Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

A SIMULATION MODEL TO EVALUATE THE PATIENT FLOW IN AN INTENSIVE CARE UNIT UNDER DIFFERENT LEVELS OF SPECIALIZATION

Andres Alban
Stephen E. Chick

Technology and Operations Management
INSEAD
Boulevard de Constance
Fontainebleau, 77305, FRANCE

Oleksandra Lvova
Danielle Sent

Department of Medical Informatics
Amsterdam UMC, University of Amsterdam
Meibergdreef 9
1105 AZ Amsterdam, THE NETHERLANDS

ABSTRACT

Intensive care units are complicated hospital departments in that they have both urgent and elective patients with a variety of specialty needs that cannot be easily treated elsewhere. Their design involves important operations strategy decisions, such as whether there are general facilities serving all patients or several specialized units for certain patient needs, or something in between. They also involve bed capacity decisions for the aggregate and potentially specialized units. This paper presents a simulation model which is used to assess trade-offs in these operational design issues with respect to three performance measures (rejection rate, rescheduling rate, and bed occupancy rate), using data and design options for the Academic Medical Center (AMC), one of two locations forming the Amsterdam University Medical Centers (UMC).

1 INTRODUCTION

The intensive care unit (ICU) is a department of the hospital that provides enhanced monitoring, treatment and nursing care to patients in an acute medical condition who require critical medical care, monitoring and supporting vital functions. Such patients may be referred from other departments of the hospital, from surgery or other wards, or directly from the emergency department. ICUs have physiologic organ support to sustain life during a period of life-threatening organ system insufficiency (Marshall et al. 2017). The design of an ICU requires several considerations (Hamilton and Shepley 2010). One such consideration is whether the ICU can serve a wide range of critical patients (general ICU), or whether it can accommodate a defined group of patients that are served by the staff specializing in the condition of need (specialized ICU) (Lott et al. 2009). Specialized ICUs must be justified with matching specialized staff and consistent volume of patients.

Specialized ICUs may provide better quality of care as compared to the general ICU. For instance, in the hospital setting Czaplinski and Diers (1998) show that specialized nursing staff reduced the mortality rate and length of stay (LOS) in 13 of the 16 diagnosis groups. Kramer and Zygun (2011) provide a meta-analysis of twelve studies investigating the effect of specialized neurological units in which they find heterogeneity of results but an overall decrease in mortality and increase in favorable outcomes. On the other hand, Lott et al. (2009) do not find a significant difference in mortality and LOS between the specialized and general ICUs. The evidence is sometimes mixed as to which is better, as discussed by Van der Sluijs et al. (2017). They outline the need for better costs, outcomes and data, then call for continuous improvement, lean principles and other process innovations. For example, the role of intensivists can support more effective ICU processes that improve patient outcomes and reduce costs. Hagen et al. (2013) study the prioritization by severity of patients with an ICU queueing model and its effect on wait times, utilization, return rates and mortality. In general, there is no consensus on the benefit of specialization in

978-1-7281-9499-8/20/$31.00 ©2020 IEEE

the ICU and Lott et al. (2009) report that, out of the 55 medical centers included in their study, 22 had only a general ICU, 17 had only a specialized ICU, and 16 had both.

The operations management literature has investigated the effect of specialization in other settings. The more prominent example is the seminal work of Skinner (1974) on "The Focused Factory" backed by evidence in several industries. KC and Terwiesch (2011) provide an econometric analysis of cardiac care in California, where they found that specialization at higher levels of the organizations (hospital level) is not useful but specialization at the department or process level can improve health outcomes and LOS. Because our focus is the ICU of an academic hospital, our study lies in the department level specialization where they found improvement. We refer the reader to the literature review of KC and Terwiesch (2011) for further reference on specialization in the healthcare setting.

In this paper, we focus on a setting in which an ICU can either serve a variety of patients (general structure) or be partitioned into several specialized units (specialized structure), each serving a more homogeneous population of patients. The main drawback of specialized units is that resources are dedicated to a particular group of patients and may not be used by other patients even when the resources are idle. This is related to the customary queueing theory result that pooling servers reduces waiting times (e.g. Wolff 1989, chapter 5.9). However, due to each queue having different arrival rates and LOS, specialized units can ensure higher chance of resource availability for any particular group of patients by blocking the use of these resources to other patients. Thus, specialized units give the hospital manager more flexibility in deciding which groups of patients receive more resources.

This paper presents a discrete-event simulation model motivated by the operation of the general ICU at Amsterdam UMC, location AMC, in Amsterdam. The model evaluates the impact on the patient flow of changing the ICU structure from general to specialized, or to a "flexible" structure that combines features of general and specialized in the spirit of Jordan and Graves (1995). The model aims to provide a tool to inform decision-makers about the patient flow under different ICU structures and capacity. The model focuses on capacity and patient flows, with quality implications assessed by the decision maker separately.

Bruin et al. (2010) use a $M/G/c/c$ queue to analyze the capacity for 24 wards in a hospital. Although the $M/G/c/c$ is a close relative of our model, certain differences require a more nuanced treatment. Similar to Bruin et al. (2010), we distinguish patients into planned and unplanned. We assume here that unplanned patients arrive in a memoryless fashion yet might be routed to another facility, while planned patients, who do not arrive in a memoryless fashion, may be rescheduled (e.g., for a postponed surgery) if capacity constraints warrant it, rather than redirecting them to other hospitals or departments.

