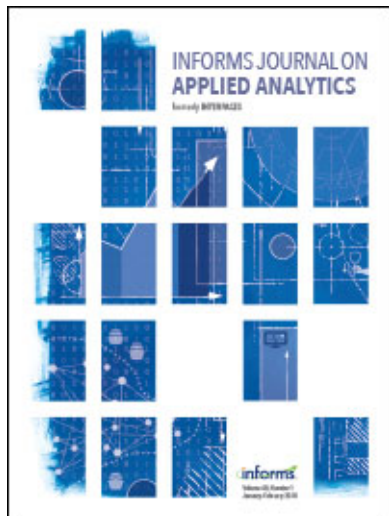




Interfaces

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

Allocating Emergency Beds Improves the Emergency Admission Flow

A. J. Thomas Schneider, P. Luuk Besselink, Maartje E. Zonderland, Richard J. Boucherie, Wilbert B. van den Hout, Job Kievit, Paul Bilars, A. Jaap Fogteloo, Ton J. Rabelink

To cite this article:

A. J. Thomas Schneider, P. Luuk Besselink, Maartje E. Zonderland, Richard J. Boucherie, Wilbert B. van den Hout, Job Kievit, Paul Bilars, A. Jaap Fogteloo, Ton J. Rabelink (2018) Allocating Emergency Beds Improves the Emergency Admission Flow. *Interfaces* 48(4):384–394. <https://doi.org/10.1287/inte.2018.0951>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Interfaces*. Copyright © 2018 The Author(s). <https://doi.org/10.1281/inte.2018.0951>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2018, The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Allocating Emergency Beds Improves the Emergency Admission Flow

A. J. Thomas Schneider,^{a,b} P. Luuk Besselink,^c Maartje E. Zonderland,^b Richard J. Boucherie,^b Wilbert B. van den Hout,^d Job Kievit,^d Paul Bilars,^e A. Jaap Fogteloo,^e Ton J. Rabelink^e

^a Department of Quality & Patient Safety, Leiden University Medical Center, Leiden 2333 ZA, Netherlands; ^b Center for Healthcare Operations Improvement and Research, University of Twente, Enschede 7500 AE, Netherlands; ^c ORTEC Consulting, Houston, Texas 77027; ^d Department of Medical Decision Making, Leiden University Medical Center, Leiden 2333 ZA, Netherlands; ^e Department of Internal Medicine, Leiden University Medical Center, Leiden 2333 ZA, Netherlands

Contact: a.j.schneider@lumc.nl,  <http://orcid.org/0000-0002-1326-2240> (AJTS); luuk.besselink@ortec.com (PLB); m.e.zonderland@utwente.nl,  <http://orcid.org/0000-0003-4888-9163> (MEZ); r.j.boucherie@utwente.nl,  <http://orcid.org/0000-0002-1046-2044> (RJB); w.b.van_den_hout@lumc.nl,  <http://orcid.org/0000-0002-6425-0135> (WBvdH); j.kievit@lumc.nl,  <http://orcid.org/0000-0002-9710-3799> (JK); p.bilars@lumc.nl (PB); a.j.fogteloo@lumc.nl (AJF); a.j.rabelink@lumc.nl,  <http://orcid.org/0000-0001-6780-5186> (TJR)

Received: April 20, 2016

Revised: November 11, 2016; June 30, 2017; December 13, 2017

Accepted: February 25, 2018

Published Online in Articles in Advance: September 12, 2018

<https://doi.org/10.1287/inte.2018.0951>

Copyright: © 2018 The Author(s)

Abstract. The increasing number of admissions to hospital emergency departments (EDs) during the past decade has resulted in overcrowded EDs and decreased quality of care. The emergency admission flow that we discuss in this study relates to three types of hospital departments: EDs, acute medical unit (AMUs), and inpatient wards. This study has two objectives: (1) to evaluate the impact of allocating beds in inpatient wards to accommodate emergency admissions and (2) to analyze the impact of pooling the number of beds allocated for emergency admissions in inpatient wards. To analyze the impact of various allocations of emergency beds, we developed a discrete event simulation model. We evaluate the bed allocation scenarios using three performance indicators: (1) the length of stay in the AMU, (2) the fraction of patients refused admission, and (3) the utilization of allocated beds. We develop two heuristics to allocate beds to wards and show that pooling beds improves performance. The partnering hospital has embedded a decision support tool based on the simulation model into its planning and control cycle. The hospital uses it every quarter and updates it with data on a 1-year rolling horizon. This strategy has substantially reduced the number of patients who are refused emergency admission.

History: This paper was refereed.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Interfaces. Copyright © 2018 The Author(s). <https://doi.org/10.1281/inte.2018.0951>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Keywords: acute medical unit • emergency department • inpatient wards • hospitals • emergency admissions • systems optimization • discrete-event simulation • length of stay • operations efficiency • decision support

Introduction

To avoid overcrowded emergency departments (EDs), congestion, and fluctuations in downstream resources, sophisticated (process) analysis is required (Hoot et al. 2008, Lee et al. 2015). Demographic changes and improved patient survival rates have contributed to the increasing number of hospitalizations of patients with complex and/or chronic conditions (Aminzadeh and Dalziel 2002). In addition, society demands cost-effective health care delivery, which puts pressure on the available resources. This results in bed utilization

rates above 85% in inpatient wards (Capewell 1996), leaving little slack in the admission flow and resulting in the refusal of patients for hospital admission (Bagust et al. 1999, Zhou et al. 2012). A major side effect of a decreasing number of available beds is the increasing number of boarders. Boarders are emergency patients waiting for admission or placed outside of their designated medical treatment specialty (which we call specialty in the remainder of this paper) ward because of bed unavailability (McMurdo and Witham 2013). In general, boarders have a significantly longer length of stay (LOS),

experience a decreased quality of care, are less satisfied, have increased mortality rates, and are associated with patient safety issues (Bernstein et al. 2009, Zhou et al. 2012, McMurdo and Witham 2013, Conway et al. 2014).

