OR Spectrum
https://doi.org/10.1007/s00291-024-00800-z

ORIGINAL ARTICLE

Integrated patient-to-room and nurse-to-patient assignment in hospital wards

Tabea Brandt[1] · **Tom Lorenz Klein**[2] · **Melanie Reuter-Oppermann**[4] · **Fabian Schäfer**[5] · **Clemens Thielen**[2,6] · **Maartje van de Vrugt**[3,7] · **Joe Viana**[8,9,10]

Received: 6 September 2023 / Accepted: 5 November 2024
© The Author(s) 2025

Abstract

Assigning patients to rooms and nurses to patients are critical tasks within hospitals that directly affect patient and staff satisfaction, quality of care, and hospital efficiency. Both patient-to-room assignments and nurse-to-patient assignments are typically agreed upon at the ward level, and they interact in several ways, such as jointly determining the walking distances nurses cover between different patient rooms. This provides the motivation to consider both problems jointly in an integrated fashion. This paper presents the first optimization models and algorithms for the integrated patient-to-room and nurse-to-patient assignment problem. We provide a mixed integer programming formulation of the integrated problem that considers the typical objectives from the single problems and additional objectives that can only be properly evaluated when integrating both problems. Moreover, motivated by the inherent complexity that results from integrating these two NP-hard and already computationally challenging problems, we devise an efficient heuristic for the integrated patient-to-room and nurse-to-patient assignment problem. We conduct extensive computational experiments on both artificial and real-world instances to evaluate the runtime and quality of the solution obtained with the heuristic. The artificial instances are generated by a parameterized instance generator for the integrated problem that is made freely available.

Keywords Integrated planning · Hospital · Patient-to-room assignment · Nurse-to-patient assignment · Heuristic

1 Introduction

For many years, an ever-rising demand for healthcare and increasing healthcare expenditures challenge hospitals to increase the efficiency of their operations (Drupsteen et al 2013). This results in a need for advanced managerial planning approaches

Extended author information available on the last page of the article

Published online: 23 January 2025

that help to utilize the available scarce resources as efficiently as possible. Consequently, a wide range of methods and approaches have been developed in the Operations Research (OR) literature that aims at improving resource utilization through efficient planning (Rais and Viana 2011; Hulshof et al 2012; Jha et al 2016). In particular, quantitative decision support has been proposed for important resources such as operating rooms (Van Riet and Demeulemeester 2015; Guerriero and Guido 2011), intensive care units (Bai et al 2018), inpatient beds (He et al 2019), physicians (Erhard et al 2018), and nurses (Benazzouz et al 2015; Clark et al 2015).

While efficient planning of single resources can already lead to improved resource utilization and large efficiency gains, it ignores the inherent complex interactions between different resources (Hulshof et al 2012) and, as a consequence, might lead to suboptimal decisions on a system level. Therefore, a need for OR models and methods for *integrated planning* of several resources has been identified (Hulshof et al 2012; Jun et al 1999; Vanberkel et al 2010). This need is particularly apparent in hospitals, where different resources are typically required and used for treating patients. A recent literature review on integrated planning of multiple resources in hospitals is provided in Rachuba et al (2023).

Rooms and beds are critical assets of hospitals since they account for a considerable part of a hospital's infrastructure and large financial investments are necessary for equipping them with medical devices that facilitate patient care (Vancroonenburg et al 2016). Additionally, a shift in demographics, the growing number of patient admissions, and rising inpatient units costs lead to high overall bed occupancy levels and require an increased focus on efficient bed management (Schäfer et al 2019; He et al 2019). On the operational level, an important planning problem typically referred to as *patient-to-room assignment (PRA)* consists of assigning patients to suitable rooms such that a variety of constraints concerning, for instance, the patient's medical needs (e.g., required medical equipment) and preferences (e.g., concerning age and gender of roommates) are satisfied, while available room capacities are respected and transfers of patients between rooms are avoided (Demeester et al 2010; Ceschia and Schaerf 2011, 2012; Schäfer et al 2019). Different variants of the PRA problem have been studied extensively in the literature both in the static setting, when all information about patients and their admission and discharge times is known in advance, and in dynamic settings, when new patients may arrive unexpectedly – see Sect. 2.1 for a detailed overview.

Medical staff also represent a particularly critical resource in hospitals since medical staff are (1) involved in most patient-related activities in a hospital, and (2) are a particularly scarce resource due to a general shortage of nurses (Aiken et al 2002) and physicians (Bodenheimer and Smith 2013; Erhard et al 2018; Thielen 2018). This has led to increasing workloads for the staff over the last decades and makes effective planning of medical staff a central concern for hospitals (Benazzouz et al 2015). In particular, aspects such as a fair distribution of workloads among staff members have a large impact on employee satisfaction and the efficient operation of a hospital. Concerning the nursing staff, distributing work fairly among nurses is considered essential for optimal quality of care (Mullinax and Lawley 2002). Here, the workload of each single nurse is mostly determined by the patients the nurse is assigned to and their care requirements. Consequently, determining a suitable *nurse-to-patient assignment (NPA)* that balances the workloads of the nurses represents an important operational problem that has received considerable attention in the literature – see Sect. 2.2 for an extensive literature review.

With a few exceptions detailed in Sect. 2.3, the PRA problem and the NPA problem have mostly been considered separately in the literature – although there are important interactions between them. For instance, studies show that the walking distances that result from traveling between their assigned patients' rooms and other locations such as the nearest nursing station have a substantial impact on a nurse's workload during a shift (Acar and Butt 2016; Butt et al 2004). These walking distances, however, can only be determined and optimized when considering PRAs and NPAs jointly. Moreover, it has been observed that assigning the minimum possible number of nurses to patients in the same room can help to minimize negative effects such as the spread of infections between rooms by nurses or the disturbance of patients by other nurses entering their room (Halwani et al 2006; Cohen et al 2012; Eveillard et al 2009; Dancer 2009). This provides strong motivation for integrating the PRA problem and the NPA problem by considering them jointly in one optimization model.

While several publications motivate and discuss the integration of patient-to-room assignment and nurse assignment or staffing decisions (see Sect. 2.3), this paper explicitly considers decisions on PRAs and NPAs in one integrated optimization problem for the first time. Besides the objectives classically considered in the two separate problems, this integrated problem also allows the evaluation of additional objectives that rely on the interaction of PRAs and NPAs. Based on existing studies on nurse workloads (Acar and Butt 2016; Butt et al 2004), these objectives include the nurses' walking distances between assigned patients' rooms and additional relevant rooms such as the nearest nursing station. Moreover, also motivated by findings from the literature (Halwani et al 2006; Cohen et al 2012; Eveillard et al 2009), assigning the minimum possible number of nurses to patients in the same room is considered as an objective in order to mitigate negative effects such as the spread of infections between rooms by nurses or the disturbance of patients by other nurses entering their room.

We provide a detailed model of the integrated PRA and NPA problem as a mixed integer program (MIP). Due to the computationally challenging nature of the integrated problem, however, this MIP can only be used as a baseline comparison for small instances, while instances of a realistic size require other solution methods in order to achieve reasonable solution times. Therefore we also provide an efficient heuristic for the problem. The heuristic extends the heuristic for the PRA problem presented in Schäfer et al (2019) to the integrated problem and additionally employs a new heterogeneity check between patient admission and discharge times for the room assignment part. Both the MIP and the heuristic are evaluated in extensive experimental results on real-world instances obtained from a ward of our partner hospital Amsterdam University Medical Center (Amsterdam, The Netherlands) as well as artificial instances. The artificial instances are generated by a parameterized instance generator for the integrated problem that is made freely available. While the MIP only addresses the static version of the integrated problem in which all information about patients is known in advance, the heuristic can be easily adapted to dynamic settings where new patients arrive after some assignments have already been fixed.

The remainder of the paper is structured as follows. Section 2 summarizes the existing work on the PRA and the NPA problems and outlines our contribution in relation to the existing literature. Section 3 introduces the integrated PRA and NPA problem, while Sect. 4 presents our MIP formulation of the problem. Section 5 presents a sequential solution approach based on a natural decomposition of the MIP formulation as well as our heuristic for the problem. Section 6 then describes both the developed instance generator for generating artificial test instances and the real-world instances obtained from a ward of our partner hospital. Our experimental results obtained on both instance types are presented in Sect. 7. The paper concludes in Sect. 8 with a summary and an outlook on future research.

2 Related literature and contribution

In this section, we first summarize the state of the art concerning PRA and NPA separately before discussing existing work on the integration of the two problems and outlining our contribution.

2.1 Patient-to-room assignment

The static version of the PRA problem has been formally introduced by Demeester et al (2010) under the name of patient admission scheduling problem.[1] In this version, all information about patients is known in advance, and the task is to assign patients to suitable rooms such that room capacity and gender policy are respected while minimizing both patient transfers and penalty costs for undesirable PRAs. One important characteristic in this definition is that not every patient can be assigned to every room and patients may also have preferences towards specific rooms based on, e.g., available equipment or the number of beds.

For the static PRA problem, mostly heuristic solution approaches are proposed in the literature, e.g., a tabu search algorithm (Demeester et al 2010), a local search algorithm (Ceschia and Schaerf 2011), a destroy and repair matheuristic (Guido et al 2018), and algorithms based on the Hungarian algorithm (Borchani et al 2021), column generation (Range et al 2014), or MIP (Thuran and Bilgen 2017). Currently the best solutions for the benchmark instances provided by Demeester et al (2010) are found by the heuristic of Guido et al (2018) and by the exact, MIP-based solution approach proposed by Bastos et al (2019). However, the exact approach uses considerably more computation time.

The complexity of the static PRA problem is studied by Vancroonenburg et al (2014) using its correspondence to the red-blue-transportation problem. They show that the PRA problem is NP-hard in general and even if all rooms have a capacity

[1] We refer to the problem exclusively as the PRA problem in the following since this is the most common term used in the recent literature. Moreover, the original term *patient admission scheduling problem* is also used with a different meaning in the literature.

of three. This result is strengthened by Ficker et al (2021), who prove that the PRA problem is also NP-hard if all rooms have capacity two.

Ceschia and Schaerf (2011, 2012) propose a dynamic version of the PRA problem that includes the handling of emergency patients and uncertainty in the patients' length of stay (LOS). They propose a metaheuristic based on simulated annealing and an instance generator as well as a set of benchmark instances. Vancroonenburg et al (2016) studied a similar problem version using two online integer linear programming (ILP) models. Lusby et al (2016) propose a large neighborhood search heuristic for the dynamic PRA problem as proposed by Ceschia and Schaerf.

A different approach for incorporating emergency patients is taken by Schäfer et al (2017, 2019, 2023), who use a rolling horizon approach, i.e., recomputation of the solution whenever a new event occurs, using a fast heuristic. In their problem definition, they consider objectives for three stakeholders (patients, nurses, and physicians) and they are also the first to consider interdependencies between patients in the same room.

For a more detailed overview of the different versions and solution approaches for the PRA problem, we refer to Zhu et al (2019), who study the compatibility of short-term and long-term objectives in the context of dynamic PRA.

2.2 Nurse-to-patient assignment

The NPA problem is also considered by many different authors, where the most common objective is balancing the workloads of the nurses. For instance, Mullinax and Lawley (2002) use this objective in the daily assignment of newborn infants to nurses in an intensive care nursery that is divided into several zones (rooms). Here, each infant might yield a different workload depending on their acuity and each nurse can be assigned a certain maximum number of infants to take care of at the beginning of the shift, but all of these need to be from the same zone. Since they find the problem to be too hard to solve using an integer program, they present a two-step heuristic approach that exploits the subdivision into zones by first computing the number of nurses allocated to each zone before assigning patients to nurses in each zone independently.

The problem introduced by Mullinax and Lawley (2002) is also considered in Sir et al (2015); Ku et al (2014); Schaus and Régin (2014); Marzouk and Kamoun (2020) – each time with a slightly different objective function. Based on the formulation proposed in Mullinax and Lawley (2002), Sir et al (2015) formulate four MIP models for NPA that model the workloads of nurses by either the patient acuity indicators from a patient classification system (PCS), survey-based nurse-specific workload scores, or a combination of the two. Ku et al (2014) focus on minimizing the variance of the nurses' workloads using mixed integer quadratic programming (MIQP) and constraint programming (CP), while Schaus and Régin (2014) minimize the variance using a two-step decomposition approach that first computes the number of nurses allocated to each zone (which is done optimally by solving a resource allocation problem) before assigning patients to nurses in each zone independently using CP. Finally, Marzouk and Kamoun (2020) formulate a binary integer program that assigns nurses to zones and individual patients with the objective of minimizing the total number of nurses used in a shift.

Other work on NPA includes Punnakitikashem et al (2008), who present a stochastic integer programming model with the objective of minimizing excess workload for nurses and compare their approach to several other assignment policies (random assignment without considering workload, a simple heuristic, and solving the mean value problem using a deterministic integer program). Sundaramoorthi et al (2009) then use three of the assignment policies from Punnakitikashem et al (2008) as well as a clustered assignment policy to test their developed simulation model for evaluating NPAs.

While most of the literature on NPA mentioned above uses patient acuity as the main factor influencing nursing workloads, Acar and Butt (2016) perform a detailed study to identify the activities that comprise a nurse's workload. They find that nurses spend a substantial part of their time traveling (walking) between locations, where travel between patient rooms and the nursing station is the most common type of travel. Here, according to Butt et al (2004), the distance traveled by nurses is correlated to their assigned patient load and location, and key distances influencing the total travel distance of a nurse are the distances between assigned patients' rooms and (1) the nearest nursing station, (2) the nearest supply room, and (3) other assigned patients' rooms. However, nurses' walking distances have not yet been considered explicitly as an objective in the NPA literature since their minimization requires the simultaneous optimization of PRAs. This also holds for the objective of assigning the minimum possible number of nurses to patients in the same room, although it is known that assigning all patients in the same room to the same nurse or a small pool of nurses helps to avoid the transfer of hospital-acquired infections (Halwani et al 2006; Cohen et al 2012; Eveillard et al 2009), in particular *Methicillin-resistant Staphylococcus aureus* (MRSA) (Dancer 2009).

