Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation

Journal Pre-proof

Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation

Roberto Aringhieri, Davide Duma, Paolo Landa, Simona Mancini

PII: S0377-2217(21)00628-7
DOI: https://doi.org/10.1016/j.ejor.2021.07.033
Reference: EOR 17362

To appear in: *European Journal of Operational Research*

Received date: 26 February 2020
Accepted date: 14 July 2021

Please cite this article as: Roberto Aringhieri, Davide Duma, Paolo Landa, Simona Mancini, Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation, *European Journal of Operational Research* (2021), doi: https://doi.org/10.1016/j.ejor.2021.07.033

Highlights

- Considering stakeholders' preferences through different performance criteria.
- Patient priority maximisation and workload balance as performance criteria.
- Understanding the solution quality combining the two criteria in the OR planning.
- OR planning defined by a hierarchical multi-objective optimisation model.
- Matheuristic Solution framework based on multi-neighbourhood local search

Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation

Roberto Aringhieri[a,*], Davide Duma[a,b], Paolo Landa[c], Simona Mancini[d,e]

[a] *Dipartimento di Informatica, Università degli Studi di Torino*
Corso Svizzera 185, 10149 Torino, Italy
[b] *Dipartimento di Matematica "Felice Casorati", Università degli Studi di Pavia*
Via Adolfo Ferrata 5, 27100 Pavia, Italy
[c] *Département d'Opérations et Systèmes de Décision, Université Laval*
Pavillon Palasis-Prince, 2325, rue de la Terrasse, Québec(QC) G1V 0A6, Canada
[d] *Dipartimento di Matematica ed Informatica, Università degli Studi di Cagliari*
Palazzo delle Scienze, Via Ospedale 72, 09124 Cagliari, Italy
[e] *Department of Operations, Energy, and Environmental Management*
University of Klagenfurt, Universität Straße 65-67, 9020 Klagenfurt, Austria

Abstract

Previous analysis suggested the opportunity to consider the preferences of different stakeholders (hospital, patients, doctors and nurses) through the adoption of both patient priority maximisation and workload balance as performance criteria. The aim of this paper is to develop an effective and efficient solution approach for the operating room planning and scheduling capable to take into account the patient priority maximisation and workload balance criteria at the same time. This work is inspired by the need of a deeper understanding of the quality of the solutions obtained when a combination of the two criteria leads the OR planning decisions. Starting from a hierarchical multi-objective optimisation model for the combined master surgical scheduling and surgical cases assignment problems, we develop a class of new multi-neighbourhood local search based matheuristic algorithms, whose main feature is to exploit an ad hoc neighbourhood to generate better solutions in a significant shorter running time. A broad quantitative analysis on new realistic instances proves the effectiveness and the efficiency of the proposed matheuristic algorithms as well as to evaluate the quality of the computed solution from an operating room management perspective.

Keywords: OR in health services, Operating room planning and scheduling, Matheuristics

*Corresponding author
Email addresses: roberto.aringhieri@unito.it (Roberto Aringhieri), davide.duma@unipv.it (Davide Duma), paolo.landa@fsa.ulaval.ca (Paolo Landa), Simona.Mancini@aau.at (Simona Mancini)

1. Introduction

The ageing of the population and the increase of need of care, given by the presence of chronic diseases and multi-morbidity in elder people, is rising the demand of health services within public and private funded national health care systems (Organisation for Economic Co-operation and Development, 2017). The increase of the demand implies the modification of the supply in terms of services provided and resources available in primary and secondary care. The World Health Organisation, during an assembly in 2015 highlighted the importance of the role of surgery care in order to achieve health care universally covered (World Health Organization, 2017).

Surgical care is one the main activity where health care providers (i.e. hospital managers) have to face challenges in terms of resource use, costs and revenues. Operating Room (OR) is the core of this activity since it can account about the 60% of the hospital admissions (Peltokorpi, 2011). OR represents one of the most expensive units in hospital budgets, where costs account about the 40% of the overall hospital costs (Denton et al., 2007). However, this activity can provide to hospital the largest income in a broad sense, making it the most important strategic area. The issues arising from OR management are mostly related to planning and scheduling activities, where combinations of resources available and patients waiting for the surgery, can provide different results with a large variation of costs and revenues.

Cardoen et al. (2010) reported different performance criteria adopted in the literature to lead and to evaluate the OR planning and scheduling decisions, which are usually defined at strategical, tactical and operational levels. Therefore different performance criteria can reflect different stakeholders' preferences (Marques & Captivo, 2017). Addressing the above mentioned preferences is challenging, as there could be conflicts amongst them (Cappanera et al., 2018).

Taking into account a patient–centred viewpoint, a preliminary comparison between two criteria – patient priority maximisation and workload balance – has been reported in Aringhieri & Duma (2017) over a set of realistic instances with a scheduling horizon of one week. Such a comparison confirmed the ability of both criteria to ensure a high level of OR utilisation dealing with long waiting lists, which is a common situation in many hospitals belonging to publicly funded health care systems. The two criteria provided different results. The patient priority maximisation is a fairness criterion among patients that allowed us to have an *ex-ante* OR utilisation close to 100% in all cases. This proves that patient priority maximisation is an excellent proxy of the OR utilisation maximisation. Conversely, the workload balance is a criterion to have a smooth workload along the week at the price of having a lower OR utilisation with respect to the patient priority.

From this perspective, a counter-intuitive result is the non-compliance between the OR utilisation and the number of the planned patients, which is higher with respect to the patient priority. A further preliminary comparison has been discussed in Aringhieri et al. (2018b) in which the two above criteria have been compared using real data on a longer horizon (one year) in order to evaluate

the impact on the OR scheduling decisions over time. From the analysis, the use of workload balance seems to be a good compromise between the need of minimising the waiting time of the patients and taking into account the quality of the doctors and nurses work in the ward.

Such a preliminary analysis suggested the opportunity to consider the more relevant preferences of different stakeholders (that is hospital, patients, doctors and nurses) at the joint tactical and operational level through the adoption of both patient priority maximisation and workload balance as performance criteria. The aim of this paper is to develop an effective and efficient solution approach for the OR planning and scheduling capable to take into account the patient priority maximisation and workload balance criteria at the same time in order to investigate and evaluate their impact on the resulting solutions on a more complex operative context.

We place our study at the joint tactical and operational level under a patient–centred viewpoint in accordance with the main insights reported in Aringhieri et al. (2015a) assuming deterministic surgery times and lengths of stay. We propose a modified version of the hierarchical multi-objective optimisation model (reported in Aringhieri et al. (2017)) for the combined master surgical scheduling and surgical cases assignment problems in order to adopt the more realistic patient priority model introduced in Valente et al. (2009) and deeply tested at S. Martino University Hospital of Genoa, Italy. In accordance with this priority model, we generate 80 new instances – more complex in terms of operative context – exploiting the work done by Leeftink & Hans (2018) to have instances with a realistic case mix. We develop a class of new algorithms for solving the modified hierarchical optimisation model in an efficient way since the preliminary tests in Aringhieri et al. (2017) showed the limitation of a general purpose solver. The proposed algorithm framework is a multi-neighbourhood local search based matheuristic in which several large neighbourhoods are sequentially addressed by means of integer programming models capable to exhaustively explore large neighbourhoods in small computational times: the main novelty of our approach is the use of an *ad hoc* neighbourhood to better balance the workload. A broad quantitative analysis proves the effectiveness and the efficiency of the proposed class of new algorithms.

The paper is organised as follows. A literature review is reported in Section 2 in order to place our work in the literature context. After describing the problem statement, we outline the modified version of the hierarchical multi-objective optimisation model in Section 3. Section 4 is devoted to the description of the proposed solution approach consisting in a multi-neighbourhood local search based matheuristic. The quality of the solutions computed by our algorithms is discussed in Section 5 in which a comprehensive analysis is presented: after describing the computational environment (5.1), we discuss the effectiveness and the efficiency of the proposed algorithms addressing also the analysis of the Pareto optimal frontier of the solutions (5.2), and then we analyse the computed solutions from an operating room management point of view (5.3). Finally, Section 6 closes the paper.

2. Literature review

Operational researchers provided in the last three decades several decision models and optimisation tools aimed at addressing the complexity of the OR management. In literature there are several reviews that systematically analysed the operational research approaches to OR planning and scheduling problems (Guerriero & Guido, 2011; Van Riet & Demeulemeester, 2015; Samudra et al., 2016; Hof et al., 2017; Zhu et al., 2019): such a large production of academic papers during the last decades proves the increasing interest in the OR management by the operational research community, and the increasing need by the hospital management of solutions and tools based on advanced techniques to face the high complexity of these problems.

OR planning and scheduling are defined by three main decision levels: strategical, tactical and operational that are respectively based on long, medium and short-term objectives (Hulsof et al., 2012). Apart from recent integrated approaches (see, e.g. Siqueira et al. (2018)), usually these levels are studied or analysed separately, given by the complexity of the problem and the time horizon (e.g. the larger is the horizon, the harder is the problem). The strategic level faces the problem of resource allocation, assigning to each surgery the resources available (e.g. staff), while the tactical level faces the problem of OR block assignment to surgical specialties in accordance with the master surgical schedule. The operational level, usually called surgery process scheduling, faces two specific problems: a) the selection of the patients waiting for surgery from a waiting list, assigning them to a specific OR session (an OR block in a specific day), and b) the definition of the sequence of the surgical procedures and the resource allocation in each OR session. Such problems can be further challenged by the inherent stochasticity of their main parameters, such as the surgery duration, the length of stay, the arrival of non-elective patients, and often combined with staffing and scheduling decisions (Beaulieu et al., 2012; Duma & Aringhieri, 2015, 2018; Wang et al., 2018a; Breuer et al., 2020).

The availability and the management of resources is one of the key factors in the development of the planning and scheduling strategies. In literature the main resources taken into account are OR session, ward, intensive care unit and post-anaesthesia care unit beds. The OR session becomes a critical resource when it can be shared with non-elective patients (Duma & Aringhieri, 2019). The availability of beds is one of the main bottlenecks in the scheduling process, especially when beds are shared between elective and emergency patients (Landa et al., 2018). The bed availability can affect the performance and the overall activity of the operating theatre. Ward beds are a resource which elective and non-elective patients compete for: an unpredictable flow of patients from the emergency department can stop or limit temporarily the OR activities of elective admissions given by the shortage of beds, or the unpredictability of ward length of stay in wards can affect OR schedules (Neyshabouri & Berg, 2017). In the literature, beds are usually considered as a constraint affecting the main objective (e.g. maximisation of patients treated), as bed is a limited resource available, and it is a fixed cost that is difficult to variate in its quantity

in the short-medium term (see, e.g. Aringhieri et al. (2015a)). The impact of the managed resources in the OR planning and scheduling can be seen in the long period (see, e.g. Siqueira et al. (2018)), as the impact of workload balance is verified in the overall activity and within long periods as it consider the seasonality and the variation of service demand given by exogenous factors. More generally, the problem of patient scheduling considering downstream resources is receiving increasing attention (see, e.g., van den Broek d'Obrenan et al. (2020); Schneider et al. (2020); Zhu et al. (2020).

In order to lead and to evaluate the OR planning and scheduling decisions, different performance criteria have been reported (Cardoen et al., 2010). The maximisation of OR utilisation rate is aimed at reducing hospital marginal and fixed costs on surgeries, as it is a cost also the OR idle time, while the maximisation of patient priority is aimed at reducing the patient waiting time in waiting lists, giving a higher priority to urgent patients with specific clinical needs. Considering the limited resources availability and the daily variation of patient access to hospital ward, the workload balance is aimed at improving the work quality provided by doctors and nurses, reducing peaks of workload and the psycho and physical stress. Usually, patient priority maximisation (Dios et al., 2015) and OR utilisation maximisation (Hans et al., 2008) are the most used, but also the minimisation of delays and cancellations (Landa et al., 2016), the maximisation of patient satisfaction (Min & Yih, 2010) and the minimisation of hospital fixed costs or societal costs (Tànfani & Testi, 2010) were considered as objective function for OR planning and scheduling. On the contrary, very few studies consider the workload balance criterion to improve the quality of care provided to the patients (see, e.g. van Oostrum et al. (2008); Cappanera et al. (2014); Aringhieri et al. (2015b); van den Broek d'Obrenan et al. (2020)). The workload balance criteria is usually assessed through the levelling of stay bed occupancy (bed levelling) (Beliën & Demeulemeester, 2008).