Bed capacity management focuses on allocating beds (and thus, staff) among patient types (i.e., emergency versus elective patients or patients whose treatments are within different specialties).

To improve the emergency admission flow, some hospitals use acute medical units (AMUs) (Li et al. 2010). “An AMU is a designated hospital ward specifically staffed and equipped to receive medical inpatients presenting with acute medical illness from EDs and outpatient clinics for expedited multidisciplinary and medical specialist assessment, care and treatment for up to a designated period (typically between 24 and 72 hours) prior to discharge or transfer to medical wards” (Scott et al. 2009, p. 389). From an operations management perspective, an AMU operates as a buffer. A buffer can operate as either an inflow buffer or an outflow buffer. An inflow buffer transforms a (highly) variable inflow into manageable outflow by accommodating all arrivals in the buffer before they are transferred farther downstream. An outflow buffer is used only if the initial server is fully occupied. In this study, we analyze AMUs that operate as inflow buffers, where the timing of transfers can be managed (between 24 and 72 hours) so that inpatient wards have time to make capacity available. As a result, downstream inpatient wards can attain higher bed utilization without increasing the number of patients refused. An AMU initially reduces pressure on ED utilization. However, in the case of structural lack of coordination between the AMU and downstream hospital wards, the AMU cannot transfer patients to the downstream hospital wards (Abraham and Reddy 2010), again resulting in an overcrowded AMU and ED (Scott et al. 2009). Ultimately, this increases the number of aforementioned boarders and contributes to the downward spiral of more emergency admission refusals.

We organized the remainder of this paper as follows. In the *Objectives* section, we explain the objectives of this study. The *Process Description* section presents the emergency admission flow process in more detail. In

the *Methods* section, we discuss our modeling approach and key performance indicators (KPIs). In the *Data* section, we discuss the data analysis required for the model input, and we follow with a description of the model in the *Model Implementation* section. We present our results in the *Results* section, and we discuss the implementation of our results in practice in the *Implementation in Practice* section. Finally, we discuss additional managerial implications, limitations, and potential extensions of our study in the *Discussion* section.

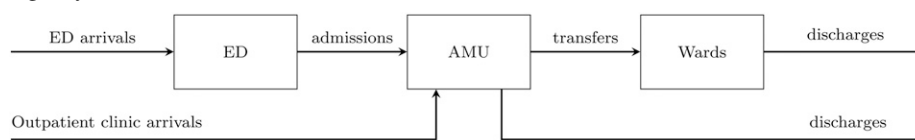
Objectives

To increase efficiency, the partnering hospital introduced an AMU. However, management of both the AMU and inpatient wards was still spending significant time in transferring patients from the AMU to downstream inpatient wards because of an insufficient number of beds in the wards and a lack of organizational guidelines and/or protocols. To simplify the transfer process, the hospital introduced the concept of allocated emergency beds; that is, each inpatient ward allocates a part of its bed capacity to accommodate patient transfers from the AMU. The first objective of this study is thus to evaluate the effect of allocating inpatient bed capacity for emergency admission patients. Because the board of directors of the partnering hospital had recently decided to restructure its inpatient wards into care units based on liaison specialties (i.e., pooled specialties, such as nephrology and endocrinology, that cooperate with each other), we formulated a second objective for this study: evaluate the effect of pooling wards on the required number of allocated emergency beds. The concept of pooling resources has been studied extensively (Mandelbaum and Reiman 1998, Cattani and Schmidt 2005); de Bruin et al. (2010) and Vanberkel et al. (2012) specifically studied it in a hospital setting. Ultimately, we wanted to structurally improve the emergency admission flow by implementing the model into the partnering hospital’s planning and control cycle.

Process Description

Figure 1 shows the basic patient flow that we are analyzing. This process flow includes three types of hospital

Figure 1. The Emergency Admission Flow



beds: (1) in the ED, (2) in the AMU, and (3) in the various downstream inpatient wards. It also includes two arrival streams: (1) from outside the hospital to the ED and (2) from the hospital outpatient clinic, thus circumventing the ED in the AMU. After a stay in the AMU, patients can either leave the hospital (i.e., be discharged) or be transferred to an inpatient ward. After a stay in the inpatient ward, a patient is discharged.

Below, we describe this process in more detail, including its logistical characteristics and decision points. Patients who require immediate medical care visit the ED. Therefore, ED arrivals come from outside the hospital at unpredictable rates. Arriving patients are seen and assessed provisionally. ED personnel identify patients who require a bed; those not needing a bed are treated and discharged. If a bed in the ED is available, the patient is assigned one, and the patient's priority status is determined by triage. Based on this priority status, an emergency room physician sees the patient immediately or at a later time. The attending physician determines an initial diagnosis and the primary medical specialty to which patient will be assigned. If no bed is available in the ED, the patient will be refused hospital admission and referred to a nearby hospital. ED arrivals who require an overnight stay are admitted to the AMU and if necessary, transferred later to an inpatient ward. In this study, we focus on emergency patients who are admitted to the hospital; therefore, we address only arrivals who are admitted to the ED.

After a random LOS in the ED, patients are admitted to the AMU when a bed becomes available. Otherwise, they wait in a bed in the ED. Admissions to the AMU can also originate from the hospital's outpatient clinics, thus circumventing the ED, when a physician at the outpatient clinic indicates that immediate hospitalization is required. Patients from the outpatient clinics are refused admission and referred to nearby hospitals when the AMU is occupied fully. In the AMU, patients are observed, diagnosed further, and given treatment if necessary. The LOS in the AMU is limited to 48 hours. This limit is a management decision of the partnering hospital, and it is applied with some flexibility. For example, a patient who has already spent 48 hours in the AMU and is expected to be discharged within 24 hours will remain in the AMU, because a transfer is time consuming for the AMU staff and stressful for patients. Additionally, patients are not transferred between 9 p.m. and 9 a.m., because staffing levels in the downstream inpatient wards are minimal during these hours. If further treatment is necessary, patients are

transferred to other inpatient wards depending on their treatment specialty and stay there for a random LOS also depending on the treatment specialty, after which they are discharged. Patients can only transfer to their destination inpatient ward if an emergency bed is available in that ward. We assume that beds that are allocated for emergency admissions cannot be used by elective patients and vice versa.