2.3 Integration of PRA and NPA

While we are not aware of any papers that explicitly integrate PRA and NPA decisions in one optimization model, some literature considering the interplay between the two problems or related problems exist. For instance, Thomas et al (2013) present a mixed-integer goal programming model for the PRA problem that takes nurses into account via required nurse-to-patient ratio constraints in each unit of a hospital. Bilgin et al (2012) develop a general, high-level hyper-heuristic approach that can be used for both the PRA problem and the nurse rostering problem. Pesant (2016) addresses the integration of the nurse staffing problem (assigning an appropriate number of nurses to each unit within a ward given a nurse roster) and the NPA problem in a neonatal intensive care unit by solving CP models for the two problems consecutively, and Punnakitikashem et al (2013) extend the stochastic programming model from Punnakitikashem et al (2008) by integrating nurse staffing decisions over multiple units into the NPA problem. Moreover, several recent papers consider patient appointment planning in outpatient chemotherapy clinics while simultaneously assigning nurses to patients or taking constraints on nurse availability into account (Liang and Turkcan 2016; Hesaraki et al 2019, 2020; Bouras et al 2021; Heshmat et al 2018). Schmidt et al (2013) integrate patient appointment planning into a patient-to-ward assignment problem, where they assume that rooms in the same ward are equal. They propose a binary integer program for this problem and compare exact and heuristic solution approaches. Ceschia and Schaerf (2016) consider an integrated planning of PRA with operating room constraints where they allow the postponement of patient appointments, for which they also provide an instance generator and a set of benchmark instances.

2.4 Our contribution

We introduce the first optimization models and algorithms for the integrated patient-to-room-and-nurse assignment (IPRNPA) problem. While several related problems have been considered in the literature as outlined in the previous section, our approach explicitly integrates decisions on patient-to-room assignments and nurse-to-patient assignments for the first time. This joint consideration of the two assignment problems permits optimization of additional objectives that rely on the interaction of the two assignment problems such as minimizing the nurses' walking distance or assigning the minimum possible number of nurses to patients in the same room to mitigate negative effects including inter-room infection spread associated with nurses and patient disturbance.

We first formalize the integrated problem by providing a detailed MIP formulation, which also serves as a baseline comparison for small problem instances. Due to the computational challenging nature of the problem, however, instances of a realistic size cannot be solved by the MIP. Therefore, we also provide an efficient heuristic for the IPRNPA problem that extends the heuristic for the PRA problem presented in Schäfer et al (2019) to the integrated problem and additionally employs a new heterogeneity check between patient admission and discharge times for the room assignment part. The runtimes and solution quality of the heuristic are demonstrated on both artificial and real-world instances. The artificial instances are generated by a parametrized instance generator for the integrated problem, which is made freely available together with a corresponding solution checker to foster future research on the integrated problem.

3 Problem definition

In this section, we formally introduce the *integrated patient-to-room and nurse-to-patient assignment (IPRNPA)* problem and the sets and parameters that are used to represent the input of the problem.

The IPRNPA problem integrates PRA and NPA on the ward level. Thus, the problem consists of assigning patients to rooms and nurses to patients on a hospital ward over a given planning period (typically one or several weeks). This section describes the static version of the problem where, similar to the static version of the PRA problem (Demeester et al 2010), all information about the patients (in particular, each patient's admission and discharge times) are known at the beginning of the planning period. The dynamic version of the problem differs from the static version in that information about new patients only becomes known either when they are admitted or a fixed time span before admission, which is analogous to the existing literature on dynamic PRA (Ceschia and Schaerf 2011, 2012).

We are given a set $\mathcal{P}$ of patients, a set $\mathcal{N}$ of nurses, and a set $\mathcal{R}$ of rooms. Moreover, there exists a (usually small) set $\mathcal{A}$ of additional rooms (such as the nursing station) with $\mathcal{R} \cap \mathcal{A} = \emptyset$. These additional rooms cannot be used for assigning patients and will only be relevant when computing the nurses' walking distances.

The considered planning period consists of a set $\mathcal{S} = \{1, \dots, S\}$ of shifts, which is partitioned into the subsets $\mathcal{S}^{\text{early}}$ of early shifts, $\mathcal{S}^{\text{late}}$ of late shifts, and $\mathcal{S}^{\text{night}}$ of night shifts. The shifts are numbered chronologically starting with an early shift and ending with a night shift. Hence, the first early, late, and night shift are numbered 1, 2, and 3, respectively, and belong to the first day, whereas the second early, late, and night shift belong to the second day and so on.

A feasible PRA demands that each patient $p \in \mathcal{P}$ is assigned to exactly one room $r \in \mathcal{R}$ during each shift between their admission shift $\text{ad\_shift}(p) \in \mathcal{S}^{\text{early}}$ and their discharge shift $\text{di\_shift}(p) \in \mathcal{S}^{\text{night}}$, which denotes the first and the last shift, respectively, of the patient's stay on the ward. In particular, this means that patients are always admitted and discharged in the morning between a night shift and the following early shift, as is the case in many real hospital wards. If patient p has already been on the ward during the last shift of the previous planning period, the value $\text{ad\_shift}(p)$ is set to $0 \notin \mathcal{S}$, and if patient p will still be on the ward after the last shift S of the current planning period, the value $\text{di\_shift}(p)$ is set to $S + 1 \notin \mathcal{S}$. Patient transfers between rooms are possible and are assumed to take place at most once a day for each patient between a night shift and an early shift. Transfers are, however, undesirable for both patients and nurses, and should, thus, be minimized. If patient p has already been on the ward during the last shift of the previous planning period, the room $y^{\text{prev}}(p) \in \mathcal{R}$ that the patient has been assigned to during this shift is also given as an input. This allows the evaluation of transfers that happen between the last (night) shift of the previous planning period and the first (early) shift of the current planning period.

Requirements concerning the PRA include respecting the capacity of each room $r \in \mathcal{R}$, which is given as a shift-independent number of available beds denoted by $\text{num\_beds}(r)$ (usually between 1 and 4) that defines the maximum number of patients that can be assigned to room r during any single shift. Moreover, depending on their specific condition, a patient might benefit from certain types of equipment in their room during certain shifts, so, during these shifts, they should be assigned to a room that features this type of equipment if possible. The set of possible equipment types is denoted by $\mathcal{E}$. The types of equipment that are present in room r are represented by the subset $\mathcal{E}(r) \subseteq \mathcal{E}$, and the shift-specific types of desired equipment of patient p during shift s are represented by the subset $\mathcal{E}(p, s) \subseteq \mathcal{E}$. Additionally, gender-mixed rooms should be avoided if possible. To this end, the set of patients is partitioned into the subsets $\mathcal{F}$ of female patients and $\mathcal{M}$ of male patients (i.e., $\mathcal{P} = \mathcal{F} \dot{\cup} \mathcal{M}$) and the number of gender-mixed rooms should be minimized across all shifts. Finally, each patient $p \in \mathcal{P}$ has an associated age group computed as

$\text{age\_group}(p) = \lfloor \frac{\text{age}(p)}{10} \rfloor$ with $\text{age}(p)$ denoting the age of the patient in years. Age groups are relevant since large age differences between patients who are simultaneously assigned to the same room are known to result in inconvenience for the patients and should, thus, be avoided.

Concerning the NPA part of the problem, the nurse roster for the planning period is given as an input. Here, for each nurse $n \in \mathcal{N}$, we are given the subset $\mathcal{S}(n) \subseteq \mathcal{S}$ of shifts that the nurse is assigned to. A feasible NPA should assign each patient $p \in \mathcal{P}$ to exactly one nurse $n \in \mathcal{N}$ with $s \in \mathcal{S}(n)$ during each shift $s \in \mathcal{S}$ between the patient's admission shift and discharge shift. Since nurses work at most one shift per day, any patient staying on the ward for at least two shifts must necessarily be assigned to different nurses during different shifts. To improve continuity of care, however, the number of different nurses who treat a single patient should be minimized.

Further requirements on the NPA include respecting nurse skill level requirements of the patients. Each nurse $n \in \mathcal{N}$ has a skill level $\text{skill\_level}(n)$ and each patient $p \in \mathcal{P}$ requires a certain minimum skill level $\text{skill\_req}(p, s)$ during each shift $s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}$ between their admission shift and their discharge shift. The set of possible skill levels is denoted by $\mathcal{L} = \{0, 1, 2\}$, where $0 =$ trainee, $1 =$ regular, and $2 =$ experienced. While an experienced nurse (skill level 2) can take care of any patient, the assignment should ensure that regular nurses (skill level 1) and trainees (skill level 0) are only assigned patients whose required skill level during a specific shift does not exceed the nurse's skill level.

Moreover, patients can induce different, shift-dependent workloads for nurses, and these workloads should be distributed fairly among the nurses - both during each single shift and overall. The workload resulting from taking care of patient $p \in \mathcal{P}$ during shift $s \in \mathcal{S}$ is expressed by a nonnegative number $\text{w\_load}(p, s)$ and depends on the age group of the patient, on their specific condition, on the time since admission, and on whether the shift is a day shift or a night shift. A fair distribution of workload among the nurses is then achieved by defining a maximum desired workload $\text{max\_load}(n, s)$ for each nurse $n \in \mathcal{N}$ during each shift $s \in \mathcal{S}(n)$ and ensuring that this maximum workload is not exceeded if possible whilst ensuring that the assigned workloads relative to the respective maxima do not differ considerably between nurses during single shifts as well as overall.

Important considerations that are influenced by both the PRA and the NPA involve trying to assign all patients in a room to the same nurse during each shift (which minimizes the spread of infections across rooms by nurses and reduces patient disturbance due to other nurses entering their room) and minimizing the nurses' walking distances between different rooms. For the evaluation of the walking distances, we are given a nonnegative number $\text{dist}(r, r')$ for each two rooms $r, r' \in \mathcal{R}$ that specifies the walking distance between rooms r and r'. Similarly, for each additional room $a \in \mathcal{A}$ and each regular room $r \in \mathcal{R}$, there is a nonnegative number $\text{dist}(a, r)$ that specifies the walking distance between additional room a and room r. The nurses' walking distances during a shift $s \in \mathcal{S}$ then depend on the patients they are assigned to during shift s and on how frequently nurses are expected to walk in a *circular pattern* directly from patient to patient during shift s

(e.g., during early shift rounds) and on how frequently they are expected to walk in a *star-like pattern* directly from additional rooms (such as the nursing station) to patients and back (e.g., when a patient calls for a nurse). These expected (absolute) frequencies are represented by two nonnegative parameters $\text{walk\_pat}^{\circ}(s)$ (*circular*) and $\text{walk\_pat}^{\star}(s)$ (*star-like*), respectively. Here, the inclusion of additional rooms, such as storage rooms, in the star-like pattern signifies scenarios where nurses may need to access these rooms from patient rooms and return, which models tasks like fetching supplies.

4 Mixed integer programming formulation

In order to model the IPRNPA problem introduced in Sect. 3 mathematically, we now present a formulation of the problem as an MIP.

The following lists summarize the sets and parameters introduced in the previous section and the decision variables used in the MIP:

Sets:

$\mathcal{P}$	Set of patients (index p)
$\mathcal{F}$	Subset of female patients
$\mathcal{M}$	Subset of male patients
$\mathcal{N}$	Set of nurses (index n)
$\mathcal{N}^{\text{prev}}(p)$	Subset of nurses that patient $p \in \mathcal{P}$ has already been assigned to during at least one shift in a previous planning period
$\mathcal{R}$	Set of rooms (index r)
$\mathcal{A}$	Set of additional rooms (e.g., the nursing station) (index a)
$\mathcal{S} = \{1, \dots, S\}$	Set of shifts (index s)
$\mathcal{S}^{\text{early}}$	Subset of early shifts
$\mathcal{S}^{\text{late}}$	Subset of late shifts
$\mathcal{S}^{\text{night}}$	Subset of night shifts
$\mathcal{S}(n)$	Subset of shifts that nurse $n \in \mathcal{N}$ is assigned to

$\mathcal{E}$	Set of possible equipment types in the rooms (index e)
$\mathcal{E}(r)$	Subset of equipment types present in room $r \in \mathcal{R}$
$\mathcal{E}(p, s)$	Subset of desired equipment types of patient $p \in \mathcal{P}$ during early shift $s \in \mathcal{S}^{\text{early}}$
$\mathcal{L} = \{0, 1, 2\}$	Set of possible skill levels of nurses (index l), where 0 = trainee, 1 = regular, 2 = experienced

Parameters:

ad_shift (p)	Shift $s \in \mathcal{S}^{\text{early}}$ during which patient $p \in \mathcal{P}$ is admitted (first shift in which a bed is required for patient p). The value is set to 0 if patient p has already been on the ward during the last shift of the previous planning period
di_shift (p)	Shift $s \in \mathcal{S}^{\text{late}}$ during which patient $p \in \mathcal{P}$ is discharged (last shift in which a bed is required for patient p). The value is set to $S + 1$ if patient p will still be on the ward after the last shift S of the planning period
$y^{\text{prev}}(p)$	Room $r \in \mathcal{R}$ that patient $p \in \mathcal{P}$ with ad_shift $(p) = 0$ has been assigned to during the last shift of the previous planning period
num_beds(r)	Nonnegative integer (most likely from $\{1, 2, 3, 4\}$) specifying the number of beds in room $r \in \mathcal{R}$
age_group (p)	Age group of patient $p \in \mathcal{P}$ computed as age_group $(p) = \lfloor \frac{\text{age}(p)}{10} \rfloor$ with age(p) denoting the age of the patient in years
skill_level (n)	Skill level of nurse $n \in \mathcal{N}$ (possible values are 0, 1, 2)
skill_req (p, s)	Minimum skill level of a nurse required by patient $p \in \mathcal{P}$ during shift $s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}$ (possible values are 0, 1, 2)
w_load (p, s)	Nonnegative number specifying the workload resulting from taking care of patient $p \in \mathcal{P}$ during shift $s \in \mathcal{S}$
max_load (n, s)	Nonnegative number specifying the maximum workload allowed for nurse $n \in \mathcal{N}$ during shift $s \in \mathcal{S}(n)$
dist (r, r')	Nonnegative number specifying the walking distance between rooms $r, r' \in \mathcal{R}$ (where dist $(r, r') =$ dist (r', r) for all $r, r' \in \mathcal{R}$, i.e., distances are symmetric)

