



Strategic room type allocation for nursing wards through Markov chain modeling

Andersen, Anders Reenberg; Vancroonenburg, Wim; Vanden Berghe, Greet

Published in:
Artificial Intelligence in Medicine

Link to article, DOI:
[10.1016/j.artmed.2019.101705](https://doi.org/10.1016/j.artmed.2019.101705)

Publication date:
2019

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Andersen, A. R., Vancroonenburg, W., & Vanden Berghe, G. (2019). Strategic room type allocation for nursing wards through Markov chain modeling. *Artificial Intelligence in Medicine*, 99, Article 101705. <https://doi.org/10.1016/j.artmed.2019.101705>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Strategic room type allocation for nursing wards through Markov chain modeling[☆]

Anders Reenberg Andersen^{a,1,*}, Wim Vancroonenburg^{b,c}, Greet Vanden Berghe^c

^aTechnical University of Denmark, Department of Applied Mathematics and Computer Science

^bFlanders Research Foundation – FWO-Vlaanderen

^cKU Leuven, Department of Computer Science, CODES & imec

Abstract

Providing patients with the best possible care is the most essential function of any hospital. In an increasing number of countries, hospitals are governed by the number of patients they are able to attract and the corresponding services they provide for patients. One such service, which is often of significant importance for patients, is the option to choose their room type.

Hospital decision makers would benefit from a strategic method for optimizing the configuration of room types among nursing wards by distinguishing between patients who prefer private rooms and those who have no preference concerning whether they are assigned to a private or shared room. Such a decision support method is currently non-existent, therefore the goal of this study is to provide a methodology for hospital management. Specifically, a mixed modeling approach is proposed which evaluates the patient flow behavior by applying a Continuous-Time Markov Chain within a heuristic search procedure. This procedure recursively improves a configuration of rooms among the wards by sampling from a gradually improved interpolation of an objective function.

Based on patient data obtained from both a Danish and Belgian hospital, the performance and robustness of the proposed approach is validated through various numerical experiments, demonstrating that solutions within a relative gap of 1% from the optimum are attained in most cases.

Keywords: Room allocation, Patient flow, Markov chains, Heuristics

1. Introduction

Rising public expenditure on health care systems has led many governments to apply budgetary pressure on hospitals to rationalize their spending [1]. At the same time, competition in hospital services is employed in many countries as a mechanism to motivate hospitals to reduce costs in order to remain competitive [2].

[☆]*Declarations of interest:* none

*Corresponding author

Email address: arean@dtu.dk (Anders Reenberg Andersen)

¹Address: Anker Engelunds Vej 1, 2800 Kgs. Lyngby, Denmark

Since patients may have freedom-of-choice regarding their hospital admission, hospitals therefore compete against one another to attract more patients. Not only do they compete on the basis of the medical services they provide, but also with regard to amenities that increase patient comfort during their stay. One such service of importance is the option for admitted patients (also referred to as *inpatients*) to choose their preferred room *type*. Many different room types may be distinguished in nursing wards, varying in capacity (examples include ward room, double room and private room) and amenities (in maternity wards some rooms may provide a shower or an extra bed for a spouse). There are significant financial incentives for hospitals to meet patients' room type preferences: for example, hospitals in Belgium may charge room supplements when meeting private room demands and physicians may even charge honorarium supplements. Hospital bills may be up to five times more expensive for private rooms than for shared, if such preferences are met [3]. However, when a patient does not have such a room preference, but is still admitted to a private room due to lack of different cheaper room types, such supplements cannot be charged. A survey in Belgian hospitals by Verhelst [4] further shows how preferred room type unavailability may even be a cause for postponing admissions. It is therefore of considerable importance for hospital administrators to address these concerns by matching the availability of different room types with the respective demand by patients in order to maximize revenue.

This study focuses on the decision problem of hospital administrators who wish to address this issue by reallocating existing room infrastructure between different nursing wards belonging to different hospital units such as different surgical disciplines. These units may have different patient arrival patterns, length-of-stay (LOS) distributions and room preference profiles that, for historical or organizational reasons, do not match their currently-allocated infrastructure. Hence, reallocation may be necessary to match current patient preferences. Currently, a methodology for finding a suitable reallocation is non-existent.

To this end, an approach is presented which accounts for the patient flow behavior using a Continuous-Time Markov Chain (CTMC) model. This model assesses the allocation of rooms in a heuristic search procedure, where the solution is gradually improved by sampling randomly from an interpolation of a suitable objective function. Based on hospital data obtained from both a Danish and Belgian hospital, the performance of the proposed approach is validated with a range of numerical experiments.

The remainder of this paper is organized as follows. First, the present work is positioned in the context of the relevant literature in Section 1.1. In Section 2, the specific assumptions and problem details of this study are elaborated upon. Section 3 presents the proposed solution approach followed by Section 4 which applies the approach by way of computational experimentation. Finally, in Section 5 conclusions and future research directions are presented.

1.1. Literature Review

The number of operations research methods that specifically address room capacity optimization in nursing wards, as opposed to solely considering bed capacity, is limited. Those studies that do consider this specific aspect generally concern decision problems at an operational decision level, where such details regarding room infrastructure cannot be ignored. Notably, Demeester et al. [5] formulated and studied a patient admission scheduling problem that addresses the assignment of admitted patients to beds over a given, short-term, planning horizon, considering room type and equipment. The consideration of gender conflicts in shared rooms and room type preferences by patients requires explicit modeling of room infrastructure. Given that Demeester et al. [5] formulated a challenging combinatorial optimization problem along with problem instances. They triggered a series of different studies further investigating algorithm development [6–8], different modeling aspects [9–11], and complexity [12]; all of which include the explicit consideration of room infrastructure. Most of these studies apply meta-heuristic optimization techniques to deal with problems of realistic size. Nevertheless, three studies apply Mixed Integer Linear Programming (MILP) to models of reduced size [12], in a dynamic setting [10] (where sub-problems are typically smaller) or combined with column generation [8] to improve lower and upper bounds on the sum of the patient-allocation penalties. Other studies in this area, though not derived from the formulation of Demeester et al. [5], demonstrate that considering room infrastructure is necessary for the practical implementation of systems. Bachouch et al. [13] presented a hospital bed management problem where patient admissions are scheduled while considering no-mixed gender rooms, isolation of contagious patients in single rooms or alone in double rooms, and incompatibilities between pathologies. A MILP model is formulated which is subsequently applied to different solvers in a computational comparative study. Schmidt et al. [14] also presented a decision support model for admission planning and assignment to wards. Their model also explicitly accounts for the availability of beds in either private or shared rooms, depending on the planned patients' preferences. Both an exact approach, using a MILP formulation, and heuristic approaches are compared in a computational study.

The application of Markov Chains to model patient flow is an uncommon approach compared to, for example, simulation-based modeling methods [15, 16]. Nonetheless, Markov Chains have been successfully applied in a variety of different cases in the last few years. Bartolomeo et al. [17] applied a Discrete-Time Markov Chain (DTMC) model to assess the re-admission probability of patients. Further, Broyles et al. [18] applied a DTMC to predict the number of inpatients, demonstrating how their model attains superior predictability compared to a seasonal Autoregressive Integrated Moving Average model.

Concerning CTMCs, both He et al. [19] and Shao et al. [20] developed a model to assess and identify bottlenecks with regard to the colonoscopy screening process and surgical operations in an emergency department, respectively. Furthermore, Wang et al. [21] applied a CTMC to model care delivery to patients in rooms by modeling the system as a closed network. Lastly, Shaw and Marshall [22] evaluated the LOS

for heart-failure patients, demonstrating how the Coxian phase-type distribution is adequate in this regard. The focus of all these studies is on modeling patient flow, while none of them utilize their approach to optimize the system. Only Andersen et al. [23] model patient hospitalization and relocation to multiple wards and employ a heuristic to optimize ward capacity. However, their study only considers capacity on an aggregated level (# beds per ward) and does not account for room infrastructure. The present study elaborates on Andersen et al. to consider room infrastructure as well.

1.2. Contribution

Interestingly, most studies considering room infrastructure availability have done so only at an operational decision level. E.g. when admitting patients to wards, one must account for the available room types and patients' preferences. However, in the context of hospital competition, matching infrastructure availability to demand is also of strategic importance for maximizing revenue and providing enhanced service to patients. Currently, no existing methodology for strategic room/bed (re)allocation considers the aspect of room types. This study proposes a CTMC model combined with a heuristic search procedure to address this aspect at the strategic decision level, where capacity and infrastructure can be reallocated between hospital units to better match individual demand patterns. To our knowledge, this is the first analytical approach that accounts for patient arrivals, relocation, and room type preferences; and, furthermore, where room configuration is optimized.

2. Problem Description

The decision problem studied in this paper involves the allocation of room types to nursing units of different medical specialisms. In this setting, the most differentiating characteristic between room types, namely being either private (one bed per room) or shared (two or more beds per room), is scrutinized. It is assumed that the total availability of private and shared rooms is fixed, but that room types may be reassigned between units. Such situations may occur, for example, when different nursing units occupying a single, physical area are reorganized or when patient characteristics such as LOS distributions or private room preferences have changed, necessitating a reallocation in order to realign available room infrastructure with demand.

Patients are assumed to arrive at the hospital according to a time-homogeneous process where both inter-arrival time and LOS are random. Furthermore, patients can be grouped into types such that each type prefers admission to a specific nursing ward. However, when capacity is insufficient, patients must not be made to wait for a bed, but be relocated to a ward where capacity is available. In addition, a certain fraction of the patients prefer admission to a private room, whereas the remaining patients have no preference concerning whether their room is private or shared. These assumptions are elaborated upon at greater length later in Section 3.2.

Formal Definition

Formally, a hospital setting is considered which features a set of wards \mathcal{W} , $|\mathcal{W}|$ types of patients, with each type preferring admission to a unique ward $i \in \mathcal{W}$, and a set of room types \mathcal{R} . Let $u_{ir} \in \mathbb{N}_0$ define the number of rooms of type $r \in \mathcal{R}$ that have been allocated to ward $i \in \mathcal{W}$ and $b_r \in \mathbb{N}_0$ define the capacity associated with each room type. Further, let set \mathcal{R} feature a subset of room types for which $b_r > 1$, and a *private* room type where $b_r = 1$. Finally, let $N_r \in \mathbb{N}_0$ define the available number of rooms of type $r \in \mathcal{R}$, and $M_i \in \mathbb{N}_0$ define the aggregated capacity of each ward $i \in \mathcal{W}$. Then,

$$\sum_{i \in \mathcal{W}} u_{ir} = N_r \quad \forall r \in \mathcal{R} \quad (1)$$

and,

$$\sum_{r \in \mathcal{R}} u_{ir} b_r = M_i \quad \forall i \in \mathcal{W} \quad (2)$$

Now, let \mathbf{u} define a matrix of the elements $u_{ir} \forall i \in \mathcal{W}, r \in \mathcal{R}$, and define:

- $f(\mathbf{u})$ yields the expected total number of patients relocated to an alternative ward per day, an alternative ward being defined as a ward having spare capacity.
- $g(\mathbf{u})$ yields the expected total number of patients who prefer a private room *and* are correspondingly assigned to one.