The ED and AMU dedicate beds for the sole use of emergency patients, whereas wards also have beds for elective patients. Given the objectives of our study, one could be tempted to focus on the effect of the number of allocated emergency beds in inpatient wards and model only the wards; however, this would not capture the patient flow through the AMU. In addition, to explain the process to the stakeholders, we want to show the effects of the allocated emergency beds on all departments involved. In this study, we, therefore, include the impact on the ED, AMU, and inpatient wards.

Methods

In this section, we explain our modeling approach and KPIs for analyzing scenarios.

Model Approach

We use discrete event simulation (DES) to analyze the emergency admission flow, because analytical modeling of the nonhomogeneous interarrival times, the different lengths of service per ward and specialty, and the time interval in which patients can be transferred from the AMU are analytically intractable. DES also provides a visual representation of the process for implementation purposes. This approach is widely used for decision making and planning in health care; for example, see the online reference database described in Hulshof et al. (2011) and the systematic reviews of Güral and Pidd (2010), Paul et al. (2010), and Hulshof et al. (2012).

Performance Indicators

We formulate the following KPIs as output of our simulation model: (1) the relative LOS in the AMU, (2) the fraction of patient arrivals refused, and (3) the utilization of the beds allocated in the inpatient wards. The first KPI is an accurate parameter to measure the level of throughput (Yoon et al. 2003). When patients cannot be transferred to an inpatient ward, the LOS will increase immediately. The relative LOS is defined as the ratio of the average LOS in the AMU divided by the average LOS in the AMU in the case of unlimited capacity. We define the term as relative LOS so that we

can directly interpret the factor that causes the scenario to improve or worsen and compare it with an unlimited capacity scenario; in this scenario, the waiting time is marginal when patients complete their LOSs between 9 p.m. and 9 a.m. and therefore, must wait until 9 a.m. to be transferred. The second KPI is the fraction of patients refused admission (in relationship to the total number of patients), and it is an accurate measure of a full system (i.e., no beds are available). The third KPI is the average utilization of the allocated beds for each inpatient ward and the beds in the AMU, and it provides information about potential bottlenecks.

Data

The model requires the following input data:

- Arrival rates in the ED and AMU per hour;
- LOS in the ED, AMU, and hospital wards based on medical specialty;
- Distribution of the number of admissions per specialty;
- Transfer rates per specialty from the AMU to the inpatient wards or of patients who have been discharged; and

- Number of emergency beds allocated in the ED, AMU, and wards (in the ED and AMU, all beds are available for emergency admissions).

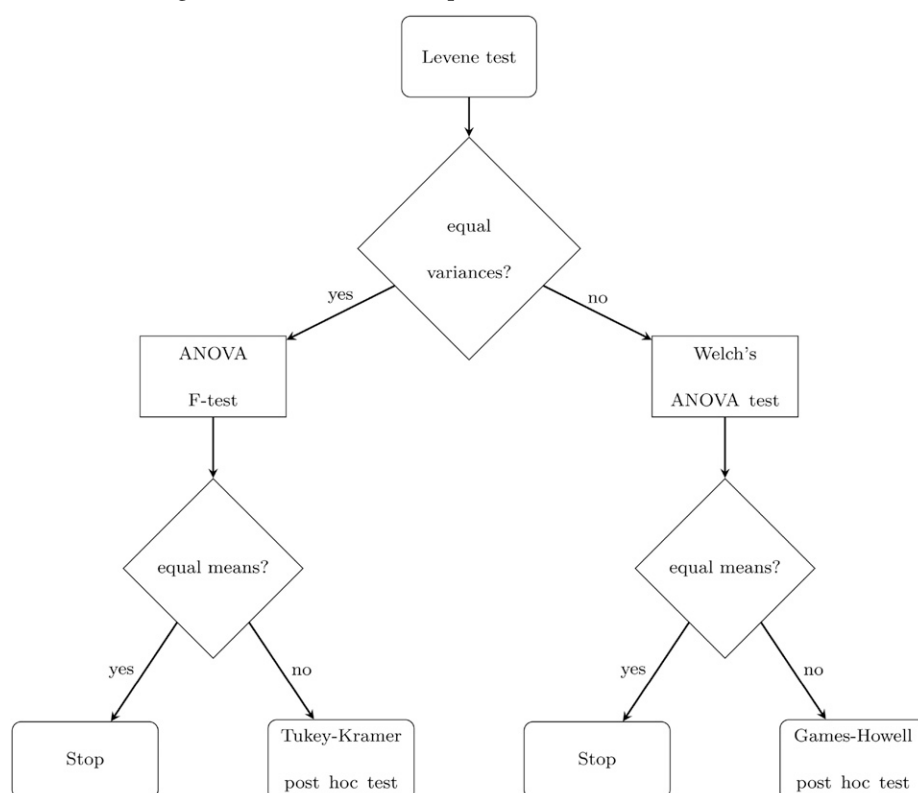
We obtained patient data from the hospital's management information system. The data set consists of 4,446 admissions in the AMU between 2012 and 2014. To overcome the high diversity in specialties and patient flows, we considered only the top 99% of admissions and excluded the remaining 1%, which are atypical cases in terms of specialty and/or ward. This resulted in seven hospital wards (i.e., a 67% reduction) and six specialties (i.e., a 50% reduction), thus significantly simplifying our calculations.

Data Analysis

The data analysis serves the following purposes: (1) finding the distribution of specialties and patient flows (to which ward patients are transferred from the AMU), (2) clustering patient groups, (3) fitting clustered patient groups to probability distributions for modeling the LOS, and (4) determining the arrival patterns.