$\mathrm{dist}(a, r)$	Nonnegative number specifying the walking distance between additional room $a \in \mathcal{A}$ and room $r \in \mathcal{R}$
$\mathrm{walk\_pat}^{\circ}(s)$	Nonnegative weight for different walking patterns depending on the shift $s \in \mathcal{S}$. A high value of $\mathrm{walk\_pat}^{\circ}(s)$ indicates that most nurses walk directly from patient to patient during shift s (circular pattern)
$\mathrm{walk\_pat}^{\star}(s)$	Nonnegative weight for different walking patterns depending on the shift $s \in \mathcal{S}$. A high value of $\mathrm{walk\_pat}^{\star}(s)$ indicates that most nurses walk directly from additional rooms such as the nursing station to patient rooms and back during shift s (star-like pattern)

Decision variables:

$y_{p,r,s}$	Binary variable indicating whether patient $p \in \mathcal{P}$ is assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\mathrm{early}}$(only defined if $\mathrm{ad\_shift}(p) \le s \le \mathrm{di\_shift}(p)$, i.e., if patient p is on the ward during early shift s)
$\mathrm{f\_in\_room}_{r,s}$	Binary variable indicating whether at least one female patient is assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\mathrm{early}}$
$\mathrm{m\_in\_room}_{r,s}$	Binary variable indicating whether at least one male patient is assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\mathrm{early}}$
$\mathrm{vio}^{\mathrm{gender}}_{r,s}$	Binary variable indicating whether more than one gender is in room $r \in \mathcal{R}$ during shift $s \in \mathcal{S}^{\mathrm{early}}$
$\mathrm{trans}_{p,s}$	Binary variable indicating whether patient $p \in \mathcal{P}$ is transferred to a different room after night shift $s \in (\mathcal{S}^{\mathrm{night}} \cup \{0\}) \setminus \{S\}$ (and before early shift $s+1$) (only defined if $\mathrm{ad\_shift}(p) \le s \le \mathrm{di\_shift}(p) - 1$)
$\mathrm{age\_group}^{\max}_{r,s}$	nonnegative fractional variable representing the maximum age group among all patients $p \in \mathcal{P}$ assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\mathrm{early}}$
$\mathrm{age\_group}^{\min}_{r,s}$	Nonnegative fractional variable representing the minimum age group among all patients $p \in \mathcal{P}$ assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\mathrm{early}}$

$x_{p,n,s}$	Binary variable indicating whether patient $p \in \mathcal{P}$ is assigned to nurse $n \in \mathcal{N}$ during shift $s \in \mathcal{S}$,(only defined if $s \in \mathcal{S}(n)$ and ad_shift$(p) \le s \le$ di_shift(p), i.e., if nurse n is assigned to shift s and patient p is on the ward during shift s)
$\text{vio}^{\text{skill}}_{p,s}$	Binary variable indicating whether patient $p \in \mathcal{P}$ is assigned to a nurse with a lower skill level than required during shift $s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}$ (only defined if ad_shift$(p) \le s \le$ di_shift(p) and skill_req$(p, s) \ge 2$, i.e., if patient p is on the ward during shift s and requires at least an experienced nurse during shift s)
$\text{ever\_assigned}_{p,n}$	Binary variable indicating whether patient $p \in \mathcal{P}$ is assigned to nurse $n \in \mathcal{N}$ during at least one shift $s \in \mathcal{S}$
$\text{vio}^{\text{load}}_{n,s}$	Nonnegative fractional variable representing the excess workload assigned to nurse $n \in \mathcal{N}$ during shift $s \in \mathcal{S}$(only defined if $s \in \mathcal{S}(n)$, i.e., if nurse n is assigned to shift s)
$\text{vio}^{\text{fair}}_{n,n',s}$	Nonnegative fractional variable representing the excess in relative workload (relative to the desired maximum) of nurse $n \in \mathcal{N}$ compared to nurse $n' \in \mathcal{N}$ during shift $s \in \mathcal{S}$
$\text{vio}^{\text{fair}}_{n,n'}$	Nonnegative fractional variable representing the overall excess in relative workload (relative to the desired maximum) of nurse $n \in \mathcal{N}$ compared to nurse $n' \in \mathcal{N}$
$\text{in\_room}_{n,r,s}$	Binary variable indicating whether nurse $n \in \mathcal{N}$ is assigned at least one patient in room $r \in \mathcal{R}$ during shift $s \in \mathcal{S}$(only defined if $s \in \mathcal{S}(n)$, i.e., if nurse n is assigned to shift s)
$\text{dist}_{n,s}$	nonnegative fractional variable representing the total walking distance for nurse $n \in \mathcal{N}$ during shift $s \in \mathcal{S}$ (only defined if $s \in \mathcal{S}(n)$, i.e., if nurse n is assigned to shift s)
$\text{both\_rooms}_{n,r,r',s}$	Binary variable indicating whether nurse $n \in \mathcal{N}$ is assigned patients in both room $r \in \mathcal{R}$ and room $r' \in \mathcal{R}$ during shift $s \in \mathcal{S}$ (only defined if assign$(n, s) = 1$, i.e., if nurse n is assigned to shift s)

The objective function of the MIP to be minimized consists of a weighted sum of several objectives. These objectives include those classically considered in the PRA problem (objectives (1)–(4)) and the NPA problem (objectives (5)–(7)). Moreover, two additional objectives are considered that rely explicitly on the interaction of the two problems: objective (8) considers assigning the minimum number of nurses per room during each shift, while objective (9) minimizes the nurses' walking distances.

The weights of the different objectives in the weighted sum are based on the existing literature and discussions with our partner hospital (see Sect. 7 for the values used in our computational experiments).

Patient transfers objective

(1) Minimization of the number of patient transfers across all patients and shifts (could be weighted differently for different patients and / or different shifts):

$$\min \sum_{\substack{p\in\mathcal{P},s\in(\mathcal{S}^{\text{night}}\cup\{0\})\setminus\{S\}:\\ \text{ad\_shift}\,(p)\le s\le\,\text{di\_shift}\,(p)-1}} \text{trans}_{p,s}$$

Patient inconvenience objective

(2) Minimization of the age group difference across all rooms and shifts:

$$\min \sum_{r\in\mathcal{R},s\in\mathcal{S}^{\text{early}}} (\,\text{age\_group}_{r,s}^{\max} - \text{age\_group}_{r,s}^{\min})$$

Gender mixing objective

(3) Minimization of gender mixing across all rooms and shifts:

$$\min \sum_{r\in\mathcal{R},s\in\mathcal{S}^{\text{early}}} \text{vio}_{r,s}^{\text{gender}}$$

Equipment violation objective

(4) Minimization of required equipment violation across all rooms and shifts:

$$\min \sum_{\substack{p\in\mathcal{P},r\in\mathcal{R},s\in\mathcal{S}^{\text{early}}:\\ \text{ad\_shift}\,(p)\le s\le\,\text{di\_shift}\,(p)\\ \text{and }\mathcal{E}(p,s)\setminus\mathcal{E}(r)\neq\emptyset}} y_{p,r,s}$$

Continuity of care objective

(5) Minimization of the number of different nurses that treat each patient across all patients (could be weighted differently for different patients):

$$\min \sum_{p\in\mathcal{P},n\in\mathcal{N}\setminus\mathcal{N}^{\text{prev}}(p)} \text{ever\_assigned}_{p,n}$$

Penalization of skill level requirements objective

(6) Minimization of violations of skill level requirements of patients:

$$\min \sum_{\substack{p \in \mathcal{P}, s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}: \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p) \\ \text{and } \text{skill\_req}(p,s) \geq 1}} \text{vio}^{\text{skill}}_{p,s}$$

Penalization of undesired workload distributions objective

(7) Minimization of undesired workload distributions for nurses:

$$\min \sum_{n \in \mathcal{N}, s \in \mathcal{S}(n)} \text{vio}^{\text{load}}_{n,s} + \sum_{n,n' \in \mathcal{N}} \text{vio}^{\text{fair}}_{n,n'} + \sum_{\substack{n,n' \in \mathcal{N}, \\ s \in \mathcal{S}(n) \cap \mathcal{S}(n')}} \text{vio}^{\text{fair}}_{n,n',s}$$

Assigning the minimum number of nurses per room objective

(8) Minimization of the number of nurses assigned to rooms across all shifts:

$$\min \sum_{n \in N, r \in R, s \in S} \text{in\_room}_{n,r,s}$$

Walking distances objective

(9) Minimization of the walking distances across all nurses and shifts

$$\min \sum_{n \in \mathcal{N}, s \in \mathcal{S}(n)} \text{dist}_{n,s}$$

The constraints of the MIP can be formulated as follows:

(I) **Assignment of patients to rooms**

(10) Each patient $p \in \mathcal{P}$ is assigned to exactly one room $r \in \mathcal{R}$ during each early shift $s \in \mathcal{S}^{\text{early}}$ between their admission and discharge:

$$\sum_{r \in \mathcal{R}} y_{p,r,s} = 1 \quad \forall p \in \mathcal{P},\ s \in \mathcal{S}^{\text{early}} \ : \ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

(11) No room $r \in \mathcal{R}$ can be assigned more than num_beds(r) patients during any early shift $s \in \mathcal{S}^{\text{early}}$:

$$\sum_{p \in \mathcal{P}:\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)} y_{p,r,s} \leq \text{num\_beds}(r) \quad \forall r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}}$$

(12) The variable $\text{f\_in\_room}_{r,s}$ ($\text{m\_in\_room}_{r,s}$) is set to one if at least one female (male) patient is assigned to room r during early shift $s \in \mathcal{S}^{\text{early}}$:

$$y_{p,r,s} \leq \text{f\_in\_room}_{r,s} \quad \forall p \in \mathcal{F},\ r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}} \ : \ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

$$y_{p,r,s} \leq \text{m\_in\_room}_{r,s} \quad \forall p \in \mathcal{M},\ r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}} \ : \ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

(13) No room $r \in \mathcal{R}$ should be assigned both female and male patients during any early shift $s \in \mathcal{S}^{\text{early}}$:

$$\text{f\_in\_room}_{r,s} + \text{m\_in\_room}_{r,s} \leq 1 + \text{vio}^{\text{gender}}_{r,s} \quad \forall r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}}$$

(II) **Patient transfers**

(14) The patient transfer variables $\text{trans}_{p,s}$ are set correctly for each patient $p \in \mathcal{P}$ and each night shift $s \in \mathcal{S}^{\text{night}}$ between their admission and discharge:

$$y_{p,r,s+1} - y_{p,r,s-2} \leq \text{trans}_{p,s} \quad \forall p \in \mathcal{P},\ r \in \mathcal{R},\ s \in \mathcal{S}^{\text{night}} \setminus \{S\} : \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p) - 1$$

(15) The patient transfer variables $\text{trans}_{p,0}$ that indicate a transfer between the last shift of the previous planning period (shift 0) and the first (early) shift of the current planning period (shift 1) are set correctly for each patient $p \in \mathcal{P}$ with $\text{ad\_shift}(p) = 0$:

$$y_{p,r,1} \leq \text{trans}_{p,0} \quad \forall p \in \mathcal{P},\ r \in \mathcal{R} \setminus \{y^{\text{prev}}(p)\} : \text{ad\_shift}(p) = 0$$

(III) **Patient inconvenience**

(16) The variable $\text{age\_group}^{\max}_{r,s}$ is restricted by the maximum age group of patients in room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\text{early}}$:

$$\text{age\_group}^{\max}_{r,s} \geq \text{age\_group}(p) \cdot y_{p,r,s} \quad \forall p \in \mathcal{P},\ r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}} : \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

(17) The variable $\text{age\_group}^{\min}_{r,s}$ is restricted by the minimum age group of patients in room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\text{early}}$:

$$\text{age\_group}^{\min}_{r,s} \leq \text{age\_group}(p) + 12 \cdot (1 - y_{p,r,s}) \quad \forall p \in \mathcal{P},\ r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}} : \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

Note that the coefficient 12 on the right-hand side needs to be increased if patients with ages 130 or older (age group 13 or higher) are present.

(18) The variable $\text{age\_group}^{\min}_{r,s}$ is set to zero if no patients are assigned to room $r \in \mathcal{R}$ during early shift $s \in \mathcal{S}^{\text{early}}$:

$$\text{age\_group}^{\min}_{r,s} \leq 12 \cdot \sum_{p \in \mathcal{P}:\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)} y_{p,r,s} \quad \forall r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}}$$

Note that the coefficient 12 on the right-hand side needs to be increased if patients with ages 130 or older (age group 13 or higher) are present.

(19) The value of the variable $\text{age\_group}^{\max}_{r,s}$ must not be smaller than the value of the variable $\text{age\_group}^{\min}_{r,s}$ for any room $r \in \mathcal{R}$ and any early shift $s \in \mathcal{S}^{\text{early}}$:

$$\text{age\_group}_{r,s}^{\min} \leq \text{age\_group}_{r,s}^{\max} \quad \forall r \in \mathcal{R},\ s \in \mathcal{S}^{\text{early}}$$

Note that this constraint is not required for the correctness of the model and is only used to improve solution times.