The importance of staff workload (clinicians and nurses) is reported in the literature of the last ten years and it affected national guidelines. Nurse guidelines were renewed in US, Australia, Sweden and other countries to include boundaries for the staff workload to reduce patient mortality following the indications provided by recent studies on workload (Aiken et al., 2008, 2014). As reported by Aiken et al. the odds on patients dying in hospitals with an average workload of 8 patients per nurse is 1.26 times greater than in hospitals with mean workloads of 4 patients per nurse. These results are confirmed in other recent studies (Lasater et al., 2020, 2021). In addition the benefits of workload balance has a direct impact on patient outcomes and on return on investment derived by preventing additional LOS and readmissions (McHugh et al., 2021).

Considering multiple objectives in bed levelling contest, the development of the master surgical schedule (MSS) presents several objectives to manage (usually in contrast), such as smoothing bed occupancy distribution with resources capacity and demand (Beliën et al., 2006) or minimising the expected total bed shortage under demand and capacity constraints (Beliën et al., 2007). MSS development can present multiple objectives together with bed levelling such as the allocation of ORs exclusively to group surgeons belonging to the same

speciality and defining a MSS simple and repetitive (Beliën et al., 2009), or considering a large set of objectives based on surgical case sequencing for patient characteristic (Cardoen et al., 2009).

More recently, a number of new contributions are characterised by the use of optimisation models with hierarchical or a multi-objective functions but in a different operative context with respect to that used in this paper. The impact of earliness and tardiness costs of surgeries on OR planning and scheduling was explored by Otten et al. (2019). The authors analyse the earliness/tardiness problem applied to OR planning, providing an optimisation model based on multiple objectives considering both hospital and patient perspectives. The minimisation of the makespan was faced by Behmanesh & Zandieh (2019). The authors proposed a bi-objective surgical case scheduling problem under uncertain service time, aimed at minimising the number of unscheduled surgical cases and the makespan. The analysis of OR schedules considering the impact of overlapping surgeries policy on OR schedules was analysed by Freeman et al. (2019). The optimisation model is aimed at minimising cost of the idle time of the surgeon and maximising the total OR time to increase the income, evaluating when the overlapping procedure policy provides better results with respect to the traditional policy. A different setting of objective functions is reported by Roshanaei et al. (2019), where the objective function is aimed at firstly balancing distributed OR scheduling to imbalance ORs among collaborating hospitals in terms of number of allocated ORs, and daily micro imbalance among open ORs in each hospital in terms of the total caseload assigned. Kamran et al. (2019) propose stochastic mixed integer programming model for OR planning and scheduling problem with a modified scheduling policy. The study aims at minimising the waiting time of patients and the tardiness for both patients who are planned to have the surgery and unplanned patients, the number of rejected and cancelled patients, the number of surgeon's day within the planning horizon, and the number of OR block in overtime. Finally, McRae & Brunner (2019) focused their analysis on the impact of uncertainty on long-term strategies for OR planning. The authors proposed a methodological framework based on an mixed integer programming model for case mix planning that can be extended to include the stochasticity of several parameters. More generally, we would remark that the objective functions are usually related to cost minimisation, maximisation of number of patients treated, maximisation of OR use, minimisation of overtime and idle time, optimal resource allocation. Further, the solution approaches are mainly based on solving the corresponding model with a general purpose solver and/or with a metaheuristic algorithm.

In the context of classical metaheuristics, one of the most critical issues is the efficient exploration of the neighbourhoods in order to guarantee the overall algorithm efficiency. This issue is more evident when adopting the Large Neighbourhood Search framework. In this context, a widely used approach to limit the running time is performing an implicit neighbourhood exploration by a sequence of destroy and repair operators (Pisinger & Ropke, 2010). The destroy operators may be defined in different ways. In routing problems, for example, a destroy operator could consist in breaking down k routes and leaving the

others unvaried. The repair operator builds up a feasible solution starting from the partially destroyed one. Generally, a greedy construction heuristic is used to implement the repair operator leading to solutions with a possible loss of quality.

Under the term *Matheuristics* are grouped all methods in which heuristic or metaheuristic techniques are hybridised with exact methods (Maniezzo et al., 2010). Between this broad family, it can be identified a specific subset of methods in which a (mixed) integer programming (MIP or IP) model is exploited to analyse large neighbourhoods. Differently from repair heuristics, the usage of MIP models allow to explore the whole large neighbourhood exhaustively, i.e. to find the best way to reconstruct the destroyed solution. This could yield to a much faster convergence toward high quality solutions.

Given a MIP (or IP) problem formulation, a neighbourhood generated by a destroy operator can be described by adding constraints that fix the value of variables, not involved in the destruction, equal to the value they assumed in the starting solution, while letting the other variables free to assume any value in their domain. The resulting over-constrained version of the model is used to identify the optimal value for the remaining variables. The solution of this model plays the role of the repair operator. This approach has been proved to be very effective on nurse rostering (Della Croce & Salassa, 2014), jobs scheduling (Della Croce et al., 2014), rich vehicle routing problems (Mancini, 2016, 2017b), and tourist cruises itinerary planning (Mancini & Stecca, 2018a).

Local Search matheuristics are a framework which exploit the mathematical model to efficiently explore large neighbourhoods, by fixing some of the variables and letting the solver optimizing the resulting over constrained model with a short time limit. The selection of the variable to be fixed is performed starting from a current solution and applying a destroy operator that implicitly defines the neighbourhood to be analysed. Most of the application of from the literature which are based on a single neighbourhood (see e.g. Mancini (2017a); Mancini & Stecca (2018b); Marques et al. (2020); Gansterer et al. (2021); Mancini et al. (2021) or on an LNS framework in which, at each iteration is drawn which neighbourhood to apply among a set of available neighbourhoods (Mancini, 2016). Although at the heuristic level the multi-neighbourhood local search is broadly applied in the literature, at the matheuristic level this is the first work in which it is proposed and applied to OR planning and scheduling.

As reported in the introduction, our preliminary analysis suggested to adopt patient priority maximisation and workload balance performance criteria since they are capable to address the more relevant stakeholders' preferences at the joint tactical and operational level when considering a patient–centred viewpoint. The literature review reported in this section indicates that the problem of considering multiple and different performance criteria in the OR planning and scheduling deserve to be studied. Although its relevance in the clinical practice, the workload balance is the less studied criteria, probably because fairness problems are usually more complex to model and solve (Nicosia et al., 2017). For this reasons, the aim of our paper is to develop an effective and efficient solution approach for the OR planning and scheduling driven by the two above

performance criteria. The idea of using a hierarchical objective function represents an attempt to lead the planning and scheduling decisions considering different stakeholders' preferences. Further, the use of a hierarchical objective function seems perfectly suited when the objectives have very different units of measurement as in our situation as opposed to a "weighted sum" approach, which works better when the objectives that can be easily reduced to the same unit of measurement (e.g. costs).

3. Problem Statement and Mathematical Models

Different performance criteria reflect different stakeholders' preferences: OR utilisation, patient priority, workload balance are the preferences of hospitals, patients, and doctors and nurses, respectively. Our preliminary analysis (Aringhieri & Duma, 2017; Aringhieri et al., 2018b) highlighted similar and different behaviours of the patient priority maximisation and the workload balance criteria. In addition, it shows that patient priority maximisation is an excellent proxy of the OR utilisation maximisation.

The main goal of our research is to address different stakeholders' preferences (hospital, patients, doctors and nurses) by leading the optimisation process using both patient priority maximisation and workload balance performance criteria. In this section we introduce the OR planning and scheduling problem that we consider in the remainder of the paper.

One of the main insight from our previous work (Aringhieri et al., 2015a) is the importance of exploiting the inherent hierarchy between tactical and operational levels in order to achieve better overall solutions and developing more efficient solution algorithms. Starting from this insight, the problem addressed in this paper consists in simultaneously (i) assigning the OR sessions to a set of surgical specialties and (ii) scheduling the patient surgeries (i.e., selecting the patients from the specialty's waiting list assigning them to the OR session of the same specialty). The objective of this problem is to optimise both patient priority and workload balance criteria by leading the solution process by a hierarchical objective function.

The hierarchical objective function is composed of the two following objective functions. The patient priority maximisation is defined as the sum of the scores of the patients selected for their surgery within the planning horizon in such a way that the greater is the score, the higher is the priority. The score of each patient is computed as *need adjusted waiting days*, that is his/her priority level multiplied by the waiting time between the diagnosis and the surgery, as introduced in Valente et al. (2009). Such an approach avoids the case in which patients with high priority level are always operated on before those with low priority level, which could be possible in the publicly funded health care systems characterised by long waiting lists. This case was numerically discussed in Aringhieri et al. (2015a). Priority levels are "coefficient, representing the speed at which the clinical need is assumed to increase along with the passing of time" as stated in Valente et al. (2009). The coefficient is not a constant value but it depends on the value of the maximum time before treatment, that is such

a period of time within which the admission to the hospital should be provided in accordance with the national guidelines. In accordance with the literature, the workload balance is assessed through the levelling of stay bed occupancy, that is the bed levelling. Such a levelling is implemented by maximising the number of beds occupied in a surgical specialty ward, in the day in which the occupation is minimum, which works as maximin fairness (Nicosia et al., 2017).

The main assumptions we made for our problem are the following. For each patient are known the surgical specialty to which he/she is assigned, the priority level, the expected length of stay (LOS), the number of days elapsed from the diagnosis until the day at the beginning of the planning horizon, the expected surgery duration. For each specialty is known the number of beds available on each day. Furthermore, the length of each OR session is supposed to be known. A patient is assigned to an OR session only if that session has been assigned to the surgical specialty which the patient belongs (block scheduling paradigm). The total expected duration of surgeries scheduled in an OR session can not exceed its length. Each scheduled patient occupies a bed in the day of his/her surgery and for a number of following days equal to his/her LOS. We assume that, as is the case in many publicly funded health systems, the number of patients on the waiting lists is greater than the maximum number of patients that can be operated on during the planning horizon considered. In addition, deterministic surgery times and lengths of stay are assumed.

3.1. Notation

In order to formalise the addressed problem and the proposed solution approach, we introduce the following notation. Let I, J and K be respectively the sets of patients, surgical specialties and operating rooms, each indexed by i, j and k. Let $T = \{1, \ldots, N_T\}$ be the set of days in the planning horizon, indexed by t. Let I_j be the subset of patients that belongs to the specialty $j \in J$.

For each patient $i \in I$, the expected duration of the surgery p_i expressed in minutes, the priority level π_i, and the expected LOS μ_i expressed in days from the day after the surgery, are given. Let ϕ_{it} be the number of elapsed day between diagnosis of patient i and day t. The score of the patient i at the day t is then computed as $\pi_i \phi_{it}$.

Each OR session in the planning horizon is uniquely defined by the pair of indices (k, t). We denote by s_{kt} the time capacity of the OR session (k, t). Let λ_{jt} be the number of beds available for specialty j on day t.

3.2. The hierarchical multi-objective optimisation model

In order to formulate our hierarchical multi-objective optimisation model, we introduce the following binary decision variables:

$$x_{ikt} = \begin{cases} 1 & \text{if patient } i \text{ is assigned to OR session } (k,t) \\ 0 & \text{otherwise} \end{cases}, \tag{1a}$$

$$y_{it} = \begin{cases} 1 & \text{if patient } i \text{ surgery is scheduled on day } t \\ 0 & \text{otherwise} \end{cases}, \tag{1b}$$

$$w_{it} = \begin{cases} 1 & \text{if patient } i \text{ occupies a bed on day } t \\ 0 & \text{otherwise} \end{cases}, \tag{1c}$$

$$z_{jkt} = \begin{cases} 1 & \text{if OR session } (k,t) \text{ is assigned to specialty } j \\ 0 & \text{otherwise} \end{cases}. \tag{1d}$$

Let be also O_1 and O_2 two dummy decision variables representing the primary and the secondary objective functions. The hierarchical multi-objective optimisation model can be formulated as follows.

$$\mathcal{H}_1: \quad \max\ z_{bw} = O_1 + MO_2 \tag{2a}$$

$$\text{s.t.} \quad \sum_{k\in K}\sum_{t\in T} x_{ikt} \le 1\,, \qquad i \in I \tag{2b}$$

$$\sum_{i\in I_j} x_{ikt} \le |I_j| z_{jkt}\,, \qquad j \in J, k \in K, t \in T \tag{2c}$$

$$\sum_{j\in J} z_{jkt} \le 1\,, \qquad k \in K, t \in T \tag{2d}$$

$$\sum_{i\in I} p_i x_{ikt} \le s_{kt}\,, \qquad k \in K, t \in T \tag{2e}$$

$$\sum_{k\in K} x_{ikt} = y_{it}\,, \qquad i \in I, t \in T \tag{2f}$$

$$\sum_{\tau=t}^{\min(t+\mu_i;N_T)} w_{i\tau} \ge \min(\mu_i+1;\tau(t))y_{it}\,, \qquad i \in I, t \in T \tag{2g}$$

$$\sum_{\tau=\max(t-\mu_i,1)}^{t} y_{i\tau} \ge w_{it}\,, \qquad i \in I, t \in T \tag{2h}$$

$$\sum_{i\in I_j} w_{it} \le \lambda_{jt}\,, \qquad t \in T, j \in J \tag{2i}$$

$$O_1 \le \sum_{i\in I_j} w_{it}\,, \qquad t \in T, j \in J \tag{2j}$$

$$O_2 = \sum_{i\in I}\sum_{t\in T}\sum_{k\in K} \pi_i \phi_{it} x_{ikt}\,. \tag{2k}$$