The distribution of specialties and patient flows are based on the historical data using frequency tables.

Figure 2. The Statistical Clustering Process of Patient Groups



Patients whose treatment is within the same specialty can be transferred to different wards; however, this could increase the specialty and ward options and thus, the complexity of our model. To keep our model as simple as possible in terms of options and increase the statistical power of the samples for fitting the probability distributions, we cluster patient groups from two perspectives: (1) the LOS of patients who are in different wards but whose care is within the same medical specialty and (2) the LOS of patients in the same ward but whose care is under a different specialty. Patients treated under the same specialty could have a similar LOS because of the similar nature of their diseases or injuries and/or because they are treated by the same staff. Our clustering process is based on the logic in Figure 2.

First, we execute a Levene test (Levene 1960) to identify differences between sample variances. If the results of the Levene test show unequal variances between samples, an analysis of variance (ANOVA) F test (Box 1953) with a Welch test statistic (Welch 1947) is required; otherwise, we perform a normal ANOVA F test to determine unequal means between samples. If the ANOVA F test shows significantly unequal means, the final step in the clustering process

is a post hoc test to analyze which samples are significantly different compared with the samples that share the same mean and variance. The post hoc test used depends on the results of the Levene test. If we find unequal variances and unequal means between samples, a Games–Howell test (Games and Howell 1976) is required to determine significant differences between the samples. For samples with equal variances but unequal means, we perform a Tukey range test (Tukey 1949). All tests are performed using SPSS, an IBM statistical software package. The results of this clustering process showed that two hospital wards have equal LOS, independent of the specialties on these wards, and that one specialty has the same LOS, independent of its designated wards. All other wards and specialties have significantly different means.

Using the outcomes of the clustering process, we fit probability distributions to each clustered patient group. For this, we use Rockwell’s ARENA Input Analyzer based on goodness-of-fit tests (i.e., the chi-squared test). The following probability distributions are used for fitting: Gamma, Erlang, Exponential, and Lognormal. Because of the limit (i.e., 48 hours in the AMU), the historical data on the LOS in the AMU display a specific gradient, where the

Figure 3. Frequencies of ED Arrivals per Hour

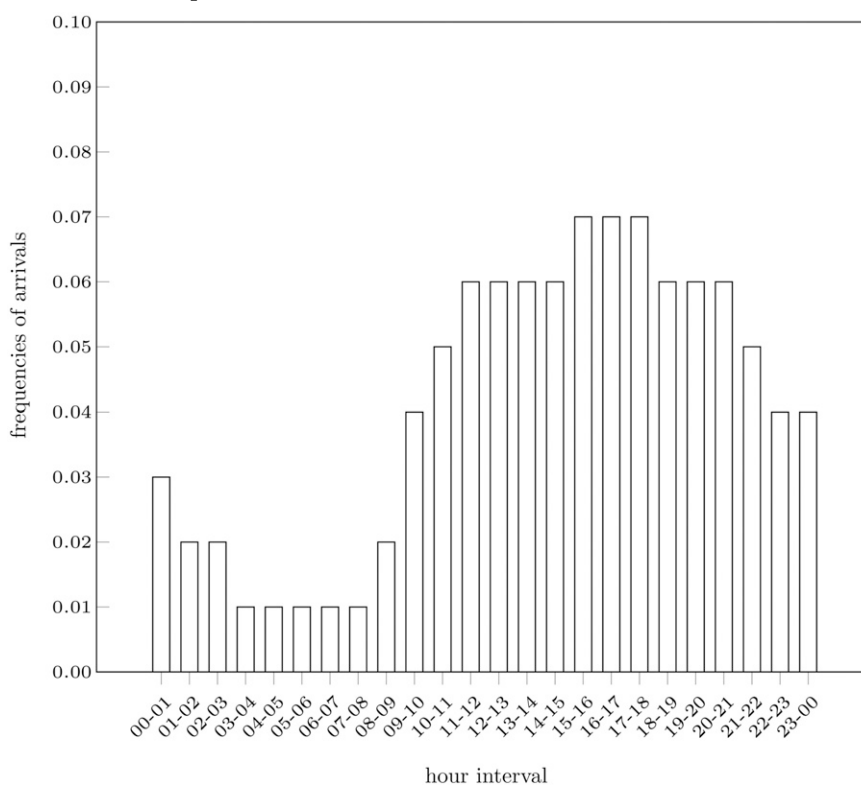


Table 1. Heuristic 1 Locates a Feasible Allocation of Emergency Beds in Inpatient Wards Using the Process Outlined

Step	Phase	Explanation
1	Initialization phase	Set allocated emergency bed capacity at 100 for each ward (approximating unlimited capacity)
2	Base phase	Set capacity to average occupied beds per ward from Step 1
3	Optimization phase	Increase capacity of ward with highest utilization rate
4	Iteration phase	Repeat Step 3 until outcomes of the initialization phase are approached sufficiently (i.e., arbitrary maximum deviation of 3% from the relative LOS at AMU)

probability mass is centered around 24 and 48 hours. Therefore, we used the empirical distribution derived from the historical data to model the LOS in the AMU.

Because the arrival of emergency patients cannot be scheduled, we want to find the arrival patterns. We can see from Figure 1 that the process has two arrival streams: (1) arrival in the ED and (2) arrival in the AMU from the hospital's outpatient clinics. Data analysis shows that patients arrive according to a nonhomogeneous process. For example, peak hours are between 3 p.m. and 8 p.m. We, therefore, determine hourly arrival rates based on the historical data and assume that arrivals occur according to a nonhomogeneous Poisson process, such as is common practice in modeling unscheduled patient arrivals (Swartzman 1970). We do not consider seasonality or differences in weekdays and weekends in the arrival rates. As an example, Figure 3 shows the daily ED arrival frequencies.

