Compiling a set of benchmark instances for operating room planning and scheduling

Catarina Martins Santos
catarina.martins.santos@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2021

Abstract

The increasing demand for healthcare coupled with insufficient resource and service levels in hospital units results in large inefficiencies in hospital management. Operating Rooms generate the highest revenue – but also the highest operating costs, representing 40% of them – and its cost-effective management has been widely debated and studied over the past couple of years. However, the existing literature only addresses this problem from either a theoretical approach or when experimental, the models are only developed for specific case studies. Additionally, the lack of methodologies tested with the same data base hinders the comparison between them, hampering the selection of the best methodology and its effective application at hospitals.

In this context, the objective of this work is to compile several benchmark instances for operating room planning and scheduling, empowering the scientific community with high-value data. For that, several researchers were contacted in order to obtain data which was analysed and converted to the format defined for the instances.

Thus, it was possible to compile a set of 247 instances formed with real surgical data and generated data. The diversified parameters that are included focus on lists of performed and planned surgeries and waiting lists. Lastly, a website was developed to publish the benchmark instances so that the scientific community can use it in their own experiments. This work paves the avenue to a greater cooperation between hospitals and researchers, allowing the evolution in the field of surgery planning optimization.

Keywords: Operating Room; Optimization; Planning; Scheduling; Benchmark Instances

1. Introduction

Nowadays, with the fast technological evolution, the access to health care services exponential increased in the last years. Evidences show that this fact besides having positive impacts on the development of the society, also leads to an higher demand for hospital care [1]. However, the rise in demand hasn't been followed with the resources availability on hospital units.

The surgical activity has a major impact on hospitals. Around 40% of the hospital costs derive from the operating room (OR) and 60% of the hospital admissions are surgical procedures [2]. Therefore, the OR is responsible for a considerable fraction of the hospital costs but also for the revenue. Being one of the major units, it is necessary to mobilize a higher number of material and human resources so that its performance is not compromised [3].

Due to the OR complexity, its planning and scheduling is crucial for a effective management. A minor error like a bad estimate of the surgery duration, leads to an inefficient utilization of the unit and their related resources and, as a consequence, the increase of the unit costs [4]. The optimization of the surgical activity is the key for a good hospital performance, so it has been the subject of many studies.

Between various OR related problems, the Master Surgical Scheduling Problem (MSSP) that is associated with assigning surgical specialties to OR time and defining a cyclical timetable for each OR and the Elective Patient Scheduling Problem (EPSP) that is related to assigning a room and a date to the surgery of patients in the waiting list are the most referenced in the literature [5].

Several approaches with different objectives have been proposed for the optimization of the MSSP and the EPSP. Nevertheless, there isn't a recognize model that is more efficient and has a higher implementation on real hospital scenarios, giving rise to a certain interest in the field. It should however be noted that less than 7% of the articles are related to specific case studies [6]. This way, the existing literature addresses them only conceptually, testing them based on specific practical cases, being their application extremely limited.

When performing a literature review focused on MSSP and EPSP, is rare to find two or more articles using the same set of surgical data to test their methodologies. This fact derive from the lack of data sharing by hospitals but also by the scientific community. As a consequence, it doesn't exist a possible comparison between developed methodologies and so, each hospital manager has a difficult decision to take when trying to optimize planning and scheduling of their ORs.

The need for developing a compilation of surgical data is quite promissory and urgent. The objective of the present work it to develop a public platform with surgical sets of data organized in benchmark instances. The focus will be on having real surgical data provided by hospitals or researchers and getting a large mixture of parameters that work as inputs for the methodologies being developed.

2. Methodology

To achieve the underlying objectives of this work, a methodology was defined that has three major phases: data collection, development of benchmark instances and publication of benchmark instances. The three phases are explained below.

2.1. Data collection

2.1.1 Parameter setting

Through a literature review of 91 articles, that focus on studying the optimization of OR planning and scheduling, the parameters used as inputs on the developed methodologies were listed. Those parameters serve as filters to analyze the data that will be collected and decided to include or not in the instances that will be created. However, any parameter that is not referred on those articles but that is collected will be reviewed and included if makes sense.

2.1.2 Contact with entities of interest

In order to collect the higher number of data sets as possible, a procedure to contact researchers and other entities of interest was developed. This was carried out as shown below :

- List the 91 reviewed articles and contact the researchers that are associated with articles that use real surgical data provided by hospitals;
- Analyze a recent literature review [6] in which articles using real surgical data are mentioned, list them and contact the correspondent author of those articles;

- From the 91 reviewed articles, contact the correspondent authors that worked with generated data;
- Resent emails to authors from which no response was received.

The contact made with different authors was preferably via email. In this way, contact addresses were obtained in several ways: through articles written by the authors, through google scholar or the researchers' personal pages. Additionally, some meetings were held to clarify the purpose of the project or in some cases to demonstrate the data set that was sent by the author in question.