Constraint (2b) states that a patient can be assigned to at most one OR session, which implies that only a subset of patients can be selected from the

long waiting list. A patient can be assigned to an OR session only if it is assigned to the surgery specialty to which he/she belongs, as stated in constraint (2c). Constraint (2d) implies that each OR sessions must be assigned to at most one specialty. Constraint (2e) imposes that the sum of the surgery times of the patients scheduled in each OR session (k, t) may not exceed the time capacity s_{kt}. Constraint (2f) allows to detect whether the surgery of the patient i is scheduled on day t. Setting $\tau(t) = N_T - t + 1$, constraints (2g) and (2h) imply that, if a patient i is scheduled on day t, he/she will occupy a bed for the next μ_i days as depicted in Figure 1 (with $\mu_i = 2$), which also illustrates the hierarchy among the decision variables x_{ikt}, y_{it} and w_{it}. For each specialty,

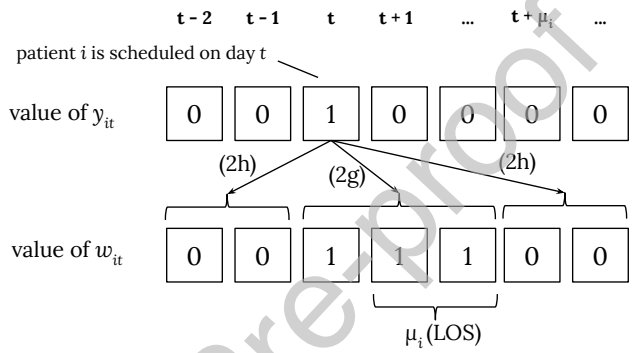


Figure 1: Hierarchy among the decision variables x_{ikt}, y_{it} and w_{it}: variables x_{ikt} determines the values of y_{it} through constraints (2f); in turn, the values of variable y_{it} determines the values of the variables w_{it} (which measure the occupancy of stay beds) through the constraints (2g) and (2h).

constraint (2i) limits the number of beds that each day are occupied by a patient to the maximum number of available beds.

The hierarchical objective function is reported in (2a) in which M is a constant set to $\frac{1}{1+\sum_i \pi_i}$. The role of the multiplier M is to ensure that if a solution S_1 has a higher value of O_1 with respect to S_2, it would be preferred whichever the corresponding value of O_2. In other words, when comparing two different solutions, the secondary objective becomes relevant only in the case of the two solutions have the same value of O_1.

The primary objective function (2j) implements the workload balance criterion by a bed levelling approach, that is levelling the stay bed occupancy. In our approach this is implemented by the maximisation of the number of beds used in the day and the specialty ward with the minimal bed usage. The max min objective function tends (instead of min max) also to implicitly fill as much as possible the OR sessions, avoiding their under utilisation. The secondary objective (2k) implements the patient priority maximisation criterion maximising the sum of scores of the patients scheduled for the surgery.

In order to complete the analysis regarding the two performance criteria, we consider also the case in which the roles of the primary and secondary objective

function swaps. This can be obtained modifying the objective function (2a) as reported the following model:

$$\mathcal{H}_2: \quad \max\ z_{wb} = O_2 + \frac{O_1}{\lambda_{min}+1} \text{ s.t. } (2a), \ldots, (2k). \tag{3}$$

The value of $\lambda_{\min}$ is a strict upper bound on O_1 computed as

$$\lambda_{min} = \min_{(j \in J, t \in T)} \lambda_{jt}. \tag{4}$$

In this way, the second term of the objective function takes values always between 0 and 1. Therefore, a solution S_1, which has a higher value of O_2 with respect to another solution S_2 and the lowest value for O_1, would result always better than S_2, whichever is the value of O_1 for S_2.

4. Multi-Neighbourhood Local Search based Matheuristic

As reported in Aringhieri et al. (2017), a general purpose solver can provide solutions only for small instances in a reasonable amount of running time. For this reason, we propose a multi-neighbourhood local search matheuristic, which consist in sequentially exploring a list of large neighbourhoods. To the best of our knowledge, this is the first work in which a multi-neighbourhood local search matheuristic is proposed and applied to OR planning and scheduling. When compared to previous matheuristic literature, the key difference of our approach is the adoption of *ad hoc* suited neighbourhood to force the improvement of the objective function.

To solve the problems defined by the constraints (2b)-(2k), we introduce the three different neighbourhoods. The first one is the following.

- N_1**: patients assignment re-optimisation**.
  The basic idea is to improve the quality of the solution regarding the patient assignment component (O_2). To this end, we keep fixed all the OR assignments to specialties (variables z_{jkt}) and re-optimise only the patient selection and assignment to surgery sessions (variables x_{ikt} and implicitly variables y_{it} and w_{it}).

Exploiting the inherent hierarchy between the specialty assignment decision and the patient selection decision, Aringhieri et al. (2015a) showed the efficacy of a neighbourhood working on the specialty assignment instead of only considering the patient selection. In accordance with this insight, two additional neighbourhoods are introduced.

- N_2**: 2-days re-optimisation**.
  We re-optimise two randomly selected days (variables z_{jkt} for two different values of t) while keeping fixed the scheduling for the other days (all the remaining variables).

- N_3: **2-specialties re-optimisation**.
  We re-optimise two randomly selected specialties (variables z_{jkt} for two different values of j) while keeping fixed the scheduling for the other specialties (all the remaining variables).

We propose two multi-neighbourhood local search matheuristics $\mathbb{A}_{z_{bw}}$ and $\mathbb{A}_{z_{wb}}$ which sequentially explore the N_1, N_2 and N_3 neighbourhoods: $\mathbb{A}_{z_{bw}}$ and $\mathbb{A}_{z_{wb}}$ solve the problem with the objective function (2a) and (3), respectively. Both algorithms are repeated for a fixed number of iterations $N_{\max}$. We remark that neighbourhoods N_2 and N_3 are repeated $\alpha_{N_2} \geq 1$ and $\alpha_{N_3} \geq 1$ within the same iteration. In order to reduce the running times, we set a different time limit L_{N_1}, L_{N_2} and L_{N_3} respectively for neighbourhoods N_1, N_2 and N_3, after which the best obtained solution is reported. Every time we find an improving solution we keep it as current solution.

When $\mathbb{A}_{z_{bw}}$ is not able to improve the primary objective (bed levelling), the algorithm tries to improve the secondary objective (patients score sum). A larger improvement of the secondary objective can consist in the insertion of a large number of patients in the scheduling. Unfortunately, such a condition makes more difficult to improve the primary objective in the subsequent neighbourhood explorations. To avoid such a situation, we introduce an *ad hoc* suited neighbourhood N_4 for $\mathbb{A}_{z_{bw}}$ to force the improvement of the primary objective: the neighbourhood N_4 is explored after N_1, N_2 and N_3, and is repeated $\alpha_{N_4} \geq 1$ times with a time limit L_{N_4}.

- N_4: **bed levelling improvement**.
  In order to push the first objective O_1 to be increased, we first identify the specialty j and day t in the current solution for which the bed occupation is minimum (and equal to O_1), and then we artificially increase its bed availability λ_{jt}; the idea is to force the selection of further patients through a slightly modified version of the O_2, which becomes $O_2 = \sum_{i \in I} \sum_{t \in T} \sum_{k \in K} x_{ikt}$. Note that this version of O_2 is equivalent to state that all the patients have the same priority. With this setting, the run the model obtained by applying neighbourhood N_1: if the best obtained solution is not feasible with respect to the original value of λ_{jt}, we reset λ_{jt} to its actual value and apply again N_1 to regain feasibility.

Similar approaches to that underlying neighbourhood N_4 have been already reported in the metaheuristic literature for the solution of SONET network design problems (Aringhieri & Dell'Amico, 2005a,b) and, more recently, for dealing with a waste disposal problem (Aringhieri et al., 2018a), the equitable coloring problem (Wang et al., 2018b), and the 2-constraint bin packing problem (Aringhieri et al., 2018c). To the best of our knowledge, this approach is never applied in the context of matheuristics and for the solution of bi-objective optimisation problems.

We observed that the bed levelling function tends to have lower OR utilisation because it does not take advantage from the insertion of patients which have no impact on the bed levelling maximisation. For this reason, at the end

of the $N_{\max}$ iterations, $\mathbb{A}_{z_{bw}}$ performs an additional exploration of N_1 in order to improve the OR utilisation by adding further patients, if possible.

Finally, Algorithms 1 and 2 report the pseudocode of the algorithms $\mathbb{A}_{z_{wb}}$ and $\mathbb{A}_{z_{bw}}$, respectively: each algorithm takes as input an instance D of the problem and the list of parameters; after computing an initial solution S' using the function CIS(D), the algorithm starts a cycle of $N_{\max}$ iterations; at the n_{iter}-th iteration, the algorithm tries to improve the current solution S' exploring a list of of large neighbourhoods recording the solution computed when it improve the current best; at the end, the algorithm returns the best computed solution S^*. Algorithm 2 differs from algorithm 1 by the use of the *ad hoc* suited neighbourhood N_4 and the post-optimisation using N_1. In our implementation the function CIS(D) returns the best solution computed by the general purpose solver on the whole model after L_{CIS} seconds.

Algorithm 1: $\mathbb{A}_{z_{wb}}$

Data: $D, N_{\max}, \alpha_{N.}, L_{N.}$;
Result: S^*;

$S' = \text{CIS}(\ D\ )$;
$S^* = S'$; $n_{iter} = 0$;

while $n_{iter} \leq N_{\max}$ **do**
  $S' = \text{IS}(S', N_1, L_{N_1})$;
  if $z_{wb}(S') \geq z_{wb}(S^*)$ **then**
    $S^* = S'$
  $S' = \text{IS}(S', N_2, \alpha_{N_2}, L_{N_3})$;
  if $z_{wb}(S') \geq z_{wb}(S^*)$ **then**
    $S^* = S'$
  $S' = \text{IS}(S', N_3, \alpha_{N_3}, L_{N_3})$;
  if $z_{wb}(S') \geq z_{wb}(S^*)$ **then**
    $S^* = S'$
  $n_{iter} = n_{iter} + 1$;

return (S^*);

Algorithm 2: $\mathbb{A}_{z_{bw}}$

Data: $D, N_{\max}, \alpha_{N.}, L_{N.}$;
Result: S^*;
$S' = \text{CIS}(\ D\ )$;
$S^* = S'$; $n_{iter} = 0$;
while $n_{iter} \leq N_{\max}$ **do**
  $S' = \text{IS}(S', N_1, L_{N_1})$;
  if $z_{bw}(S') \geq z_{bw}(S^*)$ **then**
    $S^* = S'$
  $S' = \text{IS}(S', N_2, \alpha_{N_2}, L_{N_3})$;
  if $z_{bw}(S') \geq z_{bw}(S^*)$ **then**
    $S^* = S'$
  $S' = \text{IS}(S', N_3, \alpha_{N_3}, L_{N_3})$;
  if $z_{bw}(S') \geq z_{bw}(S^*)$ **then**
    $S^* = S'$
  $S' = \text{IS}(S', N_4, \alpha_{N_4}, L_{N_4})$;
  if $z_{bw}(S') \geq z_{bw}(S^*)$ **then**
    $S^* = S'$
  $n_{iter} = n_{iter} + 1$;
$\text{IS}(S', N_1, L_{N_1})$;
if $z_{bw}(S') \geq z_{bw}(S^*)$ **then**
  $S^* = S'$
return (S^*);

5. Quantitative analysis

The main objective of this section is to analyse the effectiveness and the efficiency of the proposed matheuristic algorithms as well as to evaluate the quality of the computed solution from an operating room management perspective.

We describe the computational environment and the benchmark instances in Section 5.1. A wide computational analysis of the proposed matheuristics $\mathbb{A}_{z_{wb}}$

and $\mathbb{A}_{z_{bw}}$ is reported in Section 5.2 discussing (i) the null impact of the seed selected and other parameter tuning (5.2.1), (ii) the comparison of the solutions provided by the models $\mathcal{H}_1$ and $\mathcal{H}_2$, and the positive impact of adopting the neighbourhood N_4 in $\mathbb{A}_{z_{bw}}$ (5.2.2), (iii) an analysis of the Pareto frontier of the solutions obtained by fixing the bed levelling value in $\mathbb{A}_{z_{bw}}$ (5.2.3). Finally, we investigates the quality of the solutions from an operating room management perspective in Section 5.3 providing (i) an analysis of the main parameters to evaluate the management of a set of ORs in a hospital, and (ii) a sensitivity analysis through a probabilistic evaluation of the surgery duration s_{kt}.

