

Needy boarding patients in emergency departments: defining a control policy for the physicians

De Boeck K, Carmen R, Vandaele N.



Needy boarding patients in emergency departments: defining a control policy for the physicians

Kim De Boeck¹, Raïsa Carmen, Nico Vandaele

Research Centre for Operations Management,

Faculty of Economics and Business,

KU Leuven, Naamsestraat 69, Leuven, Belgium

E-mail: kim.deboeck@kuleuven.be, raisa.carmen@kuleuven.be, nico.vandaele@kuleuven.be

Abstract. Boarding patients and the extra workload they introduce are a major concern in emergency departments. Not in the least because they confront the physicians with a challenging task: prioritizing between boarding patients and patients currently under treatment in the emergency department. The main contribution of this paper is the examination of different control policies for the physicians when needy boarding patients are added to the analysis. Using discrete-event simulation, three static control policies (first-come, first-served and always prioritizing either boarding patients or the other patients) and two dynamic control policies (using threshold values and accumulating priorities) are studied. For operational system performance, the recommended control policy is simple and straightforward: never prioritize boarding patients. However, in an emergency department setting, health-related performance measures also need to be considered: physicians cannot disadvantage one type of patients in favour of operational system performance. The result is a trade-off between operational system performance measures and health-related performance measures. Furthermore, we conclude that applying a first-come, first-served policy performs extremely well in a wide range of situations.

Key words: OR in health services, decision support for practice, simulation, needy boarding patients, prioritization

¹ Corresponding author

1. Introduction

A serious issue often encountered in hospital emergency departments (EDs) is crowding. Crowding occurs when the demand for emergency resources exceeds the resources available in the ED (Moskop, Sklar, Geiderman, Schears & Bookman, 2009). This can cause several negative effects including longer waiting time and length of stay (LOS), lost hospital revenue, increased ambulance diversion, more patients that leave without being seen (LWBS), and even higher mortality rate (Hoot & Aronsky, 2008).

One of the main contributors to crowding is inpatient boarding which results from the inability to transfer patients from the ED to the inpatient ward (IW) (Moskop et al., 2009). As a consequence, these boarding patients need to wait in the ED. While waiting, they require treatment from the ED staff and occupy beds which, as a consequence, cannot be used by newly arriving patients. In this way, boarding patients congest the ED.

Although the importance of inpatient boarding is stated multiple times in the operations research (OR) and operations management (OM) literature, very little research focuses on this topic (Saghafian, Austin & Traub, 2015). Furthermore, when boarding is included in the analysis, it is often assumed that boarding patients occupy beds in the ED, but do not require any other resources. However, in reality, boarding patients still need additional check-ups by ED physicians, can require time from ED nurses or occupy other ED equipment than beds (Armony, Israelit, Mandelbaum, Marmor, Tseytlin & Yom-Tov, 2015).

This paper intends to investigate the effect of this simplification by incorporating the additional check-ups required by boarding patients into the analysed system. Adding these required check-ups increases the workload of the physicians, but also raises an additional question. That is, after finishing a treatment, the physician needs to decide whether he or she will see a boarding patient or one of the other patients in the system. Since the treatment characteristics of these two kinds of patients differ, this decision has an important impact on system performance. Moreover, in an ED setting, prioritization decisions also have an outspoken health aspect that should not be neglected. Indeed, a hospital cannot ignore (a certain type of) patients in favour of operational system performance. This paper therefore considers two types of performance measures while examining different control policies for the physicians: operational system performance measures and health-related performance measures.

The remainder of this paper is organized as follows. Section 2 provides an overview of the currently existing literature on this topic and Section 3 presents the methodology applied in this paper. Section 4 describes the results by analysing and comparing different control policies. Finally, Section 5 presents the main conclusions and opportunities for future research.

2. Literature review

Inpatient boarding is also referred to in the literature as “access block” (Au et al., 2008; Luo, Cao, Gallagher & Wiles, 2013) or “bed block” (Rashwan, Abo-Hamad, & Arisha, 2015; Saghafian et al.,

2015). It is one of the key factors contributing to overcrowding and has a significant influence on patient's LOS in the ED (Shi, Chou, Dai, Ding & Sim, 2016; Carmen, Defraeye & Van Nieuwenhuysse, 2015). Furthermore, these longer ED stays increase hospital LOS, risk of hospital-acquired infections and mortality rates (Armony et al., 2015). Additionally, inpatient boarding is defined as the main cause of ambulance bypass which occurs when ambulances are forced to take their patients to another hospital because the ED is too crowded; e.g. Au et al. (2008) observe that access block is responsible for 76% of all ambulance bypass. Also non-boarding patients may suffer from delays in the transfer process; since boarding patients still require treatment while waiting to be admitted to the IW, the additional workload they put on ED staff may be significant as boarding patients take up to 11% of physician time in the ED (Armony et al., 2015).

Surprisingly, however, when analysing the ED, inpatient boarding is often neglected or not the main focus of the study. Indeed, articles in the OR/OM literature concentrating on this specific topic are notably scarce (Saghafian et al., 2015). Relevant inpatient boarding articles are categorized according to two distinguishing features: the articles' focus and methodology (Table 1). Firstly, although ED boarding occurs at the intersection of the ED and the IW, articles that take into account ED boarding usually focus either on the ED or on the IW. Secondly, given the wide range of OR/OM tools currently available, articles are categorized according to the methodology used in the analysis.

Table 1: Overview of relevant inpatient boarding articles.

Article	Focus		Methodology			
	ED	IW	Simulation	Queueing theory	Statistical methods	Data analysis
Bair et al. (2010)	X		X			
Carmen et al. (2014)	X		X			
Carmen et al. (2015)	X		X			
Crawford et al. (2014)	X		X			
Saghafian et al. (2012)	X		X	X		
Au et al. (2008)	X			X		
Bagust et al. (1999)		X	X			
Mustafee et al. (2012)		X	X			
Rashwan et al. (2015)		X	X			
Shi et al. (2016)		X	X	X		
Luo et al. (2013)		X			X	
Armony et al. (2015)		X				X

Root causes of inpatient boarding are usually sought in the IW. Shortage in the number of IW beds is often identified by hospital staff to be the number one cause of boarding (Mustafee et al., 2012). Other reasons for inpatient boarding are the inability to discharge IW patients in a timely manner (Rashwan et al., 2015; Crawford, Parikh, Kong & Thakar, 2014; Luo et al., 2013) and ED "batch admitting", which occurs when several boarding patients are queued and admitted to the IW altogether at a time convenient

to the staff (Luo et al., 2013). On the one hand, Armony et al. (2015) group plausible causes for ED-to-IW delays into four main categories: inadequate synchronization between the ED and the IW, bad work methods, shortage in staff availability and lack of equipment availability. Shi et al. (2016), on the other hand, attribute boarding times to a mismatch between the daily number of arrivals and discharges and a mismatch between the discharge timing and hourly arrival pattern. Depending on which process or resource that is defined to be the main cause of inpatient boarding, the proposed improvement strategies differ.

Articles that study the consequences of inpatient boarding tend to model the system from an ED perspective. Performance measures considered in these studies are the rate of LWBS (Bair, Song, Chen & Morris, 2010) and ambulance bypass (Au et al., 2008). Even if boarding is not the main focus, ED studies should take this phenomenon into account to obtain realistic results (Carmen, Defraeye, Celik Aydin & Van Nieuwenhuysse, 2014; Carmen et al., 2015; Saghafian, Hopp, Van Oyen, Desmond & Kronick, 2012).

Depending on the focus of the article, studies model either the ED or the IW in detail, but to the best of our knowledge, none model both the ED and the IW. With boarding being located at the intersection of the ED and IW, modelling it is complicated and it is often neglected. Furthermore, even when it is included in the study, it is usually modelled in a very basic way. On the one hand, when focusing on the IW, boarding patients are typically modelled as IW admissions which arrive according to a certain arrival process (Rashwan et al., 2015; Bagust, Place & Posnett, 1999; Luo et al, 2013). On the other hand, when using an ED focus, it is often assumed that boarding patients only occupy ED beds and do not need any further treatment in the ED (Crawford et al., 2014; Carmen et al., 2014; Carmen et al., 2015). However, data analysis (e.g. Armony et al., 2015) shows that boarding patients take up to 11% of physician time in the ED. Neglecting this can considerably underestimate the staff's workload. This paper aims to contribute to the existing literature by adding these additional treatments required by boarding patients to an ED-focused analysis.

Another way to categorize the literature that concentrates on inpatient boarding is to classify articles based on the OR/OM methodologies used for investigating this topic (Table 1). Simulation is used for the main analysis (Saghafian et al., 2015; Zeltyn et al., 2011), or as an additional methodology to verify an analytical model by comparing the simulation output with empirical estimates or to conduct a sensitivity analysis on the outcome of the study (e.g. Shi et al., 2016). Simulation has been the preferred OR/OM tool to study ED operations for years (e.g. Bagust et al., 1999) since it is able to simulate a complex and stochastic environment and therefore it can capture the real system's behaviour in a much better way (Saghafian et al., 2015; Rashwan et al., 2015). Also, it is cheaper and faster than performing real-time experimentation (Mustafee et al., 2012). Furthermore, using simulation gives the opportunity to run multiple replications of the same configuration, making the conclusions derived from the simulation outcomes more statistically reliable. However, the main drawback is the lack of generality over the wide range of different simulation studies (Günal & Pidd, 2010). Most of the time, simulation

is used for modelling specific units and facilities, and these models are rarely reused in other similar studies.

Alternative OR/OM methodologies used in the reviewed articles are queueing theory (Shi et al., 2016; Saghafian et al., 2012; Au et al., 2008), statistical methods, including regression and functional principal component analysis, (Luo et al., 2013) and data analysis (Armony et al., 2015). An interesting observation is that two out of three articles using queueing theory do not rely solely on queueing theory, but combine it with simulation. So, it seems that taking into account inpatient boarding and solely relying on queueing theory for conducting the analysis is a rare combination. Saghafian et al. (2015) concurs that ignoring blocking issues in the ED is one of the four main deficiencies in queueing theory models commonly used when analysing ED patient flow.