Model Implementation

We modeled the DES using Tecnomatix plant simulation software from Siemens. The model has a generic setup; therefore, various configurations (e.g., the number of wards and beds, medical specialties, patient flows, and LOS) can be analyzed without changing the core design of the model.

Arrivals in the ED or AMU are assigned a specialty and a specific destination (e.g., after a stay in the AMU, the patient will be discharged or transferred to a hospital ward for further treatment) according to a single random Bernoulli trial using probabilities derived from the frequency tables mentioned in the *Data Analysis* section. The LOS at each department is based on the

department's medical specialties. When new patients arrive in the ED or AMU and all beds are occupied, they are refused and leave the system. When patients are ready to transfer from the ED to the AMU or from the AMU to an inpatient ward and the destination is occupied, they wait at their current location.

Simulation Initialization

To obtain statistically reliable results from our simulation, we need to initialize our model. We start initializing the simulation with all parameters derived from the data analysis: frequency tables for specialties and destinations, the probability distributions for the LOS, and the arrival patterns. We also need to dimension the ED, AMU, and wards according to the partnering hospital's practices. The ED and the AMU have eight and 24 allocated beds, respectively, and a patient can be transferred to one of seven wards (the number of allocated beds in wards will vary for each scenario).

Because the simulation starts with an empty system (i.e., no patients are present), a warm-up period is required to reach steady state. We, therefore, exclude all results from the warm-up period. The length of the warm-up period is determined using the Welch method. This method plots moving averages of the means from the i th observation for a number of replications and an arbitrarily long run length per KPI. The mean of multiple replications of the i th observation smooths the variability over individual observations and therefore, gives insights into the dependency on the initial state. We then use moving averages over these means to smooth out high-frequency oscillations. The warm-up period is determined through a graphical interpretation of the

Table 2. Heuristic 2 Locates a Feasible Allocation of Emergency Beds in Care Units (i.e., Pooled Inpatient Wards) Using the Process Outlined

Step	Phase	Explanation
1	Initialization phase	Set allocated emergency bed capacity of care units equal to capacity of pooled wards (Table 3)
2	Optimization phase	Decrease capacity (i.e., number emergency beds) of care unit with lowest utilization rate
3	Iteration phase	Repeat Step 2 until outcomes of separate wards are approached sufficiently (arbitrary percentage of patients refused < 0.01)

Table 3. Input Parameters and Results for Each Scenario for Allocated Emergency Beds in Individual Wards

Scenario	AMU			W1		W2		W3		W4		W5		W6		W7		Total beds
	<i>B</i>	Rel LOS	ρ	Beds	ρ	Beds	ρ	Beds	ρ	Beds	ρ	Beds	ρ	Beds	ρ	Beds	ρ	
Init	0.00	1.00	0.35	100	0.09	100	0.04	100	0.02	100	0.02	100	0.01	100	0.02	100	0.01	700
Avg	0.16	2.58	0.95	9	0.81	4	0.78	2	0.97	2	0.85	1	0.72	2	0.75	1	0.80	21
A	0.15	2.30	0.89	9	0.86	4	0.82	3	0.68	2	0.89	1	0.76	2	0.79	1	0.85	22
B	0.12	2.17	0.87	9	0.88	4	0.84	3	0.70	3	0.61	1	0.78	2	0.81	1	0.87	23
C	0.09	2.11	0.85	10	0.80	4	0.86	3	0.70	3	0.61	1	0.78	2	0.83	1	0.88	24
D	0.07	1.97	0.80	10	0.81	4	0.88	3	0.71	3	0.61	1	0.79	2	0.84	2	0.45	25
E	0.05	1.77	0.76	10	0.83	5	0.71	3	0.73	3	0.62	1	0.80	2	0.86	2	0.45	26
F	0.01	1.48	0.64	10	0.85	5	0.74	3	0.75	3	0.65	1	0.84	3	0.60	2	0.48	27
G	0.01	1.19	0.61	11	0.80	5	0.71	3	0.77	3	0.67	1	0.85	3	0.67	2	0.49	28
H	0.01	1.09	0.54	11	0.80	5	0.71	3	0.78	3	0.68	2	0.43	3	0.67	2	0.50	29
I	0.00	1.07	0.52	12	0.74	5	0.71	3	0.78	3	0.68	2	0.43	3	0.67	2	0.68	30
J	0.00	1.05	0.49	12	0.74	5	0.71	4	0.59	3	0.68	2	0.43	3	0.67	2	0.50	31
K	0.00	1.04	0.48	13	0.68	5	0.71	4	0.59	3	0.68	2	0.43	3	0.67	2	0.50	32
L	0.00	1.03	0.46	13	0.68	6	0.59	4	0.59	3	0.68	2	0.43	3	0.67	2	0.50	33

Notes. Bold indicates adjusted bed capacity compared with the previous scenario. Avg, average; *B*, percentage of refused patients; Init, initial; rel LOS, relative LOS in the AMU; ρ , bed utilization; *W_i*, ward *i*.

plotted moving averages per KPI, resulting in a length of 365 days.

We determine the run length using the convergence method of Robinson (2004). This method implies convergence of the cumulative means of KPIs over multiple replications as the run length increases. The convergence level is measured as the ratio of the positive difference between the maximum and the minimum of the cumulative mean of all replications until day *t* divided by the same maximum of the cumulative mean of all replications until day *t*. We set the convergence level to 0.01, resulting in a run length of 3,250 days. We then rounded up the run length to 10 years (i.e., 3,650 days).

The final step is to determine the required number of replications. We use a stopping criterion on the relative error of the aforementioned KPIs. We set the relative error bound to 1%, and five replications proved to be sufficient.

Heuristics

In the simulation model, we analyze different distributions of emergency beds allocated in inpatient wards. To locate a feasible solution, we developed one heuristic per objective (Tables 1 and 2).