5.1. Computational environment

All the mathematical models $\mathcal{H}_1$, $\mathcal{H}_2$, and the matheuristics $\mathbb{A}_{z_{wb}}$ and $\mathbb{A}_{z_{bw}}$ were programmed in Mosel within the FICO Xpress-optimizer 8.3 platform. All the tests for $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathbb{A}_{z_{wb}}$, $\mathbb{A}_{z_{bw}}$ were performed on a 2.1 gigahertz AMD Opteron 8425HE with 12 cores with 16 gigabytes of memory.

We create a benchmark set B of realistic instances in order to test our algorithms. The set B is created starting from the different patient case mix generated by the tool proposed in Leeftink & Hans (2018), which are based on real-life case mixes (11 surgical specialties). Such surgical specialties are: General surgery (CHI), Otolaryngology (ENT), Ophthalmic surgery (EYE), Obstetric and gynaecologic surgery (GYN), Neurological surgery (NEU), Surgical oncology (ONC), Orthopaedic surgery (ORT), Plastic surgery (PLA), Thoracic surgery (THO), Urology (URO), and remaining specialties, such as colorectal surgery, paediatric surgery, trauma surgery, vascular surgery, etc. (MIX).

For each specialty, we generated a waiting list made of 300 patients, each one characterised by the expected operating time p_i, the expected LOS μ_i, the priority level π_i, and the number of elapsed days between the diagnosis of patient i and the day t of planning ϕ_{it}: p_i is generated by the tool developed in Leeftink & Hans (2018) in accordance to the selected case mix; μ_i is an integer random number uniformly distributed in $[1, \ldots, 5]$; π_i is a value in $\{1, 2, 6, 12, 45\}$ in accordance with the case study discussed in Valente et al. (2009); ϕ_{it} is an integer number in $[1, \ldots, 360]$ distributed according to Table 1.

The benchmark set B has been generated starting from the above lists of patients: for each instance, patients have been randomly selected from the list of the specialty in such a way to keep unchanged the distribution of the patients with respect to π_i. We generated 80 instances varying the main instance parameters as follows: $|I| = \{50, 100\}$ patients from each one of the $|J| = \{4, 8\}$ specialties. The number of ORs is set to $|K| = \{|J|, |J| + 2\}$, and each OR sessions in the 5 days has a time capacity set to $s_{kt} = 480$ minutes. Finally the number of beds are set to $\lambda_{jt} = \{20, 40\}$, for each specialty j and day t. With these parameters, we can obtain 20 different combinations of feasible settings for which we generated 5 different instances varying the $|J|$ specialties, as reported in Table A.7 and Table A.8.

Table 1: Patient distribution: for each specialty, the two columns report the number and the percentage of patients having ϕ_{it} discrete uniformly distributed in $[1, \ldots, \frac{360}{\pi_i}]$.

π_i	CHI		ENT		EYE		GYN		MIX		NEU	
45	15	5%	30	10%	15	5%	30	10%	30	10%	15	5%
12	90	30%	75	25%	45	15%	30	10%	75	25%	45	15%
6	60	20%	90	30%	120	40%	120	40%	90	30%	120	40%
2	75	25%	75	25%	105	35%	90	30%	75	25%	90	30%
1	60	20%	30	10%	15	5%	30	10%	30	10%	30	10%
π_i	ONC		ORT		PLA		THO		EYE			
45	60	20%	30	10%	15	5%	45	15%	15	5%		
12	150	50%	60	20%	45	15%	105	35%	60	20%		
6	75	25%	120	40%	120	40%	120	40%	90	30%		
2	15	5%	75	25%	90	30%	30	10%	105	35%		
1	0	0%	15	5%	30	10%	0	0%	30	10%		

5.2. Performances and quality of the proposed algorithms

In this section, we report a comprehensive analysis of our matheuristic algorithms. Section 5.2.1 reports about the tuning of the algorithm parameters while Section 5.2.2 reports the results of our algorithms compared with the model solutions. Finally, an analysis of the Pareto optimal frontier is discussed in Section 5.2.3.

5.2.1. Parameter tuning

Preliminary computational experiments have been conducted in order to determine the setting of the two main parameters of our algorithms. Since neighbourhoods N_2, N_3 and N_4 have a random component, we would evaluate the possible dependency of the algorithm results from the initial random seed. Further, we would evaluate the impact of different values of the time limit L_{N_4}.

Table 2 reports the results of the random seed dependency analysis performed on the problem $\mathcal{H}_1$ using the algorithm $\mathbb{A}_{z_{bw}}$ with $L_{N_4} = 100$ seconds over the benchmark B. The first three columns report the sum of the average objective function values, O_1 and O_2, respectively, while the last column reports the average running time in seconds. We consider the sum instead of an average value since we are interested in evaluating the dependency over all instances. The first part of the table shows the results obtained with ten different random seeds while the second one shows several statistics over the values in the first part. Finally, the last row reports the average results obtained directly solving the model with a time limit of 3600 seconds.

The results prove that our algorithm is systematically better than the model solution, independently from the random seed used: as a matter of fact, the worst results – reported in the $\min_{\mathbb{A}^{100}_{z_{bw}}}$ – row are better than those of the model. Further, the standard deviation analysis – reported in the $\sigma_{\mathbb{A}^{100}_{z_{bw}}}$ row – proves a minimal and acceptable level of random seed dependency corresponding to a variation of the 0.73%, 0.74% and 0.92% of the objective function, O_1 and O_2, respectively. Therefore, in the following we use a single seed set to 99110 for

Table 2: Dependency from the random seed (algorithm $\mathbb{A}_{z_{bw}}$).

		$\sum z_{bw}$	$\sum O_1$	$\sum O_2$	avg. time
seeds	18963	980.38	971	9378286	640.4
	28189	972.48	963	9481542	692.8
	38255	989.58	980	9581636	676.2
	45924	982.40	973	9400172	680.6
	61260	989.61	980	9605517	738.3
	82581	974.63	965	9628829	642.3
	84467	988.49	979	9488099	676.8
	87409	981.58	972	9575808	626.7
	97317	987.45	978	9454495	699.9
	99110	995.57	986	9569741	682.9
stats	$\mu_{\mathbb{A}_{z_{bw}}}$	984.22	974.70	9516412.5	675.7
	$\sigma_{\mathbb{A}_{z_{bw}}}$	7.23	7.21	87967.9	32.7
	$\max_{\mathbb{A}_{z_{bw}}}$	995.57	986.00	9628829.0	738.3
	$\min_{\mathbb{A}_{z_{bw}}}$	972.48	963.00	9378286.0	626.7
	$\mathcal{H}_1$	965.13	957.00	8128346.0	3600.0

$\mathbb{A}_{z_{bw}}$. We use the same seed also for $\mathbb{A}_{z_{wb}}$, whose dependency analysis is less relevant.

Regarding the setting of L_{N_4} we conducted several tests with values belonging to $\{30, 50, 80, 100, 130, 150\}$. The results do not show a clear dominance, and by consequence we decided to consider three different versions of our algorithm $\mathbb{A}_{z_{bw}}$ varying the time limit L_{N_4} in $\{30, 100, 150\}$. In the following, we denote such versions as $\mathbb{A}^{30}_{z_{bw}}$, $\mathbb{A}^{100}_{z_{bw}}$ and $\mathbb{A}^{150}_{z_{bw}}$, respectively. The remaining parameters are set as follows: $\alpha_{N_2} = \alpha_{N_3} = \alpha_{N_4} = 3$, $L_{N_1} = L_{N_2} = L_{N_3} = L_{CIS} = 5$ seconds, and $N_{\max} = 4$.

5.2.2. Performance of the algorithms

In this section, we provide a detailed performance analysis of our proposed algorithms, comparing the results with the solution computed by the XPress general purpose solver.

The following three tables share the same structure. The columns are grouped in two sets. The first set concerns the case in which the workload balance is the primary objective function, reporting respectively the results for the model, for the matheuristic $\mathbb{A}_{z_{bw}}$ with the three different time limits, for the matheuristic $\mathbb{A}^*_{z_{bw}}$ (which is the combination of the three previous versions returning the best solution computed), and for the matheuristic $\mathbb{A}^0_{z_{bw}}$ (which is the algorithm $\mathbb{A}_{z_{bw}}$ without the use of N_4 for the problem $\mathcal{H}_1$). The second set concerns the case in which the patient priority maximisation is the primary objective function, reporting respectively the results for the model and for the matheuristic $\mathbb{A}_{z_{wb}}$. The rows of the tables report the average results concerning the objective function, the gap of the algorithm with respect to the model, and the running time, respectively. The last three rows provide an analysis of the bed levelling value reporting the number of time in which the algorithm give a

better (wins) or the same (draws) result of the model, and the sum of the O_1 values. Finally, the rows report the results computed over the whole benchmark set B or a part of it.

Table 3: Performance of the algorithms: analysis over the whole benchmark set B.

	workload balance						patient priority	
	$\mathcal{H}_1$	$\mathbb{A}^{30}_{z_{bw}}$	$\mathbb{A}^{100}_{z_{bw}}$	$\mathbb{A}^{150}_{z_{bw}}$	$\mathbb{A}^{*}_{z_{bw}}$	$\mathbb{A}^{0}_{z_{bw}}$	$\mathcal{H}_2$	$\mathbb{A}_{z_{wb}}$
o.f.	12.1	12.2	12.4	12.4	12.6	11.0	11749.6	12251.1
gap	–	3.75%	5.96%	5.56%	7.63%	-6.72%	–	3.63%
time	3600.0	526.4	682.9	874.8	2084.1	274.5	3600.0	224.4
wins	–	14	17	15	18	9	–	57
draws	–	43	49	50	54	31	–	22
$\sum O_1$	957	967	986	982	1002	867	201	522

Regarding the workload balance case, the results provided in Table 3 prove the superiority of the three matheuristics $\mathbb{A}^{30}_{z_{bw}}$, $\mathbb{A}^{100}_{z_{bw}}$ and $\mathbb{A}^{150}_{z_{bw}}$ with respect to the solution computed by the general purpose solver for $\mathcal{H}_1$, reducing the average running time from about 4 up to 7 times. The unsatisfactory results of $\mathbb{A}^{0}_{z_{bw}}$ prove the need and the positive impact of the neighbourhood N_4 to deal with the bed levelling component of the hierarchical objective function. The results (especially those regarding O_1) seem to indicate that $\mathbb{A}^{100}_{z_{bw}}$ is the best matheuristic, even if the results of $\mathbb{A}^{*}_{z_{bw}}$ are better on average but requiring a larger running time.

Regarding the patient priority maximisation case, the results provided in Table 3 prove the superiority of the matheuristic $\mathbb{A}_{z_{wb}}$ with respect to the solution computed by the general purpose solver for $\mathcal{H}_2$, reducing the average running time of about 16 times. It is worth noting that the matheuristic $\mathbb{A}_{z_{wb}}$ computes better solutions than those for $\mathcal{H}_2$ also in terms of O_1 with 57 wins, and 65 wins if we compare the value of O_2.

Table 4: Performance of the algorithms: analysis over the whole benchmark set B clustering the instance on the number of patients for each specialty.

		workload balance						patient priority	
		$\mathcal{H}_1$	$\mathbb{A}^{30}_{z_{bw}}$	$\mathbb{A}^{100}_{z_{bw}}$	$\mathbb{A}^{150}_{z_{bw}}$	$\mathbb{A}^{*}_{z_{bw}}$	$\mathbb{A}^{0}_{z_{bw}}$	$\mathcal{H}_2$	$\mathbb{A}_{z_{wb}}$
$\lvert I\rvert = 50$	o.f.	11.5	11.0	11.2	11.0	11.4	9.9	9586.1	9905.1
	gap	–	-3.77%	-2.49%	-4.20%	-0.89%	-13.19%	–	2.75%
	time	3600.0	407.8	477.8	610.4	1496.0	227.2	3600.0	206.2
	wins	–	0	1	0	1	0	–	32
	draws	–	26	31	29	34	17	–	8
	$\sum O_1$	457	438	444	436	452	394	89	260
$\lvert I\rvert = 100$	o.f.	12.6	13.4	13.7	13.8	13.9	12.0	13913.1	14597.0
	gap	–	11.27%	14.41%	15.32%	16.15%	-0.25%	–	4.51%
	time	3600.0	645.0	888.0	2672.2	1139.2	321.8	3600.0	242.6
	wins	–	14	16	15	17	9	–	25
	draws	–	17	18	21	20	14	–	14
	$\sum O_1$	500	529	542	546	550	473	112	262

Table 4 reports the results of the instances in B clustered with respect to the number of patients belonging to each specialty. The results prove a large superiority of the solutions computed by the $\mathbb{A}_{z_{bw}}$ algorithms with respect to those computed solving directly the model $\mathcal{H}_1$ on larger instances while, on the contrary, the solutions of $\mathcal{H}_1$ are still better on the smallest ones. However the results for $\mathbb{A}^{100}_{z_{bw}}$ show that the differences in terms of the bed levelling (objective function O_1) is only on 8 instances out of 40 while in the remaining 32 instances the value of O_1 is equivalent. The results confirm the positive impact of the neighbourhood N_4, and the superiority of the solutions computed by $\mathbb{A}_{z_{wb}}$ with respect to those computed solving directly $\mathcal{H}_2$, though the differences are smaller for the smaller instances.