In contrast to the numerous articles that apply simulation, studies that use other OR/OM methodologies than simulation are remarkably rare. This is not surprising since adding inpatient boarding to the analysis makes it more realistic and accurate, but also more complex. With this rise in complexity, the need for an OR/OM methodology that can capture all these complex aspects of the ED also grows. As discussed before, simulation is particularly suitable for modelling such environments. Consequently, we will also rely on simulation in this paper when investigating and comparing different control policies for ED physicians.

3. Methodology

3.1. Problem setting

Based on Carmen (2017), the ED is analysed from a patient flow perspective. In our flow network, patients can return to the physician for additional treatment several times (re-entrant patients) and boarding patients generate additional check-ups and thus extra workload (needy boarding patients). A schematic overview of this network is provided in Figure 1.

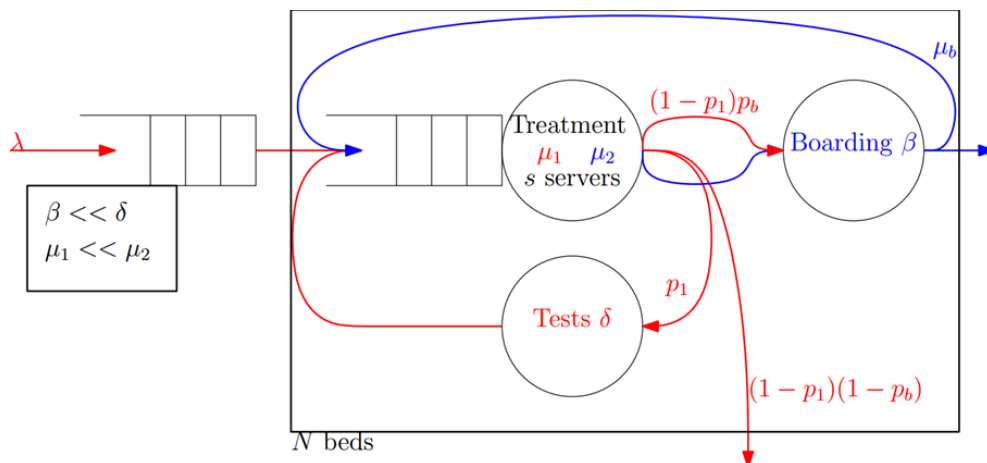


Figure 1: Overview of ED model.

A homogenous flow of patients arrives in the ED according to a Poisson distribution with rate λ . Looking at the relevant inpatient boarding articles in section 2, it is common to assume a Poisson arrival

rate in an ED setting. Before starting the actual treatment process, they wait in the external queue until they obtain a bed, of which there are N . We assume all arriving patients have equal priority and no life-threatening cases are present in the external queue; such cases are handled outside our model. After obtaining a bed, patients enter the internal queue in order to see one of the s physicians. This initial treatment process is exponentially distributed with mean $\frac{1}{\mu_1}$ and it is assumed that any of the s physicians can treat any patient.

After this treatment step, there are three ways in which patients can continue their treatment process. First, with probability p_1 , patients need to return to the physician for additional treatment after a delay which is exponentially distributed with mean $\frac{1}{\delta}$. This delay represents any treatment step that does not require the physician's presence like nurse treatments, lab tests, or scans. We assume that this treatment step has infinite capacity. Second, with probability $(1 - p_1)p_b$, patients need admission into the hospital, but there is no IW bed available for them. Therefore, they have to wait in the ED until they can be hospitalized in the IW. This boarding time is exponentially distributed with mean $\frac{1}{\beta}$. While waiting, boarding patients generate additional check-ups according to a Poisson distribution with rate μ_b . The durations of these check-ups by the physician are exponentially distributed with mean $\frac{1}{\mu_2}$. When a patient's boarding time is over, he or she can go to the IW if, at that time, he or she is not seeing the physician or waiting for a check-up. Otherwise, the patient will be transferred to the IW immediately after returning from the physician. Last, with probability $(1 - p_1)(1 - p_b)$, patients leave the ED after treatment.

With needy boarding patients, there are three kinds of patients waiting in the internal queue: patients that have just started their treatment process (thus patients that have just obtained a bed), patients that are returning from tests, and boarding patients. In this paper, the former two will be referred to as test patients. Consequently, two types of patients are analysed in our model: test and boarding patients. This distinction is made clear in the flow model displayed in Figure 1 by using different colours: red for test patients and blue for boarding patients.

Given the different characteristics of test and boarding patients, the choice whether to give priority to one or another can have a significant impact on system performance. This paper determines control policies for the physician, thereby investigating and comparing several policies while taking into account various performance measures. The determined control policies will specify which patients should be prioritized in what situations. A distinction is made between static and dynamic control policies. In static control policies, physicians apply a prioritization rule independent from the system state. In dynamic control policies however, the system state dictates which type of patients gets priority in the internal queue.

3.2. Scenarios

Five scenarios, classified into static and dynamic control policies, are analysed and compared (Table 2).

1. In the first scenario, physicians apply a first-come, first-served (FCFS) policy.
2. In the second scenario, physicians always give priority to boarding patients.
3. Alternatively, in the third scenario, physicians always give priority to test patients.
4. The fourth scenario contains the first dynamic control policy, which uses a threshold policy similar to van Dijk and van der Sluis (2009). In this scenario, the physicians' priority choice depends on a threshold rule. With m_{nb} the number of needy boarding patients in the internal queue, θ_{nb} the threshold value of these needy boarding patients, m_b the number of boarding patients in the system and θ_b the threshold value of these boarding patients, the threshold rule is as follows: *Thr*(θ_{nb}, θ_b) = *priority is given to boarding patients if $m_{nb} \geq \theta_{nb}$ or $m_b \geq \theta_b$, otherwise, test patients get priority.* For combinations of θ_{nb} and θ_b where $\theta_{nb} \geq \theta_b$, the physicians' prioritization choice depends solely on θ_b . Consequently, combinations for which $\theta_{nb} > \theta_b$ are not relevant (they generate the same outputs as combinations for which $\theta_{nb} = \theta_b$) and we only investigate combinations where $\theta_{nb} \leq \theta_b$.
5. The fifth and final scenario implements an accumulating priority queue where a patient's priority is a linear function of his or her waiting time in the internal queue (Fajardo & Drekić, 2017; Li & Stanford, 2016; Sharif, Stanford, Taylor & Ziedins, 2014; Stanford, Taylor & Ziedins, 2014). The rate z_i at which the priority increases depends on the patient type i ($i \in \{t, b\}$ where t stands for test patients and b for boarding patients). The physician prioritizes the patient with the highest accumulated priority in the internal queue.

Henceforth, the abbreviations in Table 2 will be used to refer to the different scenarios.

Table 2: Overview of analysed scenarios.

Scenario	Control Policy	Static (S)/ Dynamic (D)
FCFS	Test and boarding patients are served on a FCFS basis	S
B	Boarding patients always get priority	S
T	Test patients always get priority	S
TRH	Priority depends on threshold rule	D
AP	Accumulating priority	D

3.3. Modelling approach

Arena Simulation Software® is used to build the flow network introduced in Section 3.1 and displayed in Figure 1. Since our aim is to study the ED in a steady-state, the graphical procedure of Welch (Mahajan & Ingalls, 2004) is used to determine the length of the warm-up period. After making 100 replications for the FCFS scenario, each with a length of 150 hours, we calculate the cross-replication averages of the utilization rates of physicians ρ_{phys} and beds ρ_{bed} for every 10 hours. Next,

we calculate the moving averages over these periods using a window of 3. This means that we calculate the moving average \overline{M}_t at period t as follows:

$$\overline{M}_t = \begin{cases} \frac{\sum_{i=1}^{2t-1} \overline{Y}_i}{2t-1} & \text{for } t \leq 3 \\ \frac{\sum_{i=t-3}^{t+3} \overline{Y}_i}{7} & \text{for } 3 < t \leq 12 \end{cases} \quad \text{with } \overline{Y}_i \text{ the cross-replication average of period } i$$

As it is apparent from Figure 2, the moving averages of both utilization rates become relatively stable after 5 periods or 50 hours. Following the graphical procedure of Welch, the simulation model enters the steady state at that point. To build in extra safety, since the dynamic scenarios apply more complex priority rules, a warm-up period of 100 hours is used.

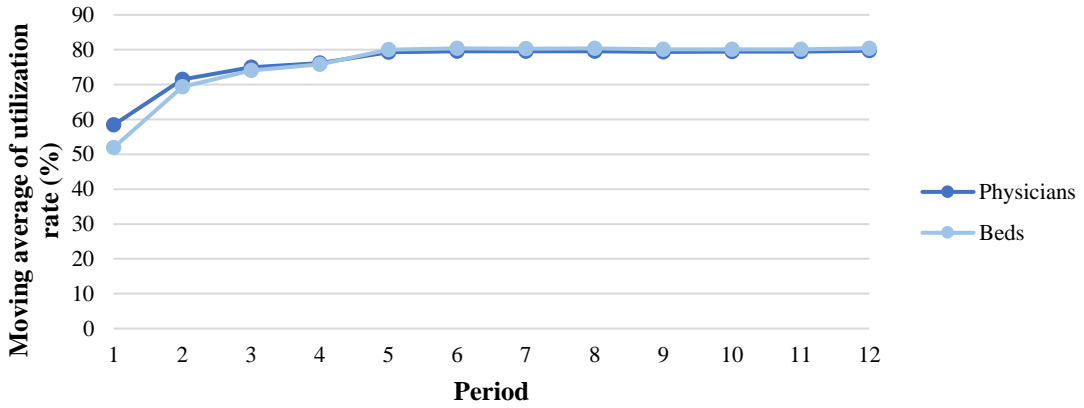


Figure 2: Moving averages of ρ_{phys} and ρ_{bed} .