Let $\tau \in \mathbb{R}_{>0}$ denote an upper bound on $f(\mathbf{u})$ ensuring that a substantial number of patients will receive their preferred care. The objective of this study is therefore to derive a configuration of the room types, u_{ir} , that fulfills,

$$\text{Maximize} \quad g(\mathbf{u}) \quad (3)$$

$$\text{Subject to} \quad f(\mathbf{u}) \leq \tau \quad (4)$$

$$\sum_{i \in \mathcal{W}} u_{ir} = N_r \quad \forall r \in \mathcal{R} \quad (5)$$

$$\sum_{r \in \mathcal{R}} u_{ir} b_r \geq 1 \quad \forall i \in \mathcal{W} \quad (6)$$

$$u_{ir} \in \mathbb{N}_0 \quad \forall i, r \in \mathcal{W}, \mathcal{R} \quad (7)$$

The aim of formulation (3)-(7) is to attain the maximum expected number of patient-room preference matches, subject to a limited number of relocated patients (Constraint 4), a fixed capacity of each room type (Constraint 5) and an assignment of minimum one room per ward (Constraint 6).

3. Modeling & Solution Approach

The evaluation of $g(\mathbf{u})$ and $f(\mathbf{u})$ depends on an explicit modeling of the admission process' related queueing system that arises from the random arrival and room occupation of patients in nursing wards. Due to the complexity of this queueing system (cf. Section 3.2), optimization problem (3)-(7) cannot be solved to optimum without complete enumeration. Therefore, this study proposes a heuristic search procedure. An outline of our approach is presented in Figure 1 with references to the sections associated with each model component. The Randomized & Interpolated Search (RIS) heuristic applies an iterative procedure to sample good solutions from the solution space of (3)-(7), where in each iteration a new solution is selected based on an interpolation of scattered samples from the solution space. As we will show later, instead of selecting a room configuration, the RIS heuristic only needs to consider the aggregated ward capacity. The structure of this approach is detailed in Section 3.1. To evaluate the behavior of the queueing system associated with each solution, a time-homogeneous CTMC model proposed by Andersen et al. [23] is employed to derive the expected room occupancy. The CTMC model will be presented in Section 3.2. Since the CTMC is computationally expensive, Section 3.3 presents a core element of our heuristic search procedure, an approximate, fast, surrogate objective function. Finally, given that patient behavior is considered to be exclusively dependent upon aggregated ward capacity, the room configuration can be derived with Integer Linear Programming (ILP). The precise means by which this is achieved is detailed throughout Section 3.4. Lastly, note that an overview of mathematical notation is presented in Table 9 of Appendix A.

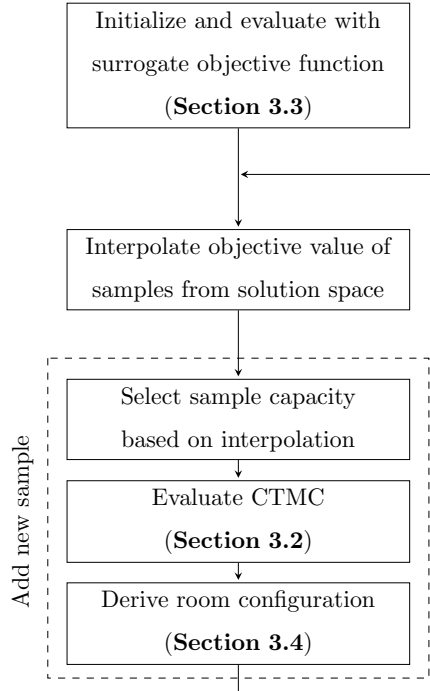


Figure 1: Outline of the RIS heuristic with references to the associated sections.

3.1. Randomized & Interpolated Search (RIS) heuristic

Consider a given room configuration \mathbf{u} and recall capacity constraints (5) and (6). Now, consider the solution space U resulting from these constraints, and let $Y_f(\mathbf{u})$ and $Y_g(\mathbf{u})$ yield estimates of $f(\mathbf{u})$ and $g(\mathbf{u})$ based on an interpolation of some known solutions in this space. Let \mathbf{x} define the set of these known solutions and $Z(\mathbf{u})$ define a probability mass function that corresponds proportionally to $Y_g(\mathbf{u})$ and sums to unity. Then, in order to approach the configuration of rooms that attains the maximum of $g(\mathbf{u})$, the following stepwise procedure is considered:

1. Select a range of initial solutions for \mathbf{x} .
2. Calculate $g(\mathbf{u})$ and $f(\mathbf{u})$ based on \mathbf{x} .
3. Derive $Y_f(\mathbf{u})$ and $Y_g(\mathbf{u})$ by interpolating between the known solutions in \mathbf{x} .
4. Derive $Z(\mathbf{u})$ in accordance with $Y_g(\mathbf{u})$.
5. Add a new configuration, \mathbf{u}' , to \mathbf{x} by sampling from $Z(\mathbf{u})$, constrained by $Y_f(\mathbf{u}) \leq \tau$ in accordance with (4), and calculate $g(\mathbf{u}')$ and $f(\mathbf{u}')$.
6. If the elapsed time exceeds the fixed time limit then **stop**; otherwise return to step 3.

The procedure is initialized by requiring that \mathbf{x} contains the $|\mathcal{W}|$ extreme points in which all room types have been moved to a single ward, respecting lower capacity bound (6), and thus ensuring that all room configurations are included in the interpolation. Next, \mathbf{x} is expanded and a basis for the interpolation is created by sampling uniformly from U .

By applying this procedure, $Y_g(\mathbf{u})$ recursively approaches $g(\mathbf{u})$ in the space constrained by (4)-(7). Notice how the interpolations $Y_g(\mathbf{u})$ and $Y_f(\mathbf{u})$ are gradually improved based on the solution samples \mathbf{x} where $Y_g(\mathbf{u})$ is employed to focus the search through probability mass function $Z(\mathbf{u})$ and $Y_f(\mathbf{u})$ is employed to estimate the feasible space.

When the sampling based on $Y_g(\mathbf{u})$ is rather widespread, the probability mass is concentrated upon the promising regions by performing the conversion $Y'_g(\mathbf{u}) = Y_g(\mathbf{u})^\beta$, where $\beta \in \mathbb{R}_{\geq 1}$, thereby amplifying the curvature of the interpolation. However, this still requires an initialization of $Y_g(\mathbf{u})$ based on uniformly distributed solution-evaluations throughout U . In other words, runtime is potentially wasted in regions that are not relevant to the objective. To overcome this, let $\tilde{f}(\mathbf{u})$ and $\tilde{g}(\mathbf{u})$ define surrogates of functions $f(\mathbf{u})$ and $g(\mathbf{u})$ that have similar optima, but shorter evaluation times. Thus, by conducting the initialization using the surrogate $\tilde{g}(\mathbf{u})$ and then switching to the true objective function, $g(\mathbf{u})$, for the remaining steps, the *true*, and slower, solution-evaluations are *only* performed in the most promising region of the search space.

Let $\tilde{\mathbf{x}}$ define the set of configurations that have been evaluated using the aforementioned surrogate function. Then, as the search procedure progresses, the interpolation will be derived on the basis of $\tilde{\mathbf{x}}$ as well

as the gradually increasing set \mathbf{x} . Now, to ensure that \mathbf{x} can replace $\tilde{\mathbf{x}}$ in a limited number of iterations, a proximity tolerance ξ is defined such that if the euclidean distance $\sqrt{\sum_{i \in \mathcal{W}} (M_i - \tilde{M}_i)^2}$ is smaller than or equal to ξ , where M_i and \tilde{M}_i are the aggregated capacities of an element in \mathbf{x} and $\tilde{\mathbf{x}}$, respectively, then the surrogate solution associated with \tilde{M}_i is removed from $\tilde{\mathbf{x}}$. The final search procedure is documented in Algorithm 1. The implications of varying β and ξ are elaborated upon in Section 4.

Algorithm 1 The RIS heuristic.

```

1:  $\tilde{\mathbf{x}} \leftarrow \text{uniformSampling}()$  ▷ Initialize and evaluate surrogate sampling
2:  $\mathbf{x} \leftarrow \emptyset$ 
3: while  $\text{elapsedTime} < \text{timeLimit}$  do
4:    $Y \leftarrow \text{interpolate}(\tilde{\mathbf{x}}, \mathbf{x})$ 
5:    $Y' \leftarrow \text{exponentiate}(Y)$  ▷ Exponentiate the interpolation
6:    $Z \leftarrow \text{convertToPMF}(Y')$ 
7:    $\mathbf{x} \leftarrow \text{addNewSample}(Z, \mathbf{x})$  ▷ Add and evaluate new sample
8:    $\tilde{\mathbf{x}} \leftarrow \text{remove}(\mathbf{x}, \tilde{\mathbf{x}})$  ▷ Check and remove proximate surrogate samples
9: end while
10:  $\mathbf{u}^* \leftarrow \text{getBest}(\mathbf{x})$ 
    return  $\mathbf{u}^*$ 

```

3.2. Evaluating $g(\mathbf{u})$ and $f(\mathbf{u})$

Recall the system presented in Section 2, featuring a set of patient types requiring assignment to a set of wards, where in case of insufficient capacity patients are either moved to an alternative ward or admitted to a location that is not included in the model, i.e. lost from the system. To model this behavior, a CTMC approach [23] is employed.

Consider a time-homogeneous CTMC with state space S and state definition $s = (w_{11}, w_{21}, \dots, w_{ij}, \dots, w_{|\mathcal{W}||\mathcal{W}|})$, where w_{ij} is the number of type i patients hospitalized in ward j , with $i, j \in \{1, 2, \dots, |\mathcal{W}|\}$. Let f'_j define the number of free beds in ward j , so $f'_j = M_j - \sum_{i \in I} w_{ij}$, where M_j is the aggregated capacity of each ward. Additionally, let λ_i define the arrival rate of patient type i and that all patient arrivals, regardless of their type, are generated according to a Poisson process. Moreover, let μ_i define the service rate of patient type i , assuming that inter-service times are exponentially distributed. Furthermore, let $p(f'_1, f'_2, \dots, f'_{|\mathcal{W}|})_{ij}$ define the fraction of type i patients routed to ward j , governed by the number of free beds in each ward; f'_1, f'_2, \dots and $f'_{|\mathcal{W}|}$.

Let q_{ss^*} define the rate at which the system transitions from a current state $s \in S$ to a new state $s^* \in S$.