(IV) **Assignment of patients to nurses**

(20) Each patient $p \in \mathcal{P}$ is assigned to exactly one nurse $n \in \mathcal{N}$ with $s \in \mathcal{S}(n)$ during each shift $s \in \mathcal{S}$ between their admission and discharge:

$$\sum_{n \in \mathcal{N}:\, s \in \mathcal{S}(n)} x_{p,n,s} = 1 \quad \forall p \in \mathcal{P},\ s \in \mathcal{S}:\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

(21) For each patient $p \in \mathcal{P}$ and each shift $s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}$ with $\text{skill\_req}(p,s) \geq 1$, the variable $\text{vio}_{p,s}^{\text{skill}}$ is set to one if the patient is not assigned to a nurse with the required skill level:

$$\sum_{\substack{n \in \mathcal{N}:\, s \in \mathcal{S}(n) \text{ and} \\ \text{skill\_level}(n) \geq \text{skill\_req}(p,s)}} x_{p,n,s} = 1 - \text{vio}_{p,s}^{\text{skill}} \quad \forall p \in \mathcal{P},\ s \in \mathcal{S} \setminus \mathcal{S}^{\text{night}}: \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p) \\ \text{and } \text{skill\_req}(p,s) \geq 1$$

(22) The variable $\text{ever\_assigned}_{p,n}$ is set to one if and only if $n \in \mathcal{N}^{\text{prev}}(p)$ or patient $p \in \mathcal{P}$ is assigned to nurse $n \in \mathcal{N}$ during at least one shift:

$$x_{p,n,s} \leq \text{ever\_assigned}_{p,n} \quad \forall p \in \mathcal{P},\ n \in \mathcal{N},\ s \in \mathcal{S}:\ s \in \mathcal{S}(n) \\ \text{and } \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)$$

$$\text{ever\_assigned}_{p,n} = 1 \quad \forall p \in \mathcal{P},\ n \in \mathcal{N}^{\text{prev}}(p)$$

$$\text{ever\_assigned}_{p,n} \leq \sum_{\substack{s \in \mathcal{S}:\, s \in \mathcal{S}(n) \text{ and} \\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)}} x_{p,n,s} \quad \forall p \in \mathcal{P},\ n \in \mathcal{N} \setminus \mathcal{N}^{\text{prev}}(p)$$

(V) **Nurses' workload**

(23) For each nurse $n \in \mathcal{N}$ and each shift $s \in \mathcal{S}(n)$, any workload exceeding $\text{max\_load}(n,s)$ leads to a corresponding increase of the variable $\text{vio}_{n,s}^{\text{load}}$:

$$\sum_{p \in \mathcal{P}:\ \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)} x_{p,n,s} \cdot \text{w\_load}(p,s) \leq \text{max\_load}(n,s) + \text{vio}_{n,s}^{\text{load}} \\ \forall n \in \mathcal{N},\ s \in \mathcal{S}(n)$$

(24) Fair workload distribution per shift: For any two nurses $n, n' \in \mathcal{N}$ assigned to a shift $s \in \mathcal{S}$, if the relative workload (relative to the desired maximum) of nurse n during shift s exceeds the relative workload of nurse n' during shift s, the variable $\text{vio}_{n,n',s}^{\text{fair}}$ must be increased correspondingly:

$$\sum_{p\in\mathcal{P}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)} x_{p,n,s}\cdot\frac{\mathrm{w\_load}(p,s)}{\mathrm{max\_load}(n,s)}$$
$$\le \sum_{p\in\mathcal{P}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)} x_{p,n',s}\cdot\frac{\mathrm{w\_load}(p,s)}{\mathrm{max\_load}(n',s)}+\mathrm{vio}^{\mathrm{fair}}_{n,n',s}$$
$$\forall n,n'\in\mathcal{N},\ s\in\mathcal{S}(n)\cap\mathcal{S}(n')$$

(25) Fair distribution of workload overall: For any two nurses $n, n' \in \mathcal{N}$, if the relative workload (relative to the desired maximum) of nurse n exceeds the relative workload of nurse n', the variable $\mathrm{vio}^{\mathrm{fair}}_{n,n'}$ must be increased correspondingly:

$$\sum_{s\in\mathcal{S}(n)}\ \sum_{p\in\mathcal{P}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)} x_{p,n,s}\cdot\frac{\mathrm{w\_load}(p,s)}{\mathrm{max\_load}(n,s)}$$
$$\le \sum_{s\in\mathcal{S}(n')}\ \sum_{p\in\mathcal{P}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)} x_{p,n',s}\cdot\frac{\mathrm{w\_load}(p,s)}{\mathrm{max\_load}(n',s)}+\mathrm{vio}^{\mathrm{fair}}_{n,n'}$$
$$\forall n,n'\in\mathcal{N}$$

(VI) **Assignment of all patients in the same room to the same nurse**

(26) The variables $\mathrm{in\_room}_{n,r,s}$ are set correctly for each nurse $n \in \mathcal{N}$, each room $r \in \mathcal{R}$, and each shift $s \in \mathcal{S}(n)$:

$$\mathrm{in\_room}_{n,r,s} \ge x_{p,n,s} + y_{p,r,s} - 1 \qquad \forall p\in\mathcal{P},\ n\in\mathcal{N},\ r\in\mathcal{R},\ s\in\mathcal{S}(n)\cap\mathcal{S}^{\mathrm{early}}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)$$
$$\mathrm{in\_room}_{n,r,s} \ge x_{p,n,s} + y_{p,r,s-1} - 1 \qquad \forall p\in\mathcal{P},\ n\in\mathcal{N},\ r\in\mathcal{R},\ s\in\mathcal{S}(n)\cap\mathcal{S}^{\mathrm{late}}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)$$
$$\mathrm{in\_room}_{n,r,s} \ge x_{p,n,s} + y_{p,r,s-2} - 1 \qquad \forall p\in\mathcal{P},\ n\in\mathcal{N},\ r\in\mathcal{R},\ s\in\mathcal{S}(n)\cap\mathcal{S}^{\mathrm{night}}:\ \mathrm{ad\_shift}(p)\le s\le \mathrm{di\_shift}(p)$$

(VII) **Nurses' walking distance**

(27) The variables $\mathrm{both\_rooms}_{n,r,r',s}$ are set correctly for each nurse $n \in \mathcal{N}$, each two rooms $r, r' \in \mathcal{R}$, and each shift $s \in \mathcal{S}(n)$:

$$\mathrm{both\_rooms}_{n,r,r',s} \ge \mathrm{in\_room}_{n,r,s} + \mathrm{in\_room}_{n,r',s} - 1\ \forall n\in\mathcal{N},\ r,r'\in\mathcal{R},\ s\in\mathcal{S}(n)$$

(28) The walking distance variables $\mathrm{dist}_{n,s}$ are set correctly for each nurse $n \in \mathcal{N}$ and each shift $s \in \mathcal{S}(n)$:

$$\mathrm{dist}_{n,s} = \mathrm{walk\_pat}^{\circ}(s)\cdot\frac{1}{2}\cdot\sum_{r\in\mathcal{R}}\ \sum_{r'\in\mathcal{R}\setminus\{r\}} \mathrm{both\_rooms}_{n,r,r',s}\cdot \mathrm{dist}(r,r')$$
$$+\mathrm{walk\_pat}^{\star}(s)\cdot\sum_{a\in\mathcal{A}}\sum_{r\in\mathcal{R}} \mathrm{in\_room}_{n,r,s}\cdot\mathrm{dist}(a,r)\ \forall n\in\mathcal{N},\ s\in\mathcal{S}(n)$$

5 Solution methods

Addressing PRA and NPA challenges in hospitals reveals complexities that hinder timely optimal solutions. The PRA problem, even in isolation, is known to be NP-hard and only unrealistically small instances can be solved to optimality in reasonable time. Thus, it not surprising that the integrated problem is very difficult to solve on instances of a realistic size and solving the MIP provided in the previous section takes a prohibitive amount of time on such instances (see also Sect. 7). Therefore, we now present two different methods for generating good solutions in reasonable computation times. The first method, which is mainly presented as a point of comparison to the completely integrated MIP from Sect. 4, is a sequential approach based on a natural decomposition of the integrated MIP into models for the two interacting subproblems. The second method is an efficient greedy heuristic for the integrated problem, which also allows easy adaptions to various dynamic problem versions.

5.1 Sequential solution approach

As a point of comparison to the completely integrated MIP presented in the previous section and to potentially achieve faster computation times, we consider the sequential solution of the two submodels for the PRA and NPA parts. Here, we first solve the PRA part and then solve the NPA part given the PRA. Note that this approach of solving the two assignment problems sequentially will not yield optimal solutions for the integrated problem in general.

The PRA part of the model consists of the patient transfers objective (1), the patient inconvenience objective (2), the gender mixing objective (3), and the equipment violations objective (4) as well as the constraints concerning the assignment of patients to rooms (I), the patients transfers (II), and the inconvenience of patients (III) and the associated decision variables.

The NPA part of the model contains the continuity of care objective (5), the penalization of skill level requirements objective (6), undesired workload distributions objective (7), the objective for assigning the minimum number of nurses per room (8), and the walking distance objective (9). As constraints, the NPA part contains those concerning the assignment of patients to nurses (IV), the nurses' workload (V), the assignment of all patients in the same room to the same nurse (VI), and the nurses' walking distance (VII). Concerning variables, the NPA model contains all variables appearing in these objectives and constraints, but the $y_{p,r,s}$ are not decision variables anymore since their values are carried over from the solution obtained for the PRA part of the model.

5.2 Heuristic solution approach

Due to the computationally challenging nature of the PRA problem, heuristic solution methods are often used to tackle this problem in the recent literature (Guido et al 2018; Schäfer et al 2019). Naturally, the computational challenges become even greater when considering PRAs and NPAs jointly in an integrated optimization problem as we do in this paper. Therefore, we now introduce a heuristic solution approach for the IPRNPA problem, which extends the PRA heuristic suggested in Schäfer et al (2019). This heuristic has been selected as the basis for our extension since it successfully balances computational effort and solution quality in a practically applicable solution method. Moreover, even though the original heuristic neglects transfers and does not incorporate NPA considerations, the basic idea of the algorithm is still suitable for the IPRNPA problem, as we demonstrate in this section.

Our heuristic is initially tailored to the static version of the IPRNPA problem, but its scope extends beyond the static setting. In fact, the heuristic does not require to have all information about patients available in advance and can easily be extended to a variety of dynamic settings. As an illustrative example, instead of restricting patient transfers between rooms to happen between a night shift and the following early shift as in the static IPRNPA problem, the heuristic could handle patient transfers whenever new information about patients is obtained. The same applies for NPAs, which could alternatively be allowed also during shifts whenever new information becomes available. Thus, even complex dynamic problem settings based on dynamic PRA problems such as the one considered by Ceschia and Schaerf (2011, 2012) can be handled without difficulties.

5.2.1 Algorithm description

The heuristic assigns patients to rooms and nurses to patients in a greedy fashion. To this end, the heuristic considers the days of the planning period chronologically, where each day is represented by the corresponding early shift. For each day, the heuristic iteratively fixes both the PRA and the NPA for only this day jointly for a single patient in a way that yields the lowest current contribution to the objective function. After each such assignment, the current objective function contributions of the remaining possible assignments for the day are updated before the assignments for the next patient are fixed. Once the room and nurse assignments have been fixed for all patients that are on the ward during the considered day, the heuristic moves on to the next day. During this iterative process, we have to take into account that decisions for the two kinds of assignments are made on different time scales. Concerning PRAs, a decision about the room the patient is assigned to is made only once per day before the start of the early shift, whereas, concerning NPAs, three different nurse must be assigned to a patient for each day (for the early, late, and night shift) since each nurse works at most one shift per day.

We now describe our heuristic, whose pseudocode is shown in Algorithm 1, more formally. Here, we first describe the algorithm without the heterogeneity check between patient admission and discharge times that is represented by the heterogeneity matrix HetMatrix in the pseudocode and is motivated and explained afterwards in Sect. 5.2.2.

Upon initialization of the heuristic, all relevant assignment variables $x_{p,n,s}$ and $y_{p,r,s}$ are set to 0 to indicate that no assignments have been made so far. Afterwards, the days of the planning period are considered in chronological order represented by their early shifts. When considering a day of the planning period represented by the corresponding early shift $s \in \mathcal{S}^{\text{early}}$, the set of relevant patients for which room and nurse assignments are to be made for this day is denoted by $\mathcal{P}(s) := \{p \in \mathcal{P} : \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)\}$. Note that $\mathcal{P}(s)$ also includes patients that have been on the ward already on the previous day if they have not been discharged yet since the room and nurse assignments for such patients have so far only been fixed up to the previous day. At each point during the assignment process for the day, the set of rooms that are still available (i.e., not yet fully occupied) is denoted by $\mathcal{R}(s) \subseteq \mathcal{R}$. The set of nurses who are on duty on the corresponding day is partitioned according to their assigned shifts into $\mathcal{N}^{\text{early}}(s)$, $\mathcal{N}^{\text{late}}(s)$, and $\mathcal{N}^{\text{night}}(s)$. Consequently, the Cartesian product $\mathcal{N}^{\text{comb}}(s) := \mathcal{N}^{\text{early}}(s) \times \mathcal{N}^{\text{late}}(s) \times \mathcal{N}^{\text{night}}(s)$ corresponds to all possible ordered triples of nurses that can potentially be assigned to a patient during the three shifts of the day that starts with the early shift $s \in \mathcal{S}^{\text{early}}$. The potential contributions to the objective function for each triple (p, n^{comb}, r) consisting of a patient $p \in \mathcal{P}(s)$, a triple $n^{\text{comb}} \in \mathcal{N}^{\text{comb}}(s)$ of nurses, and an available room $r \in \mathcal{R}(s)$ are stored in a contribution table denoted by ContribTable.

Whenever the heuristic starts considering some day of the planning period starting with early shift $s \in \mathcal{S}^{\text{early}}$, the contribution table is filled with the current contributions that correspond to the potential assignments of patients $p \in \mathcal{P}(s)$ to nurse combinations $n^{\text{comb}} \in \mathcal{N}^{\text{comb}}(s)$ and rooms $r \in \mathcal{R}(s)$. These current contributions are computed based on the changes to the objective function value that would currently be induced by the corresponding assignments. Here, the objective function is the same weighted sum of the objectives (1)–(9) as in the MIP described in Sect. 4.