We use a replication length of 1100 hours of which the first 100 hours count as a warm-up period. In each scenario, the number of replications equals 100. Using these settings, the half-widths of ρ_{phys} and ρ_{bed} never exceed 0.36% in the analysed scenarios.

In order to compare the different scenarios of Table 2, we rely on the 95% confidence intervals: scenarios significantly differ from each other when their 95% confidence intervals do not overlap.

3.4. Input data

An overview of the input used in the simulation model is given in Table 3. The estimates for μ_1 , δ and p_1 are taken from Yom-Tov and Mandelbaum (2014), who obtained their parameters from real-life data. Based on data analysis of Armony et al. (2015), we set the boarding probability p_b equal to 35%, the average boarding rate β to $\frac{1}{3}$ patients per hour, the average rate at which boarding patients generate check-ups μ_b to 4 check-ups per hour and the average treatment rate of boarding patients μ_2 to 40 patients per hour. λ , s and N are determined such that ρ_{phys} and ρ_{bed} are close to 80 % for the FCFS scenario.

Table 3: Input used in the simulation model.

μ_1	δ	p_1	p_b	β	μ_b	μ_2	λ	s	N
12.37	1.31	0.7268	0.35	$\frac{1}{3}$	4	40	6.3	3	28

4. Numerical results

This section presents the obtained numerical results. The dynamic policies are analysed in Sections 4.1, 4.2 and 4.3. We do not provide a separate section for the static control policies since these are included in the analysis of the dynamic control policies. The AP scenarios where $\frac{z_b}{z_t}$ is set to 1, infinite and 0 equal the FCFS, B and T scenarios respectively. Also, TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where θ_{nb} and θ_b are both set sufficiently high (so that they are never exceeded) equal the B and T scenarios respectively. Next, in Section 4.4, a comparison is made between the different scenarios introduced in Section 3.2.

4.1. Effect of dynamic control policies on operational system performance measures

We analyse the effect of the TRH scenario by varying threshold values θ_{nb} and θ_b . As explained in Section 3.2, only combinations for which $\theta_{nb} \leq \theta_b$ are looked into. Also, since no further impact on performance measures is observed for $\theta_{nb} \geq 10$ or $\theta_b \geq 15$, these combinations are not shown in the graphs. The effect of the AP scenario is analysed by varying $\frac{z_b}{z_t}$. This is done by altering z_b , while keeping z_t constant and equal to 1. An overview and explanation of the analysed values of $\frac{z_b}{z_t}$ are given in Table 4.

Table 4: Overview and explanation of analysed values of $\frac{z_b}{z_t}$.

$\frac{z_b}{z_t}$	Explanation ²
0	T scenario
0.2	$z_b = 0.2 * z_t$
0.4	$z_b = 0.4 * z_t$
0.6	$z_b = 0.6 * z_t$
0.8	$z_b = 0.8 * z_t$
1	FCFS scenario
1.25	$z_t = 0.8 * z_b$
1.67	$z_t = 0.6 * z_b$
2.5	$z_t = 0.4 * z_b$
5	$z_t = 0.2 * z_b$
9999	B scenario

² The values of $\frac{z_b}{z_t}$ displayed on the x-axis in the graphs in this section are chosen such that on the left of $\frac{z_b}{z_t} = 1$ the formula $z_b = \alpha * z_t$ applies and on the right of $\frac{z_b}{z_t} = 1$ the formula $z_t = \alpha * z_b$. Furthermore, α goes from 0 to 1 in steps of 0.2. Consequently, values of $\frac{z_b}{z_t}$ on the left and right of $\frac{z_b}{z_t} = 1$ which are based on the same α (for example $\frac{z_b}{z_t} = 0.2$ and $\frac{z_b}{z_t} = 5$) are located equally far from $\frac{z_b}{z_t} = 1$ on the x-axis in the graphs.

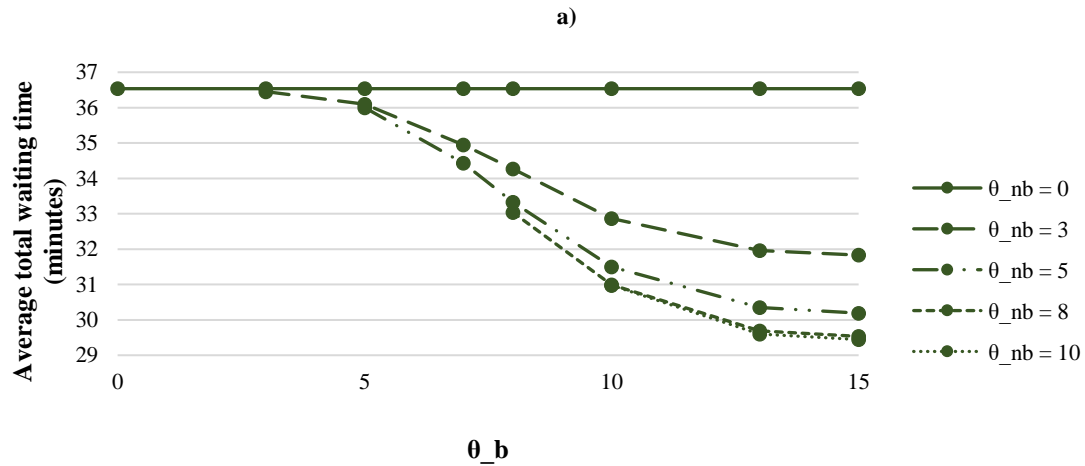
4.1.1. Waiting times

From a patient's operational perspective, the most important performance measure is the ED waiting time. The different waiting times are defined as follows.

- Average total waiting time = average external queue waiting time + average total internal queue waiting time
- Average total internal queue waiting time = (0.65 * average total internal queue waiting time of test patients) + (0.35 * average total internal queue waiting time of boarding patients) ³
- Average total internal queue waiting time of test patients = average number of physician visits of test patients * average single internal queue waiting time of test patients⁴
- Average total internal queue waiting time of boarding patients = average total internal queue waiting time of test patients + (average number of physician visits of boarding patients * average single internal queue waiting time of boarding patients)

Furthermore, all calculated average waiting times are unconditional averages; even patients that did not have to wait and hence have a waiting time of 0 minutes are taken into account.

Figure 3 shows that prioritizing test patients (high threshold values θ_{nb} and θ_b or low values for $\frac{z_b}{z_t}$) minimizes total waiting time. This impact is a combination of two opposite movements: a positive effect on the external queue waiting time and a negative effect on the total internal queue waiting time (Figure 4).



³ Probability of boarding $p_b = 0.35$.

⁴ The average single internal queue waiting time is the average internal queue waiting time of one physician visit.

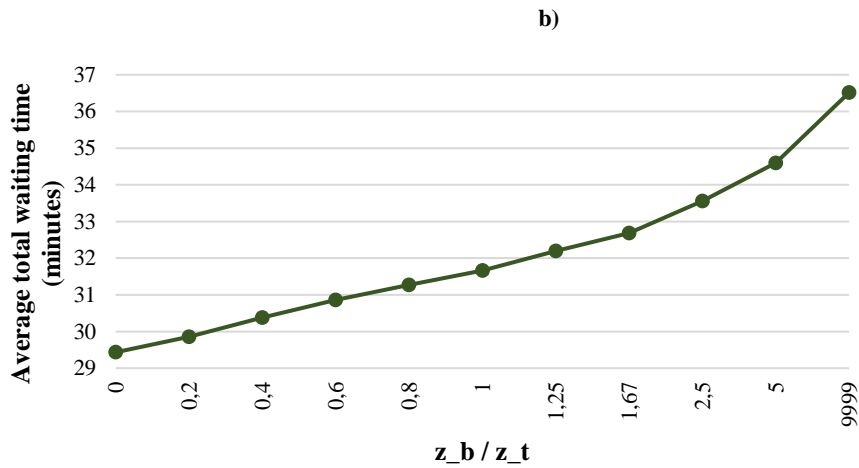


Figure 3: Average total waiting time for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