Table 5: Performance of the algorithms: analysis over the whole benchmark set B clustering the instance on the number of stay beds allocated to each specialty.

		workload balance						patient priority	
		$\mathcal{H}_1$	$\mathbb{A}^{30}_{z_{bw}}$	$\mathbb{A}^{100}_{z_{bw}}$	$\mathbb{A}^{150}_{z_{bw}}$	$\mathbb{A}^{*}_{z_{bw}}$	$\mathbb{A}^{0}_{z_{bw}}$	$\mathcal{H}_2$	$\mathbb{A}_{z_{wb}}$
$\lambda_{jt} = 20$	o.f.	11.3	11.8	12.1	12.0	12.3	10.7	11124.8	11871.7
	gap	–	8.79%	11.71%	11.00%	13.67%	-1.09%	–	5.66%
	time	3600.0	528.5	652.5	744.6	1925.6	282.8	3600.0	226.0
	wins	–	11	12	10	13	9	–	28
	draws	–	16	20	22	23	13	–	12
	$\sum O_1$	448	467	479	475	488	424	95	255
$\lambda_{jt} = 40$	o.f.	12.8	12.6	12.8	12.8	13.0	11.2	12374.4	12630.4
	gap	–	-1.29%	0.21%	0.13%	1.59%	-12.35%	–	1.60%
	time	3600.0	524.2	713.4	1005.0	2242.6	266.3	3600.0	222.8
	wins	–	3	5	5	5	0	–	29
	draws	–	27	29	28	31	18	–	10
	$\sum O_1$	509	500	507	507	514	443	106	267

Table 5 reports the results of the instances in B clustered with respect to the number of stay beds allocated to each specialty. The gaps reported seem to suggest that the instances with less number of stay beds are harder than the others, both considering bed levelling and patient priority. For the instances with a smaller number of stay beds, all the $\mathbb{A}_{z_{bw}}$ algorithms outperform the results obtained by solving directly $\mathcal{H}_1$, while for those with a larger number of stay beds the results are almost equivalent. As in the previous tables, the results confirm the positive impact of the neighbourhood N_4, and the superiority of the solutions computed by $\mathbb{A}_{z_{wb}}$ with respect to those computed solving directly $\mathcal{H}_2$, though the differences are smaller for the instances with a larger number of stay beds.

Summing up, the large running time reduction proves the efficiency of the proposed algorithms while the average improvement of the solution quality proves their effectiveness. Detailed results are reported in Tables A.9 and A.10 respectively for $\mathcal{H}_1$ and $\mathbb{A}_{z^{100}_{bw}}$, and for $\mathcal{H}_2$ and $\mathbb{A}_{z_{wb}}$.

5.2.3. Pareto optimal frontier analysis

In many real-life multi-objective optimisation problems, a common study is to analyse the Pareto optimal frontier in order to determine a solution corre-

sponding to the best compromise according to the decision-maker's preferences. This can be quite difficult to determine due to the large number of possible solutions belonging to the frontier. In our problem, this analysis has been made easier by the discrete nature of one of the two components of our objective function, that is the value O_1 regarding the bed levelling.

In this section we provide an analysis of the Pareto optimal frontier in order to evaluate the possible contrast between the two criteria, workload balance and patient priority maximisation. This analysis is performed by running the $\mathbb{A}_{z_{wb}}$ fixing the value of O_1 equals to f belonging to the discrete interval $[b_{wb}, b_{bw}]$, where b_{wb} and b_{bw} are respectively the value of O_1 computed by $\mathbb{A}_{z_{wb}}$ and $\mathbb{A}^{100}_{z_{bw}}$. We denote such a algorithm version with $\mathbb{A}_{z_{wb}}(f)$.

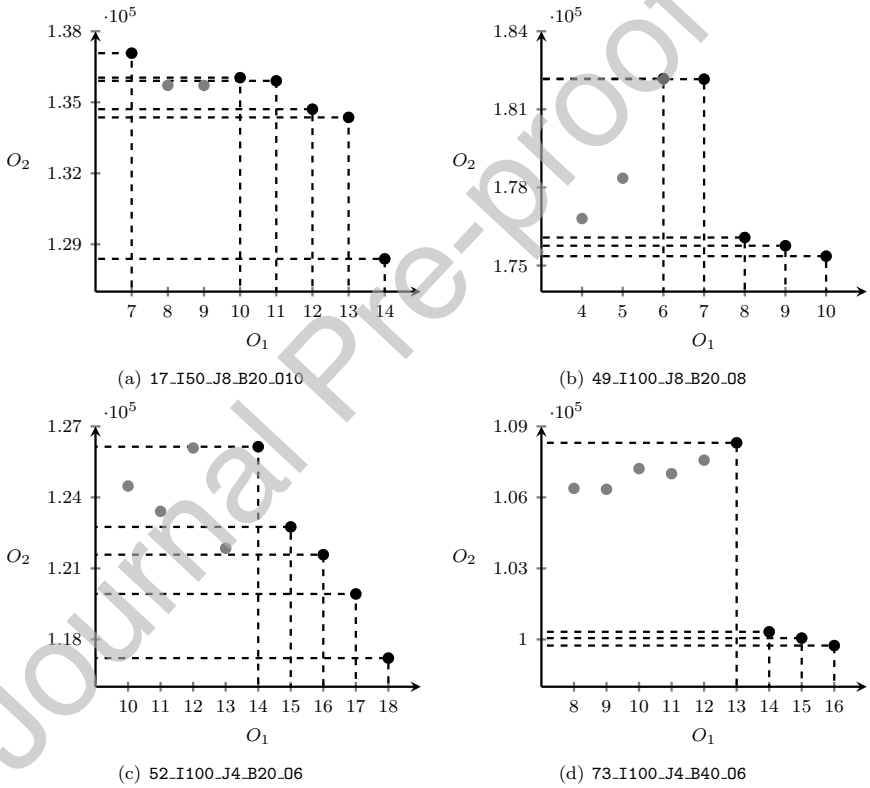


Figure 2: Pareto frontier analysis: four selected instances whose characteristics and parameters are depicted in Tables A.7 and A.8.

The Figure 2 draws the Pareto frontier of four selected instances. The plots show a general dominance of the solutions having a larger value of O_1: for the instance 73_I100_J4_B40_06 the solution with $O_1 = 13$ dominates all the solutions with a lower value of O_1; similar remarks hold also for the other three instances plotted. Therefore the workload balance criterion seems to be a proxy

for the patient priority maximisation, at least for values not so far from the optimal one. In practice, this means that it seems possible to have solutions that are a good compromise between the two criteria.

Further, we would point out an interesting remark which comes out from this analysis. Especially for the instances having less patients for each specialty, the algorithm $\mathbb{A}_{z_{wb}}(f)$ with $f = b_{bw}$ is often unable to find a feasible solution. Since such a value of f derives from the solution computed by $\mathbb{A}^{100}_{z_{bw}}$, this further proves the importance and the positive impact of the neighbourhood N_4.

5.3. Operating room management considerations

In this section, we investigates the quality of the solutions from an operating room management perspective providing (i) an analysis of the main parameters to evaluate the management of a set of ORs in a hospital, and (ii) a sensitivity analysis through a probabilistic evaluation of the surgery duration s_{kt}, which is the most relevant parameter for hospital managers to evaluate a solution at the operative level (see, e.g. Landa et al. (2016)).

We first analyse the solutions produced by the algorithms $\mathbb{A}^{100}_{z_{bw}}$ and $\mathbb{A}_{z_{wb}}$ by considering (i) the number P of scheduled patients, (ii) the utilisation U_{OR} of the operating sessions, (iii) the utilisation U_B of the stay beds, and (iv) the difference Δ between the maximum and the minimum number of occupied stay beds. Table 6 reports the average μ and the standard deviation σ of such values for the benchmark B classified by the corresponding values of $|I|$, $|J|$ and λ_{jt}. We remark that each group of instances considers a number of patients equals to $|I|\,|J|$.

Table 6: Comparing the solutions from a managerial point of view.

				$\mathbb{A}_{z_{wb}}$				$\mathbb{A}^{100}_{z_{bw}}$			
$\|I\|$	$\|J\|$	λ_{jt}		P	U_{OR}	U_B	Δ	P	U_{OR}	U_B	Δ
50	4	20	μ	137.40	84.93%	68.45%	7.93	138.40	93.91%	68.88%	6.80
			σ	21.02	5.15%	9.87%	2.12	19.20	3.14%	9.16%	2.01
		40	μ	142.00	87.08%	35.74%	14.13	142.70	94.52%	35.65%	11.30
			σ	24.10	2.60%	11.95%	3.31	24.15	2.17%	11.92%	2.43
	8	20	μ	255.80	85.25%	63.86%	9.23	264.70	92.53%	66.30%	7.33
			σ	21.89	1.80%	11.09%	0.89	19.79	3.02%	9.08%	0.71
		40	μ	267.30	86.42%	33.97%	14.83	270.90	93.54%	34.61%	11.60
			σ	23.35	1.96%	11.22%	1.62	20.77	1.61%	9.85%	2.30
100	4	20	μ	149.30	74.08%	71.08%	8.83	165.78	92.71%	80.14%	4.81
			σ	16.51	6.32%	7.34%	1.45	19.53	3.89%	8.43%	1.27
		40	μ	183.50	85.23%	46.64%	17.73	190.50	95.25%	48.10%	14.45
			σ	32.66	3.15%	15.75%	3.63	34.56	1.62%	16.60%	2.00
	8	20	μ	291.80	80.29%	73.96%	8.66	318.50	91.23%	79.98%	6.34
			σ	32.52	7.10%	16.91%	1.07	12.99	2.29%	6.01%	0.58
		40	μ	338.44	86.25%	43.18%	18.89	357.60	92.76%	45.60%	14.38
			σ	35.13	1.36%	18.63%	3.26	33.82	3.97%	16.43%	1.75

We observe that the values for the algorithm $\mathbb{A}^{100}_{z_{bw}}$ are generally better than those for the algorithm $\mathbb{A}_{z_{wb}}$. Considering the instances with $|I| = 50$, we observe that the values of Δ show the capability of $\mathbb{A}^{100}_{z_{bw}}$ to determine a better

overall workload balance while keeping a similar stay bed utilisation, as shown by the values of U_B.

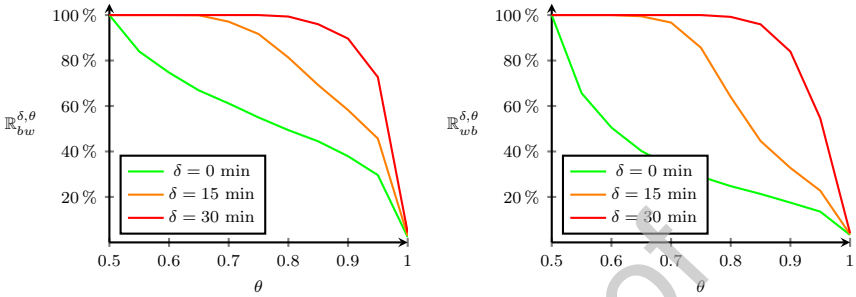
We observe a substantial better utilisation of the OR sessions for $\mathbb{A}^{100}_{z_{bw}}$ with respect to $\mathbb{A}_{z_{wb}}$, which significantly increases for the instances with $|I| = 100$. This is due by the fact that our instances are characterised by several specialties with a different case mix: when the set $|I|$ of a specialty is characterised by patients having a large different operating time, it could happen that patients with high priority have short operating time, and vice versa. For example, this the case of those instances that include specialty MIX (the list is reported in Tables A.7 and A.8) whose surgeries are quite different (e.g. colorectal, paediatric, trauma, ...). In this situation, $\mathbb{A}_{z_{wb}}$ fosters the scheduling of high priority patients which could be not enough to fill in the OR session because of the bottleneck effect of the limited number of stay beds, while the search for a bed levelling of $\mathbb{A}^{100}_{z_{bw}}$ implies a good utilisation of the OR sessions, as discussed in Aringhieri & Duma (2017). Finally, the remarks on the workload balance and stay bed utilisation for the instances with $|I| = 50$ hold also for those with $|I| = 100$ but highlighting a slightly larger difference between the values of U_B caused by the bottleneck effect discussed above.