The heuristic then fixes the assignments of rooms and nurse combinations for the day represented by early shift s for all patients in $\mathcal{P}(s)$ in a greedy fashion. This is done by iteratively identifying the triple (p, n^{comb}, r) that has the lowest contribution value in ContribTable and setting the values of the corresponding assignment variables $x_{p,n,s}$ and $y_{p,r,s}$ to 1, which means that patient p is assigned to the nurse combination n^{comb} and the room r. Subsequently, the allocated patient p is removed from $\mathcal{P}(s)$ and the room r is removed from $\mathcal{R}(s)$ in case that the assignment has resulted in full occupancy of room r on the corresponding day. Moreover, the contribution table ContribTable needs to be updated by removing all entries corresponding to patient p and possibly room r. In addition, the update involves adjusting the objective contributions for the remaining possible assignments of the day to ensure that they accurately represent the new state of PRA and NPAs. For example, the previously-calculated contribution to the gender mixing objective for room r is updated for all remaining patients in $\mathcal{P}(s)$ with a different gender than the patient p just assigned. After the updated contribution table has been computed, the heuristic continues with the next iteration based on the updated contribution table until $\mathcal{P}(s) = \emptyset$ (i.e., all assignments for the current day have been fixed) and the procedure continues with the next day.

The heuristic terminates once the last day of the planning period has been considered, i.e., once all PRA and NPAs have been fixed for the whole planning period.

5.2.2 Heterogeneity check between patients

A point that is not considered in the heuristic as described so far is the coordination of arrivals and discharges of patients that are assigned to the same room. If occupancy levels are high, this could lead to the unnecessary creation of gender-mixed rooms or avoidable patient transfers. For example, assigning several male patients with similar arrival and discharge shifts to different rooms instead of the same room might leave no other rooms available for female patients admitted shortly afterward unless some of the male patients are transferred. To avoid such unfavorable incidents explicitly, we next describe a new patient-to-patient heterogeneity measure, whose values are stored in a heterogeneity matrix. This heterogeneity matrix is used in the heuristic to assign patients who are discharged at similar times to the same room by explicitly incorporating discharge shifts during the computation of possible objective function contributions.

The proposed patient-to-patient heterogeneity measure is based on the difference between two patients' discharge shifts. Formally, for two patients $p, p' \in \mathcal{P}$, we let $\text{di\_shift\_diff}(p, p')$ denote the absolute difference between the discharge shifts $\text{di\_shift}\,(p)$ and $\text{di\_shift}\,(p')$. The heterogeneity for the two patients is then calculated by the following formula[2]:

$$\text{het}(p, p') := \ln \text{di\_shift\_diff}(p, p')$$

This means that the heterogeneity value $\text{het}(p, p')$ is the natural logarithm of the difference between $\text{di\_shift}\,(p)$ and $\text{di\_shift}\,(p')$. Here, the logarithm is used in order to limit the growth of the values in cases where the planning period is long, where large values would otherwise occur. Note that, since the logarithm is a strictly increasing function, a lower heterogeneity value for two patients signifies more similar discharge dates and, thus, a better fit between the two patients to be assigned to the same room.

The heterogeneity values of all patient pairs are calculated upon initialization of the heuristic and stored in a heterogeneity matrix denoted by HetMatrix. Here, since the heterogeneity values are symmetric (i.e., $\text{het}(p, p') = \text{het}(p', p)$ for all $p, p' \in \mathcal{P}$), it suffices to compute the values above the main diagonal of the matrix to improve efficiency. The heterogeneity values of patient pairs are then used whenever the current objective contributions are computed for an assignment that involves assigning a patient to a room to which other patients have already been assigned on the same day. In this case, the appropriately weighted maximum of the heterogeneity values between the new patient and the already assigned patients is added as an additional summand to the current objective contributions.

[2] In case that $\text{di\_shift}\,(p) = \text{di\_shift}\,(p')$, where the argument of the logarithm equals zero, we set $\text{het}(p, p')$ to zero. In all other cases, the value is strictly positive.

Algorithm 1 IPRNPA heuristic

Input: An instance of the IPRNPA problem
Output: patient-to-nurse assignments x, patient-to-room assignments y

$x_{p,n,s} \leftarrow 0$ for all $p \in \mathcal{P}, n \in \mathcal{N}, s \in \mathcal{S}$
$y_{p,r,s} \leftarrow 0$ for all $p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{S}^{\text{early}}$
HetMatrix $\leftarrow$ calculateHeterogeneityMatrix
for $s \in \mathcal{S}^{\text{early}}$ in increasing order **do**
  $\mathcal{P}(s) \leftarrow \{p \in \mathcal{P} : \text{ad\_shift}(p) \leq s \leq \text{di\_shift}(p)\}$
  $\mathcal{N}^{\text{comb}}(s) \leftarrow \mathcal{N}^{\text{early}}(s) \times \mathcal{N}^{\text{late}}(s) \times \mathcal{N}^{\text{night}}(s)$
  $\mathcal{R}(s) \leftarrow \mathcal{R}$
  ContribTable $\leftarrow$ calculateContributionTable($\mathcal{P}(s), \mathcal{N}^{\text{comb}}, \mathcal{R}(s)$, HetMatrix)
  while $\mathcal{P}(s) \neq \emptyset$ **do**
    $(p, n^{\text{comb}}, r) \leftarrow$ argmin(ContribTable)
    $y_{p,r,s} \leftarrow 1$
    $(n^{\text{early}}, n^{\text{late}}, n^{\text{night}}) \leftarrow n^{\text{comb}}$
    $x_{p,n^{\text{early}},s}, x_{p,n^{\text{late}},s}, x_{p,n^{\text{night}},s} \leftarrow 1$
    $\mathcal{P}(s) \leftarrow \mathcal{P}(s) \setminus \{p\}$
    if room r is fully occupied **then**
      $\mathcal{R}(s) \leftarrow \mathcal{R}(s) \setminus \{r\}$
    end if
    ContribTable $\leftarrow$ updateContributionTable(ContribTable, p, n^{comb}, r, HetMatrix)
  end while
end for

6 Instance generator and case study

We now present a detailed description of the parameterized instance generator that we developed to create realistic test instances of the IPRNPA problem. Afterwards, we present the structure and key data of the real-world instances obtained from our partner hospital. The concrete parameter values used for generating the artificial instances as well as the numbers of instances considered will be described in Sect. 7.

6.1 Instance generator

In order to generate larger numbers of realistic test instances, we developed a parameterized IPRNPA problem instance generator. The source code of the instance generator is publicly available on GitHub at https://doi.org/10.5281/zenodo.12750420. Additionally, a solution checker for these instances can be found at https://doi.org/10.5281/zenodo.12750359.

The instance generator offers the possibility to create a specifiable number of test instances based on user-defined parameters. These parameters include the number of instances to be created and the option to specify the length of the planning period in weeks. The remaining input parameters can be divided into two main categories: room-related parameters and nurse-related parameters. In the context of room-related parameters, aspects such as the number of patient rooms, their capacity (single, double, triple, or quadruple rooms), occupancy rate, presence of additional rooms, and possible types of room equipment play a central role. In the context of nurse-related parameters, significant options for fine-tuning are available. This includes setting the maximum desired workloads for nurses and specifying the skill levels to be considered. Concerning the total number of available nurses and the nurse roster, the generator operates in two modes: manual and automatic. In the manual mode, the user specifies the number of nurses explicitly, based on which the generator strives to create a feasible nurse roster using the binary integer program (BIP) outlined in Appendix A. In the automatic mode, on the other hand, the generator increases the number of nurses until a feasible nurse roster can be generated using the BIP.

The output of the generator is a set of instances of the specified cardinality. In each instance, a defined number of rooms is described, which is divided into single, double, triple, and quadruple rooms according to the specified distribution of room sizes. The rooms are assigned randomly selected equipment from the specified types of possible equipment. Additional rooms such as nursing stations or storage rooms can be added, and a minimum of one such room is required to calculate walking distances based on the star-like walking pattern. The weighting factors that determine the importance of the circular versus the star-like walking pattern depend on the type of shift. The circular pattern is favored during early shifts, a more equal split during late shifts, and the star-like pattern during night shifts.

The number of nurses is derived from the number of patient rooms, the distribution of the number of beds per room, and the number of nurses of each skill level required per shift, all of which are specified when creating an instance. In instances where three skill levels are specified, we assume that 20% are experienced nurses (skill level 3), 60% are regular nurses (skill level 2) and that 20% are trainees (skill level 1). For instances with two skill levels, we assumed that 80% are skill level 2 and 20% are skill level 1. Additionally, the maximum desired workload associated with each skill level is set to 10 for skill level 1, 12.5 for skill level 2, and 15 for skill level 3. This represents an average nurse-to-patient ratio of 1:4, 1:5, and 1:6, respectively, during each shift.

The patients are generated based on the room configuration and the desired occupancy level. Each patient is assigned a ten-year age group uniformly sampled from 20–30 to 90–100, and, an admission shift, based on the number of rooms and the rooms' capacities. A patient's discharge shift is set as the minimum of the admission shift plus a sampled length of stay (LOS) in days, drawn from a discrete uniform distribution on $\{1,\dots,5\}$, or the last shift of the planning period. Gender is currently assigned based on a 50-50 female-male split. The nurse skill level required for each shift of a patient's LOS is assumed to decrease monotonously. The workload generated by a patient p during each shift is based on a gamma distribution with $\alpha = 3$, $\beta = 0.5 + \text{age\_group}\,(p)/10$, with a minimum of 1, a maximum of 5, and an exponential smoothing parameter of 0.1 that describes the monotonous decrease. The equipment required by the patient is sampled from the types of possible room equipment and assumes monotonously decreasing requirements over the shifts of a patient's LOS.

6.2 Real-world instances

To investigate the potential of our methods using real-world data, we also test them on real-world instances from a Short Stay Unit of our partner hospital. As with our instance generator, these instances are publicly available on GitHub at https://doi.org/10.5281/zenodo.12750420.

The considered ward is not restricted to a single medical specialty, but only patients requiring care that can be delivered according to a strict protocol are admitted. Consequently, there are no acute admissions to this ward and admission and discharge dates of patients as well as their care requirements are known in advance. The ward is closed on weekends, so each patient's ward LOS is at most five days.

During the week, the nursing staff operates in three shifts (early, late and night) to ensure continuous day and night care. Due to the protocol-based care, the nurses who work on the ward do not specialize in a single medical specialty, but require a broad skill set. There are two nurse skill levels: experienced and trainee. One experienced nurse can take care of four to six patients simultaneously during a shift depending on the patient's care requirements and the nurse's experience, while trainee nurses can take care of about two patients in parallel. There is no particular nurse-to-patient ratio during night shifts, but at least two nurses must always be present.

The ward consists of 17 patient rooms of varying sizes: four single rooms, 10 double rooms, two triple rooms, and one quadruple room. Therefore, the ward has a total capacity of 34 beds. Additionally, there is one nursing station where the nurses are usually located when they are not attending to patients. We were provided with the floor plan to estimate the walking distances between the rooms, which we calculated according to the shortest walking path between the centers of each room pair.

In our numerical experiments, we use real admission data and nurse rosters of 40 weeks (about nine months), from before the COVID-19 pandemic. Because the ward closes on the weekend, each week in the data can be considered as a separate instance. We acknowledge that, despite the comprehensive dataset shared by the hospital, certain input data have been omitted due to privacy concerns in order to safeguard individual patient identities. For example, data on the skill levels required for taking care of individual patients and the resulting patient-specific workloads for nurses have not been provided. These missing data have been generated using the corresponding functions of our instance generator presented in Sect. 6.1 based on realistic parameter values that have been established in cooperation with Amsterdam University Medical Center (see Sect. 6.2). The data provided led to infeasible instances for two weeks. In one case, a patient was assigned to a shift for which no nurse was on duty, while in the second instance, a shortage in bed capacity during a particular shift rendered it infeasible to accommodate the required patient load.

7 Experimental results

This section presents our experimental results obtained by testing the MIP as well as the solution methods presented in Sect. 5 on both artificial instances generated by our instance generator from Sect. 6.1 and the real-world instances described in Sect. 6.2. Furthermore, in Sect. 7.3, we explore the impact of adopting an integrated planning approach in comparison to the sequential solution approach, while also conducting an analysis of conflicting objectives.

All computational experiments were performed on a Linux system running Ubuntu 23.04. The hardware includes an AMD EPYC 7542 processor with 32 CPU cores and 64 threads, operating at a base clock speed of 2.9GHz. The system is equipped with 512 GB of memory. The experiments were implemented using Python 3.11 and Gurobi 10.0.1. To solve the completely integrated MIP presented in Sect. 4 and the two submodels in the sequential solution approach from Sect. 5.1, we dedicated 8 threads to each instance when solving with Gurobi.

As described in Sect. 4, the objective function to be minimized consists of a weighted sum of several separate objectives. The specific weights for these objectives were determined based on the existing literature and discussion with our partner hospital. The weights from the existing literature were taken from Demeester et al (2010) and are as follows: the patient transfers objective (1) was assigned a weight of 11, the gender mixing objective (3) a weight of 5, and the equipment violation objective (4) a weight of 5. These weights are frequently employed in similar optimization studies (e.g., Ceschia and Schaerf 2011; Range et al 2014; Thuran and Bilgen 2017), contributing to consistency and comparability across research in the field. Furthermore, in consultation with our partner hospital the following weights were established: the patient inconvenience objective (2) and the continuity of care objective (5) were each given a weight of 1, the penalization of skill level requirements objective (6) and undesired workload distributions objective (7) a weight of 5, the assignment of the minimum number of nurses per room objective (8) a weight of 2, and the walking distances objective (9) a weight of 0.05. We extensively discussed the significance of the various objectives with a contact person from the hospital. This individual is not only familiar with our model but also possesses expertise in MIP modeling, which informed our decision-making process. When applying our heuristic, we also considered the heterogeneity values of patient pairs as discussed in Sect. 5.2.2. Here, we assigned a weight of 1 to incorporate this factor.