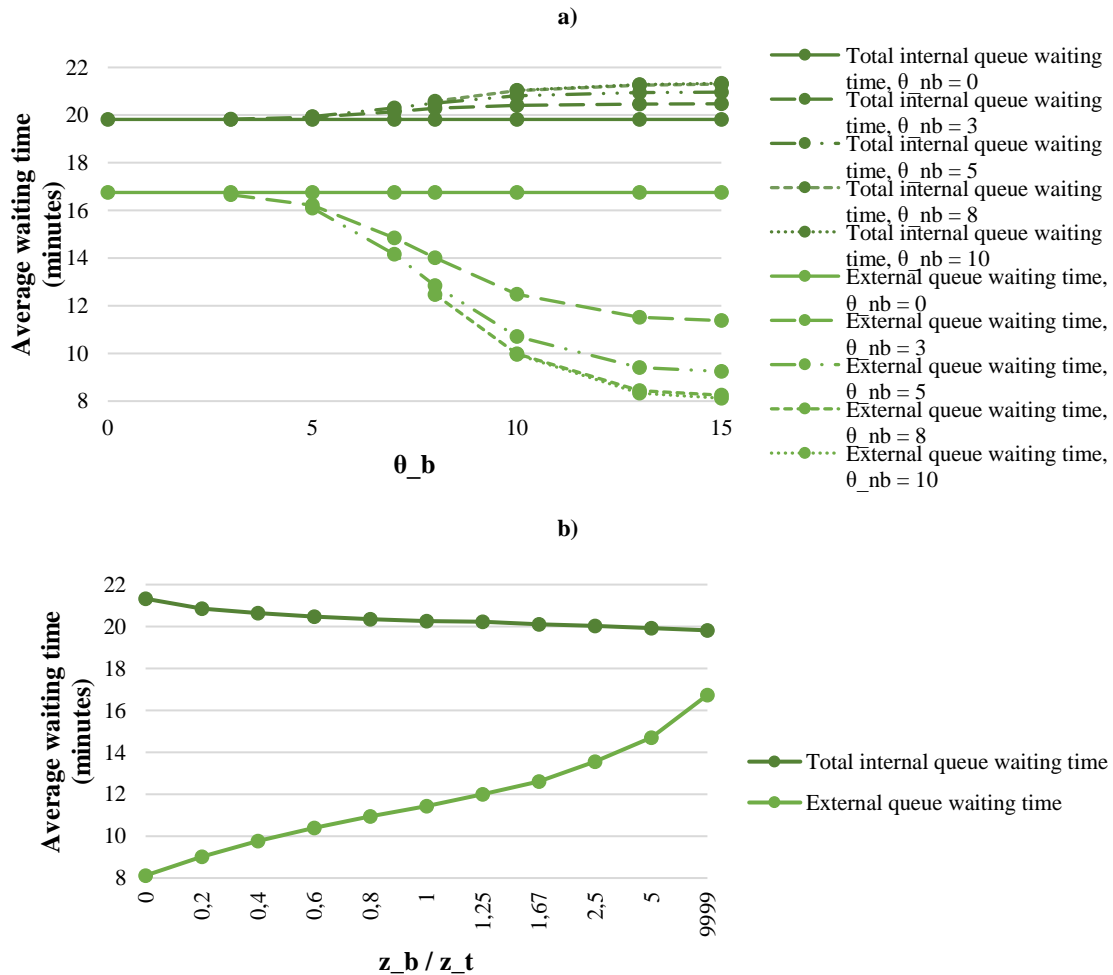


Figure 4: Average external queue and total internal queue waiting time for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

On the one hand, the average external queue waiting time is favourably impacted by giving more priority to test patients (Figure 4). The intuition behind this is that every time test patients have seen a physician, they release a bed with probability $(1 - p_1)(1 - p_b)$. Hence, prioritizing boarding patients has a negative impact on the external queue waiting time; all test patients stay longer in the internal queue and occupy beds for a longer time, preventing patients waiting in the external queue from entering the system. Boarding patients, on the contrary, have a fixed boarding time. Prioritizing test patients only increases the time a bed is occupied for boarding patients that are blocked in the internal queue. So, whereas prioritizing boarding patients increases the time a bed is occupied for all test patients, prioritizing test patients only increases this time for boarding patients that are blocked.

On the other hand, prioritizing test patients negatively impacts the average total internal queue waiting time (Figure 4). To get more insight into this impact, we must turn our attention to the average total internal queue waiting time of test and boarding patients (Figure 5), the average single internal queue waiting time of test and boarding patients (Figure 6) and the average number of physician visits of test and boarding patients (Figure 7).

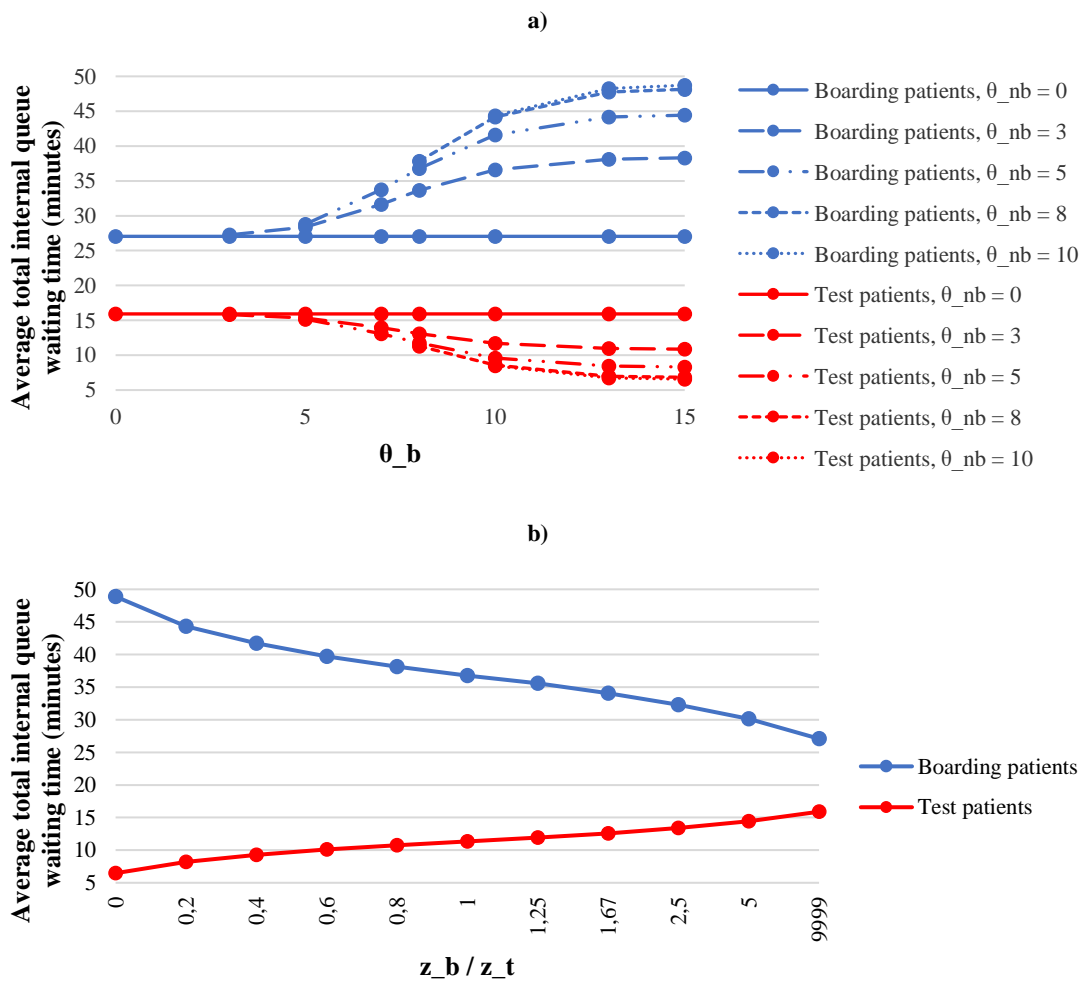


Figure 5: Average total internal queue waiting time of test and boarding patients for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

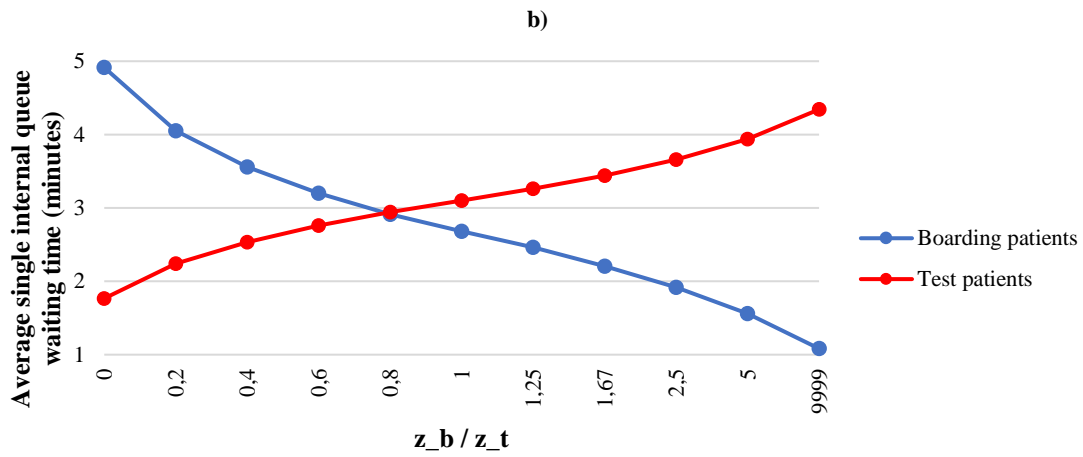
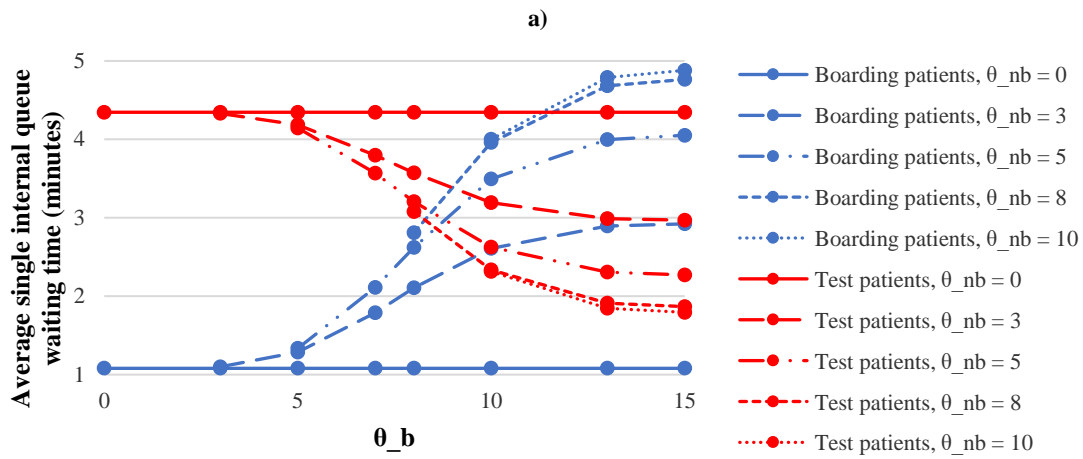
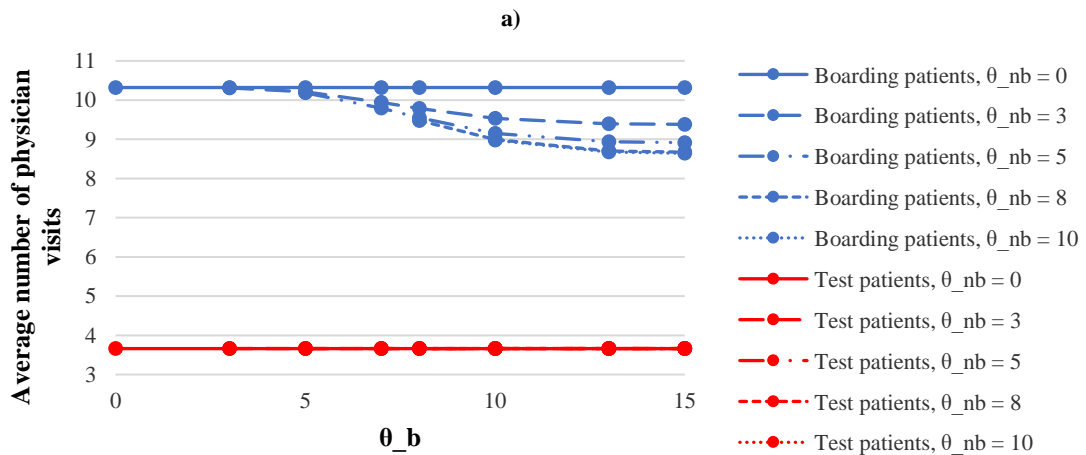