In the first heuristic, we start with an unlimited bed capacity (the initialization phase of the heuristic and scenario Init in Table 3). This scenario shows the best

Table 4. The Input Parameters and Results for Each Scenario for Emergency Beds Allocated in Care Units

Scenario	<i>B</i>	AMU		CU1		CU2		CU3		Total beds
		Rel LOS	ρ	Beds	ρ	Beds	ρ	Beds	ρ	
Init	0.00	0.79	0.36	19	0.66	7	0.59	7	0.51	33
I	0.00	0.79	0.36	19	0.66	7	0.59	6	0.60	32
II	0.00	0.81	0.37	19	0.66	6	0.69	6	0.60	31
III	0.00	0.85	0.39	19	0.66	6	0.69	5	0.72	30
IV	0.00	0.86	0.39	18	0.69	6	0.69	5	0.72	29
V	0.00	0.87	0.40	17	0.73	6	0.69	5	0.72	28
VI	0.00	0.98	0.45	17	0.73	5	0.84	5	0.72	27
VII	0.01	1.26	0.57	17	0.72	5	0.83	4	0.89	26
VIII	0.01	1.27	0.58	16	0.77	5	0.83	4	0.89	25
IX	0.01	1.30	0.59	15	0.82	5	0.82	4	0.89	24

Note. *B*, percentage of refused patients; *CU_i*, care unit *i*; Init, initial; ρ , bed utilization; rel LOS, relative LOS in the AMU.

performances given the unlimited capacity, and therefore patients have only marginal waiting times as a results of the time window between 9 p.m. and 9 a.m. when transfers cannot take place. In this scenario, we found an upper bound of our solution space. The distribution of beds among inpatient wards for the next phase of the heuristic (the base phase) is based on the average utilization of the initialization phase. The base phase provides a lower bound for our solution space. Using these averages, we do not consider the stochasticity of the process; therefore, this scenario (Avg in Table 3) is characterized by underperformance. In the next phase of the heuristic (the optimization phase), we, therefore, analyze which ward is the bottleneck (i.e., the ward with the highest utilization rate) and increase the number of allocated beds in this ward by one. The heuristic stops iterating when the stopping criterion is met. We arbitrarily chose a maximum deviation of 0.03 from the relative LOS in the AMU of the initialization phase as stopping criterion.

The second heuristic (Table 2) starts again with an initialization phase, where we use the solution of the first heuristic as input. Because we suspect that pooling resources will improve performance (e.g., the required bed capacity will be lower), we want to know which care unit has the lowest utilization rate. The heuristic now decreases the capacity of the care unit with the lowest utilization rate (i.e., optimization phase) and iterates until the stopping criterion is met (iteration phase).

Results

The first objective of this study is to evaluate the effect of allocating beds in inpatient wards for patients included in the emergency admissions flow. Using the simulation model and heuristic for this objective, we analyzed 14 scenarios. In Table 3, we list these scenarios, including their input parameters and output values for the KPIs. Per scenario (i.e., a row in the table), the bed capacity of inpatient ward i with the highest utilization is increased by one bed. We graphically show this using bold numbers.

The last row in Table 3 shows that 33 allocated emergency beds are required to achieve performance similar to that in the Init scenario. However, the bed utilization per ward does not exceed 70%, which is low.

To evaluate the second objective, we pooled the wards according to the configuration of the partnering hospital and applied the second heuristic. The heuristic starts with the initialization phase (scenario Init in Table 4). Per scenario (i.e., a row in the table), the bed capacity of the care unit with the lowest utilization rate

is decreased by one bed, which we show graphically using bold numbers, and stops when the stopping criterion is met. Care Unit 1 consists of the Wards 1 and 2, Care Unit 2 consists of Wards 3 and 4, and Care Unit 3 consists of Wards 5–7.

The results (Table 4) show that pooling resources by allocating beds among wards further improves the outcomes. In the best-performing scenario using the first heuristic for separate inpatient wards, 33 allocated emergency beds are required. When pooling inpatient wards, the required number of emergency beds decreases to 24 without a significant decrease in performance (based on the KPIs).

Implementation in Practice

This research started with a request from AMU management to analyze the bottlenecks in the AMU's patient flows. Because these flows are multidepartmental, we also involved the management of the other hospital departments (e.g., ED and inpatient wards). After we reached consensus about the definition of the problem and potential solutions, we started constructing the simulation model. AMU management was involved in every step of the simulation study, and we discussed the results with the management of all involved departments and the hospital's board of directors.

We also discussed the structure related to tactical decision making using the outcomes of the simulation model, which is now embedded in the planning and control cycle of the partnering hospital. At the beginning of every quarter, the hospital reevaluates the distribution of the emergency beds allocated in the wards. It implements the results of this reevaluation at the beginning of the next quarter; that is, ward managers adjust the number of emergency beds allocated (at the expense of beds available for elective patients). This allows management to adjust other resources (principally staffing levels) and adjust its planning for elective patients based on the reevaluation. For the evaluation process, we use a 1-year rolling horizon of data. Our experience shows that an adjustment of zero to three beds per inpatient ward is required each quarter.

As we mention above, ward management must work with and manage multiple stakeholders. In practice, management could prioritize elective patients and therefore, not completely adjust the number of allocated emergency beds that our model suggests. Overall, this model resulted in a 70% decrease in the number of patients refused admission, while elective patient admissions also increased.

Discussion

In this study we show the positive impact of allocating emergency beds on ED crowding. This allocation strategy eliminates boarders and therefore, makes a positive contribution to the quality of care that can be expected (Singer et al. 2011). It may also improve patient satisfaction, because crowding is associated with lower patient satisfaction (Pines et al. 2011). One of our primary objectives during the development phase was to create a model that could be generalized. Our model can easily be adapted to different settings by adding wards or specialties, varying the numbers of beds, and varying the LOSs. The model could, therefore, be useful for other hospitals facing similar problems. It is a user-friendly planning tool, granting (medical) management in hospitals the power to determine the optimal number of allocated emergency beds relative to their flow dynamics and resource utilization. It provides immediate input for interdepartmental alignment; that is, it provides information about bottlenecks to ensure that congestion in one department does not cause bottlenecks in other departments that are involved in the same patient flow. It also enables managers to evaluate capacity decisions on patient flow, thus simplifying the tactical capacity decisions that they must make. The DES model is a universal and powerful tool that supports the planning and control process. The partnering hospital uses the tool for tactical decision making and has completely integrated it into the hospital's practices.