To streamline the process, a standard email was initially developed and sent simultaneously to all authors chosen at each stage of the process described above. In this email, the project and its objective were presented, as well as the interest in collecting the data used as input for the developed methodologies in each article.

2.2. Development of benchmark instances

For the development of benchmark instances, the collected data was organized, edited and standardized. The various instances don't need to have the same content and parameters or to have the same size since they come from data that was sent by different authors or different hospitals. Even so, it is important to have the same name to the same type of parameters in each instance and to be presented on the same format.

A model was developed to be used for all the data sets that were analyzed and for their conversion into standardized benchmark instances. To develop it, some rules were taken: the format, the order of the parameters and the name and definition of each parameter. Since each instance will be published on a CSV file, data will be presented as a tabular format in plain text, separated by commas. For the second rule, it was used a template of one of the collected data and through this, the order in which the parameters would be listed was defined. At last, the name of each parameter was defined and each description was held with the help of a surgical glossary [7], although it is necessary to adapt it for each specific data set.

Afterwards, as the majority of the collected data sets came directly from documents shared by hospitals, it was necessary to proceed with their analysis and treatment, so that they would conform with the benchmark instance model previously defined. The entire process was carried out in *Microsoft Excel*, through its functions and tools.

After the data processing and consequent formation of the various benchmark instances, it was mandatory to define the format and the platform on which they will be presented to the scientific community.

2.3. Publication of benchmark instances

After the development of the various benchmark instances, the focus was on sharing and publishing them. But first, the instances created were sent to the authors and hospitals that provided them, so that they could confirm that all data could be shared together with the source (the article).

For this purpose, it was important to define in advance the organization and format in which each instance would be shared, as well as the type of platform that would be developed. In order to share the various instances, they were organized and placed in an accessible format. Thus, they were separated into folders in *zip* format, where each folder corresponds to a specific hospital, shared by an author or researcher.

Each folder contains the number of instances that the correspondent author or hospital provided in a CSV file and the instances are further accompanied by a *txt* file that provides the source from which the data was obtained, additional information about the hospital to it belongs, as well as information and instructions on how to read each instance.

Finally, it was necessary to define the type and create an open access platform to publish the developed instances, so they can be used by the ones in need and therefore, it was decided that it would be shared on a website developed through *WordPress*. This is an open source content management system, which allows the creation and editing of a website in a simple and practical way, without the need to use a programming language. For this to be published, it is necessary to host it in a domain that was provided by *Instituto Superior Técnico*. Thus, an easy-access website was created and developed, which will be disclosed with all benchmark instances created so far.

3. Results

This section consists of the presentation and analysis of the results obtained, through the application of the methodology described in the previous section. The results are presented by the same three phases as the methodology: data collection, development of benchmark instances and publication of benchmark instances.

3.1. Data collection

3.1.1 Parameter set

The definition of the most interesting parameters, which act as inputs in the developed methodologies and models, was carried out by the literature review that was carried. It was possible to extract 83 parameters, where the 15 most used are: real surgery duration, planned surgery duration, number of ORs, surgical specialty, number of surgeries performed per week, length of stay, timestamps, number of surgeons, arrival rates, priority, resources capacity, waiting list, block active time, allocated surgeon and resources utilization, in descending order of utilization.

With this review it was intended to achieve two objectives, the collection and inclusion of the highest possible variability of parameters, so that they can adapt to the largest number of existing methodologies or the ones to be developed. But also to give preference to the parameters that are most referenced in the articles reviewed, as probably they will also be the most wanted in the future. Although these parameters are defined as those of greatest interest, in the case of obtaining others that are not included, they will be analyzed and used if they fit into the present study.

3.1.2 Established contacts

To obtain the data sets, several correspondent authors of articles within the MSSP and EPSP context were contacted. The authors were chosen through the process explained in the methodology.

Thus, a total of 145 authors were approached. From the 145 contact attempts, 54 responses were obtained. From the responses received, 14 (26%) were from authors who sent and shared the data sets used in their articles (positive answer) and 40 (74%) were from authors who did not shared the data (negative answer), due to various reasons. This information is presented in the pie charts in the figure 1.

Positive responses are analyzed in the following subsection. As for the negative responses, several reasons were received by each author for not sending the data set used in their articles. From the 40 negative answers, 15 (38%) were emails that were forwarded by the correspondent author to other co-authors, 20 (50%) are associated with the impossibility of sharing due to confidentiality agreements with the hospital and 5 (12%) refers to other answers, such as the language in which data were presented or the unavailability to search for the data as it was used many years ago.