From a management point of view, the robustness of the solution with respect to the surgery duration is more relevant than that with respect to the LOS. Among many others, the main reason is that overtime is a scarce and expensive resource. On the contrary, the LOS is a less relevant parameter in the clinical practice: when the patient presents several morbidities and complications, she/he is transferred to another ward (usually general medicine) in such a way to offer better treatment in terms of care and better health outcomes (see, e.g. Duckitt et al. (2010); Regina et al. (2019)). To this end, we provide a sensitivity analysis of the surgery duration by measuring the probability for an OR session to exceed the total capacity s_{kt} due to an observational or measurement error of p_i. We assume therefore that the real operating time is a Gaussian distributed random variable ψ_i with average value equals to the expected operating times p_i and standard deviation equals to 10%, that is $0.1\,p_i$.

Given that a sum of Gaussian distributed random variables is itself a Gaussian distributed random variable, the real total duration of an OR session is a Gaussian distributed random variable with average value equals to the sum of the expected operating times $\sum p_i$ and standard deviation equals to $0.1\sqrt{\sum p_i^2}$, as depicted in the following examples: let us suppose that $s_{kt} = 480$ and having two different schedules, the first one composed of two surgeries of 200 minutes each, and the second one composed of one surgery of 100 minutes and one of 300 minutes; in this case, we have a different probability of exceeding the total capacity of the OR session: although we obtain the same OR session utilisation, the first schedule provides a lower standard deviation, that is 28 minutes vs. 32 minutes.

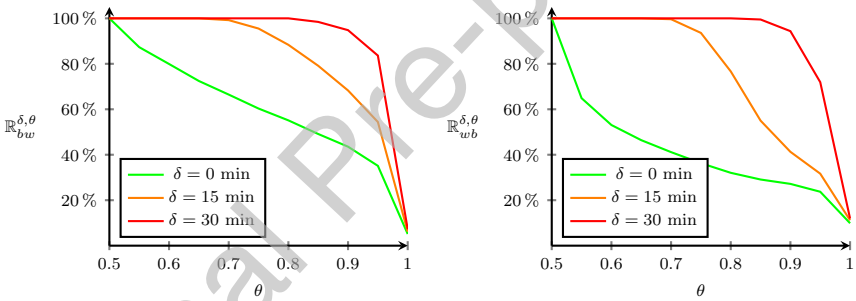
We introduce the robustness indicators $\mathbb{R}^{\delta,\theta}_{bw}$ and $\mathbb{R}^{\delta,\theta}_{wb}$ defined as the fraction of the sessions such that the probability of not exceeding $s_{kt} + \delta$ ($\delta \geq 0$) is equal to or greater than a minimum target θ, in which δ is a fixed tolerance. Such indicators are computed as follow: given the set I_{kt} of the patient scheduled in a

certain session (k,t), we compute the probability $\mathbb{P}^{\delta}_{kt} = \mathbb{P}(\sum_{i \in I_{kt}} \psi_i \leq s_{kt} + \delta)$, and then we can estimate the index as the percentage of the sessions (k,t) such that $\mathbb{P}^{\delta}_{kt} \geq \theta$.



(a) $|I| = 50$, $(\mathbb{A}^{100}_{z_{bw}})$ (b) $|I| = 50$, $(\mathbb{A}_{z_{wb}})$

Figure 3: Robustness: comparing $\mathbb{A}^{100}_{z_{bw}}$ and $\mathbb{A}_{z_{wb}}$ on instances with 50 patients per specialty.



(a) $|I| = 100$, $(\mathbb{A}^{100}_{z_{bw}})$ (b) $|I| = 100$, $(\mathbb{A}_{z_{wb}})$

Figure 4: Robustness: comparing $\mathbb{A}^{100}_{z_{bw}}$ and $\mathbb{A}_{z_{wb}}$ on instances with 100 patients per specialty.

In Figures 3 and 4 we plotted the robustness indicators for three different tolerances values $\delta = 0, 15, 30$ and ranging the minimum target $\theta \in [0.5, 1]$. The four plots show how quickly or slowly the robustness indices $\mathbb{R}^{\delta,\theta}_{bw}$ and $\mathbb{R}^{\delta,\theta}_{wb}$ decrease as soon as the value of θ increases. For instance, choosing a target $\theta = 0.9$, Figures 3(a) reports a probability of 37.91%, 58.21% and 89.65% of not exceeding the total capacity of 480 minutes plus a tolerance of 0, 15 and 30 minutes, respectively. There are two general insights provided by the robustness analysis.

The first insight is that the algorithm $\mathbb{A}^{100}_{z_{bw}}$ is significantly more robust of the algorithm $\mathbb{A}_{z_{wb}}$, especially when the tolerance is equal to 0, that is when the schedule is not allowed to use overtime to complete the session. This results is counter-intuitive with respect to the OR utilisation reported in Table 6. This remark can be explained as follow. Considering a pair of patients with the same value of μ_i, the optimal solution determined by the bed levelling criterion

could favour the patient having shorter duration p_i (and reduced impact of the robustness indicators) in order to allow the possibility of adding further patients, if needed. Further, patients having longer p_i are distributed among OR sessions in order to avoid bed levelling bottlenecks.

The second insight is that increasing the number of patients to be operated on, we have a slightly better robustness for both of the algorithms. In accordance with the above considerations, it is easier to select patients with short p_i when the number of patients to be operated on increases.

6. Conclusions

The aim of this paper is to develop an effective and efficient solution approach for the OR planning and scheduling capable to take into account the patient priority maximisation and workload balance criteria at the same time. This work is inspired by the need of a deeper understanding of the quality of the solutions obtained when a combination of the two criteria leads the OR planning decisions.

Starting from a hierarchical multi-objective optimisation model for the combined master surgical scheduling and surgical cases assignment problems, we developed a class of new multi-neighbourhood local search based matheuristic algorithms. In particular, we developed an ad hoc neighbourhood, which has been proved to generate better solutions in a significant shorter running time with respect to a general purpose solver. We would remark that this approach was never applied in the context of matheuristics and for the solution of bi-objective optimisation problems, to the best of our knowledge.

We provided an accurate computational analysis on realistic instances proving the effectiveness of the proposed approach, also providing a Pareto optimal frontier analysis. From this analysis, it clearly emerges that the optimisation of the workload balance determines good quality solutions also from a patient priority point of view, while the opposite is not true. Further we investigate the computed solutions from an OR management point of view also evaluating their robustness with respect to the uncertainty of the surgery durations. From this analysis, it emerges that the solutions obtained using the workload balance as primary criterion are better than those obtained using the priority maximisation criterion.

A future development could consist in the extension of our methodology in order to consider the stay bed as a shared resource among different wards. Another possible development could consider an evaluation of the proposed hierarchical multi-objective over time in order to evaluate the impact of the OR planning decisions on a longer time horizon, as in Aringhieri et al. (2018b). As an alternative of our approach, it could be of interest to consider a multiple objective approach in order to compare the quality of the solutions provided and, more general, to further investigate how to consider the different preferences of the stakeholders. Finally, it would certainly be of interest to integrate in our solution approach the uncertainty of the main parameters such as the surgery duration and the length of stay.

Acknowledgements

The authors wish to thank the anonymous reviewers for their valuable comments which significantly improved the quality of the article.

References

Aiken, L., Clarke, S., Sloane, D., Lake, E., & Cheney, T. (2008). Effects of hospital care environment on patient mortality and nurse outcomes. *Journal of Nurse Administration*, *38*, 223–229.

Aiken, L., Sloane, D., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kzka, M., Lesaffre, E., McHugh, M., Moreno-Casbas, M., Rafferty, A., Schwendimann, R., Scott, P., Tishelman, C., van Achterberg, T., Sermeus, W., & for the RN4CAST consortium (2014). Nurse staffing and education and hospital mortality in nine european countries: a retrospective observational study. *The Lancet*, *383*, 1824–1830.

Aringhieri, R., Bruglieri, M., Malucelli, F., & Nonato, M. (2018a). A special vehicle routing problem arising in the optimization of waste disposal: A real case. *Transportation Science*, *52*, 277–279.

Aringhieri, R., & Dell'Amico, M. (2005a). Comparing metaheuristic algorithms for sonet network design problems. *Journal of Heuristics*, *11*, 35–57.

Aringhieri, R., & Dell'Amico, M. (2005b). Solution of the sonet ring assignment problem with capacity constraints. *Operations Research/ Computer Science Interfaces Series*, *30*, 93–116.

Aringhieri, R., & Duma, D. (2017). Patient–centred objectives as an alternative to maximum utilisation: comparing surgical case solutions. In A. Sforza, & C. Sterle (Eds.), *Optimization and Decision Science: Methodologies and Applications. ODS 2017* (pp. 105–112). Springer, Cham volume 217 of *Springer Proceedings in Mathematics and Statistics*.

Aringhieri, R., Duma, D., & Faccio, E. (2018b). Ex post evaluation of an operating theatre. In *Joint EURO/ALIO International Conference 2018 on Applied Combinatorial Optimization* (pp. 157–164). volume 69 of *Electronic Notes in Discrete Mathematics*.

Aringhieri, R., Duma, D., Grosso, A., & Hosteins, P. (2018c). Simple but effective heuristics for the 2-constraint bin packing problem. *Journal of Heuristics*, *24*, 345–357.

Aringhieri, R., Landa, P., & Mancini, S. (2017). A hierarchical multi-objective optimisation model for bed levelling and patient priority maximisation. In A. Sforza, & C. Sterle (Eds.), *Optimization and Decision Science: Methodologies and Applications. ODS 2017* (pp. 113–120). Springer, Cham volume 217 of *Springer Proceedings in Mathematics and Statistics*.

Aringhieri, R., Landa, P., Soriano, P., Tànfani, E., & Testi, A. (2015a). A two level Metaheuristic for the Operating Room Scheduling and Assignment Problem. *Computer & Operations Research*, *54*, 21–34.

Aringhieri, R., Landa, P., & Tànfani, E. (2015b). Assigning surgery cases to operating rooms: A VNS approach for leveling ward beds occupancies. Electronic Notes in Discrete Mathematics (pp. 173–180).

Beaulieu, I., Gendreau, M., & Soriano, P. (2012). Operating rooms scheduling under uncertainty. In E. Tànfani, & A. Testi (Eds.), *Advanced Decision Making Methods Applied to Health Care* (pp. 13–32). Springer Milan volume 173 of *International Series in Operations Research & Management Science*.

Behmanesh, R., & Zandieh, M. (2019). Surgical case scheduling problem with fuzzy surgery time: An advanced bi-objective ant system approach. *Knowledge–Based Systems*, *186*, 104913.

Beliën, J., & Demeulemeester, E. (2008). A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research*, *189*, 652–668.

Beliën, J., Demeulemeester, E., & Cardoen, B. (2006). Visualizing the demand for various resources as a function of the master surgery schedule: A case study. *Journal of Medical Systems*, *30*, 343–350.

Beliën, J., Demeulemeester, E., & Cardoen, B. (2007). Building cyclic master surgery schedules with levelled resulting bed occupancy: A case study. *European Journal of Operational Research*, *176*, 1185–1204.

Beliën, J., Demeulemeester, E., & Cardoen, B. (2009). A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling*, *12*, 147–161.

Breuer, D. J., Lahrichi, N., Clark, D. E., & Benneyan, J. C. (2020). Robust combined operating room planning and personnel scheduling under uncertainty. *Operations Research for Health Care*, *27*, 100276.

van den Broek d'Obrenan, A., Ridder, A., Roubos, D., & Stougie, L. (2020). Minimizing bed occupancy variance by scheduling patients under uncertainty. *European Journal of Operational Research*, *286*, 336–349.

Cappanera, P., Visintin, F., & Banditori, C. (2014). Comparing resource balancing criteria in master surgical scheduling: A combined optimisation-simulation approach. *International Journal of Production Economics*, *158*, 179–196.

Cappanera, P., Visintin, F., & Banditori, C. (2018). Addressing conflicting stakeholders priorities in surgical scheduling by goal programming. *Flexible Services and Manufacturing Journal*, *30*, 252–271.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, *119*, 354–366.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*, 921–932.