Then,

$$q_{ss^*} = \begin{cases} \lambda_i & \text{if } s^* = (\dots, w_{ii} + 1, \dots, f'_i - 1, \dots) \text{ and } f'_i > 0 \quad \forall i \in I \\ \lambda_i p(f'_i = 0)_{ij} & \text{if } s^* = (\dots, w_{ij} + 1, \dots, f'_j - 1, \dots) \text{ and } f'_i = 0, f'_j > 0 \quad \forall i, j \in I, i \neq j \\ \lambda_i p(f'_i = 0, f'_k = 0)_{ij} & \text{if } s^* = (\dots, w_{ij} + 1, \dots, f'_j - 1, \dots) \text{ and } f'_i = 0, f'_k = 0, f'_j > 0 \quad \forall i, j, k \in I, i \neq j \neq k \\ \vdots & \vdots \\ \lambda_i p(f'_i = 0, f'_k = 0, \dots, f'_l = 0)_{ij} & \text{if } s^* = (\dots, w_{ij} + 1, \dots, f'_j - 1, \dots) \text{ and } f'_i = 0, f'_k = 0, \dots, f'_l = 0, f'_j > 0 \quad \forall i, j, k \dots l \in I, i \neq j \neq k \neq \dots \neq l \\ \mu_i w_{ij} & \text{if } s^* = (\dots, w_{ij} - 1, \dots, f'_j + 1, \dots) \text{ and } w_{ij} > 0 \quad \forall i, j \in I \end{cases}$$

where all other transition rates $q_{ss^*} = 0$ for $s \neq s^*$, and the diagonal elements $q_{ss} = -\sum_{s^* \in S \setminus \{s\}} q_{ss^*}$.

The fraction $p(f'_i = 0, f'_k = 0, f'_j > 0, \dots, f'_N > 0)_{ij}$ is abbreviated $p(f'_i = 0, f'_k = 0)_{ij}$, the change $s^* = (\dots, w_{ij} + 1, \dots, f'_j - 1, \dots)$ indicates the arrival of patient i to a ward j , and the change $s^* = (\dots, w_{ij} - 1, \dots, f'_j + 1, \dots)$ indicates a corresponding discharge. Let Q define the transition rate matrix of rates $q_{ss^*} \quad \forall s, s^* \in S$. To derive state distribution π , S is reduced by employing the truncation algorithm described by Andersen et al. [23] where bounds on the probability mass are used to identify the neglectable subspace of S . The global balance equations $\pi Q = 0$ are solved using successive overrelaxation [24, p. 311] with a relaxation parameter equal to 1.75. Convergence is measured on the largest relative difference between successive iterations, in accordance with Andersen et al. [23].

Let $\pi_i(n)$ define the probability of exactly $n \in \mathbb{N}_0$ patients being hospitalized in ward $i \in \mathcal{W}$; $0 \leq n \leq M_i$. $\pi_i(n)$ is consequently a marginal distribution to the state distribution of the CTMC. Recall that all patients may be categorized as one of two types: Patients who prefer a private room and patients who have no preference concerning whether they are assigned to a shared or private room. Suppose we randomly sample an arriving patient. The probability that the patient prefers a private room is $\psi \in \mathbb{R}_{0 \leq \psi \leq 1}$. Now, consider that n patients have been admitted to ward $i \in \mathcal{W}$. Let $P_i(y^a, y^b)$ define the probability that of those n patients, exactly y^a patients prefer a private room and $y^b = n - y^a$ patients have not indicated their preference. Then,

$$P_i(y^a, y^b) = B(y^a, n, \psi) \cdot \pi_i(n) \quad (8)$$

where

$$B(y^a, n, \psi) = \binom{n}{y^a} \psi^{y^a} (1 - \psi)^{y^b} \quad (9)$$

is the probability mass function of the binomial distribution with $n = y^a + y^b$ trials, and success probability ψ . Further, let $\rho_i(y^a)$ define the probability that exactly y^a beds are occupied by patients who prefer a private room, and that the occupied beds can be located in both private and shared rooms. Then,

$$\rho_i(y^a) = \sum_{y^b=0}^{M_i-y^a} P_i(y^a, y^b) \quad (10)$$

1 which from (8) results in:

$$\rho_i(y^a) = \sum_{y^b=0}^{M_i-y^a} \left(B(y^a, n, \psi) \cdot \pi_i(n) \right) \quad (11)$$

2 Function $\rho_i(y^a)$ is essential to both the definition of $f(\mathbf{u})$ and $g(\mathbf{u})$, as will be demonstrated in what
3 follows.

4
5 Assume the following ordering of patients as they are hospitalized:

- 6 1. Whenever patients who prefer private rooms are hospitalized, they will *always* be assigned to a private
7 room if one is available.
- 8 2. Patients who have no preference regarding room types, are *only* assigned to a private room if no shared
9 room capacity is available.

10 By observing an arbitrary ward $i \in \mathcal{W}$, the expected number of patients who prefer a private room and are
11 correspondingly assigned to one is $\sum_{k=0}^{u_{i, \text{private}}} (k \cdot \rho_i(k)) + \sum_{k=u_{i, \text{private}}+1}^{M_i} (u_{i, \text{private}} \cdot \rho_i(k))$, resulting in the
12 following objective function:

$$g(\mathbf{u}) = \sum_{i \in \mathcal{W}} \left(\sum_{k=0}^{u_{i, \text{private}}} (k \cdot \rho_i(k)) + \sum_{k=u_{i, \text{private}}+1}^{M_i} (u_{i, \text{private}} \cdot \rho_i(k)) \right) \quad (12)$$

13 where $u_{i, \text{private}}$ is the number of private rooms allocated to ward $i \in \mathcal{W}$. Notice that $g(\mathbf{u})$ is, in essence,
14 independent of the characteristics of the shared room types. Regarding $f(\mathbf{u})$, which ensures an upper bound
15 on the number of relocated patients through (4), the overall flow of patients into the hospital is of more
16 concern. Consider blocking probability $\pi_i^B = \pi_i(M_i)$ of ward $i \in \mathcal{W}$, then

$$f(\mathbf{u}) = \sum_{i \in \mathcal{W}} \lambda_i \pi_i^B \quad (13)$$

17 denotes the total expected number of patients who are rejected and correspondingly relocated upon arrival.
18 Figure 2 depicts the dependencies between the CTMC and Expressions (12) and (13), respectively. Notice
19 that the behavior of the system, as evaluated by the CTMC, depends only on the aggregated capacity. This
20 feature will be exploited in the search procedure using ILP modeling introduced in Section 3.4.

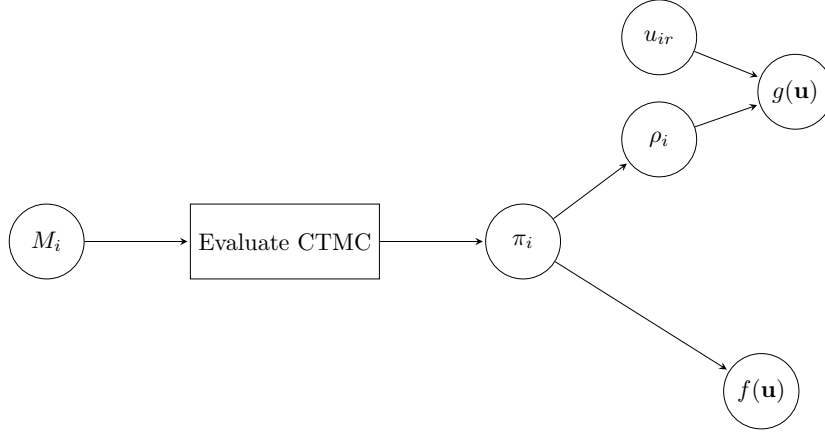


Figure 2: Dependencies in evaluating $g(\mathbf{u})$ and $f(\mathbf{u})$.

3.3. The Surrogate Functions

Consider the flow of patients into a single ward, as depicted in Figure 3. From this perspective one notices that the system approaches an $M/M/c/c$ queueing model as the number of arriving relocated patients decreases. That is, a queue where the capacity of the entire system equals the number of servers. For the $M/M/c/c$ model,

$$\tilde{\pi}_i(n) = \frac{(\lambda_i/\mu_i)^n/n!}{\sum_{k=0}^{M_i} (\lambda_i/\mu_i)^k/k!} \quad (14)$$

where $\tilde{\pi}_i(n)$ is the probability that exactly $n \in \mathbb{N}_0$ patients are hospitalized in ward $i \in \mathcal{W}$ [24, p. 434]. Equation (14) is therefore an approximation of $\pi_i(n)$, which accuracy decreases as more patients are relocated within the system. Correspondingly, if all patients are lost from the system on arrival, then Equation (14) substitutes for $\pi_i(n)$ exactly. $f(\mathbf{u})$ may therefore be approximated by:

$$\tilde{f}(\mathbf{u}) = \sum_{i \in \mathcal{W}} \lambda_i \tilde{\pi}_i^B \quad (15)$$

where $\tilde{\pi}_i^B = \tilde{\pi}_i(M_i)$. Similarly, in the surrogate for $g(\mathbf{u})$, $\tilde{\pi}_i(n)$ is employed to approximate (11) by,

$$\tilde{\rho}_i(y^a) = \sum_{y^b=0}^{M_i-y^a} \left(B(y^a, n, \psi) \cdot \tilde{\pi}_i(n) \right) \quad (16)$$

which is then used to substitute $\rho_i(y^a)$ in Equation (12), leading to $\tilde{g}(\mathbf{u})$. Notice that when (14) replaces $\pi_i(n)$ from the CTMC, the computational effort of setting-up and applying successive over-relaxation is avoided which is the proposed approach to the global balance equations, $\boldsymbol{\pi}Q = 0$ [23]. As a result, the search procedure is scoped rather quickly by creating an initial outline of both $f(\mathbf{u})$ and $g(\mathbf{u})$.

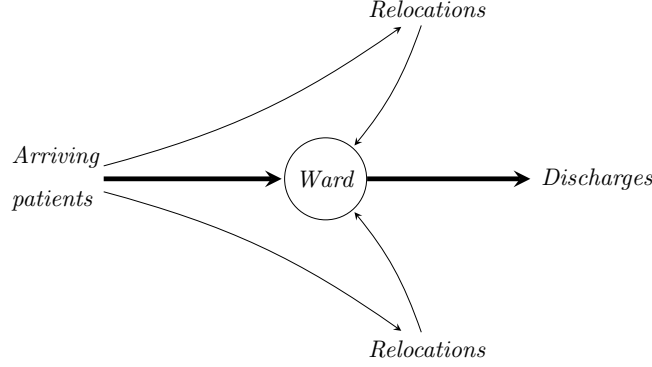


Figure 3: Flow of patients *to* and *from* a single ward.

3.4. Sub-Optimal Room Configuration

Recall the dependencies in deriving functions $f(\mathbf{u})$ and $g(\mathbf{u})$, depicted in Figure 2. We only require the aggregated capacity M_i to evaluate the system through the CTMC. Not until then is the room configuration \mathbf{u} applied to the patient occupancy distribution ρ_i to derive the objective value $g(\mathbf{u})$. Hence, by assuming a fixed aggregated capacity, M_i , the problem of maximizing $g(\mathbf{u})$ reduces to an ILP model.