The rest of the paper unfolds as follows. In Section 2 we describe the ICU of the AMC. Section 3 explains the discrete-event simulation modeling the patient flow in the ICU. In Section 4, we present results of using our model to analyze the patient flow at the AMC, illustrating how to use our model as a decision support tool. Section 5 summarizes an adaptation of the model to support ICU capacity expansion planning at the start of the COVID-19 outbreak. Section 6 is devoted to conclusions of this paper.

2 THE ICU AT THE AMSTERDAM UMC, LOCATION AMC

The Amsterdam University Medical Centers, location AMC, provides integrated patient care, fundamental and clinical scientific research, and teaching. The hospital has 45 clinical and non-clinical departments, including a multidisciplinary ICU. The ICU admits patients referred from several departments such as the emergency department, neurosurgery, and pulmonology.

Patients in the available dataset are labeled into six different specialisms depending on their health condition: cardiopulmonary surgery (CAPU), cardiology (CARD), internal medicine (INT), surgery (CHIR), neurosurgery (NEC), neurology (NEU). A specialism is a branch of medical practice that is focused on a defined condition. Patients where the referring specialism was unknown or not captured by the six main groups are labeled "Other". Patients are additionally labeled as *planned* and *unplanned* patients. Planned patients are scheduled by one of the departments at the hospital for instance after a surgical procedure, and unplanned patients arrive from the emergency room or other units due to emergency situations. We refer to this as the planning status of the patient.

We have a subset of ICU data for admissions and rejections for the years 2015 and 2016. Among the information available to us for admitted patients are the time of admission, the time of discharge, specialism, and planning status. For rejected patients, patients for whom no bed was available, we have the date of arrival and specialism. Only unplanned patients can be rejected. Planned patients will always be replanned rather than rejected.

The total number of arrivals was 1,881 patients in 2015 and 2,170 in 2016. Out of those, 1,779 and 2,043 patients were admitted, and 102 and 127 patients were rejected, respectively. The overall average arriving patients per day was therefore 5.55. In total, 66% of all arrived patients (2,668 patients) were unplanned, and out of them 9% (229 patients) were rejected. Among all patients, 82% (3,117 patients) were admitted on weekdays and 18% (705 patients) were admitted on weekends. Table 1 shows the number of patients belonging to each specialism with the respective average length of stay and arrival rate.

Table 1: Basic statistics by specialism for the years 2015 and 2016.

	CAPU	CARD	CHIR	INT	NEC	NEU	Other	Total
Arrival rate per day	1.85	0.42	0.75	1.35	0.47	0.57	0.13	5.55
Number of admitted patients	1,346	289	497	939	307	394	50	3,822
Planned patients	1,211	5	103	17	38	8	1	1,383
Unplanned patients	135	284	394	922	269	386	49	2,439
Average length of stay in days	1.89	4.08	4.47	4.58	4.31	3.75	2.60	3.45
Planned patients	1.64	2.87	2.66	6.44	2.58	7.52	0.23	1.84
Unplanned patients	4.16	4.11	4.94	4.55	4.56	3.67	2.65	4.36
Number of rejected patients	7	18	49	49	36	22	48	229

3 DISCRETE-EVENT SIMULATION MODEL OF PATIENT FLOWS IN THE ICU

Planned and unplanned patients arrive at the ICU from different referral departments and are labeled into specialisms. We assume that the arrival processes of each of the specialisms are mutually independent, and the arrival process of planned patients is independent of the arrival process of unplanned patients. We assume the LOS for each patient is independent of the rest and identically distributed for patients within each specialism and planning status. We assume that arrival patterns remain equal over time. For the simulation, we estimate arrival processes and LOS distributions based on historical data.

If an arriving patient cannot be immediately admitted to the ICU, planned patients are *rescheduled* for the following day, while unplanned patients are *rejected* and diverted to another hospital. For this simulation, we assumed that the decision to admit or reschedule a planned patient is done 12 hours in advance of the arrival of the patient. Thus, a bed is reserved in advance for planned patients, or they are rescheduled even if by the time of the arrival a bed has become available. Figure 1 summarizes the main components of the model.

We are interested in exploring how different ICU structures perform. We consider three structures that arrange distribution of beds to the specialisms and determine when to admit or reject/reschedule patients:

1. **General** ICU: a multidisciplinary unit where all patients are admitted to a centralized ICU and a patient of any specialism can be admitted to any bed. Patients are only rejected or rescheduled when the capacity of the entire ICU is being used.
2. **Specialized** ICU: an ICU partitioned into several units where each unit is only able to admit certain specialism(s). Patients are rescheduled or rejected if the unit in charge of their specialism is out of capacity, even if beds in other units are available.

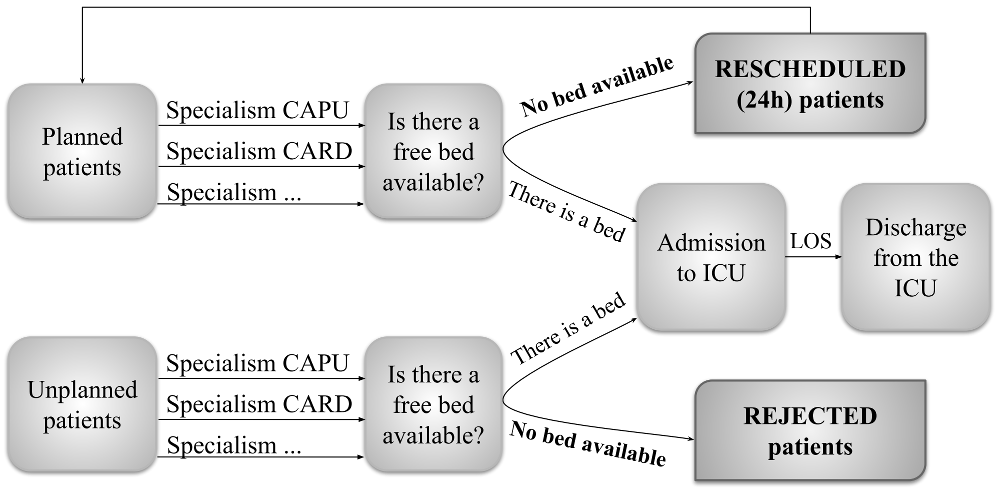

Figure 1: Model of the ICU patient flow.

