and 2 are relatively negligible. The results also match...
Table 1  Accuracy of the weekly estimated mean census in comparison with observed data

<table>
<thead>
<tr>
<th></th>
<th>ICUs</th>
<th>IMCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MICU</td>
<td>SICU</td>
</tr>
<tr>
<td>Estimated $\mu$</td>
<td>14.82</td>
<td>16.69</td>
</tr>
<tr>
<td>Observed $\mu$</td>
<td>14.55</td>
<td>16.37</td>
</tr>
<tr>
<td>% Deviation</td>
<td>1.82</td>
<td>1.91</td>
</tr>
</tbody>
</table>

closely with respect to different days of a week on a unit-by-unit basis. Figure 4 illustrates that our model performs very well in estimating both mean and standard deviation of the census for one of the ICUs in our network (MICU).

Having a reliable census process approximation, we can perform optimization to deliver operational measures such as admission thresholds given a desirable service level without depending on computationally expensive simulation models. We used the Gurobi optimization package through a Python interface to build and solve our MIP model. We applied our modeling methodology to patient data that spanned over 18 months to derive the optimal admission thresholds for different units under 97.5% service level. We consider two scenarios: 1) constant admission thresholds for different days of the week (OPT-C) and 2) dynamic admission thresholds varying by day of the week (OPT-D).

Since there is no historical record of hospital operational metrics under the optimal schedule, we compare the performance of our optimal admission policy both in constant and dynamic settings with the hospital’s current admission rule, using a high-fidelity simulation model. Our partner hospital currently follows an admission policy where patients with an MRM of 0.2 and higher are prioritized to be admitted to ICUs, and patients with an MRM higher than 0.07 are prioritized to be admitted to IMCs [9]. Although this is the guideline, in practice, some high risk patients are placed in GMB.

For the simulation modeling we follow a similar framework as discussed in [17]. However, we consider a more detailed unit transition probability metric for patients by conditioning on the arrival day and the arrival admission unit using $P_{d|k|u}(r)$ matrices as described in Section 3.2 and used in Theorems 1 and 2. Therefore, we can generate sample path of patients based on their arrival day and unit and build their stochastic trajectories until discharge. We use the historical data to estimate the transition probabilities as well as the distribution of arrivals, $\bar{F}_{d|k|}(a)$. We run the simulation for five years and report the results based on the last four years to account for a warm-up period of one year. We repeat the simulation 100 times and present the average results.

We validated our simulation model by considering the real-world hospital operations. We used almost two years of historical data to calibrate the model. To do so, we implemented the hospital’s current admission rules in our simulation model and captured the key features of the system such as the average unit’s census by day of week, standard deviation of the unit’s census by day of week, the average number of admissions for different units by day of week and the average length of stay per unit by patient class and mortality risk. We were able to see that the results of our simulation model nearly matches the actual hospital operations.
The results in Table 2 show that the optimal admission thresholds outperform the historical practice patterns. Under both constant and dynamic settings, our methodology improves the number of weekly admissions and weekly blockage for all ICU and IMC units. The OPT-C model increased the weekly average admissions to ICUs and IMCs by 11% and 7.5%, whereas the OPT-D model boosted that number by 37% and 12%, respectively. Also, the number of blocked patients in overall IMC and ICU units reduced by 33% and 34.5% in OPT-C and OPT-D models, respectively. The improvement in admission and blockage shows that our methodology successfully assigned the patients to appropriate units upon patient arrival. The advantage of admitting higher risk patients to the higher levels of a hospital is twofold. First, on average it is expected to reduce the number of unplanned transfers. The reason is that in general beds (GMB in our model), higher risk patients are more likely to get sicker and require unplanned transfers to a higher level of care unit where they can be treated more effectively. Second, it should on average decrease the overall length of stay. It may be that higher level of care units, which offer a greater intensity of nurse care, are more effective at speeding up the healing of patients; however, it is beyond our scope to rigorously test that hypothesis. Table 3 shows the average length of stay of high risk patients for different admission policies (in days).