Let \mathcal{R}' define the set of shared room types. That is, $\mathcal{R}' \subset \mathcal{R}$ and $|\mathcal{R}'| = |\mathcal{R}| - 1$. Further, let set $\mathcal{J}_i = \{0, 1, 2, \dots, M_i\}$ account for the number of beds that can be assigned to private rooms in each ward $i \in \mathcal{W}$. Additionally, let the decision variable $z_{ij} \in \{0, 1\}$ equal 1 whenever ward $i \in \mathcal{W}$ is assigned j private beds, where $j \in \mathcal{J}_i$; and otherwise 0. Further, let parameter $\omega_{ij} \in \mathbb{R}_{\geq 0}$ define the expected number of patients who both prefer *and* are also assigned to private rooms in ward $i \in \mathcal{W}$, given that j private rooms are available in this ward. That is, following the convention in Equation (12),

$$\omega_{ij} = \sum_{k=0}^j (k \cdot \rho_i(k)) + \sum_{k=j+1}^{M_i} (j \cdot \rho_i(k)) \quad (17)$$

where as before $\rho_i(k)$ is derived using Equation (11), resulting in the objective function $\sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{J}_i} z_{ij} \omega_{ij}$, which yields exactly the same result as Equation (12).

Lastly, let $N_{private}$ define the available number of private rooms, and $y_{ir} \in \mathbb{N}_0$ the number of shared room types $r \in \mathcal{R}'$ assigned to ward $i \in \mathcal{W}$. A feasible room configuration can then derived by employing the ILP model:

$$\text{Maximize} \quad \sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{J}_i} z_{ij} \omega_{ij} \quad (18)$$

$$\text{Subject to} \quad \sum_{j \in \mathcal{J}_i} z_{ij} = 1 \quad \forall i \in \mathcal{W} \quad (19)$$

$$\sum_{j \in \mathcal{J}_i} z_{ij} \cdot j + \sum_{r \in \mathcal{R}'} y_{ir} b_r = M_i \quad \forall i \in \mathcal{W} \quad (20)$$

$$\sum_{i \in \mathcal{W}} \sum_{j \in \mathcal{J}_i} z_{ij} \cdot j \leq N_{\text{private}} \quad (21)$$

$$\sum_{i \in \mathcal{W}} y_{ir} b_r \leq N_r \quad \forall r \in \mathcal{R}' \quad (22)$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j \in \mathcal{W}, \mathcal{J}_i \quad (23)$$

$$y_{ir} \in \mathbb{N}_0 \quad \forall i, r \in \mathcal{W}, \mathcal{R}' \quad (24)$$

Constraints (19)-(22) are defined as follows:

- (19) ensures that each ward receives a fixed amount of private beds.
- (20) ensures that the distribution of the aggregated capacity is maintained, keeping the parameter ω_{ij} valid.
- Finally, (21) and (22) restrict the maximum occurrence of each room type. Recall that N_r defines the total number of room types $r \in \mathcal{R}$ available to the hospital.

By evaluating $\pi_i(n)$ and $\tilde{\pi}_i(n)$, using the aforementioned CTMC and Equation (14), ILP formulation (18)-(24) can be employed to yield the optimum room configuration conditioned by the distribution of the aggregated capacity. Therefore, instead of evaluating based on the room configuration directly, this feature is exploited in our heuristic search procedure by applying aggregated capacity M_i as the decision variable.

Recall Algorithm 1, where \mathbf{x} and $\tilde{\mathbf{x}}$ contain the samples for which the true and surrogate functions have been evaluated, respectively. By applying the aggregated capacity to \mathbf{x} and $\tilde{\mathbf{x}}$, the associated room configuration is derived as follows:

1. Evaluate distributions $\pi_i(n)$ and $\tilde{\pi}_i(n)$ as per Sections 3.2 and 3.3.
2. Calculate $f(\mathbf{u})$ and $\tilde{f}(\mathbf{u})$; and derive the sub-optimal room configuration which maximises $g(\mathbf{u})$ and $\tilde{g}(\mathbf{u})$ by ILP formulation (18)-(24).

4. Numerical Study

In this section, the RIS heuristic presented in Section 3 is evaluated in a range of numerical experiments to assess its performance. These experiments are conducted on hospital data introduced by Andersen

et al. [23], and room infrastructure data from a Belgian hospital. All experiments are implemented in Java, including the CTMC from Section 3.2. To derive $Y_g(\mathbf{u})$ and $Y_f(\mathbf{u})$ natural neighbor interpolation [25] is employed using the `SibsonInterpolator2` class of the Java Mines Toolkit². Lastly, the ILP model presented in Section 3.4 is solved using IBM ILOG CPLEX 12.7.1³.

4.1. Data Description

The data for our subsequent experiments is based on three different datasets which are obtained from the study by Andersen et al. [23]. The data accounts for three different wards and consists of patient arrival rates and length of stay distributions; the respective routing probabilities in the system, and lastly the total bed capacity. No data was obtained specifically with regard to the number of room types for this case. However, data from a Belgian hospital⁴ suggests that the number of private rooms may easily constitute half of the total bed capacity. This proportion will serve as the basis for the three sets. Furthermore, even though the presented approach may be generalized to any capacity configuration for the shared room types, only a single shared room type consisting of two beds is considered, next to a single private room type (i.e. $|\mathcal{R}| = 2$).

All experiments primarily consider a dataset referred to as the *original* set, which is based solely on true patient data. Two additional sets, *high arrival rate* and *high relocation*, are derived from the original data by adjusting the arrival rate and routing probability parameters, respectively. These additional sets are included to assess the potential changes in patient characteristics. In addition, since no data was obtained concerning the proportion of private patients, a value of $\psi = 0.2$ is assumed, unless otherwise stated.

The parameters associated with each dataset are presented in Tables 1 and 2. Furthermore, the initialization of the RIS heuristic includes a minimization of the expected number of relocated patients, $f(\mathbf{u})$ from Andersen et al. [23]. Each minimization, denoted as $\min\{f(\mathbf{u})\}$, is presented in Table 3.

²Java Mines Toolkit on interpolation and gridding - <http://dhale.github.io/jtk/api/edu/mines/jtk/interp/package-summary.html>

³IBM ILOG CPLEX Optimization Studio - <https://www.ibm.com/products/ilog-cplex-optimization-studio>

⁴Data supplied by hospital AZ Maria Middelaers, based in Gent, Belgium, in the context of iMinds ICON project AORTA - <https://www.imec-int.com/nl/imec-icon/research-portfolio/aorta>

Dataset	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3	Total Beds	$N_{private}$	N_{double}
Original	5.42	3.96	2.52	0.19	0.19	0.11	74	36	19
High Arrival Rate	6.78	3.96	2.52	0.19	0.19	0.11	74	36	19
High Relocation	5.42	3.96	2.52	0.19	0.19	0.11	74	36	19

Table 1: Rates and capacities associated with each of the three datasets. All rates and the total bed capacity are obtained from [23], whereas the ratio of private to shared rooms, $\frac{N_{private}}{N_{double}}$ is based on data from a Belgian hospital³. All rates are given in number of patients per day.

Dataset	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	p_{31}	p_{32}	p_{33}
Original	-	0.05	0.23	0.10	-	0.27	0.06	0.00	-
High Arrival Rate	-	0.05	0.23	0.10	-	0.27	0.06	0.00	-
High Relocation	-	0.05	0.95	0.10	-	0.90	0.06	0.00	-

Table 2: The routing probabilities associated with each of the three datasets, respectively. All parameter values and assumptions concerning patient relocation have been obtained from [23]. Note that $p(f'_i = 0)_{ij}$ is abbreviated p_{ij} .

Dataset	$\min\{f(\mathbf{u})\}$	M_1	M_2	M_3
Original	1.592	32	24	18
High Arrival Rate	2.354	39	23	12
High Relocation	1.688	33	25	16

Table 3: Each minimization, $\min\{f(\mathbf{u})\}$, obtained from Andersen et al. [23]. The associated distribution of beds for each dataset is provided.

4.2. Error of the Surrogate Function

Prior to evaluating the heuristic search procedure, an assessment of the error of both surrogate functions was performed by conducting a full enumeration of the search space. In order to accommodate this, room availability was limited to $N_{private} = 20$ private and $N_{double} = 10$ shared double rooms for the *high arrival rate* and *high relocation* datasets. Otherwise, the full availability of rooms for the *original* dataset was employed. The enumeration was conducted using the parameters from all three datasets (Table 1 and 2) on both the *true* functions $g(\mathbf{u})$ and $f(\mathbf{u})$, and *surrogate* functions $\tilde{g}(\mathbf{u})$ and $\tilde{f}(\mathbf{u})$.

Results were evaluated by calculating the error and comparing each functions' optimum. Table 4 presents the euclidean distance between the optima of $g(\mathbf{u})$ and $\tilde{g}(\mathbf{u})$. Notice that this is measured on the distribution of the aggregated capacity given that this is the primary decision variable. Table 4 includes the relative error concerning the optimum of the surrogate function. Lastly, Figure 4 shows the error of the original

data, namely $g(\mathbf{u}) - \tilde{g}(\mathbf{u})$. The figure also illustrates the optima of both the true and surrogate objective function.

Recall from Section 3.3 that the surrogate functions are based on $M/M/c/c$ model (14) which does not account for the hospitalization of relocated patients. It is therefore expected that the surrogate functions will lose accuracy as the number of relocated patients increases. That is, either when general routing probabilities are high, or if the wards lack capacity. The latter situation is reflected in Figure 4, showing that the error is smaller when the capacity is more evenly distributed, thereby resulting in fewer relocated patients.

Regarding the routing probabilities, Table 4 demonstrates how the relative error is fairly robust with regard to the changes between the *original* and *high relocation* datasets. However, the optima have changed substantially from a euclidean distance of approximately 1.4 to 20.8 beds. This shows that when a substantial number of patients can be relocated within the system, the RIS heuristic has to adapt interpolation $Y_g(\mathbf{u})$ to objective function $g(\mathbf{u})$ despite an inaccurate surrogate function.

Dataset	$\sqrt{\sum_{i \in \mathcal{W}} (M_i^* - \tilde{M}_i^*)^2}$	$(g(\tilde{\mathbf{u}}^g)/\tilde{g}(\tilde{\mathbf{u}}^g)) - 1$	$(f(\tilde{\mathbf{u}}^f)/\tilde{f}(\tilde{\mathbf{u}}^f)) - 1$
Original	1.414	0.016	0.092
High Arrival Rate	6.164	0.010	0.050
High Relocation	20.833	0.018	0.053

Table 4: The euclidean distance between the optima M_i^* of $g(\mathbf{u})$ and \tilde{M}_i^* of $\tilde{g}(\mathbf{u})$, and the relative errors at the optima, $\tilde{\mathbf{u}}^g$ and $\tilde{\mathbf{u}}^f$, for each surrogate function $\tilde{g}(\mathbf{u})$ and $\tilde{f}(\mathbf{u})$, respectively.