Della Croce, F., Grosso, A., & Salassa, F. (2014). A matheuristic approach for the two-machine total completion time flow shop problem. *Annals of Operations Research*, *213*, 67–78.

Della Croce, F., & Salassa, F. (2014). A variable neighborhood search based matheuristic for nurse rostering problems. *Annals of Operations Research*, *218*, 185–199.

Denton, B., Viapiano, J., & Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, *10*, 13–24.

Dios, M., Molina-Pariente, J. M., Fernandez-Viagas, V., Andrade-Pineda, J. L., & Framinan, J. M. (2015). A decision support system for operating room scheduling. *Flexible Services and Manufacturing Journal*, *88*, 430–443.

Duckitt, R., Palsson, R., Bosanska, L., Dagna, L., Durusu, T., & Vardi, M. (2010). Cdime group. common diagnoses in internal medicine in europe 2009: a pan-european, multi-centre survey. *European Journal Internal Medicine*, *21*, 449–52.

Duma, D., & Aringhieri, R. (2015). An online optimization approach for the real time management of operating rooms. *Operations Research for Health Care*, *7*, 40–51.

Duma, D., & Aringhieri, R. (2018). The Real Time Management of Operating Rooms. In C. Kahraman, & I. Topcu (Eds.), *Operations Research Applications in Health Care Management* (pp. 55–79). Springer International Publishing AG volume 262 of *International Series in Operations Research & Management Science*.

Duma, D., & Aringhieri, R. (2019). The management of non-elective patients: shared vs. dedicated policies. *Omega*, *83*, 199–212.

Freeman, N., Zhao, M., Melouk, S., & Mittenthal, J. (2019). The economics of overlapping surgeries. *Production and Operations Management*, *10*, 2629–2652.

Gansterer, M., Födermayr, P., & Hartl, R. F. (2021). The capacitated multi-level lot-sizing problem with distributed agents. *International Journal of Production Economics*, *235*, 108190.

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, *14*, 89–114.

Hans, E., Wullink, G., van Houdenhoven, M., & Kamezier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*, 1038–1050.

Hof, S., Fugender, A., Schoenfelder, J., & JO., B. (2017). Case mix planning in hospitals: a review and future agenda. *Health Care Management Science*, *2*, 207–220.

Hulsof, P., Kortbeek, N., Boucherie, R., Hans, E., & Bakker, P. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, *2*, 129–175.

Kamran, M., Karimi, B., Dellaert, N., & Demeulemeester, E. (2019). Adaptive operating rooms planning and scheduling: A rolling horizon approach. *Operations Research for Health Care*, *22*, 100200.

Landa, P., Aringhieri, R., Soriano, P., Tànfani, E., & Testi, A. (2016). A hybrid optimization algorithm for surgeries scheduling. *Operations Research for Health Care*, *8*, 103–114.

Landa, P., Sonnessa, M., Tànfani, E., & Testi, A. (2018). Multiobjective bed management considering emergency and elective patient flows. *International Transactions in Operational Research*, *25*, 91–110.

Lasater, K., McHugh, M., Rosenbaum, P., Aiken, L., Smith, H., Reiter, J., Niknam, B., Hill, A., Hochman, L., Jain, S., & J.H., S. (2021). Valuing hospital investments in nursing: multistate matched-cohort study of surgical patients. *BMJ Quality & Safety*, *30*, 46–55.

Lasater, K., Sloane, D., McHugh, M., Cimiotti, J., Riman, K., Martin, B., Alexander, M., & Aiken, L. (2020). Evaluation of hospital nurse-to-patient staffing ratios and sepsis bundles on patient outcomes. *American Journal of Infection Control*, .

Leeftink, G., & Hans, E. W. (2018). Case mix classification and a benchmark set for surgery scheduling. *Journal of Scheduling*, *21*, 17–33.

Mancini, S. (2016). A real-life multi depot multi period vehicle routing problem with a heterogeneous fleet: Formulation and adaptive large neighborhood search based matheuristic. *Transportation Research Part C: Emerging Technologies*, *70*, 100–112.

Mancini, S. (2017a). A combined multistart random constructive heuristic and set partitioning based formulation for the vehicle routing problem with time dependent travel times. *Computers & Operations Research*, *88*, 290–296.

Mancini, S. (2017b). The hybrid vehicle routing problem. *Transportation Research Part C: Emerging Technologies*, *78*, 1–12.

Mancini, S., Gansterer, M., & Hartl, R. F. (2021). The collaborative consistent vehicle routing problem with workload balance. *European Journal of Operational Research*, *293*, 955–965.

Mancini, S., & Stecca, G. (2018a). A large neighborhood search based matheuristic for the tourist cruises itinerary planning. *Computers & Industrial Engineering*, *122*, 140 – 148.

Mancini, S., & Stecca, G. (2018b). A large neighborhood search based matheuristic for the tourist cruises itinerary planning. *Computers & Industrial Engineering*, *122*, 140–148.

Maniezzo, V., Stutzle, T., & Voss, S. (Eds.) (2010). *Matheuristics: Hybridizing Metaheuristics and Mathematical Programming*. Springer.

Marques, A., Soares, R., Santos, M., & Amorim, P. (2020). Integrated planning of inbound and outbound logistics with a rich vehicle routing problem with backhauls. *Omega*, *92*, in press.

Marques, I., & Captivo, M. (2017). Different stakeholders perspectives for a surgical case assignment problem: Deterministic and robust approaches. *European Journal of Operational Research*, *261*, 260–278.

McHugh, M., Aiken, L., Sloane, D., Windsor, C., Douglas, C., & Yates, P. (2021). Effects of nurse-to-patient ratio legislation on nurse staffing and patient mortality, readmissions, and length of stay: a prospective study in a panel of hospitals. *The Lancet*, .

McRae, S., & Brunner, J. (2019). Assessing the impact of uncertainty and the level of aggregation in case mix planning. *Omega*, *102086*.

Min, D., & Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, *206*, 642–652.

Neyshabouri, S., & Berg, B. (2017). Two-stage robust optimisation approach to elective surgery and downstream capacity planning. *European Journal of Operational Research*, *260*, 21–40.

Nicosia, G., Pacifici, A., & Pferschy, U. (2017). Price of fairness for allocating a bounded resource. *European Journal of Operational Research*, *257*, 933 – 943.

van Oostrum, J., van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., & Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, *30*, 355–374.

Organisation for Economic Co-operation and Development (2017). *Caring for Quality in Health*. Organisation for Economic Co-operation and Development.

Otten, M., Braaksma, A., & Boucherie, R. (2019). Minimizing earliness/tardiness costs on multiple machines with an application to surgery scheduling. *Operations Research for Health Care*, *22*, 100194.

Peltokorpi, A. (2011). How do strategic decisions and operative practices affect operating room productivity? *Health Care Management Science*, *14*, 370–382.

Pisinger, D., & Ropke, S. (2010). Handbook of metaheuristics: International series in operations research & management science. chapter Large Neighborhood Search. (pp. 399–419). Springer, Berlin.

Regina, M. L., Guarneri, F., Romano, E., Orlandini, F., Nardi, R., Mazzone, A., Fontanella, A., Campanini, M., Manfellotto, D., Bellandi, T., Gussoni, G., Tartaglia, R., & Squizzato, A. (2019). What quality and safety of care for patients admitted to clinically inappropriate wards: a systematic review. *Journal of General Internal Medicine*, *34*, 1314–1321.

Roshanaei, V., Luong, C., Aleman, D., & Urbach, D. (2019). Reformulation, linearisation, and decomposition techniques for balanced distributed operating room scheduling. *Omega*, .

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling*, *19*, 493–525.

Schneider, A. J. T., van Essen, J. T., Carlier, M., & Hans, E. W. (2020). Scheduling surgery groups considering multiple downstream resources. *European Journal of Operational Research*, *282*, 741–752.

Siqueira, C. L., Arruda, E. F., Bahiense, L., Bahr, G. L., & Motta, G. R. (2018). Long-term integrated surgery room optimization and recovery ward planning, with a case study in the brazilian national institute of traumatology and orthopedics (INTO). *European Journal of Operational Research*, *264*, 870 – 883.

Tànfani, E., & Testi, A. (2010). A pre-assignment heuristic algorithm for the master surgical schedule problem (MSSP). *Annals of Operations Research*, *178*, 105–119.

Valente, R., Testi, A., Tànfani, E., Fato, M., Porro, I., Santo, M., Santori, G., Torre, G., & Ansaldo, G. (2009). A model to prioritize access to elective surgery on the base of clinical urgency and waiting time. *BMC, Health Services Research*, *9*.

Van Riet, C., & Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, *7*, 52–69.

Wang, J., Guo, H., Bakker, M., & Tsui, K. (2018a). An integrated approach for surgery scheduling under uncertainty. *Computer and Industrial Engineering*, *118*, 1–8.

Wang, W., Hao, J.-K., & Wu, Q. (2018b). Tabu search with feasible and infeasible searches for equitable coloring. *Engineering Applications of Artificial Intelligence*, *71*, 1–14.

World Health Organization (2017). *Surgical care systems strengthening: developing national surgical, obstetric and anaesthesia plans*. World Health Organization.

Zhu, S., Fan, W., Yang, S., Pei, J., & Pardalos, P. M. (2019). Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization*, *3*, 757–805.

Zhu, T., Luo, L., Shen, W., Xu, X., & Kou, R. (2020). Admission scheduling of inpatients by considering two inter–related resources: beds and operating rooms. *Optimization*, (pp. 1–39).

Appendix A. Description of the instances and their solution values

Instances and a summary file of the results are available online at `http://www.di.unito.it/~aringhie/benchmarks.html`.

Table A.7: Benchmark instances in B: I is the number of patients for each one of the J specialities, B is the number of beds for each specialty, O is the number of operating rooms available for each of the 5 days. The bullet indicates from which specialties the patients have been selected.

instance id	CHI	ENT	EYE	GYN	MIX	NEU	ONC	ORT	PLA	THO	EYE
01_I50_J4_B20_O4	•	•	•	•							
02_I50_J4_B20_O4					•	•	•	•			
03_I50_J4_B20_O4	•								•	•	•
04_I50_J4_B20_O4			•	•	•	•					
05_I50_J4_B20_O4							•	•	•	•	
06_I50_J8_B20_O4	•	•	•	•	•	•	•	•			
07_I50_J8_B20_O4				•	•	•	•	•	•	•	•
08_I50_J8_B20_O4	•	•	•	•				•	•	•	•
09_I50_J8_B20_O4	•	•	•		•		•		•	•	•
10_I50_J8_B20_O4		•	•	•	•	•	•	•		•	
11_I50_J4_B20_O6	•	•	•	•							
12_I50_J4_B20_O6					•	•	•	•			
13_I50_J4_B20_O6	•								•	•	•
14_I50_J4_B20_O6			•	•	•	•					
15_I50_J4_B20_O6							•	•	•	•	
16_I50_J8_B20_O6	•	•	•	•	•	•	•	•			
17_I50_J8_B20_O6				•	•	•	•	•	•	•	•
18_I50_J8_B20_O6	•	•	•	•				•	•	•	•
19_I50_J8_B20_O6	•	•	•		•		•		•	•	•
20_I50_J8_B20_O6		•	•	•	•	•	•	•		•	
21_I50_J4_B40_O4	•	•	•	•							
22_I50_J4_B40_O4					•	•	•	•			
23_I50_J4_B40_O4	•								•	•	•
24_I50_J4_B40_O4			•	•	•	•					
25_I50_J4_B40_O4							•	•	•	•	
26_I50_J8_B40_O4	•	•	•	•	•	•	•	•			
27_I50_J8_B40_O4				•	•	•	•	•	•	•	•
28_I50_J8_B40_O4	•	•	•	•				•	•	•	•
29_I50_J8_B40_O4	•	•	•		•		•		•	•	•
30_I50_J8_B40_O4		•	•	•	•	•	•	•		•	
31_I50_J4_B40_O6	•	•	•	•							
32_I50_J4_B40_O6					•	•	•	•			
33_I50_J4_B40_O6	•								•	•	•
34_I50_J4_B40_O6			•	•	•	•					
35_I50_J4_B40_O6							•	•	•	•	
36_I50_J8_B40_O6	•	•	•	•	•	•	•	•			
37_I50_J8_B40_O6				•	•	•	•	•	•	•	•
38_I50_J8_B40_O6	•	•	•	•				•	•	•	•
39_I50_J8_B40_O6	•	•	•		•		•		•	•	•
40_I50_J8_B40_O6		•	•	•	•	•	•	•		•	

Table A.8: Benchmark instances in B: I is the number of patients for each one of the J specialities, B is the number of beds for each specialty, O is the number of operating rooms available for each of the 5 days. The bullet indicates from which specialties the patients have been selected.