The objective values in the following experiments are presented as relative percentages to allow for clearer comparison within and across different solution methods. Reporting absolute numbers can be ambiguous, as specific weightings influence them and lack interpretative clarity. Using relative percentages ensures a more coherent analysis and an easier understanding of the results. For transparency, we have published the absolute numbers on GitHub at https://doi.org/10.5281/zenodo.12750420.

7.1 Artificial instances

We established a structured framework involving two distinct scenarios, each subdivided into three specific variations and two different planning period lengths. This results in a total of 12 scenario-variation-planning period combinations. For each of these 12 combinations, we generated 10 artificial instances using our instance generator described in Sect. 6.1.

The two scenarios encompass configurations of 30 and 60 beds. Within each scenario, Variation 1 comprises exclusively double rooms, Variation 2 exclusively triple rooms, and Variation 3 encompasses a diverse mix of room types, including single, double, triple, and quadruple rooms. For the 30 beds scenario, this allocation translates to 3 single rooms, 5 double rooms, 3 triple rooms, and 2 quadruple rooms. These numbers are doubled for the 60 beds scenario. Moreover, our investigation includes two different planning period lengths for each scenario-variation combination, spanning either 2 or 4 weeks.

Throughout our analysis, we used input parameters for our instance generator that encompass two distinct equipment types, three possible nurse skill levels, the inclusion of a nursing station as an additional room, and a constant occupancy rate of 85%.

In addressing the given scenarios, we applied three distinct solution methods to solve the artificial instances, which we will refer to as Methods 1–3 in the following: the MIP from Sect. 4 (Method 1), the sequential solution approach from Sect. 5.1 (Method 2) and the heuristic solution approach from Sect. 5.2 (Method 3). For Method 1, the runtime was limited to a maximum of 3 h (10800 s). In the case of Method 2, the same total runtime limit was evenly distributed between the two addressed submodels. Additionally, we configured all MIPs to be terminated when the MIP gap falls below 5%. The results, including objective values and runtime data, are presented in Table 1.

The reported results present the objective values as percentages, with the Method 1 objective value serving as the baseline for each instance. These percentages represent the resulting objective proportions, averaged across the 10 instances for each scenario-variation-planning period combination. It is important to emphasize that values below 100 indicate superior performance to Method 1 since the model aims to minimize the objective function. Notably, Method 2 consistently outperforms Method 1 in all scenario-variation-planning period combinations, while Method 3 solution approach surpasses Method 1 in nearly all instances. When comparing Method 2 to Method 3, it becomes clear that Method 2 excels in the 30 beds scenario but lags behind Method 3 for the 60 beds scenario. The standard deviation of the objective values for Method 2 and Method 3 is generally low across most scenario-variation-planning period combinations. However, a few instances with higher standard deviations can be attributed to a handful of strongly deviating instances.

When comparing the runtimes of the three methods, it is evident that Method 1 consistently exceeded the runtime limit in all instances, while Method 2 did so in nearly all instances. In contrast, Method 3 displayed remarkable efficiency, requiring, on average, only 27 to 79 seconds for the 30 beds scenario instances and 874 to 1773s for the 60 beds scenarios. This disparity in runtimes is primarily due to the

Table 1 Artificial instances: Runtimes and objective values

Scenario	Variation	Planning horizon	Runtime [seconds]		Objective value[1]	
			Avg	Stdev	Avg [%]	Stdev [%]
Method 1: MIP						
30 beds	Var. 1 (double rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
	Var. 2 (triple rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
	Var. 3 (mixed rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
60 beds	Var. 1 (double rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
	Var. 2 (triple rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
	Var. 3 (mixed rooms)	2 weeks	10800	0	100	0
		4 weeks	10800	0	100	0
Method 2: Sequential solution approach						
30 beds	Var. 1 (double rooms)	2 weeks	10800	0	60	3
		4 weeks	10800	0	50	10
	Var. 2 (triple rooms)	2 weeks	6604	469	73	2
		4 weeks	10800	0	42	9
	Var. 3 (mixed rooms)	2 weeks	9855	927	65	3
		4 weeks	10800	0	47	10
60 beds	Var. 1 (double rooms)	2 weeks	10800	0	26	2
		4 weeks	10800	0	28	3
	Var. 2 (triple rooms)	2 weeks	10800	0	21	4
		4 weeks	10800	0	25	3
	Var. 3 (mixed rooms)	2 weeks	10800	0	24	2
		4 weeks	10800	0	24	1
Method 3: Heuristic solution approach						
30 beds	Var. 1 (double rooms)	2 weeks	27	1	87	4
		4 weeks	56	1	71	15
	Var. 2 (triple rooms)	2 weeks	38	1	117	3
		4 weeks	79	1	66	14
	Var. 3 (mixed rooms)	2 weeks	31	1	100	5
		4 weeks	64	2	71	15
60 beds	Var. 1 (double rooms)	2 weeks	889	42	17	1
		4 weeks	1520	64	16	0
	Var. 2 (triple rooms)	2 weeks	1065	42	19	3
		4 weeks	1773	52	19	2
	Var. 3 (mixed rooms)	2 weeks	874	53	19	1
		4 weeks	1515	116	17	0

[1] Objective values are provided as proportions (expressed as percentages) of those obtained using Method 1. Values below 100 indicate superior performance to Method 1

non-linear growth in nurse combinations for one-day NPAs, as the heuristic assigns three nurses simultaneously. In total, the 30 beds scenario had 21 nurses available to be on duty, while the 60 beds scenario had 31 nurses. Additionally, for each scenario-variation-planning period combination, the runtimes remained stable with low standard deviation, highlighting the consistency of Method 3's performance. Doubling the planning horizon from 2 to 4 weeks in both scenarios and variations resulted in an approximate doubling of the runtime, indicating a linear relationship between runtime and planning horizon length.

By factoring in the achieved MIP gaps (Table 6 in Appendix B) alongside the objective values and runtimes, we gain insights into the difficulty levels of each scenario-variation-planning period combination. Variation 2, which exclusively considers triple rooms in the 30 beds scenario, emerges as the easiest to solve for Method 1, while Variations 1 and 3, focusing on double rooms and a mixed set of room types, respectively, seem equally challenging. Notably, Method 3 exhibits greater efficiency in Variations 1 and 3 compared to Variation 2 when considering MIP gaps and runtimes. In relation to efficiency, the conclusion for Method 2 is less clear, with a higher MIP gap in the PRA subproblem resulting in a lower MIP gap of the NPA subproblem and vice versa. MIP gaps exceeding 100% for Method 1 indicate cases where the initial root relaxation of the MIP model could not be solved within the runtime limit, resulting in Gurobi finding suboptimal heuristic solutions. In summary, Method 3 consistently outperforms Method 1 and Method 2 across all cases, excelling in runtime, objective value, or both, particularly in complex scenarios where achieving optimality is notably demanding.

Since the considered objective function is a weighted sum of the separate objective functions (1)–(9), we also consider the values of these separate objective functions for a more comprehensive analysis. The detailed outcomes are presented in Table 2.

Upon examining each objective separately, several noteworthy observations emerge. Method 2 and Method 3 in both scenarios exhibit a notable advantage over Method 1 in strictly minimizing patient transfers (objective (1)). The objectives related to continuity of care (objective (5)) and the number of nurses per room (objective (8)) exhibit similar behavior across Methods 1 to 3. Gender mixing (objective (3)) is strongly avoided in Method 2 and to a moderate extent in Method 3. Interestingly, it is worth noting that equipment violations (objective (4)) are more pronounced in Methods 2 and 3 only in the 30 beds scenario, while this issue does not appear as prominently in the 60 beds scenario. In Methods 2 and 3, the skill violation objective (6) worsens with higher bed numbers due to the infrequent skill violations in Method 1, leading to substantial percentage variations. These findings underscore each method's distinct strengths and weaknesses in addressing specific objectives.

Table 2 Artificial instances: Separate objective values differentiated by method

Objective		Method 1		Method 2		Method 3	
		Obj. value[1]		Obj. value[1]		Obj. value[1]	
		30 beds [%]	60 beds [%]	30 beds [%]	60 beds [%]	30 beds [%]	60 beds [%]
(1)	Transfers	100	100	0	20	1	4
(2)	Inconvenience	100	100	30	41	64	55
(3)	Gender mixing	100	100	2	2	41	27
(4)	Equipment violation	100	100	205	24	420	75
(5)	Continuity of care	100	100	103	110	97	101
(6)	Skill	100	100	218	456	378	847
(7)	Workload	100	100	35	17	97	8
(8)	Nurses per room	100	100	79	109	75	100
(9)	Walking distances	100	100	59	35	52	26

[1] Objective values are provided as proportions (expressed as percentages) of those obtained using Method 1 (MIP)

7.2 Real-world instances

To test all developed methods using real-world data, we sourced data from a Short Stay Unit of our partner hospital as described in Sect. 6.2. The dataset comprises 40 individual instances, each spanning a planning period of one week.

To address the absence of certain input data needed for the IPRNPA problem, we generated the missing information using the corresponding functions implemented in our instance generator introduced in Sect. 6.1 based on realistic parameter values established in cooperation with our partner hospital. Specifically, we utilized the functions to generate data on nurse skill level requirements of patients and nurse workloads induced by patients. This approach allowed us to create complete instances for our analyses whilst maintaining patient privacy and confidentiality.

When comparing average (minimum, maximum) parameter values across the considered instances, we observe notable variations. There are approximately 17 (min. 13, max. 23) nurses attending to around 62 (min. 45, max. 76) patients. Patient LOS are around 4.1 (min. 3.6, max. 4.8) shifts, resulting in an average occupancy rate of 61% (min. 46%, max. 72%), while each patient causes a workload of about 2.7 (min. 2.2, max. 3.1) during a day shift, i.e., early and late shift.[3] We selected the 20 instances with the highest occupancy rates from the available dataset for further analysis, taking care to exclude holiday times and instances exhibiting unnatural utilization levels. This careful selection process ensures that the chosen instances accurately represent the most demanding and meaningful scenarios for in-depth examination.

[3] The presented LOS and workload numbers are averages over all patients of an instance.

Table 3 Real-world instances: Runtimes and objective values

Instance #	Method 1: MIP		Method 2: Sequ. sol. app.		Method 3: Heur. sol. app.	
	Runtime [seconds]	Obj. value[1] [%]	Runtime [seconds]	Obj. value[1] [%]	Runtime [seconds]	Obj. value[1] [%]
1	10800	100	5401	98	11	118
2	10800	100	3638	101	9	118
3	10800	100	3005	107	10	147
4	10800	100	3079	100	11	134
5	10800	100	4298	103	9	122
6	10800	100	3775	105	10	130
7	10800	100	1139	99	11	129
8	10800	100	2914	104	11	131
9	10800	100	831	112	10	147
10	10800	100	1491	113	13	145
11	10800	100	2856	104	10	130
12	10800	100	2740	105	10	157
13	10800	100	548	110	8	147
14	10800	100	164	121	6	140
15	10800	100	3014	112	9	124
16	10800	100	4895	109	9	151
17	10800	100	2065	108	10	127
18	10800	100	416	102	10	120
19	10800	100	1573	108	10	140
20	10800	100	5806	104	12	165
Total[2]	**10800**	**100**	**2682**	**106**	**10**	**136**

[1] Objective values are provided as proportions (expressed as percentages) of those obtained using Method 1 (MIP)

[2] Average across all instances

In tackling the real-world instances, we again applied all three solution methods referred to as Methods 1–3 as in Sect. 7.1. The termination criteria based on runtime limits and MIP gap were also adopted from Sect. 7.1. The results, encompassing objective values and runtime data, are presented in Table 3.

The results in real-world instances parallel those in the artificial ones (Sect. 7.1) concerning running time limits, where Method 1 consistently reaches the time limit in all cases. In Method 2, the running time limit is only exceeded in the PRA subproblem of instance 20. In stark contrast, Method 3 demonstrates remarkable efficiency, with an average running time of just 10 seconds, vastly outperforming Method 2, which has an average running time of 2682 seconds.

Regarding objective values, Method 1 consistently outperforms Methods 2 and 3 in almost all instances, with average objective values of 106% and 136%, respectively. The real-world instances are less complex than the artificial 30 beds scenario, evidenced by an average MIP gap of 43% (Table 7 in Appendix B). Method 2 optimally solves all PRA subproblems except for instance 20, where a 100% MIP gap occurs because the lower bound equaled 0. In NPA subproblems, the MIP gap drops below 5% during runtime, prompting the optimization to halt accordingly.

In summary, the evaluation of three methods for healthcare management optimization reveals distinct trade-offs. Method 1, while suitable for small-scale settings such as those provided by our partner hospital, suffers from prohibitively high running time costs, often taking thousands of seconds to achieve results comparable to the heuristic, which only needs a few seconds. For medium-sized problems like the 30 beds scenario, Method 2 delivers superior results compared to Method 3 at the expense of significant running time. In contrast, Method 3 consistently stands out regarding running time efficiency across problem sizes and showcases excellent objective values, particularly in larger scenarios like the 60 beds scenario. Hence, Method 3 emerges as the preferred choice for optimizing healthcare management, striking a favourable balance between computational efficiency and solution quality, especially in more extensive and complex healthcare settings.

7.3 Sensitivity analyses

In this section, we examine the consequences of adopting an integrated planning approach as opposed to the sequential solution approach (Sect. 7.3.1). Further, we conduct an analysis of conflicting objectives by successively amplifying the weights of each objective (Sect. 7.3.2).