<table>
<thead>
<tr>
<th>Table 3 Average length of stay of high risk patients for different admission policies (in days)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Average Length of Stay</td>
</tr>
</tbody>
</table>

It is worth noting that the dynamic threshold policy enables the hospital to admit more patients to ICUs and IMCs than do static thresholds. By allowing dynamic thresholds, a hospital can maintain the desired service level and further increase the weekly admission volume to the higher levels of the hospital. This happens because the constant thresholds must be set to accommodate the busiest day of the week, whereas the dynamic thresholds can manage the tradeoffs and patterns observed in the stay of arrivals on various days to gain new performance. Please note that weekly patterns set by the surgical block schedule cause such patients to exhibit weekly patterns of need for beds and for higher level of care services. Surgical specialties vary in the length of patient stay and in patient need for higher level of care beds. Therefore, employing a dynamic admission policy for different days of the weekly plan is able to not only raise the long-run expected number of admitted critical patients but also to better exploit operational asymmetries across the days of the week.

To show how admission thresholds change in different models we set the daily risk threshold for different levels of care in a conservative fashion. For instance, to assign the admission threshold for ICUs on Saturday, we pick the highest threshold among all the ICU units and patient classes on that day. Figure 5 illustrates the significance of the improvements that result from the dynamic threshold model. OPT-D reduced the threshold for ICU admission on every day of the week. The lower admission threshold for both ICU and IMC indicates the higher availability of critical units for more high risk patients in the long run without compromising the service level.

Figure 6 displays the average number of initial admissions by day of week for ICUs, IMCs and GMBs. Note that later transfers across units are not counted, although they are modeled. shows that OPT-D has selectively moved higher risk patients, who used to be admitted to general medical beds, to critical care units. Note that there is less of a difference on Tuesday, Wednesday, Thursday, and Friday. This is consistent with the fact that the units are the
most full midweek. Our methodology empowers high level hospital strategies by providing data-driven operational prescriptions. Our approach can guide care providers to better understand the value of personalized admission control and how to make trade-offs.

Figure 7 shows the Pareto curve generated by solving the OPT-D model for service levels ranging from 95 to 97.5 percent (for the resources and arrival distribution of our partner hospital, higher service levels will result in infeasibility). This curve defines a meaningful boundary between

Fig. 5 Top: Constant admission threshold based on mortality risk and day of week. Bottom: Dynamic admission threshold based on mortality risk and day of week.
utilization and accessibility that provides actionable insight for the hospital management system. Our optimization program delivered the operating curve automatically, with each point taking about 30 minutes to be generated. We should emphasize that our model can go far beyond what has been discussed in this section. For instance, our model will be very insightful in capacity planning for critical care units where care providers can see how the system behavior changes by increasing or decreasing resources in critical care units. In general, our methodology can be seen as a predictive platform for a patient’s future needs given her initial unit of admittance. Such a platform enables hospitals to test the robustness of their admission policies by easily generating and testing new arrival scenarios against their current admission rules.

5 Conclusion and future work

We introduced a novel perspective to address the admission policy for a network of critical care units. Although much work remains to be done to assess a patient’s health risk, this paper takes advantage of one of the early measures of patient mortality risk to show how patient risk metrics can be used to rationalize and optimize the admission of patients given the limited capacity of higher level of care in a hospital. We developed a methodology that generates optimal admission policies for different ICUs and IMCs in a hospital network. Considering the formidable cost of critical care beds, our model can guide hospital admission policies towards a data-informed solution that guarantees a higher service level as well as higher accessibility.

Our suggested approach helps hospitals improve the effectiveness of their care delivery by ensuring that the neediest patients are admitted to the ICUs and IMCs at all times. By using a mixed integer program, we determined different thresholds for each unit and each day of the week. This lets us exploit the network effects inherent in the hospital and allow the policy to incorporate the weekly census pattern cycle and the stochastic care pathways revealed in the historical data.

It is essential to note that our proposed methodology is not limited to our partner hospital. It can be used by any hospital with a real-time health risk metric and a suitable patient records database. We expect this approach to gain acceptance as hospitals deploy methods to assess patients’ risks. Our method will diminish wait times to admit high-risk patients to higher level of care (ICU and IMC) by selectively assigning a more appropriate initial admission unit. Therefore, it has the potential to decrease mortalities caused by delayed access or non-access to critical care units. By incorporating stochastic location random fields, the method is based on an accurate prediction of patient unit/ward needs throughout her stay in the hospital.