Figure 6: Average single internal queue waiting time of test and boarding patients for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.



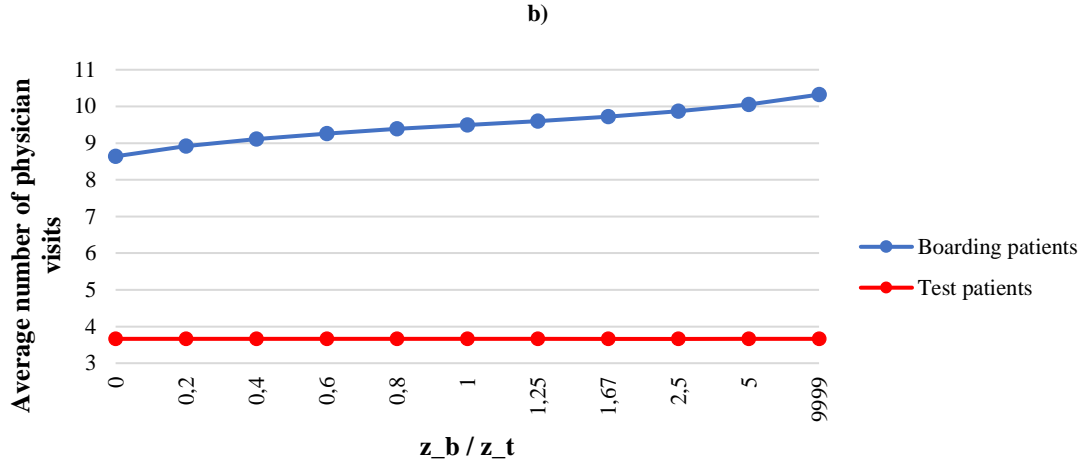


Figure 7: Average number of physician visits of test and boarding patients for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

As can be seen from Figure 6, there is a clear trade-off between the single internal queue waiting times of test and boarding patients; lowering the average waiting time of test patients unavoidably increases the average waiting time of boarding patients. This is not surprising, since evidently, giving more priority to one type of patients harms the other type.

Remarkably, in Figure 6, the average internal queue waiting time of boarding patients in their best-case scenario (B scenario) is lower than that of test patients in their best-case scenario (T scenario). Likewise, boarding patients' average internal queue waiting time in their worst-case scenario (T scenario) is higher than that of test patients in their worst-case scenario (B scenario). This can be explained by the difference in physician treatment time between the two types of patients. Test patients require more time and hence keep the physician longer occupied. Consequently, a patient waiting for a test patient to be finished has to wait longer than a patient waiting for a boarding patient. Boarding patients' waiting time is therefore extended more when priority is given to test patients than the other way around which explains the difference in average internal queue waiting time in the worst-case scenarios. A parallel reasoning can be applied to explain the difference observed in the best-case scenarios. When priority is given to boarding patients, these need to wait until the boarding patients that arrived before them in the internal queue are finished. The same holds for test patients when they are prioritized. Since boarding patients' average treatment time is shorter than that of test patients, the former benefit more from receiving priority. These differences in best- and worst-case scenarios clarify why altering θ_{nb} and θ_b or $\frac{z_b}{z_t}$ has a larger impact on the total internal queue waiting time of boarding patients than on that of test patients (Figure 5). This, in turn, explains why prioritizing test patients increases the total internal queue waiting time (Figure 4). Though, looking at Figure 6, one would expect the increase in total internal queue waiting time to be much higher than observed in Figure 4. However, we also need to consider how many times the patients visit the physician (Figure 7). On the one hand, since boarding patients have a fixed boarding time, prioritizing them increases their number of physician

visits. Indeed, every time a boarding patient returns from the physician and his boarding time is not over yet, he generates a new check-up at rate μ_b . This in turn mitigates the positive effect of prioritizing boarding patients on the total internal queue waiting time. On the other hand, since test patients return to the physician with probability p_1 , independent from the applied control policy, prioritizing them does not impact their physician visits at all. Furthermore, whereas test patients' total internal queue waiting time represents 65% of the total internal queue waiting time in Figure 4, boarding patients' total internal queue waiting time only accounts for 35%. Combining these two aspects provides an explanation for the relatively small increase in total internal queue waiting time as seen in Figure 4.

So, taking this altogether, the more priority is given to test patients, the more the total waiting time decreases. This effect is the result of a trade-off between a decrease in external queue waiting time and an increase in total internal queue waiting time. However, the effect on external queue waiting time is more outspoken and is therefore the key driver of the evolution in total waiting time.

Although overall the same effect on waiting times is observed for the two dynamic control policies, there is also a clear difference between them. In the TRH scenario, each of the curves follows an S-shape. Starting from a control policy that extremely favours one type of patients, slightly varying θ_{nb} and θ_b only has a small effect on waiting times. For instance, when using threshold values far below the average number of needy boarding patients m_{nb} and boarding patients m_b , the conditions for prioritizing boarding patients are almost always fulfilled. Only sporadically, m_{nb} and m_b go below these threshold values. Consequently, priority is rarely given to test patients, explaining the limited effect on waiting times. A similar reasoning holds when applying high threshold values (θ_{nb} and θ_b far above the averages of m_{nb} and m_b). Not surprisingly, the largest impact is observed when varying threshold values θ_{nb} and θ_b around the averages of m_{nb} and m_b . The waiting times in the AP scenario, however, follow a much more fluent pattern. This is explained by the factors on which we rely to determine who gets priority. On the one hand, in the AP scenario, the prioritization depends on two factors: a patient's priority rate z_i and his internal queue waiting time. Consequently, when patient type i has the highest priority rate, a patient of type j can nevertheless receive priority if his internal queue waiting time is sufficiently high. On the other hand, in the TRH scenario, we solely rely on the threshold values θ_{nb} and θ_b . As a result, when patients of type i are prioritized, a patient of type j will not receive priority as long as m_{nb} and m_b do not change significantly. Even though this patient's internal queue waiting time may be unacceptably high, priority is not switched. This difference between the two dynamic policies has two important consequences when implementing these policies in practice. Firstly, an AP policy can better approach the desired value of performance measures since varying the priority rates has a more fluent effect than varying the threshold values used in a TRH policy. Secondly, when extremely prioritizing one type of patients, the other patient type suffers more in terms of excessive waiting times when using a TRH policy.

4.1.2. Utilization rates

Next, we turn our attention to the utilization rates of physicians ρ_{phys} and beds ρ_{bed} (Figure 8), where a similar pattern is observed as for the average total waiting time. Giving more priority to test patients (high threshold values θ_{nb} and θ_b or low values for $\frac{z_b}{z_t}$) results in lower utilization rates. The effect on ρ_{phys} is explained by the effect of the physicians' prioritization decision on the number of physician visits of test and boarding patients. As explained in Section 4.1.1, prioritizing boarding patients increases their number of physician visits, putting more workload on the physicians. On the contrary, a test patient's number of physician visits is independent from the applied control policy. Consequently, prioritizing boarding patients increases ρ_{phys} . The effect on ρ_{bed} is explained by the impact on the time during which test and boarding patients occupy beds. As mentioned in Section 4.1.1, prioritizing boarding patients increases this time for all test patients. However, prioritizing test patients only increases the time a bed is occupied for boarding patients that are blocked in the internal queue.