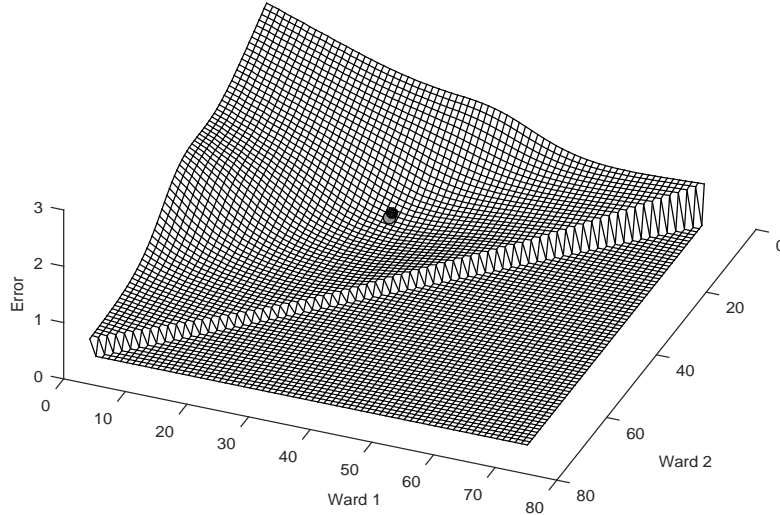


Figure 4: The error of surrogate objective function $\tilde{g}(\mathbf{u})$ on the original dataset, defined as $g(\mathbf{u}) - \tilde{g}(\mathbf{u})$, and derived by enumerating all solutions; showing, additionally, the optimum of $g(\mathbf{u})$ (black) and $\tilde{g}(\mathbf{u})$ (gray).

4.3. Evaluating the RIS Heuristic Parameters

The implications of adjusting the essential parameters are now explored. That is, the number of initial surrogate samples from the search space, the size of the exponent β , and the proximity tolerance ξ . All these parameters have been tested sequentially on the original data.

The results from adjusting the initial surrogate sampling is presented in Figure 5, showing the resulting interpolated estimate, $Y_g(\mathbf{u})$, and the associated runtimes based on 5, 20, 35 and 50 samples from the search space. Surrogate sampling is uniformly distributed and it is therefore expected that $Y_g(\mathbf{u})$ converges to the true function $g(\mathbf{u})$ when the sample size increases. Interestingly, the general shape of $g(\mathbf{u})$ can be determined fairly early, as shown in the experiment with only 5 samples.

By considering the strategic application of the RIS heuristic, we deem that the associated runtimes are fairly negligible, and since the apparent optimum does not change substantially after obtaining more than 20 samples, we deem that this is an adequate number of samples for our later optimization experiments.

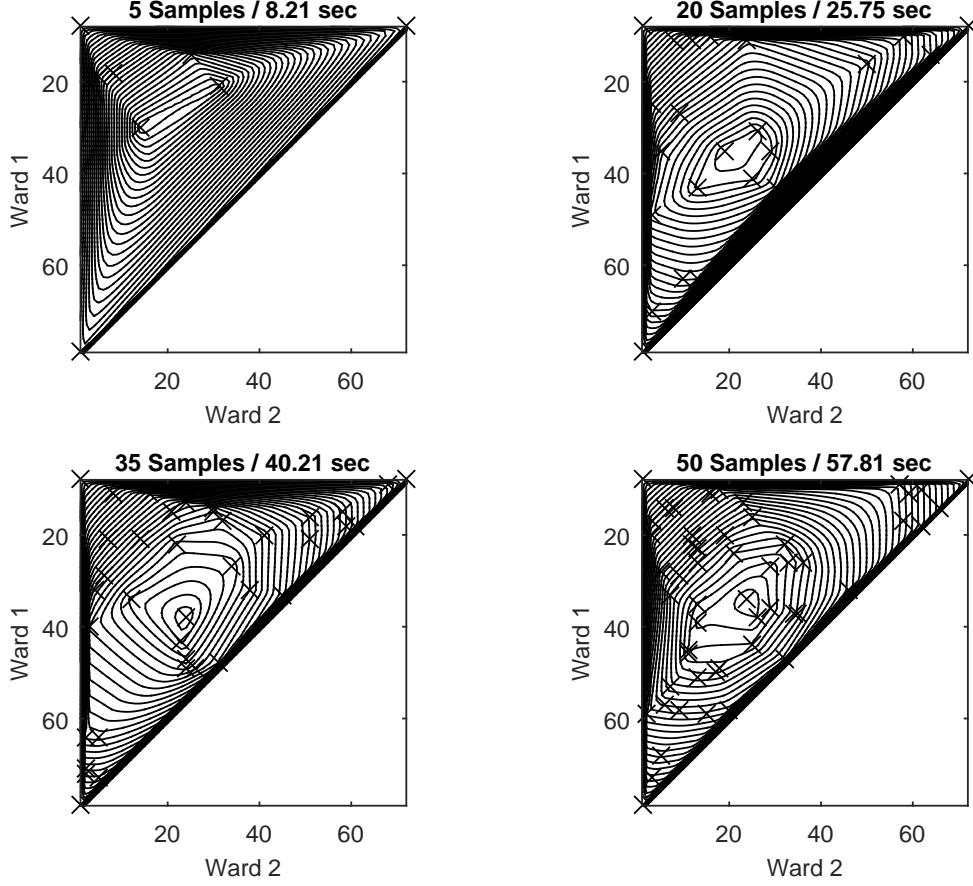


Figure 5: The result of gradually increasing the number of surrogate samples on the interpolated estimate $Y_g(\mathbf{u})$. $Y_g(\mathbf{u})$ does not change substantially upon applying more than 20 samples.

Exponent β was assessed based on values of 1, 8 and 16. In accordance with the RIS heuristic, these experiments were conducted by first obtaining 20 uniformly distributed samples, followed by 20 samples according to the recursively-updated density function $Z(\mathbf{u})$. The surrogate objective function was again employed to conserve the runtime of the experiments. Results are presented in Figure 6, showing that the search intensifies around the apparent optimum as a function of β . Notice that the experiment where $\beta = 1$, corresponding to a complete omission of the conversion, demonstrates the usefulness of this approach as the sampling is almost uniformly distributed. In the experiment where $\beta = 8$, samples are generally close to the optimum, whereas in the last experiment, where $\beta = 16$, samples are concentrated on the apparent optimum with only a few outliers. Based on these experiments, a value of $\beta = 8$ is employed to focus on the most promising region of the search space, but still attains some diversification.

Lastly, the effect of adjusting the proximity tolerance, ξ , was assessed using values of 1, 4 and 8. The

experiments were conducted by applying the full RIS heuristic using 20 initial surrogate samples, an exponent of $\beta = 8$, and an upper bound on the permitted number of relocated patients of $\tau = 1.9$. Each experiment was conducted using a runtime of 30 minutes.

Results of these experiments yielded almost identical performance in each case. This might have been caused by the choice of τ , which results in a rather limited sample space. Consequently, the proximity tolerance is rather arbitrarily set to $\xi = 4$ during the subsequent optimization experiments.

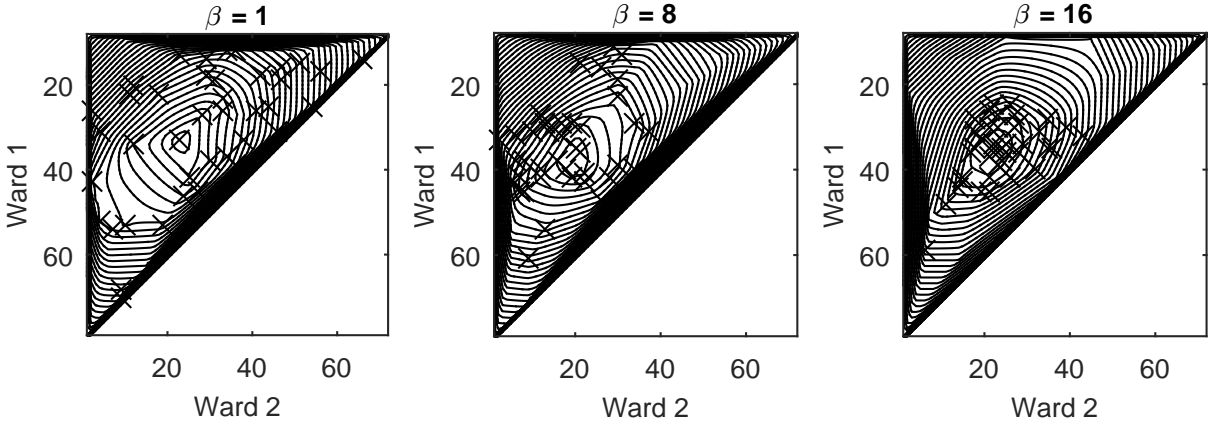


Figure 6: Interpolation $Y_g(\mathbf{u})$ resulting from a gradually increasing exponent, β . Each experiment is conducted by initializing with 20 uniform surrogate samples, followed by 20 samples based on $Z(\mathbf{u})$ (crosses).

4.4. Applying the RIS Heuristic

The full RIS heuristic is applied to the data presented in Section 4.1 based on the tests from Section 4.3. We begin by presenting an example of a single heuristic run, where gradually-obtained solutions are compared against the true optimum.

Overall performance is assessed by way of a number of experiments which compare the heuristic's solutions to the true optimum. Since no data was obtained concerning preference for private rooms, proportion ψ is investigated using three different levels. Furthermore, the robustness to changes in the patient arrival rates and relocation probabilities is of interest. Experiments will therefore be conducted on all three datasets (cf. Table 1 to 3).

4.4.1. A Single Run

Figure 7 illustrates the progression of the RIS heuristic on the *original* dataset for a runtime of 60 minutes and a bound of $\tau = \min\{f(\mathbf{u})\} \cdot 1.20 = 1.91$. During this time 21 iterations were conducted. The figure shows the interpolation $Y_g(\mathbf{u})$, samples \mathbf{x} and $\hat{\mathbf{x}}$, the optimum obtained by enumeration, and finally the *estimated* and *true* feasible search space defined by τ .