Although our model carefully incorporates a congestion network model of hospital unit capacity, the capacity violation outcome can be different for hospitals with non-normal workload distribution. However, the normality of hospital workload has been a common observation in the literature. It is also worth mentioning that our methodology is most successful when the majority of ICU and IMC beds are not occupied by mandatory patients so we have more room for the optimal assignment of discretionary patients to the right units.

For future work, one direction is to consider dynamically updated patient mortality/health risk assessments, provided such longitudinal data are available. Dynamic risk measures can provide significant information regarding the patient recovery process that is valuable for improved admission decisions.

Appendix A: Notations and MIP models

Notation: Variables

\( \theta_{udi}^k \) binary variable that is equal to 1 if we allow patients of class \( k \) and risk group \( i \) to be admitted to unit \( u \) on day \( d \), and equal to 0 otherwise (decision variable). \( \Theta \) refers to the matrix \([\theta_{udi}^k]\). 

\( \mu_{du} \) mean census in unit \( u \) on day \( d \), given \( \Theta \). 

\( v_{du} \) variance of the census in unit \( u \) on day \( d \), given \( \Theta \).

Notation: Parameters

\( C_{du} \) capacity of unit \( u \) on day \( d \).
The mathematical models are given below.

MIP Model
MIP model with the Newton Approximation

\[
\text{Maximize } \sum_{d \in D} \sum_{k \in K} \sum_{u \in \mathcal{U}} \sum_{\ell \in \ell(GMB)} \alpha_{\text{udi}}^{k} R_i \\
\text{S.t. } \mu_{du} = \frac{\gamma_u \left( V_{du} + \hat{\sigma}_{du} \right)}{\hat{\sigma}_{du}} \leq C_{du} \quad \forall d \in D, u \in \mathcal{U} \\
\mu_{du} = \sum_{i \in I} \sum_{d' \in D} \sum_{\ell' \in \ell(K)} \sum_{k' \in K} p_{d'k'}^{u_i} (d - d' + 7n) \\
\quad \cdot \hat{F}_{d'k'} (a)^{\alpha_{\text{udi}}^{k}} \\
\nu_{du} = \sum_{i \in I} \sum_{d' \in D} \sum_{\ell' \in \ell(K)} \sum_{k' \in K} \sum_{n=0}^{\infty} p_{d'k'}^{u_i} (d - d' + 7n) \hat{F}_{d'k'} (a) \\
\quad \cdot \nu_{d}^{u} (d, d' + 7n) \hat{F}_{d'k'} (a)^{\alpha_{\text{udi}}^{k}} \\
\sum_{u \in \mathcal{U}} \alpha_{\text{udi}}^{k} = 1 \\
\sum_{\ell \in \ell(GMB)} \alpha_{\text{udi}}^{k} = 1 \\
\sum_{\ell \in \ell (\text{U})} \alpha_{\text{udi}}^{k} = 1 \\
\sum_{\ell \in \ell (\text{U})} \alpha_{\text{udi}}^{k} = 1 \\
\alpha_{\text{udi}}^{k} \leq \alpha_{\text{udi}}^{k} (t+1) \\
\alpha_{\text{udi}}^{k} \leq \alpha_{\text{udi}}^{k} (t+1) \\
\alpha_{\text{udi}}^{k} \leq \alpha_{\text{udi}}^{k} (t+1) \\
\alpha_{\text{udi}}^{k} \text{ binary} \\
\mu_{du}, \nu_{du} \geq 0
\]