3. **Flexible** ICU: an ICU partitioned into several units where each unit is in charge of certain specialism(s). Unlike the specialized ICU, each unit can use available beds in only one additional unit. In the simulation, we use a long chain as recommended by Jordan and Graves (1995). For instance, in a four-unit ICU, patients of Unit 1 can be admitted by Unit 2, patients of Unit 2 can be admitted by Unit 3, patients of Unit 3 by Unit 4 and patients of Unit 4 by Unit 1. A patient that is redirected to another unit is transferred back to the specialized unit when a bed is vacated in the patient's specialisms' unit.

Our simulation model assesses three performance measures in a way which is intended to quantify trade-offs across those metrics. Those trade-offs support decision making regarding the following:

1. The structure of the ICU: general, specialized, or flexible.
2. The capacity of the ICU, i.e., how many beds to have available at any given time. In the case of specialized and flexible structures, how to distribute resources among the units.

Some hospitals may see a larger benefit of specialized units in quality of care and be willing to spend more resources to extract the gains of specialization, and there are some detailed features of ICU operations which are not modeled. We therefore propose that the model be used for insight generation and decision support, rather than as a normative decision making tool. The three performance measures assessed by the model are as follows:

1. *Occupancy rate* is the fraction of time during which beds are occupied. For instance, a two-bed unit where both beds are occupied for half of the time horizon and only one is occupied the remaining half, has an occupancy rate of 0.75.
2. *Rejection rate* is the fraction of unplanned patients that is rejected.
3. *Rescheduling rate* is the number of times patients are rescheduled divided by total number of planned patients. This fraction can be larger than one because any planned patient can potentially be rescheduled several times before being finally admitted.

Hospitals aim to minimize the rejection and rescheduling rates, while maximizing the occupancy rate to reduce the idle-time of the ICU resources. However, in making decisions there is a trade-off between such performance measures. The prevailing trade-off is that larger capacity reduces rejection and rescheduling rates but at the same time reduces the occupancy rate. An additional trade-off arises in our model when we decide how to allocate resources for different specialized units. For instance, Unit 1 may admit mostly planned patients, while Unit 2 may admit mostly unplanned patients. By shifting capacity from Unit 1 to Unit 2, the rejection rate decreases but the rescheduling rate increases.

4 RESULTS

We illustrate the benefits of our model with a study of the ICU at the AMC. Using the available data described in Section 2 we estimate the required parameters in Section 4.1. Using the estimated parameters, we evaluate the different partitions of the specialized structure (Section 4.2). We then compare the different potential structures under the same capacity (Section 4.3) and in a scenario where specialization reduces LOS by 20% (Section 4.4).

4.1 Input Modeling for Arrivals and Lengths of Stay

We simulate the ICU flow using probability distributions for the arrival and LOS estimated from the available data. Table 2 presents the parameters estimated for the unplanned patients and Table 3 presents estimates for the planned patients. We assume that specialisms CARD, NEU and patient with no referring specialism recorded, i.e. Other, do not have any planned patients because the number of observations in the data is negligible (see Table 1). Planned surgery for cardiology or for neurology would be classified for the associated surgical unit (CAPU and NEC respectively).

The arrival of the unplanned patients is assumed to be a Poisson process. This is a reasonable assumption for unplanned patients because there is no expected association between arriving patients. The arrival process is estimated independently for weekday and weekend arrivals as recommended in the literature (e.g., Bruin et al. 2010). We test the Poisson assumption using the number of unplanned arrivals per day with a chi-squared test and find p-values of 0.42 and 0.39, respectively, not rejecting our assumption. While the rate of arrival at night is lower than the rate during the day, we presume that such differences do not significantly affect the simulation output, and estimate a single Poisson process for any time of the day. The mean arrivals of the Poisson process for each of the specialisms is estimated with the inverse average interarrival time to leverage the additional information provided by the time of arrival. Because the time of arrival is not available for rejected patients ($\approx$6% of unplanned patients), we impute their arrival time as 12pm for the purpose of estimation without inducing any bias.

The arrival of planned patients is assumed to follow the following process. Each day, the number of patients that arrive at the ICU follows a categorical distribution, where the probabilities are given by the fraction of days in the two-year period that a given number of patients arrived. The time of arrival is independent for all patients and follows a beta distribution with parameters α and β. Because the planned arrivals start around 10am, the beta distribution is shifted to the right by 10/24 of a day modulo 1.