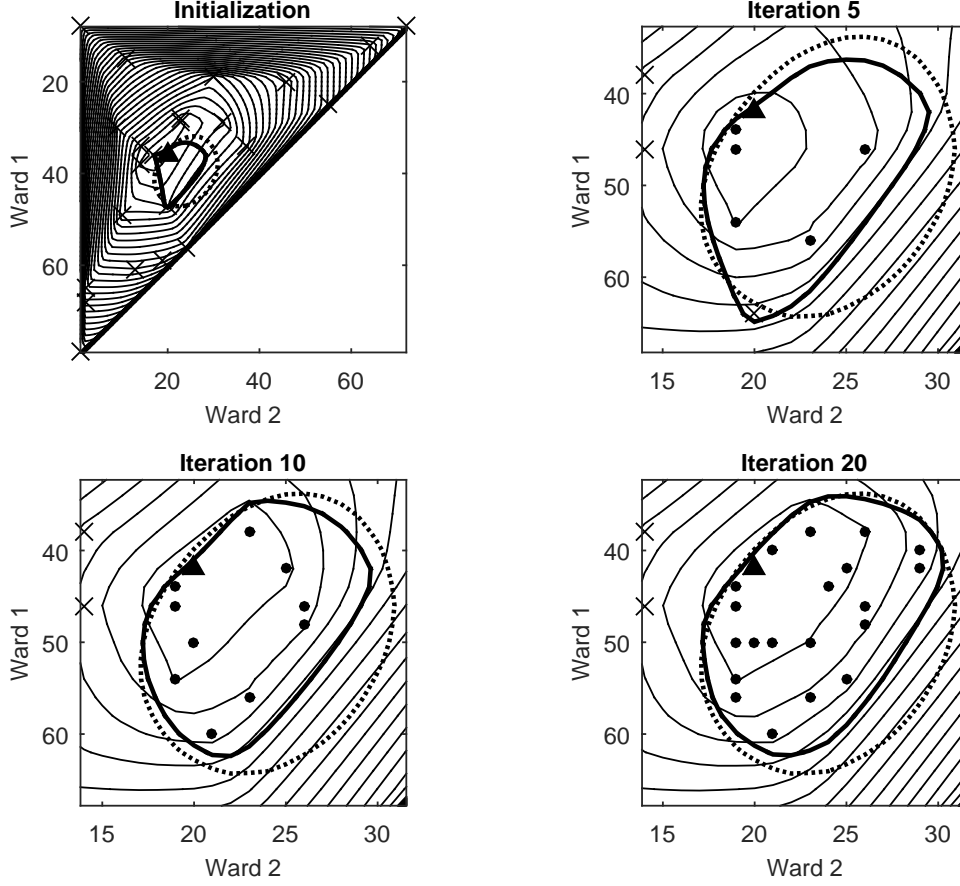


Figure 7: Progression of the RIS heuristic. The interpolation, $Y_g(\mathbf{u})$, samples \mathbf{x} (dots) and $\tilde{\mathbf{x}}$ (crosses) and the optimum (triangle) are shown. The true and estimated feasible search space is depicted with dotted and solid line, respectively.

The heuristic initializes with 20 surrogate samples, as shown in the upper left corner. At this stage, the apparent optimum is already close to the true optimum, which is immediately included within the search space. The remaining three graphs show the sampled solutions for iteration 5, 10 and 20. Notice that the estimated search space initially violates the true search space (in Figure 7, iteration 1 and 5: bottom-left solid line - estimated search space - exceeds dotted line - true search space), but then converges to the true search space near the optimum. At iteration 20 the estimated search space attains high accuracy near the true optimum, which decreases as more capacity is allocated to wards 1 and 2. Notice that samples are almost uniformly distributed due to the low slope near the optimum.

This example demonstrates the advantage of sampling from an interpolation based on a mix of both fast surrogate and slower true evaluations to determine the most promising region for an objective function of complex structure. The general performance of this approach for different parameter variations is

investigated in the following section.

4.4.2. Overall Performance

The overall RIS heuristic performance was assessed in two parts. Firstly, two runs were conducted on the *original* dataset for each of three levels of the private patient proportion $\psi = 0.2, 0.5$ and 0.7 , with a fixed runtime of 60 minutes. To properly assess the heuristic solutions, results were compared against the true optima obtained by enumerating the search space.

Furthermore, since robustness regarding changes in the patient characteristics is of particular interest, similar runs were conducted for the *high arrival rate* and *high relocation* datasets. However, in order to determine the optima for these additional tests, the room availability was limited to $N_{private} = 20$ and $N_{double} = 10$ rooms. Due to the reduction of feasible ward capacity configurations, the relocation bound to yield the maximal search space was omitted, and the runtime decreased to 10 minutes.

The results for the *original* dataset are presented in Table 5, featuring first the heuristic room configuration for private rooms, $u_{i,private}$, and shared double rooms $u_{i,double}$. Next, Table 5 presents the iterations corresponding to the number of solution-evaluations conducted using the true objective function, the heuristic objective value, the gap between the best obtained solution and the optimum, and finally the objective value for the optimum.

As expected, capacity is distributed among the wards according to the arrival rate of the three patient types. That is, with respect to both private and shared double rooms. By contrast, the fraction of private patients arriving, ψ , appears to have little effect on the distribution of private rooms, since the solutions are similar across all runs. This is potentially a result of the relative difference between the arrival rates of each patient type, which shall be assessed in the last part of this section, where the arrival rate has been increased for ward 1.

In general, the experiments presented in Table 5 yield excellent results, as the relative gap between the heuristic and true optimum is consistently below 1%. The reader should notice that these results have been obtained after a runtime of 1 hour, whereas the complete enumeration of the search space to determine the true optimum finished only after approximately 3 weeks runtime.

ψ	Rep.	Heuristic									Optimal
		$u_{1,pr.}$	$u_{2,pr.}$	$u_{3,pr.}$	$u_{1,do.}$	$u_{2,do.}$	$u_{3,do.}$	Iter.	Obj. val.	Gap (%)	Obj. val.
0.2	1	13	11	12	8	6	5	28	12.68	0.55	12.75
0.2	2	15	10	11	8	5	6	20	12.71	0.31	12.75
0.5	1	15	11	10	8	5	6	23	30.09	0.63	30.28
0.5	2	15	10	11	7	6	6	29	30.18	0.33	30.28
0.7	1	15	10	11	7	6	6	20	35.21	0.06	35.23
0.7	2	14	10	12	8	5	6	27	35.20	0.09	35.23

Table 5: Result of optimizing the room configuration by applying our RIS heuristic. Variables $u_{i,private}$ and $u_{i,double}$ are abbreviated $u_{i,pr.}$ and $u_{i,do.}$ for all wards $i \in \mathcal{W}$. Two runs were conducted for each of the three levels of the private patient fraction, ψ .

Table 6 provides further results, showing the heuristic solutions for the *high arrival rate* and *high relocation* datasets. For the high arrival rate, the system is found to be more sensitive to changes in the private patient fraction since increasing the fraction results in more private rooms to be allocated to ward 1. Otherwise, the obtained solutions attain a small relative gap that consistently stays below 1%, showing that increasing the number of arriving patients does not affect the search procedure’s performance.

Next, for a substantially larger number of relocated patients it is expected that more iterations are required to adapt the interpolation and obtain useful solutions, due to the lower accuracy of the surrogate objective function. Recall how it was previously determined that a substantial distance is present between the true and surrogate optima for the *high relocation* dataset in Section 4.2. The experiments indicate that this is only the case for a medium private patient fraction of $\psi = 0.5$ since the relative gap has increased a few percentage points. The other levels remain relatively unchanged, and even slightly improved for $\psi = 0.2$. Thus, these experiments indicate that good solutions are derived for a large number of relocated patients as well.

4.5. Validation

Finally, this section validates the assumption that an improved room configuration leads to better operational efficiency for inpatient admissions. An initial, poor quality, room configuration (RC1) is compared with an optimized room configuration (RC2, determined as a result of the RIS heuristic) in a day-to-day scheduling simulation. The simulation begins from an initial, empty set of wards with given room configuration (either RC1 or RC2). As the simulation progresses, inpatients arrive according to the arrival rates determined by Table 1. At the start of each simulated day, patients are assigned to a ward and room by solving a simplified version of the *reactive* ILP model by Vancroonenburg et al. [10]. This ILP model is ‘reactive’

Dataset	ψ	Rep.	Heuristic									Optimal
			$u_{1,pr.}$	$u_{2,pr.}$	$u_{3,pr.}$	$u_{1,do.}$	$u_{2,do.}$	$u_{3,do.}$	Iter.	Obj. val.	Gap (%)	Obj. val.
High Arr.	0.2	1	9	5	6	7	2	1	53	7.57	0.66	7.62
High Arr.	0.2	2	9	5	6	7	2	1	64	7.57	0.66	7.62
High Arr.	0.5	1	9	2	9	5	1	4	74	17.35	0.57	17.45
High Arr.	0.5	2	11	2	7	5	1	4	78	17.36	0.52	17.45
High Arr.	0.7	1	14	1	5	6	1	3	70	19.76	0.05	19.77
High Arr.	0.7	2	12	0	8	5	1	4	83	19.76	0.05	19.77
High Rel.	0.2	1	5	4	11	3	0	7	44	7.70	0.39	7.73
High Rel.	0.2	2	5	3	12	1	1	8	18	7.72	0.13	7.73
High Rel.	0.5	1	2	8	10	1	4	5	3	17.30	2.32	17.71
High Rel.	0.5	2	8	1	1	4	1	5	34	17.45	1.47	17.71
High Rel.	0.7	1	6	0	14	3	1	6	11	19.82	0.15	19.85
High Rel.	0.7	2	1	2	17	2	2	6	2	19.76	0.45	19.85

Table 6: Results of applying our RIS heuristic to the *high arrival rate* and *high relocation* datasets. Variables $u_{i,private}$ and $u_{i,double}$ are abbreviated $u_{i,pr.}$ and $u_{i,do.}$ for all wards $i \in \mathcal{W}$. The room availability was limited to $N_{private} = 20$ and $N_{double} = 10$ rooms. Due to the limited search space, all runs were conducted without the relocation bound and a runtime of 10 minutes.

in the sense that it only considers arriving patients on the present (simulated) day. This contrasts with the ‘anticipative’ ILP model [10] which employs pre-registration information to look-ahead further in the future, to consider and pre-assign (known) future arriving patients. The latter model is superior in terms of finding high-quality assignments (regarding matching patient preferences, minimizing gender policy violations, etc.). However, no such pre-registration information is considered in the scope of the present paper.

The reactive patient assignment ILP model can be defined as follows. Please note that due to the extensive use of mathematical symbols in the paper, this section redefines variables and indices used in the previous sections. We refer to Table 10 in Appendix A for a complete overview of the notation relevant to this subsection. Let binary decision variable x_{pr} equal 1 if patient p (which either arrived on the current day, or is still present from previous admission) is admitted to room r . Let binary decision variable y_{rd} equal 1 if the room is occupied by male patients or 0 by female patients on day d . Let $r = \emptyset$ denote a *dummy* room, where $x_{p\emptyset} = 1$ indicates a patient being refused (or being relocated to a ward which is not considered in the current problem). R denotes the set of rooms available from the wards in \mathcal{W} , and b_r denotes the capacity of each room $r \in R$, i.e. in the current dataset $b_r \in \{1, 2\}$ (private or shared). P denotes the set of patients currently arriving for admission or still present after admission on an earlier day in the simulation. $ad(p)$ and $elos(p)$ denote the arrival day and *expected length of stay* (available from Table 1, $1/\mu$ parameter for each patient type) in days. D denotes an upper bound on the planning horizon (in days), which is restricted by

either the length of the simulation time horizon or the maximum remaining expected length of stay among patients $p \in P$. d' denotes the current day of the simulation, whereas d is used to index all remaining days of the planning horizon. Finally, c_{pr} denotes a cost matrix, attributing a perceived penalty of admitting a patient p to a room r for one day. The elements c_{pr} , defined for each room (including the dummy room) and patient combination, are given by the sum of:

- w_{pref} , a room preference penalty if the assigned room does not meet the patient's preference (i.e. a shared room when the patient prefers private),
- $(1 - p_{ij}) \cdot w_{\text{reloc}}$, a relocation penalty if the assigned room is not in the preferred ward (refer to Table 2 for values of p_{ij}),
- w_{\emptyset} , a refusal penalty if the patient is not admitted to any room from wards \mathcal{W} , and is left unassigned. In this case, the patient is assumed to be admitted elsewhere.