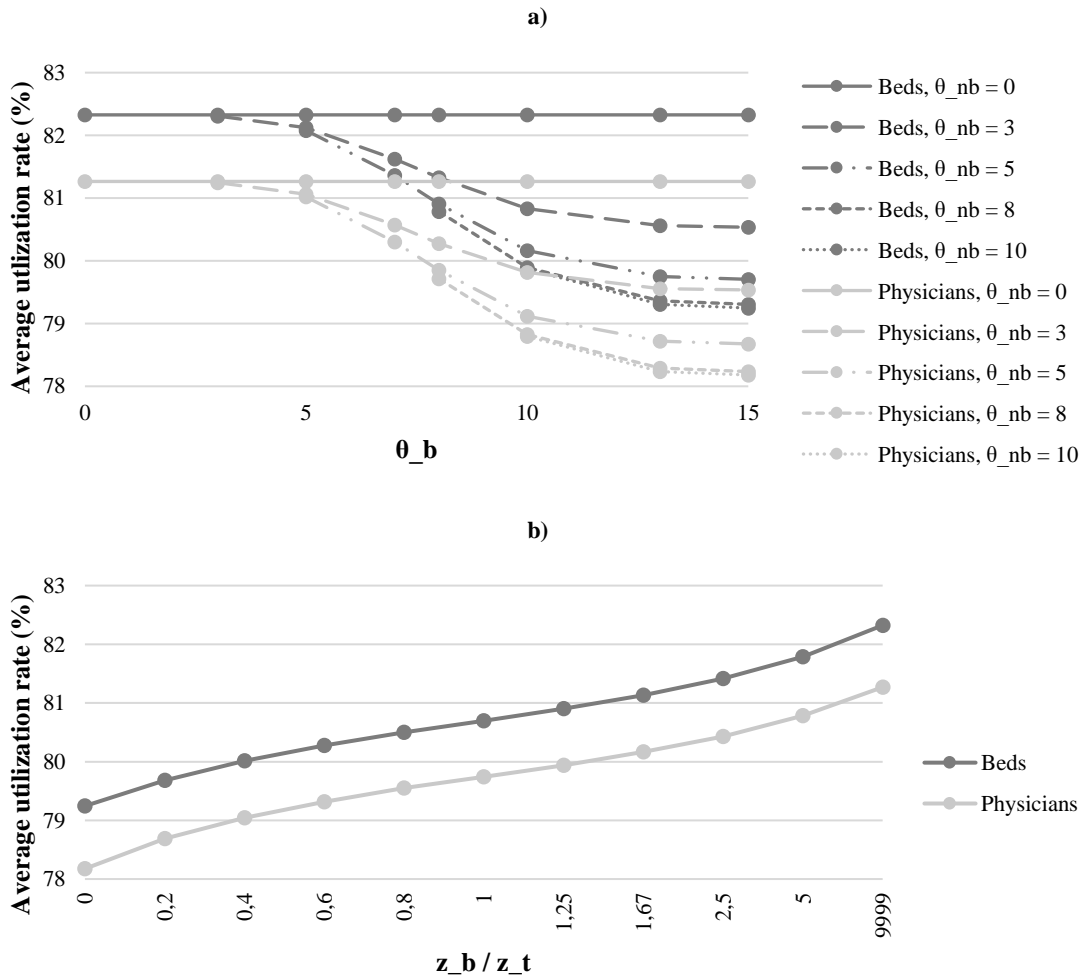


Figure 8: Average utilization rates for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

So, considering operational system performance measures like waiting times and utilization rates, we can define an optimal and easy-to-use control policy for the physicians: test patients always get priority in the internal queue. However, this disregards the fact that boarding patients who already have higher hospital mortality risk (Carr, Hollander, Baxt, Datner & Pines, 2010) are not receiving the best medical treatment for their condition and are not as closely monitored as in the IW (Armony et al., 2015). Therefore, only looking at operational system performance measures when determining a physicians' control policy may be inadequate. A similar conclusion is made by Decouttere and Vandaele (2014) who state that solely focusing on operational system performance measures is insufficient in a health care environment and that other, perhaps conflicting, performance measures need to be taken into account. An analysis of health-related performance measures concentrating on boarding patients' medical conditions seems necessary. Since the operational system performance measures incentivize giving priority to test patients, their medical risk is minimal and is therefore not explicitly looked at.

4.2. Effect of dynamic control policies on health-related performance measures

4.2.1. Boarding times

To investigate the impact of the dynamic control policies on boarding patients' medical conditions, we first look at the average boarding time and its underlying drivers: the average total non-needy time and total needy time of boarding patients (Figure 9).

The physicians' prioritization choice has a negligible effect on boarding time. For instance, altering between the two most extreme control policies only changes the average boarding time 0.0549 hours or 3.3 minutes. This is the extra average time a boarding patient spends in the ED, on top of his or her fixed boarding time, because of blocking. That is, the patient is allowed to enter the IW, but is still waiting for or receiving physician treatment, preventing him or her from leaving the ED. Hence, blocking on average increases a boarding patient's stay in the ED with only 1.8%.

The average total needy time, on the contrary, is very sensitive to changes in θ_{nb} and θ_b or $\frac{z_b}{z_t}$; it doubles when test patients are prioritized. Likewise, the total non-needy time decreases. This is an important finding, since boarding patients' medical conditions are worse when they are in a needy state. Therefore, the substantial increase in boarding patients' needy time might lead to severe health issues. However, it is very difficult to define the threshold value of total needy time above which boarding patients are at high medical risk. Indeed, the needy time both includes internal queue waiting time and physician treatment time. When seeing a physician, a boarding patient is not at risk anymore. Remember that in our model, check-ups required by boarding patients can only be executed by physicians. Also, boarding patients' medical conditions are not constantly critical during their time in the internal queue. Only when this waiting time exceeds a certain threshold, boarding patients experience high medical risk. To further explore this insight, we next turn our attention to the boarding patients' risk percentage.

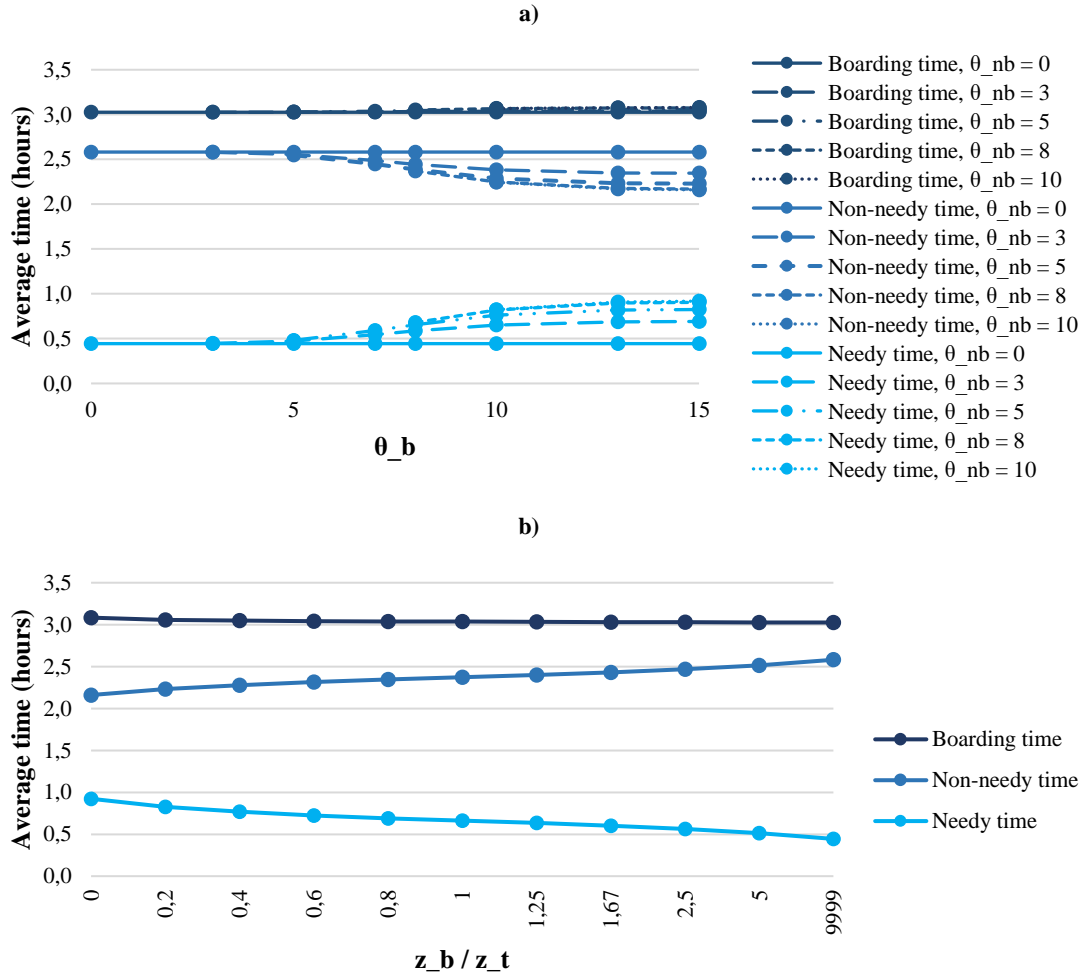


Figure 9: Average boarding time, total non-needy time and total needy time of boarding patients for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

4.2.2. Boarding patients' risk percentage

We define boarding patients' risk percentage as the percentage of boarding patients in the internal queue that has to wait longer than τ minutes before seeing a physician. Based on Bergs et al. (2014) and Freund, Vincent-Cassy, Bloom, Riou and Ray (2013), we set τ to 15 minutes.

Altering the control policy may have a disastrous effect on boarding patients' risk percentage (Figure 10). For instance, switching from always prioritizing boarding patients to always prioritizing test patients increases this percentage from 0% to 10%. Hence, boarding patients suffer badly from control policies that mainly prioritize test patients. This confirms our insight that only looking at operational system performance measures when deciding on the physicians' control policy is inadequate.