Allocating a shared resource (e.g., beds) for specific populations could result in suboptimal utilization, because it reduces flexibility. This study addresses this problem by analyzing the effect of pooling capacity (e.g., the pooling of emergency beds in inpatient wards). The results show a significant improvement in bed utilization without decreased performance in either the fraction of patients refused hospital admission or the LOS in the AMU. This supports earlier research that shows that pooling can improve both bed utilization in hospital wards and the numbers of patients who are refused hospital admission (de Bruin et al. 2010, Vanberkel et al. 2012).

Including more case studies is necessary to determine the correlation between allocated emergency beds, flow congestion, and boarders. A limitation of the model is the empirical distribution of the AMU LOSs. Substituting beds between emergency patients and elective patients is complex, because factors, such as the length of the waiting list and the seriousness of

elective patient conditions, are also important. Our study does not currently address other factors, such as staffing levels, that also influence performance. To improve both the accuracy of the results and model validation, further research is necessary on the arrival patterns, including analyzing differences between arrival days and hours and predictors for the AMU LOS. In addition, factors that prevent a patient from being discharged immediately (e.g., a potential transfer to a nursing home) are likely causes of significantly longer LOSs in inpatient wards, thus resulting in the potential risk of refusals in the ED and AMU. We looked solely at the configuration of the partnering hospital. Therefore, analyzing different configurations of pooled wards could be a future research topic. Finally, the model that we present could be used to facilitate capacity allocation decisions of emergency patients from a regional perspective. In addition, the simulation model can visually show (to management) the flow dynamics when allocating beds for emergency patients; thus, it can increase the likelihood of successful implementation.

This research shows not only that allocating beds for emergency patients at hospital wards improves the emergency admission flow but also, that its implementation into a tool helps elucidate the pros and cons of this allocation and thus, facilitates implementation.

References

- Abraham J, Reddy MC (2010) Challenges to inter-departmental coordination of patient transfers: A workflow perspective. *Internat. J. Med. Informatics* 79(2):112–122.
- Aminzadeh F, Dalziel WB (2002) Older adults in the emergency department: A systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann. Emerg. Med.* 39(3): 238–247.
- Bagust A, Place M, Posnett JW (1999) Dynamics of bed use in accommodating emergency admissions: Stochastic simulation model. *BMJ* 319(7203):155–158.
- Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, McCarthy M, et al. (2009) The effect of emergency department crowding on clinically oriented outcomes. *Acad. Emerg. Med.* 16(1):1–10.
- Box GEP (1953) Non-normality and tests on variances. *Biometrika* 40(3):318–335.
- Capewell S (1996) The continuing rise in emergency admissions. *BMJ* 312(7037):991–992.
- Cattani K, Schmidt GM (2005) The pooling principle. *INFORMS Trans. Ed.* 5(2):17–24.
- Conway R, O'Riordan D, Silke B (2014) Long-term outcome of an AMAU—a decade's experience. *QJM* 107(1):43–49.
- de Bruin AM, Bekker R, van Zanten L, Koole GM (2010) Dimensioning hospital wards using the Erlang loss model. *Ann. Oper. Res.* 178(1): 23–43.

- Games PA, Howell JF (1976) Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *J. Ed. Behav. Statist.* 1(2):113–125.
- Günl MM, Pidd M (2010) Discrete event simulation for performance modelling in health care: A review of the literature. *J. Simulation* 4(1):42–51.
- Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, Aronsky D (2008) Forecasting emergency department crowding: A discrete event simulation. *Ann. Emerg. Med.* 52(2):116–125.
- Hulshof PJH, Kortbeek N, Boucherie RJ, Hans EW, Bakker PJM (2012) Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems* 1(2):129–175.
- Hulshof PJH, Boucherie RJ, van Essen JT, Hans EW, Hurink JL, Kortbeek N, Litvak N, et al. (2011) ORchestra: An online reference database of OR/MS literature in health care. *Health Care Management Sci.* 14(4):383–384.
- Lee EK, Atallah HY, Wright MD, Post ET, Thomas C, Wu DT, Haley LL (2015) Transforming hospital emergency department workflow and patient care. *Interfaces* 45(1):58–82.
- Levene H (1960) Robust tests for equality of variances. Olkin I, Ghurye SG, Hoefding W, Madow WG, Mann HB, eds. *Contributions to Probability and Statistics* (Stanford University Press, Stanford, CA), 278–292.
- Li JYZ, Yong TY, Bennett DM, O'Brien LT, Roberts S, Hakendorf P, Ben-Tovim DT, Phillips PA, Thompson CH (2010) Outcomes of establishing an acute assessment unit in the general medical service of a tertiary teaching hospital. *Med. J. Australia* 192(7):384–387.
- Mandelbaum A, Reiman MI (1998) On pooling in queueing networks. *Management Sci.* 44(7):971–981.
- McMurdo MET, Witham MD (2013) Unnecessary ward moves. *Age Ageing* 42(5):555–556.
- Paul SA, Reddy MC, DeFlitch CJ (2010) A systematic review of simulation studies investigating emergency department overcrowding. *Simulation* 86(8–9):559–571.
- Pines JM, Batt RJ, Hilton JA, Terwiesch C (2011) The financial consequences of lost demand and reducing boarding in hospital emergency departments. *Ann. Emerg. Med.* 58(4):331–340.
- Robinson S (2004) *Simulation: The Practice of Model Development and Use* (Wiley, Chichester, UK).
- Scott I, Vaughan L, Bell D (2009) Effectiveness of acute medical units in hospitals: A systematic review. *Internat. J. Quality Health Care* 21(6):397–407.
- Singer AJ, Thode HC, Viccellio P, Pines JM (2011) The association between length of emergency department boarding and mortality. *Academic Emerg. Med.* 18(12):1324–1329.
- Swartzman G (1970) The patient arrival process in hospitals. *Health Services Res.* 5(4):320–329.
- Tukey JW (1949) Comparing individual means in the analysis of variance. *Biometrics* 5(2):99–114.
- Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, Litvak N (2012) Efficiency evaluation for pooling resources in health care. *OR Spectrum* 34(2):371–390.
- Welch BL (1947) The generalization of student's problem when several different population variances are involved. *Biometrika* 34(1–2):28–35.
- Yoon P, Steiner I, Reinhardt G (2003) Analysis of factors influencing length of stay in the emergency department. *Canadian J. Emerg. Med. Care* 5(3):155–161.
- Zhou JC, Pan K-H, Zhou D-Y, Zheng S-W, Zhu J-Q, Xu Q-P, Wang C-L (2012) High hospital occupancy is associated with