These weights are set as follows: $w_{\emptyset} = 10000$, $w_{\text{reloc}} = 1000$ and $w_{\text{pref}} = 10$.

Using this notation, the ILP model can be formulated as follows:

$$\text{Minimize } \sum_{p \in P} \sum_{r \in R} c_{pr} \cdot \text{elos}(p) \cdot x_{pr} \quad (25)$$

Subject to

$$\sum_{r \in R} x_{pr} + x_{p\emptyset} = 1 \quad \forall p \in P \quad (26)$$

$$\sum_{\substack{p \in P: \\ d < ad(p) + \text{elos}(p)}} x_{pr} \leq b_r \quad \forall r \in R, d = d', \dots, D \quad (27)$$

$$\sum_{\substack{p \in P: \\ d < ad(p) + \text{elos}(p) \\ p = \text{male}}} x_{pr} \leq b_r \cdot y_{rd} \quad \forall r \in R, d = d', \dots, D \quad (28)$$

$$\sum_{\substack{p \in P: \\ d < ad(p) + \text{elos}(p) \\ p = \text{female}}} x_{pr} \leq b_r \cdot (1 - y_{rd}) \quad \forall r \in R, d = d', \dots, D \quad (29)$$

$$x_{pr'} = 1 \quad \forall p \in P : ad(p) < d', r' = \text{AssignedRoom}() \quad (30)$$

$$x_{pr} \in \{0, 1\} \quad \forall p \in P, r \in R \cup \emptyset \quad (31)$$

$$y_{rd} \in \{0, 1\} \quad \forall r \in R, d = d', \dots, D \quad (32)$$

The model minimises the total cost penalties incurred by the assigned and admitted patients, weighted by their remaining length of stay (Equation (25)). Constraints (26) ensure that all patients are assigned to a room or to the dummy room. Constraints (27) ensure the capacity limit of each room for each day of

Dataset	Room config.	Ward 1		Ward 2		Ward 3		Obj. val
		# Private	# Shared	# Private	# Shared	# Private	# Shared	
Original	RC1	5	5	25	4	6	10	26.1
Original	RC2	15	7	10	3	11	6	35.2
High Arr.	RC1	5	5	25	4	6	10	26.3
High Arr.	RC2	18	8	9	6	9	5	35.5
High Rel.	RC1	5	5	25	4	6	10	26.0
High Rel.	RC2	14	8	11	6	11	5	35.2

Table 7: Room configurations for each dataset applied in the simulation.

the planning horizon. Constraints (28) and (29) ensure that either male or female patients are assigned to a room, i.e. it ensures a gender separation policy for shared rooms. Constraints (30) ensure that patients who were already admitted prior to the current simulation day d' stay assigned to their previously assigned room r' . Finally, (31) and (32) define the domains of the decision variables.

The simulation was run on the three datasets described in Section 4.1. Room configurations RC1 and RC2 depend on the dataset, and are constructed as described in Table 7. Furthermore, the fraction of male patients among all patients (other patients being female) is considered as an additional parameter (denoted by ϕ) since this fraction may be ward-dependent. The simulation runs over 120 simulation days during which arrivals are generated. For each combination of the considered parameters (fraction of male patients, dataset and room configuration), we ran 1000 simulation replications.

The results (averaged over 1000 replications) are summarized in Table 8, showing for each dataset and male patient-fraction, the difference between RC1 and RC2 (value larger than 0 if RC2 is better) in respectively room preference penalties, patient relocations to other wards, patient refusals, and the global objective value as defined by Equation (25) calculated over the simulation horizon of 120 simulation days. Finally, the difference is statistically significant by a p-value from the Wilcoxon Rank Sum Test [26, 27]. Notice that all p-values are indeed smaller than 0.05.

These results show that RC2, while sometimes having a *marginally* higher number of patient refusals, the penalties of mismatched room preferences (our primary concern) and patient relocations, on average, greatly improves over RC1. This validates the assumption that an improved room configuration (i.e. RC2 over RC1) leads to improved operational efficiency for inpatient admissions.

5. Conclusion

The ability to choose a private room over a shared room is becoming an increasingly important factor for patients to choose a hospital for admission. Being able to meet those requests is of strategic importance to hospitals, both in increasing patient comfort and satisfaction but also in generating extra revenue from

Dataset	ϕ	Avg. Impr. room pref. (# patient days)	Avg. Impr. reloc. (# patient days)	Avg. Impr. refusal (# patient days)	Avg. Obj. Impr. (absolute)	(%)	p-value
Original	0.3	75.08	1716.00	-0.62	1532696.68	26.03	<0.01
Original	0.4	78.93	1714.20	-0.25	1535081.748	26.12	<0.01
Original	0.45	78.94	1716.74	-1.09	1528571.86	26.04	<0.01
Original	0.5	81.15	1713.65	-0.59	1531002.102	26.07	<0.01
Original	0.55	80.63	1715.05	0.58	1543720.948	26.30	<0.01
Original	0.6	81.32	1715.66	-0.47	1533627.837	26.12	<0.01
Original	0.7	76.33	1713.56	0.24	1539232.789	26.16	<0.01
High Arr.	0.3	54.80	2328.58	0.58	2049016.12	22.16	<0.01
High Arr.	0.4	59.52	2325.88	0.55	2046362.37	22.20	<0.01
High Arr.	0.45	57.14	2328.65	-0.98	2033640.76	22.07	<0.01
High Arr.	0.5	56.83	2328.14	-0.22	2040755.74	22.15	<0.01
High Arr.	0.55	58.91	2330.96	-1.86	2026654.69	22.04	<0.01
High Arr.	0.6	57.95	2327.84	-0.31	2039361.02	22.15	<0.01
High Arr.	0.7	57.13	2327.06	-0.10	2041191.40	22.13	<0.01
High Rel.	0.3	70.40	1831.99	0.77	1084299.45	30.30	<0.01
High Rel.	0.4	73.47	1834.95	-0.26	1076126.23	30.19	<0.01
High Rel.	0.45	73.47	1832.93	0.49	1081874.07	30.37	<0.01
High Rel.	0.5	73.86	1834.16	0.43	1081416.84	30.32	<0.01
High Rel.	0.55	73.90	1832.84	0.97	1085754.38	30.50	<0.01
High Rel.	0.6	72.54	1833.36	0.04	1076919.65	30.32	<0.01
High Rel.	0.7	70.64	1831.81	0.90	1085614.61	30.40	<0.01

Table 8: Simulation results (averaged over 1000 replications) for each dataset and male patient fraction, showing the average improvement (Impr.) for room preference violations, relocated patients, refused patients (all three are denoted in # mismatched patient days) and the total and relative objective cost (Eq. 25 evaluated over the simulation horizon) improvement. P-values are all calculated to be below 0.01.

charging room/honorarium supplements. However, existing infrastructure may not be adequately allocated between nursing wards to meet the current demand.

This study sought to provide hospital decision makers with a strategic tool for improving the allocation of room types among hospital wards. More specifically, the aim was to accommodate patients who prefer private room assignments, by first assuming a fixed number of room types, and second that these room types can be reallocated among the wards.

The proposed approach is based on a continuous-time Markov chain model that derives the patient occupancy distributions, and a heuristic search procedure referred to as Randomized and Interpolated Search (RIS) that searches for the best possible room configuration. RIS recursively improves an initial solution by sampling from the search space based on a gradually improved interpolation of the objective function. The fact that occupancy distributions are fixed for an unchanged aggregated capacity is exploited in order to derive the sub-optimal room type configuration using integer linear programming. Consequently, aggregated capacity allocations form the primary decision variables for the proposed RIS heuristic to operate on. This results in reducing the search space for the RIS heuristic by omitting room type configuration decision variables.

Based on data from both a Danish and Belgian hospital, the applicability and effectiveness of the approach was demonstrated through various experiments which vary the fraction of patients who prefer private hospitalization, the overall arrival rate, and lastly the number of patients relocated within the system. In a computational study, it is shown that the RIS heuristic has the potential to derive near-optimal solutions that attain relative gaps below 1% within short runtimes which make the method applicable in practice. Moreover, it was demonstrated how configuring room resources on a strategic level benefits the day-to-day decisions of assigning patients to rooms through simulation.

Finally, the reader should notice that the proposed approach is not only applicable to the specific case of optimizing the room configurations in a hospital setup, but to any queueing problem where jobs are serviced among different nodes and may prefer a specific, but limited resource. Examples of such similar environments vary from manufacturing setups where products are processed at different stations and may require a specific tool, to call centers where customers may prefer an operator of a specific skill set.

The study comprised various experiments, including different parameter variations as well as input data-sets that demonstrate the performance and the robustness of the approach. However the analysis was restricted to a specific hospital case featuring a fairly limited number of wards. More complex cases should be assessed, preferably with a greater number of disposable rooms and room types. Additionally, future experiments should consider that patient scheduling is not only constrained by room preferences but also gender, and that this mix can be a function of the preferred ward for the patient.

2 Acknowledgments

3 Wim Vancroonenburg is a post-doctoral researcher funded by Research Foundation Flanders - FWO Vlaanderen.
 4 This research was further funded and supported by the Danish governmental organization, Region Sjælland.
 5 Especially, we thank the department of Production, Research and Innovation for providing us with insight
 6 into the operations of the Danish hospitals. Lastly, we thank Prof. Bo Friis Nielsen at the Technical Uni-
 7 versity of Denmark for providing us with advice on probability theory. Editorial consultation provided by
 8 Luke Connolly (KU Leuven).