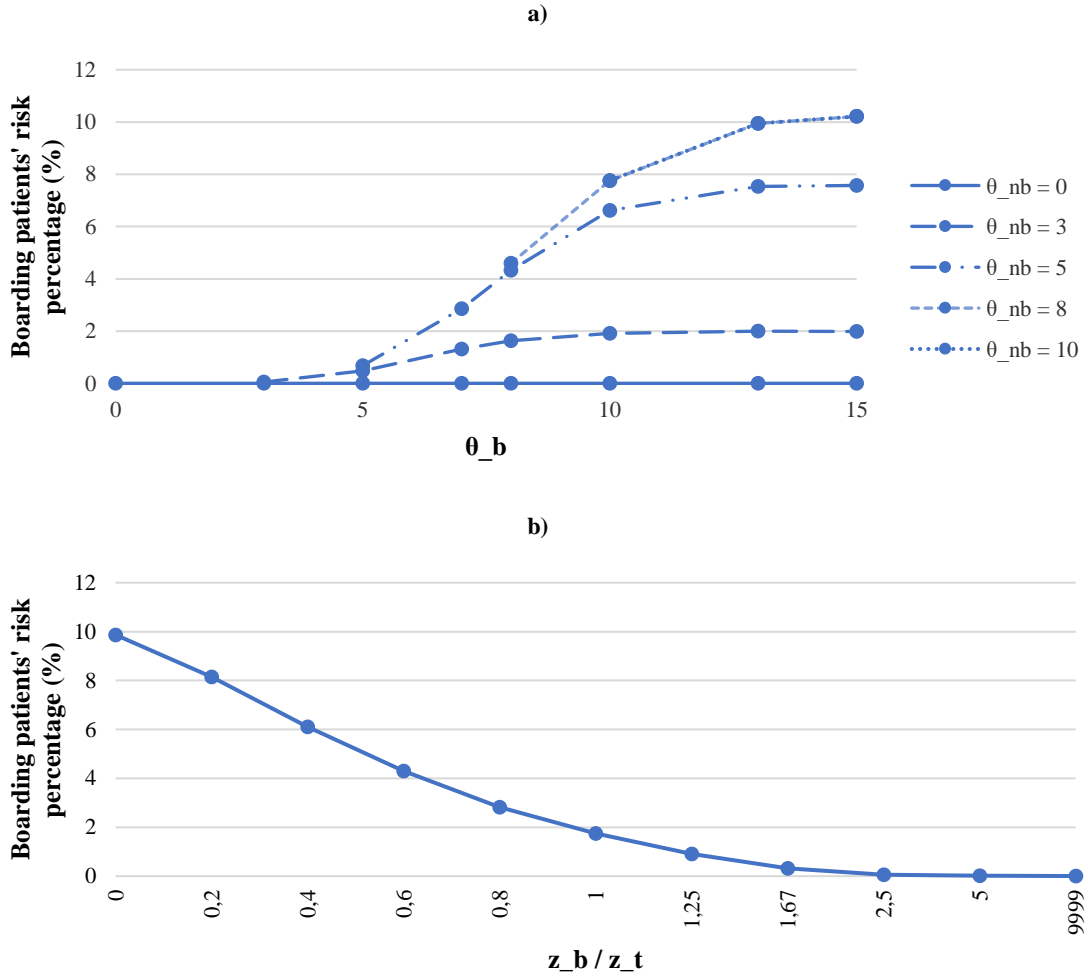


Figure 10: Boarding patients' risk percentage for a) different threshold values θ_{nb} and θ_b and b) different values of $\frac{z_b}{z_t}$. TRH scenarios where $\theta_{nb} = \theta_b = 0$ and where $\theta_{nb} = 10$ and $\theta_b = 15$ equal the B and T scenarios respectively. AP scenarios where $\frac{z_b}{z_t}$ is set to 1, 9999 and 0 equal the FCFS, B and T scenarios respectively.

4.3. Recommended dynamic control policies

To balance the trade-off between operational system performance measures and health-related performance measures, we recommend the following procedure. First and foremost, the hospital needs to decide on the target boarding patients' risk percentage it is willing to accept. In this paper, we look at four target risk percentages: 1%, 3%, 5.5% and 8%. Thereafter, using simulation, this percentage is used to set threshold values θ_{nb} and θ_b and priority rates z_b and z_t for which operational system performance measures are optimized, while the target boarding patients' risk percentage is not exceeded. An overview of the recommended TRH and AP control policies is given in Table 5.

Table 5: Overview of recommended TRH and AP control policies.

Target boarding patients' risk percentage	Recommended TRH control policy	Recommended AP control policy
1%	$\theta_{nb} = 2, \theta_b = 15$	$z_b = 1.25, z_t = 1$
3%	$\theta_{nb} = 3, \theta_b = 14$	$z_b = 0.80, z_t = 1$
5.5%	$\theta_{nb} = 4, \theta_b = 15$	$z_b = 0.50, z_t = 1$
8%	$\theta_{nb} = 5, \theta_b = 15$	$z_b = 0.25, z_t = 1$

4.4. Comparison of scenarios

As pointed out before, it is important to look at two aspects simultaneously: the system performance aspect and the health aspect. We therefore analyse the different scenarios in terms of operational system performance measures (waiting time and utilization rates) and health-related performance measures (boarding patients' risk percentage) (Figure 11). To get a better view of the best performing scenarios, we add the efficient frontiers⁵. Furthermore, the 95% confidence intervals are added in order to compare the scenarios. As mentioned in Section 3.3, scenarios differ significantly from each other if their 95% confidence intervals do not overlap. The legend used in Figure 11 is shown in Table 6.

The trade-off in Figure 11 is clear; operational system performance measures can only be improved at the expense of boarding patients' medical risk. This confirms the intuition we derived in previous sections. Moreover, even though the improvement in operational system performance may not be statistically significant, the deterioration in boarding patients' medical conditions mostly is. Or, in other words, whereas operational system performance is more robust regarding changes in control policy, boarding patients' medical risk is extremely sensitive to it.

Comparing the two dynamic control policies for each of the defined target boarding patients' risk percentages, we observe a difference in accuracy. Indeed, an AP policy manages to approach these risk percentages more precisely. The reason being that, when using a TRH control policy, threshold values must be integer numbers. Consequently, these can only be altered in discrete steps of one, which in turn also changes the risk percentage in a discrete way. In the AP scenario however, priority rates can be altered in a much more precise way, resulting in a higher level of accuracy. This implies that an AP control policy can respect the constraint on boarding patients' medical conditions while using a higher risk percentage. Or, in other words, an AP control policy can yield higher operational system performance for the same target boarding patients' risk percentage. Indeed, given the trade-off defined above, a higher risk percentage results in improved operational system performance. Following this reasoning, one might suggest that an AP control policy outperforms a TRH control policy. However, as Figure 11 shows, this is not the case. Looking at the 95% confidence intervals, we can see that the two dynamic control policies do not differ significantly from each other for any of the operational system

⁵ The efficient frontier is the set of optimal scenarios that realize the best operational system performance for a defined level of boarding patients' risk percentage.

performance measures. This leads us to conclude that both dynamic control policies can be used to reach the desired target boarding patients' risk percentage while optimizing operational system performance.

We also derive from Figure 11 that a simple, static FCFS policy can serve as an alternative for the dynamic policies in several situations. For a target boarding patients' risk percentage of 3%, the outcomes of the FCFS scenario do not significantly differ from those using dynamic control policies. Hence, here, we recommend using a FCFS policy, since it is much easier to apply in practice. Also, in the case of a target risk percentage of 5.5%, the FCFS scenario still performs well. Only one out of three operational system performance measures differs significantly from the dynamic control policies and this difference is not very large. Physicians that highly value an "easy-to-use" control policy may therefore prefer the FCFS policy despite its slightly weaker operational system performance. Only for more extreme risk percentages (that is, 1% and 8%), the FCFS policy is not adequate.

Tabel 6: Legend used in Figure 11.

Marker	Scenario	Target boarding patients' risk percentage
●	FCFS	N/A
⊘	B	N/A
●	T	N/A
■	TRH	1%
◆	TRH	3%
●	TRH	5.5%
▲	TRH	8%
▣	AP	1%
◇	AP	3%
⊘	AP	5.5%
▲	AP	8%

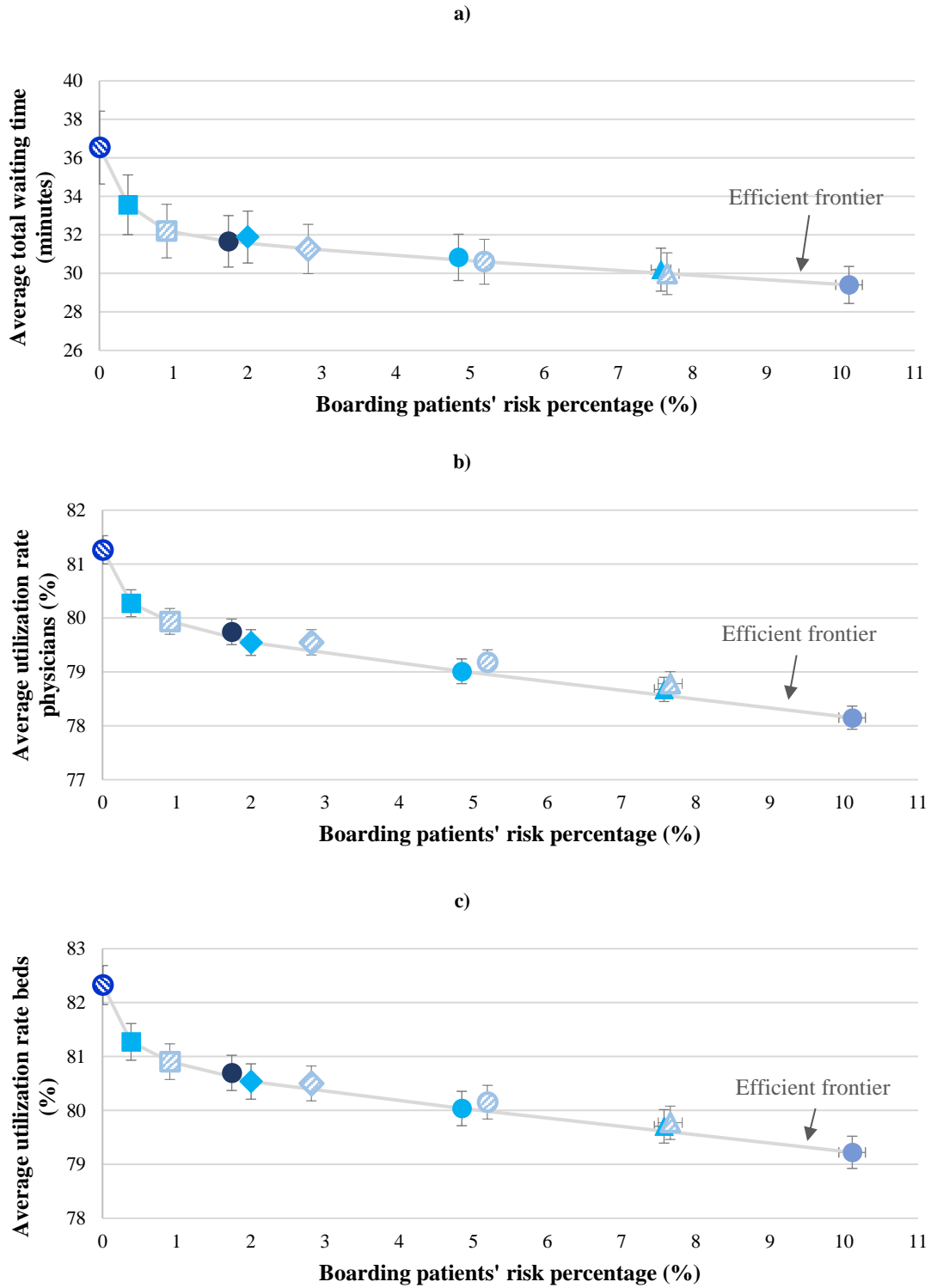


Figure 11: Overview of scenarios in terms of a) average total waiting time, b) average ρ_{phys} and c) average ρ_{bed} compared to boarding patients' risk percentage.