7.3.1 Magnitude of effect analysis

This section investigates the impact of adopting an integrated planning approach (Method 1: MIP) for the IPRNPA problem in contrast to a sequential solution approach (Method 2). In particular, we investigate whether the integration of the two assignment problems leads to improvements with respect to the objectives that rely on the interaction of PRA and NPA. These objectives include assigning the minimum possible number of nurses to patients in the same room (objective (8)) or minimizing the nurses walking distance (objective (9)). To systematically investigate these effects in various settings, we again use the three variations of double, triple, and mixed room introduced in Sect. 7.1. To achieve optimal solutions, we use small-scale problem instances, limiting the planning horizon to 2 days (6 shifts), with a fixed capacity of 18 beds for all variations. Additionally, a maximum runtime of 120 h is imposed. Variation 1 comprises 9 double rooms, Variation 2 includes 6 triple rooms, and Variation 3 encompasses 2 single rooms, 3 double rooms, 2 triple rooms, and 1 quadruple room. To ensure robustness and generalizability, 5 instances are generated for each variation using our developed instance generator. Table 4 shows the relative objective values of optimal solutions obtained using the sequential solution approach (Method 2) in relation to those of optimal solutions of the integrated MIP model (Method 1).

Given the short planning horizon of 2 days, it is natural that transfers and equipment violations do not occur for any instances, irrespective of the variation, and there is no difference in the continuity of care objective between the two methods. As is to be expected, by first solving the PRA separately before solving the NPA, Method 2

Table 4 Artificial instances: Impact of integrated planning (Method 1) vs. sequential solution approach (Method 2)

Objective		**Var. 1** (double rooms)				**Var. 2** (triple rooms)				**Var. 3** (mixed rooms)			
		Obj. value[1]				Obj. value[1]				Obj. value[1]			
		Avg [%]	Min [%]	Max [%]	Stdev [%]	Avg [%]	Min [%]	Max [%]	Stdev [%]	Avg [%]	Min [%]	Max [%]	Stdev [%]
(1)	Transfers	–	–	–	–	–	–	–	–	–	–	–	–
(2)	Inconvenience	41	30	56	9	56	40	100	22	31	14	50	12
(3)	Gender mixing	80	0	100	40	40	0	100	49	60	0	100	49
(4)	Equipment violation	–	–	–	–	–	–	–	–	–	–	–	–
(5)	Continuity of care	100	100	100	0	100	100	100	0	100	100	100	0
(6)	Skill	169	120	300	66	143	100	167	23	142	120	200	29
(7)	Workload	101	99	104	1	99	98	100	1	100	98	102	1
(8)	Nurses per room	111	108	113	2	131	118	153	13	134	129	139	4
(9)	Walking distances	132	125	139	6	165	138	201	23	154	135	177	14

[1] Objective values obtained using Method 2 are given as percentages of those obtained using Method 1

obtains better results on average for some of the PRA objectives (objectives (2) and (3)), but notably worse results for some of the NPA objectives (most notably, objective (6)). More interestingly, for the objectives that rely on the integration of the two subproblems, Method 2 results in Variation 1 (2, 3) showing, on average, 11% (31%, 34%) more assigned nurses per room and a 32% (65%, 54%) increase in nurses' walking distance. In extreme cases, walking distances double, while at minimum, they are at least 25% longer. This shows that a lack of integration in solving the IPRNPA leads to severe penalization of nursing staff. During staff shortages, valuable working time is wasted as walking time instead of being utilized to care for patients. Furthermore, the higher number of nurses assigned to the same room increases disturbances of patients and the risk of infections spreading between rooms by nurses. The overall performance evaluation reveals that, on average, Method 2 exhibits a 4% (3%, 7%) decline in its total weighted objective for Variation 1 (2, 3) instances, respectively. Notably, this decline fluctuates within a range of 3% (2%, 5%) to 5% (3%, 9%). It is crucial to acknowledge that the efficacy of Method 1's improvement is highly dependent on the assigned weights. Specifically, allocating more weight to objectives (6), (8), and (9) significantly magnifies the magnitude of the observed effects and, therefore, further increases the benefits of using an integrated planning approach.

7.3.2 Conflicting objectives analysis

To examine the trade-off impacts among objectives (1)–(9), we generated nine additional scenarios. In each scenario, we independently increased the weight associated with one of the objectives by a factor of 10, while keeping all other weights constant. In order to investigate the effect of these scenarios, we used artificial instances with 60 beds and a planning horizon of two weeks. Due to the computational constraints imposed by the size of the problem instances and planning horizon, we employed the heuristic solution approach (Method 3), which allows us to observe the trade-off impacts within realistic settings while ensuring computational feasibility. Other solution methods, such as solving the MIP (Method 1), were deemed unsuitable due to their excessive runtimes. Specifically, the MIP approach would have limit the problem instances in terms of bed capacity and planning horizon, consequently negating the significance of objective (1) (see Sect. 7.3.1).

The sensitivity analysis, as detailed in Table 5, provides valuable insights into the impact of emphasizing various objectives in the optimization process. Notably, focusing on the transfer objective (1) yields a scenario where all transfers are successfully avoided, albeit with a noticeable increase in gender mixing violations. On the other hand, prioritizing the patient inconvenience objective (2) proves effective in reducing inconvenience by an impressive 85%. However, this comes at a cost, quadrupling the number of transfers, doubling gender mixing violations, raising equipment violations by about 21%, and increasing nurses' walking distances by 10%. The gender mixing objective (3) stands out as a success in avoiding gender mixing violations and simultaneously reducing transfers by 24%, showing no significant deviations in other objectives. Conversely, prioritizing the equipment violations objective (4) results in a fivefold increase in transfers and a 39% increase in gender mixing violations.

Table 5 Artificial instances: Sensitivity analyses for amplified objective weights

Amplified objective (in rows)		**Objective value**[1]								
		(1) Transfers [%]	(2) Incon-veience [%]	(3) Gender mixing [%]	(4) Equipment violation [%]	(5) Continuity of care [%]	(6) Skill [%]	(7) Workload [%]	(8) Nurses per room [%]	(9) Walking distances [%]
	Base scenario	100	100	100	100	100	100	100	100	100
(1)	Transfers	0	104	111	103	99	103	101	100	100
(2)	Inconvenience	414	15	205	121	101	106	100	105	110
(3)	Gender mixing	76	105	0	106	100	104	98	101	102
(4)	Equipment violation	498	109	139	57	101	103	99	101	101
(5)	Continuity of care	555	103	112	96	97	113	99	100	96
(6)	Skills	795	106	121	104	103	57	96	108	115
(7)	Workload	90	99	107	99	101	103	95	104	106
(8)	Nurses per room	529	114	151	96	101	108	119	93	86
(9)	Walking distances	1126	108	135	108	114	106	116	93	88

[1] Objective values are provided as proportions (expressed as percentages) of those obtained using the base scenario in which no objective weight is amplified

Focusing on the continuity of care objective (5) leads to a scenario with five and a half times more transfers, 12% additional gender mixing violations, and a 13% increase in skill violations. Stressing the skill level violations objective (6) results in eight times more transfers, a 21% increase in gender mixing constraints, and a 15% longer walking distance for nurses. Emphasizing the workload objective (7) yields a modest 5% reduction in workloads and a 10% decrease in transfers, with minimal impacts on other objectives. Focusing on the nurses per room objective (8), however, leads to only a 5% reduction in the number of nurses per room, accompanied by a fivefold increase in transfers, a 14% rise in inconvenience, a 51% increase in gender mixing violations, and a 19% increase in workload. Finally, emphasizing walking distances objective (9) increases the number of transfers by a factor of eleven, gender mixing violations by 35%, continuity of care by 14%, and nurse workloads by 16%.

Summarizing the conflicting objectives analysis, when assessing deviations exceeding 10%, two objectives stand out as the most conflicting (see Table 5 columns): the gender mixing objective (3), influenced by 7 out of 8 other amplified objectives, and the transfers objective (1), impacted by 6. Surprisingly, no deviations exceeding 10% occur in the nurses per room objective in any scenario, indicating its stability. The transfer objective is notably sensitive to changes in most of the NPA objectives ((5), (6), (8), (9)) which highlights the significant impact of the NPA on the PRA. These findings underscore the intricate trade-offs involved in solving the IPRNPA and the need for an integrated and balanced approach to meet various objectives.

8 Conclusion and outlook

Motivated by important interactions of PRA and NPA decisions in hospital wards, this paper explicitly considers both types of assignment decisions in one integrated optimization problem for the first time. We introduce the IPRNPA problem and provide a formal mathematical description as a mixed integer program. Since the PRA problem and the NPA problem are already NP-hard and very difficult to solve for realistic instance sizes, it is unsurprising that the integrated problem is computationally challenging and cannot be solved to (near) optimality in reasonable time using the completely integrated MIP model. Therefore, we present an efficient heuristic for the integrated problem that can compute high-quality solutions quickly on both artificially generated and real-world instances obtained from our partner hospital. The managerial insights from this study are threefold. First, the heuristic solution approach highlighted its superiority in running time efficiency across various problem sizes, particularly excelling in larger and more complex scenarios. While the integrated MIP struggles with high running times, the heuristic solution approach strikes a favorable balance between computational efficiency and solution quality. Its superiority in objective values is notably pronounced in medium and large-scale scenarios, making it the preferred choice for real-world settings. Second, integrating PRA and NPA is crucial for reducing nurse walking distances and minimizing the number of nurses assigned to the same room. Without this integration, operational inefficiencies lead to longer walking times, more frequent patient disturbances, and increased risk of infection spread. Third, the study's analysis of conflicting objectives reveals significant conflicts, particularly in equipment violations and transfers objectives, both influenced by other factors. This underscores the complexity of the considered integrated planning problem and the need for a balanced, integrated approach to meet various objectives while minimizing adverse effects on patient care and operational efficiency.

We also devise a parameterized instance generator for the problem. This generator is made freely available to other researchers to foster additional investigations on the IPRNPA problem, which we believe represents a challenging and practically relevant problem to be further investigated in the future. For instance, while our heuristic solution method allows easy adaptions to some dynamic versions of the problem, explicitly investigating different dynamic extensions with increasing degrees of data uncertainty (e.g., patient no-shows or unexpected changes of patients' care requirements and/or LOS after admission) might represent a fruitful direction for future research. Moreover, while the nurse roster for the planning period is considered as an input of the problem in this paper, integrating rostering decisions into the problem formulation might represent an interesting extension.

Appendix A Nurse rostering formulation

We use a simple binary integer programming formulation to generate the nurse rosters that are part of our random instances. The formulation is based on the description presented in the first International Nurse Rostering Competition (INRC) 2010 (Haspeslagh et al 2014). As the focus of this work is not on nurse rostering, we use a simple but fast formulation instead of a perfectly detailed one. This formulation includes all constraints that are relevant concerning the use of a nurse roster as an input of the IPRNPA. However, it is very easy to substitute the used nurse rostering formulation in our code.

Similar to the INRC, we determine the roster for the planning period considering one ward. We use a subset of constraints of the INRC in order to compute a simple, yet still realistic nurse roster. These constraints include that the number of required nurses per shift must be met, not more than a given maximum allowed number of shifts can be assigned to any single nurse during the planning period, and that minimum rest times for nurses between shifts are respected.

In addition to some of the notation and parameters introduced in Sect. 4, we use the following parameters and decision variables:

Parameters:

skill_nurses(s, l)	number of nurses with at least skill level $l \in \mathcal{L}$ required during shift $s \in \mathcal{S}$. The sum of this number over all skill levels defines the minimum number of nurses needed per shift in total.
max_shifts	maximum allowed number of shifts per nurse within the time horizon.

Decision variables:

$\text{assign}_{n,s}$ binary variable indicating whether nurse $n \in \mathcal{N}$ is assigned to shift $s \in \mathcal{S}$

The focus of our formulation is on the generation of a feasible roster to be used as an input for the IPRNPA. Therefore, we consider the minimization of the total number of assigned nurses as our objective to prevent unnecessary assignments that are not required to fulfill the considered constraints:

Assignment objective

(29) Minimization of the number of assigned nurses:

$$\min \sum_{s \in \mathcal{S}, n \in \mathcal{N}} \text{assign}_{n,s}$$

The following constraints must be met by a feasible nurse roster:

(30) Each nurse can work at most one shift per day:

$$\text{assign}_{n,s} + \text{assign}_{n,s+1} + \text{assign}_{n,s+2} \leq 1 \quad \forall n \in \mathcal{N}, s \in \mathcal{S}^{\text{early}}$$

(31) For each skill level $l \in \mathcal{L}$, at least skill_nurses(s, l) nurses with at least skill level l must be assigned during each shift $s \in \mathcal{S}$:

$$\sum_{n \in \mathcal{N}:\ \text{skill\_level}\,(n) \geq l} \text{assign}_{n,s} \geq \text{skill\_nurses}(l, s) \quad \forall l \in \mathcal{L}, s \in \mathcal{S}$$

(32) The minimum total number of nurses must be assigned during each shift $s \in \mathcal{S}$:

$$\sum_{n \in \mathcal{N}} \text{assign}_{n,s} \geq \sum_{l \in \mathcal{L}} \text{skill\_nurses}(l, s) \quad \forall s \in \mathcal{S}$$

(33) No nurse $n \in \mathcal{N}$ can be assigned to more than max_shifts many shifts during the planning period:

$$\sum_{s \in \mathcal{S}} \text{assign}_{n,s} \leq \text{max\_shifts} \quad \forall n \in \mathcal{N}$$

(34) On the day after a night shift, a nurse can only have another night shift (or the day off):

$$\text{assign}_{n,s} + \text{assign}_{n,s+1} + \text{assign}_{n,s+2} \leq 1 \quad \forall n \in \mathcal{N}, s \in \mathcal{S}^{\text{night}}$$

(35) On the day after a late shift, a nurse cannot have a morning shift:

$$\text{assign}_{n,s} + \text{assign}_{n,s+2} \leq 1 \quad \forall n \in \mathcal{N}, s \in \mathcal{S}^{\text{late}}$$

Appendix B MIP gap evaluation on artificial and real-world instances

Table 6 Artificial instances: Average MIP gap

Scenario	Variation	Planning horizon	Method 1	Method 2	
			MIP gap[1]	MIP gap[1]	
			IPRNPA [%]	PRA [%]	NPA [%]
30 beds	Var. 1 (double rooms)	2 weeks	69.9	58.2	6.2
		4 weeks	78.6	63.4	9.5
	Var. 2 (triple rooms)	2 weeks	57.3	82.6	4.9
		4 weeks	77.6	79.6	5.8
	Var. 3 (mixed rooms)	2 weeks	64.9	54.9	5.1
		4 weeks	77.7	68.9	7.9
60 beds	Var. 1 (double rooms)	2 weeks	122.6	75.2	61.6
		4 weeks	125.6	89.2	65.4
	Var. 2 (triple rooms)	2 weeks	104.1	94.6	51.5
		4 weeks	128.3	97.0	59.9
	Var. 3 (mixed rooms)	2 weeks	122.6	74.3	56.8
		4 weeks	126.3	90.5	59.7

[1] The MIP gap is evaluated after either a running time limit of 10,800 seconds or 5,400 seconds for each subproblem or when the MIP gap reaches a value less than 5%. The MIP gap is shown summarized as an average value

Table 7 Real-World instances: MIP gap

Instance #	Method 1	Method 2	
	MIP gap[1]	MIP gap[1]	
	IPRNPA [%]	PRA [%]	NPA [%]
1	55	0	5
2	58	0	5
3	43	0	5
4	48	0	5
5	53	0	5
6	46	0	5
7	49	0	5
8	48	0	5
9	36	0	5
10	36	0	5
11	43	0	5
12	50	0	5
13	34	0	5
14	25	0	5
15	47	0	5
16	51	0	5
17	43	0	5
18	38	0	5
19	29	0	5
20	36	100	5
Total[2]	**43**	**5**	**5**

1 The MIP gap is evaluated after either a running time limit of 10,800 seconds or 5,400 seconds for each subproblem or when the MIP gap reaches a value less than 5%.