instance id	CHI	ENT	EYE	GYN	MIX	NEU	ONC	ORT	PLA	THO	EYE
41_I100_J4_B20_O4	•	•	•	•							
42_I100_J4_B20_O4					•	•	•	•			
43_I100_J4_B20_O4	•								•	•	•
44_I100_J4_B20_O4			•	•	•	•					
45_I100_J4_B20_O4							•	•	•	•	
46_I100_J8_B20_O4	•	•	•	•	•	•	•	•			
47_I100_J8_B20_O4				•	•	•	•	•	•	•	•
48_I100_J8_B20_O4	•	•	•	•				•	•	•	•
49_I100_J8_B20_O4	•	•	•		•		•		•	•	•
50_I100_J8_B20_O4		•	•	•	•	•	•	•		•	
51_I100_J4_B20_O6	•	•	•	•							
52_I100_J4_B20_O6					•	•	•	•			
53_I100_J4_B20_O6	•								•	•	•
54_I100_J4_B20_O6			•	•	•	•					
55_I100_J4_B20_O6							•	•	•	•	
56_I100_J8_B20_O6	•	•	•	•	•	•	•	•			
57_I100_J8_B20_O6				•	•	•	•	•	•	•	•
58_I100_J8_B20_O6	•	•	•	•				•	•	•	•
59_I100_J8_B20_O6	•	•	•		•		•		•	•	•
60_I100_J8_B20_O6		•	•	•	•	•	•	•		•	
61_I100_J4_B40_O4	•	•	•	•							
62_I100_J4_B40_O4					•	•	•	•			
63_I100_J4_B40_O4	•								•	•	•
64_I100_J4_B40_O4			•	•	•	•					
65_I100_J4_B40_O4							•	•	•	•	
66_I100_J8_B40_O4	•	•	•	•	•	•	•	•			
67_I100_J8_B40_O4				•	•	•	•	•	•	•	•
68_I100_J8_B40_O4	•	•	•	•				•	•	•	•
69_I100_J8_B40_O4	•	•	•		•		•		•	•	•
70_I100_J8_B40_O4		•	•	•	•	•	•	•		•	
71_I100_J4_B40_O6	•	•	•	•							
72_I100_J4_B40_O6					•	•	•	•			
73_I100_J4_B40_O6	•								•	•	•
74_I100_J4_B40_O6			•	•	•	•					
75_I100_J4_B40_O6							•	•	•	•	
76_I100_J8_B40_O6	•	•	•	•	•	•	•	•			
77_I100_J8_B40_O6				•	•	•	•	•	•	•	•
78_I100_J8_B40_O6	•	•	•	•				•	•	•	•
79_I100_J8_B40_O6	•	•	•		•		•		•	•	•
80_I100_J8_B40_O6		•	•	•	•	•	•	•		•	

Table A.9: Benchmark instances in B: models and matheuristic solution values for the workload balance case (instances are denoted using only the two first letters of the "*instance id*" used in Table A.7 and A.8).

	$\mathcal{H}_1$			$\mathbb{A}^{100}_{z_{bw}}$				$\mathcal{H}_1$			$\mathbb{A}^{100}_{z_{bw}}$		
	z_{bw}	O_1	O_2	z_{bw}	O_1	O_2		z_{bw}	O_1	O_2	z_{bw}	O_1	O_2
01	11.0461	11	46064	9.0458	9	45782	**41**	13.0685	13	68545	13.0682	13	68182
02	9.0759	9	75916	9.0760	9	76041	**42**	13.1048	13	104774	13.1054	13	105388
03	8.0603	8	60330	8.0606	8	60626	**43**	9.0787	9	78745	9.0785	9	78456
04	13.0567	13	56659	13.0561	13	56056	**44**	16.0927	16	92725	15.0903	15	90288
05	8.0725	8	72467	8.0720	8	71981	**45**	12.1081	12	108071	12.1076	12	107640
06	10.1104	10	110373	10.1158	10	115826	**46**	8.0493	8	49307	13.1711	13	171132
07	8.1172	8	117197	8.1282	8	128219	**47**	7.0564	7	56438	10.2008	10	200838
08	8.1025	8	102513	8.1066	8	106571	**48**	7.1185	7	118461	11.1478	11	147791
09	9.1138	9	113765	9.1212	9	121194	**49**	6.0526	6	52574	10.1583	10	158347
10	11.1103	11	110313	11.1224	11	122446	**50**	7.0574	7	57388	13.1946	13	194584
11	11.0539	11	53900	11.0554	11	55355	**51**	15.0755	15	75496	15.0862	15	86203
12	16.0797	16	79666	16.0808	16	80843	**52**	18.1137	18	113657	18.1222	18	122184
13	13.0627	13	62729	13.0679	13	67916	**53**	14.0743	14	74252	15.0908	15	90815
14	16.0637	16	63660	16.0625	16	62520	**54**	17.0739	17	73934	15.0890	15	88952
15	13.0815	13	81528	12.0837	12	83749	**55**	16.1035	16	103515	15.1253	15	125303
16	12.1204	12	120378	12.1281	12	128071	**56**	9.0602	9	60177	13.1673	13	167257
17	13.1020	13	102016	12.1347	12	134736	**57**	8.0601	8	60140	12.1831	12	183096
18	11.1213	11	121275	10.1273	10	127339	**58**	8.0568	8	56822	12.1733	12	173320
19	12.1189	12	118913	11.1326	11	132647	**59**	12.1210	12	120963	13.1651	13	165104
20	13.1254	13	125389	14.1221	14	122058	**60**	8.0631	8	63072	12.2017	12	201699
21	12.0522	12	52245	12.0517	12	51681	**61**	13.0756	13	75612	13.0749	13	74936
22	9.0863	9	86273	9.0864	9	86348	**62**	13.1160	13	116016	13.1142	13	114205
23	10.0617	10	61698	10.0610	10	61036	**63**	11.0786	11	78630	11.0785	11	78547
24	12.0582	12	58175	12.0573	12	57259	**64**	13.0841	13	84125	13.0837	13	83682
25	9.0733	9	73247	9.0732	9	73238	**65**	12.1113	12	111306	12.1124	12	112377
26	10.1199	10	119931	10.1197	10	119717	**66**	12.1747	12	174677	12.1874	12	187378
27	8.1226	8	122618	8.1286	8	128545	**67**	11.1898	11	189808	9.2001	9	200062
28	10.1083	10	108344	10.1109	10	110902	**68**	9.1245	9	124503	11.1454	11	145396
29	10.1181	10	118126	10.1198	10	119823	**69**	10.1817	10	181740	11.1947	11	194731
30	10.1295	10	129467	10.1292	10	129203	**70**	10.2036	10	203557	13.2231	13	223064
31	12.0565	12	56467	12.0562	12	56240	**71**	16.0893	16	89349	16.0904	16	90442
32	17.0843	17	84288	17.0849	17	84897	**72**	20.1211	20	121051	20.1259	20	125887
33	12.0622	12	62195	12.0619	12	61949	**73**	16.1045	16	104543	16.1053	16	105333
34	14.0624	14	62434	14.0618	14	61765	**74**	22.1000	22	99987	22.1021	22	102078
35	15.0877	15	87744	14.0895	14	89525	**75**	19.1284	19	128404	17.1316	17	131633
36	11.1180	11	118012	11.1288	11	128810	**76**	13.1708	13	170753	16.1879	16	187933
37	13.1507	13	150739	10.1539	10	153866	**77**	13.2129	13	212866	15.2239	15	223948
38	12.1174	12	117363	12.1165	12	116534	**78**	15.1661	15	166120	15.1853	15	185349
39	12.1295	12	129520	12.1177	12	117689	**79**	13.1699	13	169912	13.1934	13	193364
40	14.1363	14	136345	10.1413	10	141285	**80**	16.2060	16	206049	15.2325	15	232529

Table A.10: Benchmark instances in B: models and matheuristic solution values for the patient priority case (instances are denoted using only the two first letters of the "*instance id*" used in Table A.7 and A.8).

	$\mathcal{H}_2$			$\mathbb{A}_{z_{wb}}$				$\mathcal{H}_2$			$\mathbb{A}_{z_{wb}}$		
	z_{bw}	O_1	O_2	z_{bw}	O_1	O_2		z_{bw}	O_1	O_2	z_{bw}	O_1	O_2
01	4620.10	0	46201	4658.58	8	46585	**41**	7098.19	9	70981	7249.79	9	72497
02	7639.86	6	76398	7636.97	7	76369	**42**	10959.60	0	109596	10808.10	10	108080
03	6079.70	0	60797	6047.86	6	60478	**43**	8024.10	0	80241	8089.88	8	80898
04	5708.30	0	57083	5706.26	6	57062	**44**	9619.70	10	96196	9635.71	11	96356
05	7213.90	0	72139	7183.77	7	71837	**45**	10930.40	0	109304	10956.20	10	109561
06	10807.30	0	108073	11852.50	0	118525	**46**	14987.40	0	149874	17711.70	0	177117
07	12918.30	0	129183	13392.80	0	133928	**47**	17599.90	0	175999	19502.79	9	195027
08	10256.50	0	102565	11045.16	6	110451	**48**	13467.80	0	134678	15494.90	0	154949
09	11453.50	0	114535	12528.76	6	125287	**49**	15650.60	0	156506	16711.94	4	167119
10	11606.00	0	116060	12589.90	0	125899	**50**	16957.40	0	169574	18815.70	0	188157
11	5437.08	8	54370	5571.50	10	55714	**51**	8344.59	9	83445	8917.42	12	89173
12	8360.48	8	83604	8480.92	12	84808	**52**	12560.50	8	125604	12663.60	10	126635
13	6730.30	0	67303	6926.09	9	69260	**53**	8883.15	5	88831	9623.90	10	96238
14	6506.99	9	65069	6692.04	14	66919	**54**	8520.31	11	85202	9066.63	13	90665
15	8273.17	7	82731	8361.69	9	83616	**55**	11887.00	5	118869	12765.69	9	127656
16	12190.30	0	121903	13218.69	9	132186	**56**	15932.80	0	159328	17876.40	0	178764
17	12605.20	0	126052	13723.97	7	137239	**57**	18486.40	0	184864	20069.70	0	200697
18	12185.50	0	121855	12942.98	8	129429	**58**	15240.50	0	152405	16201.10	0	162011
19	12684.70	0	126847	13071.58	8	130715	**59**	16332.10	0	163321	18319.08	8	183190
20	12839.00	0	128390	14285.00	0	142850	**60**	17391.40	0	173914	18471.10	0	184711
21	5230.60	0	52306	5228.78	8	52287	**61**	7575.96	6	75759	7539.79	9	75397
22	8592.95	5	85929	8575.06	6	85750	**62**	11528.10	10	115280	11349.90	10	113498
23	6255.90	0	62559	6198.07	7	61980	**63**	7992.78	8	79927	8031.29	9	80312
24	5806.66	6	58066	5840.76	6	58407	**64**	8485.69	9	84856	8540.07	7	85400
25	7377.10	0	73771	7377.97	7	73779	**65**	11325.90	0	113259	11298.00	10	112979
26	12147.90	0	121479	12083.74	4	120837	**66**	18280.20	0	182802	18892.57	7	188925
27	12800.90	0	128009	13118.70	0	131187	**67**	19381.50	0	193815	20173.30	0	201733
28	11290.90	0	112909	11573.70	0	115737	**68**	14888.70	0	148887	15143.57	7	151435
29	12222.00	0	122220	12594.76	6	125947	**69**	19324.80	0	193248	20298.50	0	202985
30	13193.80	0	131938	13575.90	0	135759	**70**	21855.80	0	218558	22102.30	0	221023
31	5628.17	7	56281	5610.99	9	56109	**71**	9002.90	0	90029	9233.11	11	92330
32	8625.07	7	86250	8616.51	11	86164	**72**	12575.50	10	125754	12613.94	14	126138
33	6284.15	5	62841	6265.80	10	62657	**73**	10522.10	0	105221	10677.18	8	106771
34	6234.68	8	62346	6282.41	11	62823	**74**	10305.00	6	103049	10593.38	18	105932
35	9006.47	7	90064	9044.99	9	90449	**75**	12999.00	0	129990	13438.40	10	134383
36	12768.20	0	127682	13428.47	7	134284	**76**	19519.70	0	195197	19747.20	0	197472
37	15257.10	0	152571	15557.06	6	155570	**77**	21983.60	0	219836	23015.20	0	230152
38	11797.00	0	117970	11836.67	7	118366	**78**	17833.30	0	178333	18835.23	3	188352
39	13107.20	0	131072	13144.07	7	131440	**79**	19776.60	0	197766	20042.57	7	200425
40	13702.80	6	137027	14331.87	7	143318	**80**	22491.20	6	224911	23363.99	9	233639