The distribution family of the LOS was chosen using the tool Stat::Fit$^{\text{TM}}$. We split the estimation of LOS by the planning status, as well as the specialisms, because LOS is generally longer for unplanned patients (see Table 1). For all but one of the specialisms, either the log-normal (lognorm) or log-logistic (loglogis) distributions were not statistically rejected. For the planned CAPU patients, all the distributions available with the tool were rejected. Because we have 1211 observations for this group of patients, we use bootstraping, i.e., random sampling with replacement from the data points, to obtain LOS samples for the planned CAPU patients. Some of the estimated distributions have fat tails and the second moment does not exist. At the random sampling stage, this generates very extreme observations that are unreasonable. While this occurs in about 0.1% of the samples, it affects the results substantially. Therefore, any sample above 200 days was set to 200 days, where 200 days was the maximum observed LOS in the data.

The capacity of the ICU varies over time due to several factors such as nurse and staff availability. However, there is no available data on the evolution of capacity of each individual important ICU resource over time that could inform a realistic process for the ICU capacity. Thus, our model assumes that the

Table 2: Maximum likelihood estimates of distributional parameters for unplanned patients.

	CAPU	CARD	CHIR	INT	NEC	NEU	Other
Interarrival time in days							
Weekdays mean	4.75	2.26	1.58	0.735	2.46	1.73	6.81
Weekends mean	6.05	2.6	1.82	0.79	2.23	1.93	8.35
Length of stay in days							
Distribution family	lognorm	loglogis	loglogis	lognorm	lognorm	loglogis	lognorm
log mean / rate	0.62	0.53	0.58	0.50	0.82	0.69	0.25
log std. dev. / shape	1.38	1.30	1.18	1.48	1.31	1.18	1.28

Table 3: Maximum likelihood estimates of distributional parameters for planned patients.

	CAPU	CHIR	INT	NEC
No. patients per day (weekday/weekend)				
Prob. of no patients in %	8.8 / 87.6	82.8 / 97.1	98.1 / 97.1	93.7 / 98.1
Prob. of 1 patients in %	18.2 / 12.0	15.9 / 2.9	1.7 / 2.9	6.1 / 1.9
Prob. of 2 patients in %	30.5 / 0.5	1.3 / 0	0.2 / 0	0.2 / 0
Prob. of 3 patients in %	27.6 / 0	0 / 0	0 / 0	0 / 0
Prob. of 4 patients in %	10.5 / 0	0 / 0	0 / 0	0 / 0
Prob. of 5 patients in %	3.6 / 0	0 / 0	0 / 0	0 / 0
Prob. of 6 patients in %	0.8 / 0	0 / 0	0 / 0	0 / 0
Time of arrival				
α-parameter	2.57	1.70	1.46	2.52
β-parameter	8.64	3.85	3.58	4.35
Length of stay in days				
Distribution family	bootstrap	lognorm	lognorm	loglogis
log mean / rate		0.17	0.59	0.83
log std. dev. / shape		1.20	1.72	1.24

capacity of the ICU is constant over time to provide insights and guidance. To determine the capacity, we retrospectively chose the value that most closely matches the rejection rate in the observed data. The capacity of beds we found is 28 which was in line with the experience of the physicians during our model validation check. We next discuss the partition of this capacity for the specialized and flexible structures.

4.2 Selecting the Partition of the Specialized ICU

Based on medical similarity between patients as indicated by the domain experts in the ICU at AMC (such as required nurse expertise and medical equipment), the specialisms can be grouped into four clusters: 1) CAPU; 2) CARD, INT, Other; 3) CHIR; and 4) NEC, NEU. Our specialized ICU thus has four units specialized for each one of the clusters but the stream of patients for each specialism is still randomly generated independently. In this section, we examine the performance and trade-offs for different bed allocations to the four units.

We run the discrete-event simulation with the parameters estimated in Section 4.1. Because the specialized units are independent of each other, we can analyze the performance for each unit separately. We run the simulation for each unit with a number of beds between two and 25 for a period of 3,770 days (approx. ten years and 120 days), which generates enough samples for confidence bounds to be small enough to get the insights we sought. The first 120 days were burned-in to warm up the queue and the remaining ten years were used to evaluate the performance measures. We use common random numbers by generating a single stream of patient arrivals with respective LOS that is then processed for all the possible values of the number of beds to decide which patients are rejected or rescheduled. Therefore, there is a positive correlation between our estimates.

To compute confidence intervals, we use the method of batch means (Asmussen and Glynn 2007, chapter IV). We split the ten years of simulation data in $N = 20$ half-year periods. For each period i, we compute the performance measure x_i. To compute the half-width of a two-sided confidence interval of level $1-\alpha$, we compute the standard error of x_i and multiply it by the t-value with level $1-\alpha/2$ and $N-1$ degrees of freedom, $t_{1-\alpha/2,N-1}$:

$$t_{1-\alpha/2,N-1}\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i-\bar{x})^2},$$

where $\bar{x}$ is the average of x_i. Throughout, we report 95% confidence intervals, i.e. $\alpha = 5\%$.

The performance of each unit as a function of the number of beds is presented in Figure 2. In viewing the figures, recall that planned patients will return until a bed is available, possibly being rescheduled several times during busy periods, and that rejected unplanned patients leave the system immediately, reducing the load during busy periods. Because rejection and rescheduling rates are essentially computed with fewer samples (rejections are per unplanned patient and reschedulings are per planned patients), their confidence intervals are larger. The confidence intervals of the rescheduling rate are particularly large because rescheduled patients come in batches and the variation between sections is large.