9 References

- 10 [1] C. de la Maisonnette and J. Oliveira Martins. Public spending on health and long-term care. *OECD Publishing*, 2013.
 11 doi: <http://dx.doi.org/10.1787/5k44t7jwwr9x-en>.
- 12 [2] OECD. Competition in hospital services. *OECD Publishing*, 2012.
- 13 [3] Christelijke Mutualiteit. Twaalfde cm-ziekenhuisbarometer (*English: Twelfth CM-hospital barometer*). https://www.cm.be/media/CM-Ziekenhuisbarometer-2016_tcm47-19246.pdf. Accessed: 2018/03/23.
- 14 [4] Annemie Verhelst. Opnameplanning in ziekenhuizen : 10 jaar later (*English: Admission scheduling in hospitals: 10 years later*). Master's thesis, Universiteit Gent, 2009.
- 15 [5] P. Demeester, W. Souffriau, P. De Causmaecker, and G. Vanden Berghe. A hybrid tabu search algorithm for automatically
 16 assigning patients to beds. *Artificial Intelligence in Medicine*, 48(1):61–70, 2010.
- 17 [6] Burak Bilgin, Peter Demeester, Mustafa Misir, Wim Vancroonenburg, and Greet Vanden Berghe. One hyper-heuristic
 18 approach to two timetabling problems in health care. *Journal of Heuristics*, 18(3):401–434, June 2012. doi: 10.1007/
 19 s10732-011-9192-0.
- 20 [7] S. Ceschia and A. Schaerf. Local search and lower bounds for the patient admission scheduling problem. *Computers &*
 21 *Operations Research*, 38(10):1452–1463, October 2011.
- 22 [8] Troels Martin Range, Richard Martin Lusby, and Jesper Larsen. A column generation approach for solving the patient
 23 admission scheduling problem. *European Journal of Operational Research*, 235(1):252–264, May 2014. doi: 10.1016/j.ejor.
 24 2013.10.050.
- 25 [9] Sara Ceschia and Andrea Schaerf. Modeling and solving the dynamic patient admission scheduling problem under uncer-
 26 tainty. *Artificial Intelligence in Medicine*, 56(3):199–205, 2012. doi: 10.1016/j.artmed.2012.09.001.
- 27 [10] Wim Vancroonenburg, Patrick De Causmaecker, and Greet Vanden Berghe. A study of decision support models for online
 28 patient-to-room assignment planning. *Annals of Operations Research*, 2013. doi: 10.1007/s10479-013-1478-1. Available
 29 online.
- 30 [11] Sara Ceschia and Andrea Schaerf. Dynamic patient admission scheduling with operating room constraints, flexible horizons,
 31 and patient delays. *Journal of Scheduling*, 19(4):377–389, aug 2016. ISSN 1094-6136. doi: 10.1007/s10951-014-0407-8.
 32 URL <http://link.springer.com/10.1007/s10951-014-0407-8>.
- 33 [12] Wim Vancroonenburg, Federico Della Croce, Dries Goossens, and Frits C. R. Spieksma. The Red-Blue transportation
 34 problem. *European Journal of Operational Research*, 237(3):814–823, 2014. doi: 10.1016/j.ejor.2014.02.055.
- 35 [13] Rym Ben Bachouch, Alain Guinet, and Sonia Hajri-Gabouj. An integer linear model for hospital bed planning. *Interna-*
 36 *tional Journal of Production Economics*, 140(2):833–843, December 2012. doi: 10.1016/j.ijpe.2012.07.023.

- [14] Robert Schmidt, Sandra Geisler, and Cord Spreckelsen. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. *BMC medical informatics and decision making*, 13(1):3, January 2013. doi: 10.1186/1472-6947-13-3.
- [15] P. Bhattacharjee and P.K. Ray. Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. *Computers and Industrial Engineering*, 78:299–312, 2014. doi: 10.1016/j.cie.2014.04.016.
- [16] M.E. Lim, T. Nye, J.M. Bowen, J. Hurley, R. Goeree, and J.-E. Tarride. Mathematical modeling: The case of emergency department waiting times. *International Journal of Technology Assessment in Health Care*, 28(2):93–109, 2012. doi: 10.1017/S0266462312000013.
- [17] N. Bartolomeo, P. Trerotoli, A. Moretti, and G. Serio. A Markov model to evaluate hospital readmission. *BMC Medical Research Methodology*, 8, 2008. doi: 10.1186/1471-2288-8-23.
- [18] J.R. Broyles, J.K. Cochran, and D.C. Montgomery. A statistical Markov chain approximation of transient hospital inpatient inventory. *European Journal of Operational Research*, 207(3):1645–1657, 2010. doi: 10.1016/j.ejor.2010.06.021.
- [19] L. He, Y. Li, and S.H. Chung. Markov chain based modeling and analysis of colonoscopy screening processes. *67th Annual Conference and Expo of the Institute of Industrial Engineers 2017*, pages 740–745, 2017.
- [20] X. Shao, J. Li, and D.A. Wiegmann. A Markov chain approach to study flow disruptions on surgery in emergency care. pages 990–995, 2013. doi: 10.1109/CoASE.2013.6654059.
- [21] J. Wang, X. Zhong, J. Li, and P.K. Howard. Modeling and analysis of care delivery services within patient rooms: A system-theoretic approach. *IEEE Transactions on Automation Science and Engineering*, 11(2):379–393, 2014. doi: 10.1109/TASE.2013.2242326.
- [22] B. Shaw and A.H. Marshall. Modelling the flow of congestive heart failure patients through a hospital system. *Journal of the Operational Research Society*, 58(2):212–218, 2007. doi: 10.1057/palgrave.jors.2602240.
- [23] A.R. Andersen, B.F. Nielsen, and L.B. Reinhardt. Optimization of hospital ward resources with patient relocation using Markov chain modeling. *European Journal of Operational Research*, 260(3):1152–1163, 2017. doi: 10.1016/j.ejor.2017.01.026.
- [24] William J. Stewart. *Probability, Markov Chains, Queues, and Simulation - The Mathematical Basis of Performance Modeling*. Princeton University Press, 1st ed. edition, 2009. ISBN 978-0-691-14062-9.
- [25] Robin Sibson. A brief description of natural neighbor interpolation. In Vic Barnett, editor, *Interpreting Multivariate Data*, pages 21–36. John Wiley and Sons, 1981. ISBN 0471280399.
- [26] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80, 1945. ISSN 23261439, 00994987. doi: 10.2307/3001968.
- [27] M. J. Crawley. *Statistics – An Introduction using R*. Wiley, 2 edition, 2015. ISBN 9781118941096.

A. Symbols & Definitions

Symbol	Definition
\mathcal{W}	Set of all wards.
$\mathcal{R}, \mathcal{R}'$	Set of all room types (\mathcal{R}) and shared room types (\mathcal{R}').
$\mathbf{u}, \mathbf{u}', \tilde{\mathbf{u}}^g, \tilde{\mathbf{u}}^f, u_{ir}$	Room configuration (\mathbf{u}), as a new iteration (\mathbf{u}'), surrogate optima ($\tilde{\mathbf{u}}^g$ and $\tilde{\mathbf{u}}^f$), and as a single element (u_{ir}).
b_r	Bed capacity of room type r .
N_r	Total number of room type r that are available to the hospital.

$M_i, \tilde{M}_i, M_i^*, \tilde{M}_i^*$	Assigned bed capacity of ward i (M_i), as input to a surrogate function (\tilde{M}_i), optimal solution (M_i^*), and optimal solution to the surrogate objective function (\tilde{M}_i^*).
$f(\mathbf{u}), \tilde{f}(\mathbf{u})$	True ($f(\mathbf{u})$) and surrogate ($\tilde{f}(\mathbf{u})$) expected number of patients relocated per day.
$g(\mathbf{u}), \tilde{g}(\mathbf{u})$	True ($g(\mathbf{u})$) and surrogate ($\tilde{g}(\mathbf{u})$) expected number of patients who prefer and are assigned to private rooms.
τ	Upper bound for function $f(\mathbf{u})$.
U	Solution space of model (3)-(7).
$Y_f(\mathbf{u})$	Estimate of $f(\mathbf{u})$ based on interpolation.
$Y_g(\mathbf{u}), Y'_g(\mathbf{u})$	Estimate of obj. function $g(\mathbf{u})$ ($Y_g(\mathbf{u})$) based on interpolation, and the conversion $Y'_g(\mathbf{u}) = Y_g(\mathbf{u})^\beta$.
$\mathbf{x}, \tilde{\mathbf{x}}$	Set of all known true (\mathbf{x}) and surrogate ($\tilde{\mathbf{x}}$) solutions.
$Z(\mathbf{u})$	Probability mass function used in sampling from the solution space.
β	Exponent used in performing the conversion $Y'_g(\mathbf{u}) = Y_g(\mathbf{u})^\beta$.
ξ	Proximity tolerance used in removing surrogate solutions from $\tilde{\mathbf{x}}$.
S, s, s^*	State space (S), current state (s) and new state (s^*) of the CTMC.
w_{ij}	Number of type i patients hospitalized in ward j .
f'_j	Number of free beds in ward j .
λ_i, μ_i	Arrival and service rate for patients of type i .
$p(f'_1, f'_2, \dots, f'_{ \mathcal{W} })_{ij}$	Fraction of type i patients routed to ward j as function of the number of free beds.
Q, q_{ss^*}	Transition rate matrix (Q) and single transition rates (q_{ss^*}).
$\boldsymbol{\pi}, \pi_i(n), \tilde{\pi}_i(n), \pi_i^B, \tilde{\pi}_i^B$	State distribution of the CTMC ($\boldsymbol{\pi}$), ward marginal probability ($\pi_i(n)$), ward surrogate marginal probability ($\tilde{\pi}_i(n)$), ward blocking probability (π_i^B), and ward surrogate blocking probability ($\tilde{\pi}_i^B$).
n	Total number of patients in an arbitrary ward.
ψ	Fraction of all arriving patients who prefer a private room.
y^a, y^b	Number of hospitalized patients preferring a private room (y^a), and number of patients who do not care whether their room is shared or private (y^b).
$P_i(y^a, y^b)$	Joint probability of y^a and y^b for ward i .
$B(y^a, n, \psi)$	Probability mass function of the binomial distribution.
$\rho_i(y^a), \tilde{\rho}_i(y^a)$	Probability that y^a beds are occupied by patients who prefer a private room ($\rho_i(y^a)$), and the associated surrogate probability ($\tilde{\rho}_i(y^a)$).
z_{ij}	Binary variable equalling 1 whenever ward i is assigned j private beds.
ω_{ij}	Expected number of patients who prefer and are assigned to private rooms in ward i with j available private rooms.

\mathcal{J}_i	Set of all combinations of private beds that can be assigned to ward i .
y_{ir}	Number of shared rooms of type r assigned to ward i .

Table 9: Mathematical symbols and their associated definitions.

Symbol	Definition
x_{pr}	Binary decision variable equal to 1 when patient p is assigned to room r .
y_{rd}	Binary aid variable equal to 1 if room r is occupied by male patients on day d of the planning horizon.
d'	Current simulation day.
p	Patient index.
r	Room index.
\emptyset	Dummy room index.
R	Set of rooms available in wards \mathcal{W} .
b_r	Bed capacity of room r .
$ad(p)$	Arrival day of patient p .
$elos(p)$	Expected length of stay (in days) of patient p .
D	Upper bound on the planning horizon (in days), restricted by the length of the simulation horizon and the maximum remaining length of stay among patients currently admitted to the hospital.
c_{pr}	Cost-matrix values, attributing a perceived penalty associated with assigning patient p to room r .
w_{pref}	Room preference penalty when the assigned room type does not meet the patient's preference.
w_{reloc}	Relocation penalty if the assigned room is not in the preferred ward.
w_{\emptyset}	Refusal penalty if a patient is not admitted.

Table 10: Mathematical symbols for the simulation validation ILP model.