Appendix B: Proof of theorems

B.1 Proof of Theorem 1

Let \( L_{sk}^{u}(r) \) be the independent and identically distributed (i.i.d.) stochastic location function of the \( a_{sk}^{u}(r) \) arrival of class \( k \) and risk group \( i \) on day \( s \). The \( u \) superscript denotes that this arrival was initially admitted to unit \( u \). \( L_{sk}^{u}(r) = (l_{sk}^{u1}(r), l_{sk}^{u2}(r), \ldots, l_{sk}^{uN}(r))^{T} \), where \( l_{sk}^{u}(r) \) equals 1 if the patient is in unit \( u \) on day \( r \) after admission and equals 0 otherwise, with \( \sum_{i=1}^{N} l_{sk}^{u}(r) = 1 \), where \( N \) is the number of units. By convention, \( l_{sk}^{u}(r) = 0 \) if \( r < 0 \). Further, let \( B_{ds}^{u}(t) \) be the stochastic census function of unit \( u \) on day \( d \) in week \( t \), and let \( \mathbf{e}_{u} \) be a row vector of zeros with a 1 in the \( u \)th element. Finally, let \( A_{k}^{u} \) be the stochastic number of patients of class \( k \) and risk group \( i \) who arrive on day \( s \).

\[
B_{ds}^{u}(t) = \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in \mathcal{U}} \sum_{k' \in K} \mathbf{e}_{u'} L_{sk}^{u}(r) (d - d' + 7(t - n')) \alpha_{\text{udi}}^{k} \\
= \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in \mathcal{U}} \sum_{k' \in K} \sum_{n=0}^{\infty} \mathbf{e}_{u'} L_{sk}^{u}(r) (d - d' + 7n) \alpha_{\text{udi}}^{k}
\]

In the first equality, we establish our stochastic census function. To do this we look at all previously admitted patients (from all units) and see if they are currently in unit \( u \). We add up all such patients to get the census of unit \( u \) on day \( d \) of week \( t \). The second equality above involves a simple change of variable, from \( n' \) to \( n \). This will become very useful when calculating \( \mu_{du} \) and \( \nu_{du} \).

The derivation of the mean census is given on the next page. Both \( \mathbf{e}_{u} L_{sk}^{u}(r) \) and \( \alpha_{\text{udi}}^{k} \) can only take values of 0 or 1, meaning \( \sum_{i=1}^{N} \mathbf{e}_{u} L_{sk}^{u}(r) \alpha_{\text{udi}}^{k} = A_{k}^{u} \leq \infty \). We know that \( A_{k}^{u} \) is finite as there cannot be an infinite number of arrivals in one day. Thus (50) holds.

For (51), we simply replace a finite sum with an infinite sum that is equivalent via use of an indicator function, where \( \mathbb{I}(x) = 1 \) if \( x \) is true and \( \mathbb{I}(x) = 0 \) otherwise.

As \( \mathbb{1} \{A_{k}^{u} \geq a\} = \mathbb{1} \{A_{k}^{u} \geq a\} \), \( A_{k}^{u} \) has a binary, \( \mathbb{1} \{A_{k}^{u} \geq a\} \), \( \mathbb{1} \{A_{k}^{u} \geq a\} < \infty \), and (52) holds. For (53), we change the summation over \( a \) in order to motivate the use of the tail function of the cumulative distribution. All terms \( L_{sk}^{u}(r) \) and \( L_{sk}^{u}(r) \) are independent of each other provided that \( a, s, k, i \neq (a', s', k', i') \). This is true because each \( L_{sk}^{u}(r) \) term represents the care pathway of an individual patient, and each patient’s care pathway is independent of all others.
Also, patient care pathways and the number of arrivals in a day are independent, and these are both independent of whether we accept a given class of patient on a given day. Thus, \( e_u L_{ski}^{au}(r), 1(A_k \geq a) \), and \( \theta_{udi}^{k} \) are mutually independent, and we can treat the expectation of their product as the product of their expectations. Also, \( P_{dkui}^u(r) = f(a) \) and \( F_{dkui}(a) \) are identical between weeks, only depending on the day of the week \( d \). Recall that by convention, \( P_{dkui}^u(r) = 0 \) if \( r < 0 \). Thus, Eq. 54 holds.