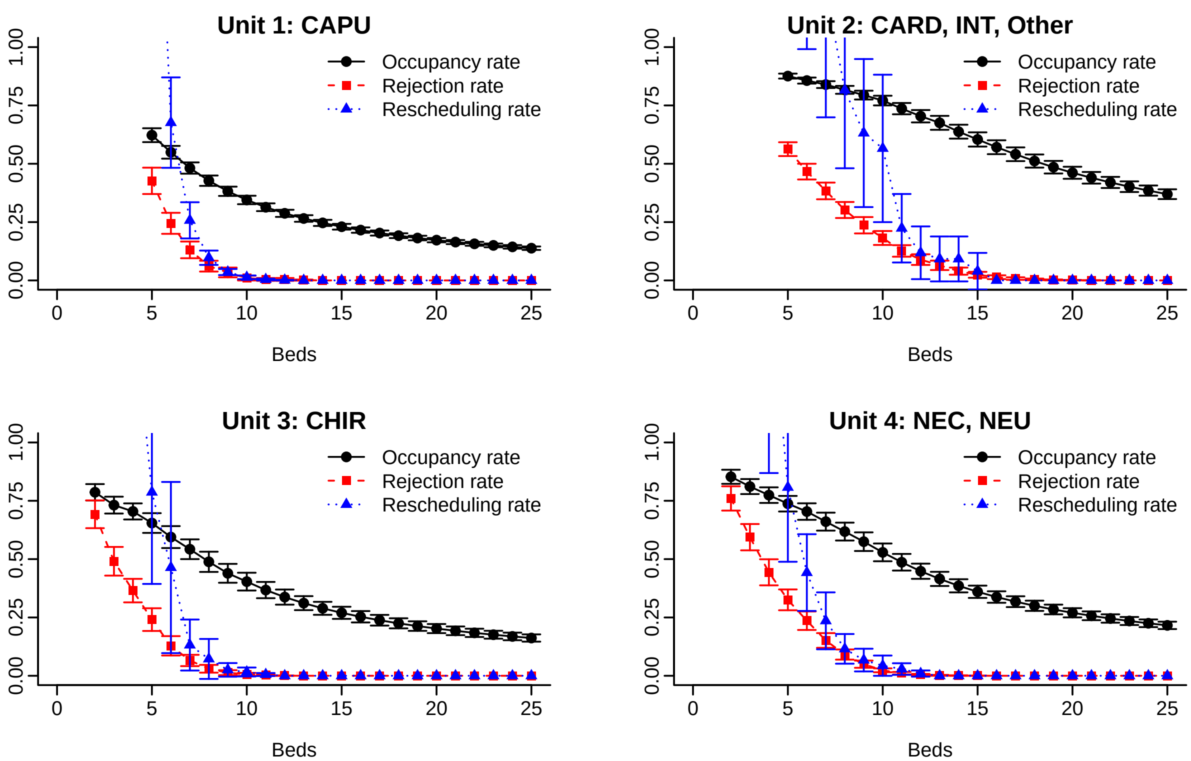

Figure 2: The performance of the specialized ICUs for 0 - 25 of beds. The fraction of patients for the four units are in order 33.6%, 34.5%, 13.3%, and 18.6%. The fraction of patients that are planned in each unit are 89.2%, 1.2%, 18.3%, and 4.2%

For the specific scenario of the ICU at the AMC, Figure 2 provides two important observations. First, the minimum number of beds required to achieve an average rejection rate below 5% (a usual target in hospitals) for each unit can be observed for 9, 14, 8, and 9 beds for each unit, respectively, which amounts to 40 beds in total (12 more than the current capacity). For this number of beds the occupancy rates are 38.2%, 63.7%, 48.8%, and 57.5%, and the rescheduling rates are 3.6%, 9.2%, 7.2%, and 6.7%, respectively. Thus, to attain a rejection rate at the target level with a specialized ICU, over 40% more beds would be necessary and around half of the available resources would have to remain idle. The general structure would require one additional bed to achieve an overall 5% rejection rate (not shown in figures).

Second, if the total current number of beds (28) were to be maintained, we find most appropriate a partition with six, eleven, five, and six beds for the four clusters (1 = CAPU; 2 = CARD, INT, Other; 3 = CHIR; and 4 = NEC, NEU), respectively. This partition maintains a reasonably balanced level for rejection rates and rescheduling rates across the clusters (while unit 2 has lower rejection and rescheduling rate, transferring one bed to any other unit would create a bigger imbalance). While other partitions may also be appropriate, we use this partition for comparison with a general ICU with the same capacity throughout the rest of this section.

Figure 2 can give insights to hospital managers who are evaluating trade-offs for partitions of specialized ICU structures. With certain goals or constraints in mind, the required resources and partition can be found. For instance, a manager may need a low rejection rate for emergency cases in the cardiology unit and prefer to have a high occupancy rate for the surgical patients. By inspecting the plots, the number of beds that strikes the best balance between performance measures can be found.

4.3 Comparison of the Different ICU Structures

In this section, we compare the performance among the three ICU structures with the current capacity of 28 beds in the AMC setting. The general ICU is expected to perform the best, followed by the flexible structure, and finally the specialized structure. While the results of this section are specific to the AMC scenario, they illustrate how our simulation model can be used by a hospital manager to identify the ideal structure for the hospital's objectives and constraints.

For each of the three considered ICU structures we again run the simulation for 3,770 days (approx. ten years and 120 days) and the first 120 days are burned-in to warm up the queue. As in Section 4.2 we use common random numbers by generating a single stream of patient arrivals that is then processed for the respective ICU structure to decide which patients are admitted, rejected or rescheduled. The performance measures and their 95% confidence intervals are computed using the same sectioning method described in Section 4.2. We do not report the comparison of the performance among the structures. We note, however, that comparisons have smaller confidence intervals than the sum because of the positive correlation induced by the common random numbers.