[2] Average across all instances

Author Contributions Tabea Brandt: Conceptualization, Methodology, Software, Writing - Original Draft. Tom Lorenz Klein: Software, Validation. Melanie Reuter-Oppermann: Conceptualization, Writing - Original Draft, Funding acquisition. Fabian Schäfer: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Project administration. Clemens Thielen: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. Maartje van de Vrugt: Conceptualization, Writing - Original Draft, Resources. Joe Viana: Conceptualization, Writing - Review & Editing.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 443158418. Joe Viana's work is primarily supported by the Norwegian Research Council, Measure for Improved Availability of medicines and vaccines – Project number 300867.

Data availibility The instance generator, instances, and solutions used for the computational experiments are available on GitHub at https://doi.org/10.5281/zenodo.12750420 along with the solution checker found at https://doi.org/10.5281/zenodo.12750359.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acar I, Butt SE (2016) Modeling nurse-patient assignments considering patient acuity and travel distance metrics. J Biomed Inf 64:192–206

Aiken LH, Clarke SP, Sloane DM et al (2002) Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 288(16):1987–1993

Bai J, Fügener A, Schoenfelder J et al (2018) Operations research in intensive care unit management: a literature review. Health Care Manage Sci 21(1):1–24

Bastos LS, Marchesi JF, Hamacher S et al (2019) A mixed integer programming approach to the patient admission scheduling problem. Eur J Operat Res 273:831–840

Benazzouz T, Echchatbi A, Bellabdaoui A (2015) A literature review on the nurses' planning problems. Int J Math Comput Sci 1(5):268–274

Bilgin B, Demeester P, Misir M et al (2012) One hyper-heuristic approach to two timetabling problems in health care. J Heuristics 18:401–434

Bodenheimer TS, Smith MD (2013) Primary care: proposed solutions to the physician shortage without training more physicians. Health Aff 32(11):1881–1886

Borchani R, Masmoudi M, Jarboui B, Siarry P (2021) Heuristics-based on the Hungarian method for the patient admission scheduling problem. In: Masmoudi M, Jarboui B, Siarry P (eds) Operations research and simulation in healthcare. Springer, Cham, pp 33–62

Bouras A, Masmoudi M, Saadani NEH et al (2021) Multi-stage appointment scheduling for outpatient chemotherapy unit: a case study. RAIRO Operat Res 55(2):589–610

Butt SE, Fredericks TK, Kumar AR, et al (2004) An evaluation of physiological work demands on registered nurses over a 12-hour shift. In: Proceedings of the XVIII Annual International Occupational Ergonomics and Safety Conference (ISOES)

Ceschia S, Schaerf A (2011) Local search and lower bounds for the patient admission scheduling problem. Comput Operat Res 38(10):1452–1463

Ceschia S, Schaerf A (2012) Modeling and solving the dynamic patient admission scheduling problem under uncertainty. Artif Intell Med 56(3):199–205

Ceschia S, Schaerf A (2016) Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays. J Sched 19:377–389

Clark A, Moule P, Topping A et al (2015) Rescheduling nursing shifts: Scoping the challenge and examining the potential of mathematical model based tools. J Nurs Manage 23(4):411–420

Cohen B, Hyman S, Rosenberg L et al (2012) Frequency of patient contact with health care personnel and visitors: implications for infection prevention. Jt Comm J Qual Patient Saf 38(12):560–565

Dancer SJ (2009) The role of environmental cleaning in the control of hospital-acquired infection. J Hosp Inf 73(4):378–385

Demeester P, Souffriau W, De Causmaecker P et al (2010) A hybrid tabu search algorithm for automatically assigning patients to beds. Artif Intell Med 48(1):61–70

Drupsteen J, van der Vaart T, van Donk DP (2013) Integrative practices in hospitals and their impact on patient flow. Int J Operat & Prod Manage 33(7):912–933

Erhard M, Schoenfelder J, Fügener A et al (2018) State of the art in physician scheduling. Eur J Operat Res 265(1):1–18

Eveillard M, Hitoto H, Raymond F et al (2009) Measurement and interpretation of hand hygiene compliance rates: importance of monitoring entire care episodes. J Hosp Inf 72(3):211–217
Ficker AMC, Spieksma FCR, Woeginger GJ (2021) The transportation problem with conflicts. Ann Operat Res 298(1):207–227
Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: a survey. Health Care Manage Sci 14(1):89–114
Guido R, Groccia MC, Conforti D (2018) An efficient matheuristic for offline patient-to-bed assignment problems. Eur J Operat Res 268(2):486–503
Halwani M, Solaymani-Dodaran M, Grundmann H et al (2006) Cross-transmission of nosocomial pathogens in an adult intensive care unit: incidence and risk factors. J Hosp Inf 63(1):39–46
Haspeslagh S, De Causmaecker P, Schaerf A et al (2014) The first international nurse rostering competition 2010. Ann Operat Res 218:221–236
He L, Madathil SC, Oberoi A et al (2019) A systematic review of research design and modeling techniques in inpatient bed management. Comput & Ind Eng 127:451–466
Hesaraki AF, Dellaert NP, de Kok T (2019) Generating outpatient chemotherapy appointment templates with balanced flowtime and makespan. Eur J Operat Res 275:304–318
Hesaraki AF, Dellaert NP, de Kok T (2020) Integrating nurse assignment in outpatient chemotherapy appointment scheduling. OR Spectrum 42:935–963
Heshmat M, Nakata K, Eltawil A (2018) Solving the patient appointment scheduling problem in outpatient chemotherapy clinics using clustering and mathematical programming. Comput & Ind Eng 124:347–358
Hulshof PJH, Kortbeek N, Boucherie RJ et al (2012) Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. Health Systems 1(2):129–175
Jha RK, Sahay BS, Charan P (2016) Healthcare operations management: a structured literature review. Decision 43(3):259–279
Jun JB, Jacobson SH, Swisher JR (1999) Application of discrete-event simulation in health care clinics: a survey. J Operat Rese Soc 50(2):109–123
Ku WY, Pinheiro T, Beck JC (2014) CIP and MIQP models for the load balancing nurse-to-patient assignment problem. In: Proceedings of the 20th International Conference on Principles and Practice of Constraint Programming (CP), pp 424–439
Liang B, Turkcan A (2016) Acuity-based nurse assignment and patient scheduling in oncology clinics. Health Care Manag Sci 19:207–226
Lusby R, Schwierz M, Range T et al (2016) An adaptive large neighborhood search procedure applied to the dynamic patient admission scheduling problem. Artif Intell Med 74:21–31
Marzouk M, Kamoun H (2020) Nurse to patient assignment through an analogy with the bin packing problem: case of a Tunisian hospital. J Operat Res Soc 72(8):1808–1821
Masmoudi M, Jarboui B, Siarry P (eds) (2021) Operations research and simulation in healthcare. Springer, Cham
Mullinax C, Lawley M (2002) Assigning patients to nurses in neonatal intensive care. J Operat Res Soc 53(1):25–35
Pesant G (2016) Balancing nursing workload by constraint programming. In: Proceedings of the 13th International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR), pp 294–302
Punnakitikashem P, Rosenberger JM, Buckley-Behan DF (2008) Stochastic programming for nurse assignment. Comput Optim Appl 40:321–349
Punnakitikashem P, Rosenberger JM, Buckley-Behan DF (2013) A stochastic programming approach for integrated nurse staffing and assignment. IEEE Trans 45(10):1059–1076
Rachuba S, Reuter-Oppermann M, Thielen C (2024) Integrated planning in hospitals: a review. OR Spectrum (online first), pp 1–54. https://doi.org/10.1007/s00291-024-00797-5
Rais A, Viana A (2011) Operations research in healthcare: a survey. Int Trans Oper Res 18(1):1–31
Range T, Lusby R, Larsen J (2014) A column generation approach for solving the patient admission scheduling problem. Eur J Oper Res 235(1):252–264
Schäfer F, Walther M, Hübner A (2017) Patient-bed allocation in large hospitals. In: Proceedings of the 3rd International Conference on Health Care Systems Engineering (HCSE), pp 299–300
Schäfer F, Walther M, Hübner A et al (2019) Operational patient-bed assignment problem in large hospital settings including overflow and uncertainty management. Flex Serv Manuf J 31:1012–1041
Schäfer F, Walther M, Grimm DG et al (2023) Combining machine learning and optimization for the operational patient-bed assignment problem. Health Care Manag Sci 26(4):785–806

Schaus P, Régin JC (2014) Bound-consistent spread constraint - application to load balancing in nurse-to-patient assignments. Eur J Comput Optimiz 2(3):123–146
Schmidt R, Geisler S, Spreckelsen C (2013) Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. BMC Medical Informatics and Decision Making 13(3)
Sir MY, Dundar B, Barker Steege LM et al (2015) Nurse-patient assignment models considering patient acuity metrics and nurses' perceived workload. J Biomed Inform 55:237–248
Sundaramoorthi D, Chen VCP, Rosenberger JM et al (2009) A data-integrated simulation model to evaluate nurse-patient assignments. Health Care Manag Sci 12:252–268
Thielen C (2018) Duty rostering for physicians at a department of orthopedics and trauma surgery. Operat Res Health Care 19:80–91
Thomas BG, Bollapragada S, Akbay K et al (2013) Automated bed assignments in a complex and dynamic hospital environment. Interfaces 43:435–448
Thuran A, Bilgen B (2017) Mixed integer programming based heuristics for the patient admission scheduling problem. Comput & Operat Res 80:38–49
Van Riet C, Demeulemeester E (2015) Trade-offs in operating room planning for electives and emergencies: a review. Operat Res Health Care 7:52–69
Vanberkel PT, Boucherie RJ, Hans EW et al (2010) A survey of health care models that encompass multiple departments. Int J Health Manage Inf 1(1):37–69
Vancroonenburg W, Croce FD, Goossens D et al (2014) The Red-Blue transportation problem. Eur J Oper Res 237(3):814–823
Vancroonenburg W, De Causmaecker P, Vanden Berghe G (2016) A study of decision support models for online patient-to-room assignment planning. Ann Oper Res 239:253–271
Zhu YH, Toffolo TAM, Vancroonenburg W et al (2019) Compatibility of short and long term objectives for dynamic patient admission scheduling. Comput & Operat Res 104:98–112

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Tabea Brandt[1] · Tom Lorenz Klein[2] · Melanie Reuter-Oppermann[4] · Fabian Schäfer[5] · Clemens Thielen[2,6] · Maartje van de Vrugt[3,7] · Joe Viana[8,9,10]

✉ Fabian Schäfer
fab.schaefer@tum.de

Tabea Brandt
brandt@combi.rwth-aachen.de

Tom Lorenz Klein
tom.klein@tum.de

Melanie Reuter-Oppermann
melanie.reuter-oppermann@dgre.org

Clemens Thielen
clemens.thielen@tum.de

Maartje van de Vrugt
n.vandevrugt@amsterdamumc.nl

Joe Viana
Joe.Viana@ntnu.no

[1] Combinatorial Optimization, RWTH Aachen University, Ahornstr. 55, 52074 Aachen, Germany

[2] TUM Campus Straubing for Biotechnology and Sustainability, Weihenstephan-Triesdorf University of Applied Sciences, Am Essigberg 3, 94315 Straubing, Germany

[3] Center for Healthcare Operations Improvement and Research, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

[4] Department of Health, Care and Public Health Research Institute (CAPHRI), Maastricht University, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands

[5] Chair of Supply and Value Chain Management, Technical University of Munich, Am Essigberg~3, 94315 Straubing, Germany

[6] Department of Mathematics, School of Computation, Information and Technology, Technical University of Munich, Boltzmannstr. 3, 85748 Garching bei München, Germany

[7] Department of Care Support, Strategy and Innovation, Amsterdam University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

[8] Department of Industrial Economics and Technology Management, Faculty of Economics and Management, Norwegian University of Science and Technology, Alfred Getz vei 3, N-7034 Trondheim, Norway

[9] Center for Service Innovation, St. Olav's Hospital, Prinsesse Kristinas gate 3, N-7030 Trondheim, Norway

[10] Department of Accounting and Operations Management, BI Norwegian Business School, Oslo, Norway