### B.2 Proof of Theorem 2

The derivation of the variance of the census is given on the next page. As previously discussed, each patient’s care pathway is independent of other patient care pathways. Also, the number of arrivals on any given day is independent of the number of arrivals on every other day. Thus, each term is independent of all other similar terms, and Eq. 59 holds. Similar to the derivation of \( \mu_{du} \), we replace the finite sum with an infinite sum and an indicator function to get (60). Each \( e_u L_{ski}^{au}(r) 1(A_k \geq a) \) term is independent of all other similar terms, and Eq. 61 holds. To get (62), we simply use the definition of variance. Similarly to the derivation of \( \mu_{du} \), we alter the summation over \( a \) in order to use the tail distribution, giving (63). As each \( e_u L_{ski}^{au}(r) 1(A_k \geq a) \theta_{udi}^{k} \) term can only take a value of 0 or 1,

\[
E \left[ \left( e_u L_{ski}^{au}(r) 1(A_k \geq a) \theta_{udi}^{k} \right)^2 \right] = \]

\[
E \left[ e_u L_{ski}^{au}(r) 1(A_k \geq a) \theta_{udi}^{k} \right].
\]

Combining this with a calculation similar to that used in our derivation of \( \mu_{du} \), we get (64). We note that \( \theta_{udi}^{k} \) is binary, which implies that \((\theta_{udi}^{k})^2 = \theta_{udi}^{k}\). Leveraging this realization, we get (65).

### Derivation of the daily mean census per unit

\[
\mu_{du} = \lim_{t \to \infty} E[ B_{du}(t) ]
\]

\[
= \lim_{t \to \infty} E \left[ \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{t-d'-7} \sum_{a=1}^{A_{k,d'+7}^{au}(t-n)} e_u L_{ski}^{au}(d-d'+7n) \theta_{udi}^{k} \right]
\]

\[
= E \left[ \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a=1}^{A_{k,d'+7}^{au}(t-n)} e_u L_{ski}^{au}(d-d'+7n) \theta_{udi}^{k} \right]
\]

\[
= E \left[ \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a_1=1}^{A_{k,d'+7}^{au}(t-n)} E \left[ \sum_{a_2=1}^{A_{k,d'+7}^{au}(t-n)} e_u L_{ski}^{au}(d-d'+7n) \theta_{udi}^{k} \right] \right]
\]

\[
= \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a_1=1}^{A_{k,d'+7}^{au}(t-n)} E \left[ e_u L_{ski}^{au}(d-d'+7n) 1(A_{k,d'+7}^{au}(t-n) \geq a) \theta_{udi}^{k} \right]
\]

\[
= \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a_2=1}^{A_{k,d'+7}^{au}(t-n)} E \left[ e_u L_{ski}^{au}(d-d'+7n) 1(A_{k,d'+7}^{au}(t-n) \geq a) \theta_{udi}^{k} \right]
\]

\[
= \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k \in K} \sum_{n=0}^{\infty} \sum_{a_2=1}^{A_{k,d'+7}^{au}(t-n)} P_{d'kui}(d-d'+7n) \theta_{udi}^{k}
\]

(55)
Data-driven optimization methodology for admission control in critical care units.

Derivation of the daily variance of census per unit

\[ v_{du} = \lim_{t \to \infty} \text{var} \left( B_{du(t)} \right) \]

\[ = \lim_{t \to \infty} \text{var} \left( \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k' \in K} \sum_{n=0}^{t} A_{k,d'+7(t-a)}^i \left( d - d' + 7n \right) \theta_{u'd'1}^k \right) \]

\[ = \text{var} \left( \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k' \in K} \sum_{n=0}^{\infty} \left( \sum_{a=1}^{\infty} \epsilon_{du}^{au'} \right) A_{k,d'+7(t-a)}^i \left( d - d' + 7n \right) \theta_{u'd'1}^k \right) \]

\[ = \text{var} \left( \sum_{i \in I} \sum_{d' \in D} \sum_{u' \in U} \sum_{k' \in K} \sum_{n=0}^{\infty} \left( \sum_{a=1}^{\infty} \epsilon_{du}^{au'} \right) A_{k,d'+7(t-a)}^i \left( d - d' + 7n \right) \theta_{u'd'1}^k \right) \]

References

32. US Census Bureau (2013)
学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具