For the specialized and flexible structures, we specify the bed partition that we found most appropriate in Section 4.2: six beds for Unit 1 (CAPU specialism), eleven for Unit 2 (CARD, INT, and Other specialisms), five for Unit 3 (CHIR) and six for Unit 4 (NEC and NEU specialisms). For the flexible structure, we additionally have to establish the long chain for redirection of patients in case of unavailable beds in the patient's specialized unit. We chose the following chain that aimed to balance the performance across units: $1 \to 4 \to 3 \to 2 \to 1$. We focus on one specific resource allocation arrangement in this section. However, a hospital manager could replicate this analysis with several arrangements of the specialized or flexible structures if warranted by the decision-making process.

Figure 3 presents the comparison of the three structures on our three performance measures. For each structure and each measure, we report the overall performance and performance by unit. Since the general ICU is not divided in units, the reported values correspond to the performance specific to the cluster of patients assigned to the respective specialized unit. We point out several observations.

First, as expected, the general structure dominates in all three performance measures, followed by the flexible, and finally the specialized. While the specialized structure does not perform particularly bad compared to the general ICU in terms of occupancy rate (67.5 vs. 75.1%), the gap in the overall rejection and rescheduling rates is substantial: 18.1 vs. 6.4% and 67.3 vs. 10.3%, respectively.

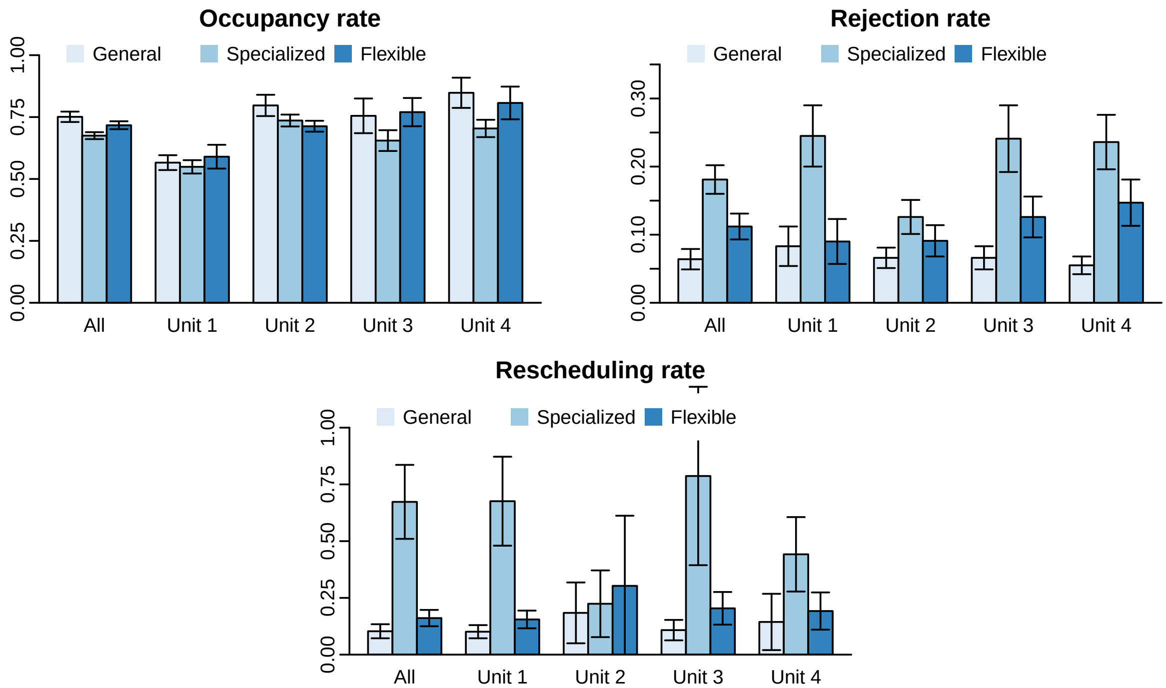

Figure 3: Comparison of the ICU structures with the parameters estimated from the data.

Second, the lowest occupancy rate for the three structures is observed in Unit 1 at 55-60%. However, the lower occupancy in Unit 1 does not translate into lower rates of rejection or rescheduling. A possible explanation for this finding is that Unit 1 is in charge of the CAPU specialism, which is mainly composed of planned patients (≈90%) with a relatively short LOS.

We observe that the flexible structure significantly improves the performance compared to the specialized structure, especially in the overall rejection and rescheduling rates: 18.1 vs. 11.2% and 67.3 vs. 16.1%, respectively. Although the flexible structure does not match the performance of the general structure, it represents a big improvement in operational performance compared to the specialized structure. In addition, the flexible structure may still capture some of the quality benefits of specializations because 77.6% of patients are treated in their specialized unit for the entire stay of the patient.

4.4 Hypothetical Scenario in Which Specialization Reduces LOS

Specialized ICUs have been shown to improve health outcomes and LOS in certain settings. Our model does not directly model health outcomes, but we do model LOS. In this section, we explore how a reduction of 20% in the LOS of patients affects the patient flow performance in the ICU. We show the results for a 20% reduction in the LOS of patients for two reasons that will be developed throughout this subsection. First, while 20% is a substantial and perhaps often even infeasible decrease, it is still not enough for the specialized structure to match the performance of the general structure. Second, it is a proof of concept that the flexible structure can achieve the right balance of specialization and the best performance among all structures. If there are associated quality gains from a reduced LOS, such as a reduction of transmission of infection within the ward, these would be modeled separately based on characteristics of a given ICU.