increased risk for patients boarding in the emergency department. *Amer. J. Med.* 125(4):416.e1–416.e7.

Verification Letter

J.C. von Rossum, Manager Care, Division 2, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, Netherlands, writes:

"Thank you for your interest in our OR/MS application efforts. I am pleased to write this letter to verify Thomas (A. J.) Schneider's paper submission of "Dedicated Emergency Beds Optimize the Emergency Admission Flow".

The use and benefits described in this paper are accurate. Thomas's use of Discrete Event Simulation has had a significant impact on our business. The model has been implemented in the planning & control cycle. It uses a one year rolling horizon, where data is updated and model settings are re-evaluated every quartile. Due to this Discrete Event Simulation model, unwanted refusals—refused patients who fit our strategy—have been reduced to marginal numbers. This research also gave many insights in the interdependencies between different departments involved.

Not only did this application reduce the number of unwanted refusals, it also—and perhaps most importantly—added value to our perception of processes. Introducing these kinds of OR/MS techniques showed us the potential of process and planning optimization, ultimately improving our quality of care.

As a final note, I would like to take this opportunity to mention that Thomas, as a part time PhD student, devotes special attention to the implementation of OR/MS techniques in hospitals. Therefore this paper fits perfectly with his ongoing research. I would absolutely recommend this work as an excellent case for publication and I feel fortunate to have Thomas as an analytic force in our organization."

Thomas Schneider is a PhD student in the Department of Quality and Patient Safety at Leiden University Medical Center (LUMC), the Netherlands, and CHOIR at the University of Twente, Enschede, the Netherlands, and is also a senior advisor of healthcare logistics at the Department of Internal Medicine at LUMC. He received his graduate degree in Industrial Engineering and Management at the University Twente in 2011. His main research interests are healthcare operations research and capacity management.

Luuk Besselink is a managing consultant of the Houston branch of ORTEC. ORTEC specializes in providing advanced analytics and optimization solutions for companies to innovate and outperform. In 2015, Luuk received his graduate degree in Econometrics and Operations Research at the Vrije Universiteit, Amsterdam, the Netherlands. Ever since, he has been working closely with clients to deliver custom-made decision support solutions using operations research and advanced analytics techniques.

Maartje Zonderland holds a PhD degree in operations research and statistics from the University of Twente, Enschede,

the Netherlands, and is senior researcher at CHOIR at the University of Twente. She is manager at the Dutch healthcare insurer Menzis, focusing on improving patient flow and healthcare quality. Her research interests are outpatient clinic optimization and semiurgent patient planning.

Richard Boucherie is professor of Stochastic Operations Research, chair of the Center for Healthcare Operations Improvement and Research, and founder of the spin-off consultancy Rhythm of the University of Twente. He received graduate degrees in mathematics and physics from the Leiden University, the Netherlands, and a PhD degree in econometrics from the Vrije Universiteit, Amsterdam. His research interests are queueing theory, Petri nets, and random walks and healthcare operations research.

Wilbert van den Hout is a health economist at LUMC, the Netherlands. He graduated with a degree in econometrics at the University of Amsterdam, the Netherlands, and obtained his PhD at the University of Tilburg, the Netherlands. He is a senior researcher at the Biomedical Data Sciences Department of the LUMC. His main research interests are methodological issues, statistical analysis and decision making in healthcare economic evaluation, and mathematical modeling of disease processes.

Job Kievit is emeritus professor of Quality in Healthcare, former professor of Clinical Decision Analysis and was head of Endocrine and Head and Neck Surgery at LUMC, the Netherlands. He obtained his MD and his PhD from the Erasmus University Rotterdam, the Netherlands. Throughout

his career, his interest has focused on the tension between the macro and micro levels of care.

Paul Bilars is managing director of the Department of Internal Medicine, LUMC, the Netherlands. He obtained his graduate degree in Law and Public Administration at Leiden University, the Netherlands. He is also managing director of the Cardiology Center Voorschoten, a board member of the Meteor Foundation, and CEO of the Leiden Regenerative Medicine Platform. His main interest is translating innovative research into actual treatments.

Jaap Fogteloo is section head of Acute Medicine of the Department of Internal Medicine, LUMC, the Netherlands, and is supervisor of the residency training program in acute internal medicine. He is also medical manager of the acute medical unit. He received his MD degree at the University of Groningen, the Netherlands, and his PhD degree at LUMC. His current research focuses on acute elderly patients and clinical decision making in acute medicine using electronic vital function monitoring.

Ton Rabelink is professor and Chair of Nephrology and Division Head of Internal Medicine at the LUMC, the Netherlands. He received his MD and PhD degree at Utrecht University, the Netherlands, and also completed fellowships in internal medicine and nephrology. His main research interests are vascular biology for renal function, organ transplantation, and cardiovascular diseases in patients with renal disease, as well as regenerative medicine with the use of stem cells.