5. Conclusion and insights

This paper analyses the prioritization decision of physicians between test and boarding patients in an ED. It does so by analysing and comparing three static (FCFS, B and T) and two dynamic (TRH and AP) control policies for various performance measures using discrete-event simulation.

The main contribution of this paper is that it quantitatively demonstrates the relevance of incorporating needy boarding patients into the analysis and the importance of defining a control policy for the ED physicians. These are two areas often neglected in current literature. When incorporating needy boarding patients, the prioritization decision has significant consequences for system performance. We claim that an important trade-off exists between operational system performance measures (like waiting times and utilization rates) and health-related performance measures (like boarding patients' risk percentage). This trade-off illustrates that defining a control policy for the ED physicians is far from trivial and that no single optimal control policy exists. Rather, the optimal control policy depends on the target boarding patients' risk percentage the hospital is willing to accept. Theoretically speaking, the two dynamic control policies can be used to obtain any desired target boarding patients' risk percentage while optimizing operational system performance. However, important differences exist between these two control policies. Firstly, an AP policy is able to match the desired value of performance measures better since the priority rates can be altered in a much more precise way than the thresholds used in the TRH policy. Next, when extremely prioritizing one type of patients, the other patient type suffers more in terms of excessive waiting times when using a TRH policy. Another important finding is that it is not always necessary to use a complicated dynamic control policy. Indeed, a simple static FCFS policy turns out to perform extremely well for a wide range of imposed targets on boarding patients' risk percentage. Moreover, a FCFS policy is extremely easy to use in practice. Consequently, it is only worthwhile to put effort into the implementation of a dynamic control policy when an extremely high or low value of target boarding patients' risk percentage is desired.

Following these results, several options for future research emerge. First of all, when studying the TRH scenario, we assume that priority is switched immediately when the threshold rule dictates so. However, this is only possible in practice when the number of needy boarding patients in the internal queue and the total number of boarding patients in the system are constantly monitored, for instance by an IT system. This brings along high implementation costs and the hospital may therefore prefer to check these numbers manually at predefined time intervals. The effect of the length of these checking intervals on the performance of the TRH policy might be an interesting research topic. Secondly, we observed the importance of considering health-related performance measures in our analysis. Although neglecting these measures might lead to disastrous actions, including them is often a challenging task. Therefore, future research might address such performance measures in EDs in more detail and explore how to make the trade-off between these measures and other system performance measures. Finally, several

assumptions are made in this paper (see Section 3.1). Further research might study systems that relax some of these assumptions.

Acknowledgements

We gratefully acknowledge the financial support from the GlaxoSmithKline Research Chair on Re-Design of Healthcare Supply Chains in Developing Countries to increase Access to Medicines.

References

- Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., & Yom-Tov, G. B. (2015). On patient flow in hospitals: a data-based queueing-science perspective. *Stochastic systems*, 5(1), 146-194.
- Au, L., Byrnes, G.B., Bain, C.A., Fackrell, M., Brand, C., Campbell, D.A., & Taylor, P. G. (2008). Predicting overflow in an emergency department. *IMA Journal of management mathematics*, 20(1), 39-49.
- Bair, A.E., Song, W.T., Chen, Y.-C., & Morris, B.A. (2010). The impact of inpatient boarding on ED efficiency: A discrete-event simulation study. *Journal of Medical Systems*, 34(5), 919-929.
- Bagust, A., Place, M., & Posnett, J.W. (1999). Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. *British Medical Journal*, 319(7203), 155-158.
- Bergs, J., Verelst, S., Gillet, J.-B., Deboutte, P., Vandoren, C., & Vandijck, D. (2014). The number of patients simultaneously present at the emergency department as an indicator of unsafe waiting times: a receiver operated curve-based evaluation. *International Emergency Nursing*, 22(4), 185-189.
- Carmen, R. (2017). Bed block in emergency departments: Impact on patient delays and system capacity. *PhD dissertation: Resource efficiency improvements in hospitals*, 69-101. KU Leuven.
- Carmen, R., Defraeye, M., Celik Aydin, B., & Van Nieuwenhuysse, I. (2014). Modelling emergency departments using discrete-event simulation: A real-life case study including patient boarding. *FEB Research report KBI_1420, KU Leuven*.
- Carmen, R., Defraeye, M., & Van Nieuwenhuysse, I. (2015). A decision support system for capacity planning in emergency departments. *International Journal of Simulation Modelling*, 14(2), 299-312.
- Carr, B.G., Hollander, J.E., Baxt, W.G., Datner, E.M., & Pines, J.M. (2010). Trends in boarding of admitted patients in US emergency departments 2003-2005. *Journal of emergency medicine*, 39(4), 506-511.
- Crawford, E.A., Parikh, P.J., Kong, N., & Thakar, C.V. (2014). Analyzing discharge strategies during acute care: A discrete-event simulation study. *Medical decision making*, 34(2), 231-241.
- Decouttere, C., & Vandaele, N. (2014). A broader view on health care system design and modelling. *Springer Proceedings in Mathematics and Statistics*, 61, 215-225.
- Fajardo, V.A., & Drekić, S. (2017). Waiting time distributions in the preemptive accumulating priority queue. *Methodology and Computing in Applied Probability*, 19, 255-284.

- Freund, Y., Vincent-Cassy, C., Bloom, B., Riou, B., & Ray, P. (2013). Association between age older than 75 years and exceeded target waiting times in the emergency department: a multicenter cross-sectional survey in the Paris metropolitan area, France. *Annals of Emergency Medicine*, 62(5), 449-456.
- Günel, M.M., & Pidd, M. (2010). Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1), 42-51.
- Hoot, N. R., & Aronsky, D. (2008). Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine*, 52(2), 126-136.
- Li, N., & Stanford, D.A. (2016). Multi-server accumulating priority queues with heterogenous servers. *European Journal of Operational Research*, 252(3), 866-878.
- Luo, W., Cao, J., Gallagher, M., & Wiles, J. (2013). Estimating the intensity of ward admission and its effect on emergency department access block. *Statistics in Medicine*, 32(15), 2681-2694.
- Mahajan, P.S., & Ingalls, R.G. (2004). Evaluation of methods used to detect warm-up period in steady state simulation. *Proceedings of the 2004 Winter Simulation Conference*, 663-671.
- Moskop, J.C., Sklar, D.P., Geideman, J.M., Schears, R.M., & Bookman, K.J. (2009). Emergency Department Crowding, Part 1—Concept, Causes, and Moral Consequences. *Annals of Emergency Medicine*, 53(5), 605-611.
- Mustafee, N., Lyons, T., Rees, P., Davies, L., Ramsey, M., & Williams, M. D. (2012). Planning of bed capacities in specialized and integrated care units: Incorporating bed blockers in a simulation of surgical throughput. *Proceedings of the 2012 Simulation Conference*, 1-12.
- Rashwan, W., Abo-Hamad, W. & Arisha, A. (2015). A system dynamics view of the acute bed blockage problem in the Irish healthcare system. *European Journal of Operational Research*, 247(1), 276-293.
- Saghafian, S., Austin, G., & Traub, S.J. (2015). Operations research/management contributions to emergency department patient flow optimization: review and research prospects. *IIE Transactions on Healthcare Systems Engineering*, 5(2), 101-123.
- Saghafian, S., Hopp, W.J., Van Oyen, M.P., Desmond, J.S., & Kronick, S.L. (2012). Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research*, 60(5), 1080-1097.
- Sharif, A.B., Stanford, D.A., Taylor, P., & Ziedins, I. (2014). A multi-class multi-server accumulating priority queue with application to health care. *Operations Research for Health Care*, 3(2), 73-79.
- Shi, P., Chou, M.C., Dai, J.G., Ding, D., & Sim, J. (2016). Models and insights for hospital inpatient operations: time-dependent ED boarding time. *Management Science*, 62(1), 1-28.
- Stanford, D.A., Taylor, P., & Ziedins, I. (2014). Waiting time distributions in the accumulating priority queue. *Queueing Systems*, 77(3), 297-330.
- van Dijk, N.M., & van der Sluis, E. (2009). Pooling is not the answer. *European Journal of Operational Research*, 197(1), 415-421.

Yom-tov, G.B., & Mandelbaum, A. (2014). Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management*, *16*(2), 283-299.

Zeltyn, S., Marmor, Y.N., Mandelbaum, A., Carmeli, B., Greensphan, O., Mesika, Y., et al. (2011). Simulation-based models of emergency departments: Operational, tactical, and strategic staffing. *ACM Transactions on Modelling and computer simulation (TOMACS)*, *21*(4), 1-25.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be