To this end, we run the discrete-event simulation for the same number of days, bed partition and with the same parameters as in section 4.3 with the only difference of 20% LOS reduction for the patients in specialized and flexible structures. Figure 4 replicates the performance measures of the general ICU

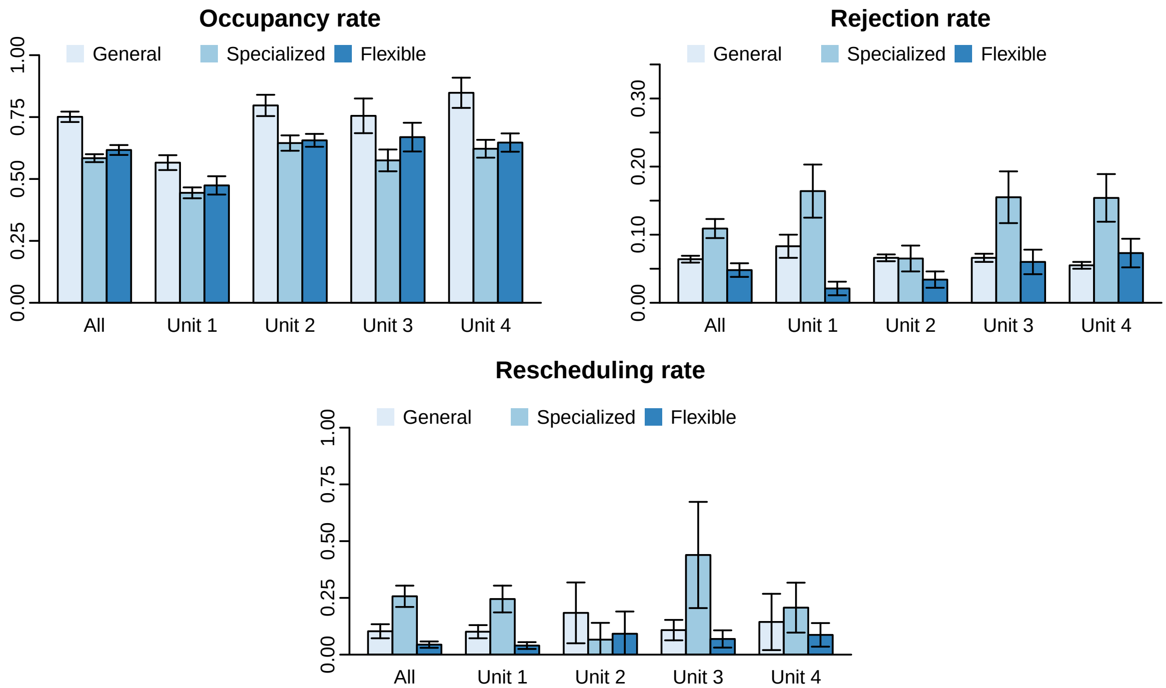

Figure 4: Performance of ICU structures when the LOS is reduced by 20% in the specialized and flexible structures.

presented in Figure 3 and modifies the results for specialized and flexible structures by showing their performance measures corresponding to the LOS reduced by 20%.

With shorter LOS the occupancy rate deteriorates significantly from 67.5% to 58.4% and from 71% to 61.7% with the specialized and flexible structures, respectively. We also observe a substantial improvement in rejection and rescheduling rates for both specialized (from 18.1% to 10.9% and from 67.3% to 25.7% respectively) and flexible structures (from 11.2% to 4.8% and from 16.1% to 4.4% respectively). Even though the improvement with the specialized structure is significant, the performance in terms of rejection and rescheduling rates are still inferior to those of the general structure. When comparing the flexible structure to the general we observe rejection and rescheduling rates that are both lower. Thus, the flexible structure with 20% reduced LOS would perform better than the general ICU in terms of rejection and rescheduling rates but would continue to have a lower occupancy rate.

5 COVID MODEL

Our model was developed prior to the surge in ICU admissions experienced during the first half of 2020 which resulted from the COVID-19 pandemic. After submitting the initial version of this paper, many hospitals, including the AMC, responded to the dramatic increase in incidence of COVID-19 through a variety of techniques to expand capacity. This planning process involves developing an expanded bank of ICU beds to manage COVID-19, and another bank of beds for non-COVID-19 ICU patients. The latter was primarily urgent cases, as elective procedures were largely dropped through this period.

We adapted the model to allow for hospitals to assess their ICU capacity expansion plans. The source code was made publicly available online at https://github.com/sechick/icu-covid-sim with an associated link to enable hospital planners to assess the performance of a range of bed capacity decisions for COVID-19 and non-COVID-19 ICU patients. Arrival rate, length of stay, and bed capacity parameters can be input, and the following performance measures were reported: average number of patients which can be handled, fraction of beds occupied, fraction of patients which need referral to other facilities due to bed blocking.

The model was used by a number of hospitals to assess their bed capacity plans. One initially unexpected usage of the model was reported by a large European hospital who used the model to help anticipate the number of referrals it might expect from other regional facilities. For further details on the model, see Alban et al. (2020).

6 CONCLUSION

We have modeled the patient flow in an ICU with patients requiring different care (specialisms) and different priority (planning status). The problem was motivated by the patient flow at the ICU of the AMC and discussions whether or not to shift from a general ICU to four dedicated smaller ICU's. We used their electronic medical records to illustrate the benefits of our model.

Our simulation model is aimed at helping hospital managers make decisions to structure and allocate resources in the ICU. In particular, our model evaluates the effect on patient flow of splitting the ICU in a set of specialized units. Furthermore, we propose a flexible structure that makes a compromise between the general ICU and specialized units.

By estimating the probability distributions for the arrival and LOS processes using data of the ICU at the AMC, we evaluated the effect of partitioning the ICU into four specialized units. We found that the patient flow would deteriorate significantly due to an increasing number of rejected and rescheduled patients. To achieve a target rejection rate of 5% in each of the specialized units, an increase in available beds of 40% would be required.

We also find that the flexible structure is a potential solution that may exploit the benefits of specialization and maintain a better patient flow. If the specialization were effective in reducing the LOS of the patients by 20%, the flexible structure would perform better than a general ICU in terms of rejected and rescheduled patients. The model was flexible for rapid adaptation to support decision making for ICU capacity expansion decisions during the COVID-19 pandemic.

ACKNOWLEDGMENTS

The authors thank Dr. Dave Dongelmans and Dr. Alexander Vlaar, intensivists, for helpful discussions about the operations of the ICU at the AMC and providing the admission data. Alban and Chick acknowledge the support of the European Union through the MSCA-ESA-ITN project (676129).

REFERENCES

Alban, A., S. E. Chick, D. A. Dongelmans, A. Van der Sluijs, W. J. Wiersinga, A. P. Vlaar, and D. Sent. 2020. "ICU Capacity Management during the COVID-19 Pandemic Using a Process Simulation". *Intensive Care Medicine* 46:1624–1626.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York, New York: Springer Science & Business Media.

Bruin, A., R. Bekker, L. Zanten, and G. Koole. 2010. "Dimensioning Hospital Wards Using the Erlang Loss Model". *Annals of Operations Research* 178:23–43.

Czaplinski, C., and D. Diers. 1998. "The Effect of Staff Nursing on Length of Stay and Mortality". *Medical Care* 36(12):1626–1638.

Hagen, M. S., J. K. Jopling, T. G. Buchman, and E. K. Lee. 2013. "Priority Queuing Models for Hospital Intensive Care Units and Impacts to Severe Case Patients". In *AMIA Annual Symposium Proceedings 2013*, 841–850. Bethesda, Maryland: American Medical Informatics Association.

Hamilton, D. K., and M. M. Shepley. 2010. *Design for Critical Care: An Evidence-Based Approach*. Abingdon, United Kingdom: Routledge.

Jordan, W. C., and S. C. Graves. 1995. "Principles on the Benefits of Manufacturing Process Flexibility". *Management Science* 41(4):577–594.

KC, D. S., and C. Terwiesch. 2011. "The Effects of Focus on Performance: Evidence from California Hospitals". *Management Science* 57(11):1897–1912.

Kramer, A. H., and D. A. Zygun. 2011. “Do Neurocritical Care Units Save Lives? Measuring the Impact of Specialized ICUs”. *Neurocritical Care* 14:329–333.

Lott, J., T. Iwashyna, J. Christie, D. Asch, A. Kramer, and J. Kahn. 2009. “Critical Illness Outcomes in Specialty versus General Intensive Care Units”. *American Journal of Respiratory and Critical Care Medicine* 179(8):676–683.

Marshall, J. C., L. Bosco, N. K. Adhikari, B. Connolly, J. V. Diaz, T. Dorman, R. A. Fowler, G. Meyfroidt, S. Nakagawa, P. Pelosi, J.-L. Vincent, K. Vollman, and J. Zimmerman. 2017. “What Is an Intensive Care Unit (ICU): A Report of the Task Force of the World Federation of Societies of Intensive and Critical Care Medicine”. *Journal of Critical Care* 37:270–276.

Skinner, W. 1974. “The Focused Factory”. *Harvard Business Review*:113–121.

Van der Sluijs, A. F., E. R. van Slobbe-Bijlsma, S. E. Chick, M. B. Vroom, D. A. Dongelmans, and A. P. Vlaar. 2017. “The Impact of Changes in Intensive Care Organization on Patient Outcome and Cost-Effectiveness–A Narrative Review”. *Journal of Intensive Care* 5(13):1–8.

Wolff, R. W. 1989. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall.

AUTHOR BIOGRAPHIES

ANDRES ALBAN is a PhD student in Technology and Operations Management at INSEAD. He has a BS degree from New Jersey Institute of Technology in Applied Mathematics and Physics. His PhD research is supported by the European Union through the Marie Skłodowska-Curie Innovative Training Network project European Sepsis Academy (676129). His email address is andres.alban@insead.edu.

STEPHEN E. CHICK is Professor of Technology and Operations Management and the Novartis Chair of Healthcare Management at INSEAD. He works in the areas of simulation analysis, sequential optimization, health care management, and Bayesian inference. He is an Associate Editor of the Simulation and Healthcare Management areas of the journal Management Science. His email address is stephen.chick@insead.edu.

OLEKSANDRA LVOVA is a MSc student in Medical Informatics at University of Amsterdam. She also has a MSc degree in Molecular Biology from the Kiev National University in Ukraine. Her email address is o.lvova@amsterdamumc.nl.

DANIELLE SENT is Assistant Professor of Medical Informatics at the Amsterdam UMC, University of Amsterdam. She studies health data science, more specifically prediction and causal models, and (clinical) decision support systems. Her email address is d.sent@amsterdamumc.nl.