Regarding the responses received, although only 26% of them were positive showing availability for cooperation and data sharing, this number was higher than expected. As for the negative answers, those corresponding to "forwarded emails" were not taken ahead by other co-authors, despite the interest showed by the correspondent author. Those included in the category "others" represents for example logistic problems, making it impossible to share the data sets. The 50% that represents

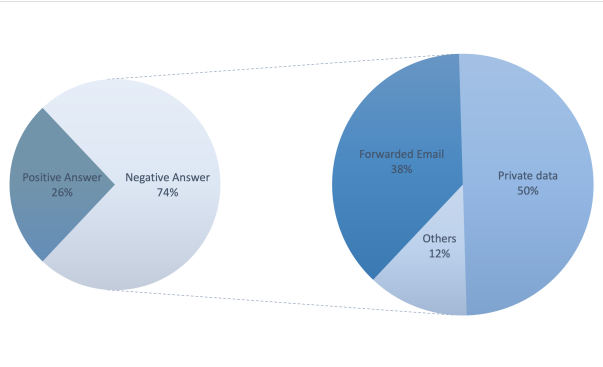

Figure 1: Answers obtained from the contacts made. The pie chart on the left shows the percentage of positive and negative answers. The pie chart on the right shows, within the negative answers, the percentage of each reason associated with each response.

"private data", prove the nowadays existing problem in the lack of sharing and publishing surgical data. This authors justified their answer due to privacy policies agreed between the hospital and the researchers, because it is a mandatory procedure asked by the hospital and only with that the data is provided. So it is possible to say that one obstacle found in the development of a platform like this is mainly made by hospitals and not by the scientific community that studies the subject, as they signed those confidentiality agreements that forbids the publication of data, even if the data is anonymous. Accordingly, hospitals prefer not to share their data (not cooperate) even though long-term cooperation can bring advantages in the development of new methodologies that enables the optimization of their processes, like the OR planning.

From a different perspective, when analyzing the existing literature, such a high number of positive responses was not expected, as until this moment any platform with compilation of surgical data hasn't been developed. Therefore, it was possible to continue the project with a significant amount of data sets.

3.1.3 Collected data

As mentioned above, it was possible to obtain 14 positive responses related with different articles, in which the data sets used were shared. Table 1 lists the articles and related references of the articles that provided the data sets. Each article is associated with a letter, from A to N, that will be used in the remaining analysis.

Data set I is associated with an anonymous author that allowed the sharing of the parameters used, as long as they were not associated with any information that could identify either the hospital or the article.

Whilst the development of the data collection process, it was also possible to get in touch with three researchers in the area who developed or are still associated with projects related to benchmarking or just the compilation of data concerning surgical activity. The person in charge of the unique public website was contacted. The platform contains a vast benchmark set with generated data, as well as an instance generator, based on real surgical data from a hospital. Although it has the same objective as the present platform, it is complementary since the objective is to focus on diversified data and parameters, with a greater focus on real surgical data. Thus, the platform already made available will be shared on the developed website, as they are complementary.

Additionally, the corresponding author of several interesting articles in the field was reached, whom despite not having real hospital data sets, shared a drive consisting of a generated database, including a list of surgeries with various parameters: allocated OR, real surgery duration, surgical specialty, information regarding the patient comorbidity level and information on whether the request for each surgery was performed before or after the closing hour of the daily schedule for each OR. Like the previous one, this will also be shared in the developed platform, as it consists of an instance, that despite having few parameters is composed by a huge number of surgeries.

Finally, through the contact with a third researcher, it was possible to know about the existence of a benchmark with real surgical data from cardiovascular surgical interventions. It contains 150 to 200 instances that are not public nor was possible to obtain, because is only available for the group of researchers associated with the hospitals that provided the data. Even so, a meeting was held to share experience in data collection and development of instances. In this benchmark, few parameters were used (usage of OR, number of overtime hours, cancellation rate and length of stay) also due to its specificity, being only for the cardiovascular specialty. Two advice were obtained for the process of creating the instances: collecting as much information as possible from the hospital that provides the data and taking into account the countries where they are based. This researcher gave a lot of emphasis on those two aspects so that, when anyone wants to use the instances, they have the possibility of knowing the size and functioning of the hospital as they affect several characteristics of the data, for example the size of the waiting lists or the resources capacity and utilization. Thereby, importance will be given to the collection of all the available information regarding each hospital and the instances will be separated according to the hospital from they derived.

At this moment, there are only three compilations of surgical data regarding the OR planning but only two of them are public. Both of them contains generated data and a small number of parameters, limiting their use for comparing various methodologies. As they are constituted only by generated data, the existence of symmetries between the various instances is more likely, compared to instances with real data. These two platforms were developed by authors and researchers who study the topic and noted the existing gap in data sharing and compilation and, therefore, their creation came to fill this gap. Analyzing the number of articles and methodologies already developed on the subject and the importance given to it, there is a very small and insufficient number of instances that can be used for the study and testing of the developed models and methodologies. This opinion was also denoted by the authors contacted, who encouraged the collection and formulation of this platform, including the two aforementioned, who gave importance to its creation to complete the two existing platforms.

In addition to the collected data throughout the contact with correspondent authors, two data sets were added to the compilation shared by two Portuguese hospitals, *Centro Hospitalar Lisboa Norte* (CHLN) and *Hospital do Espírito Santo de Évora* (HESE). Each one of these was associated with the letter O and P, respectively.

In table 2 some characteristics of the collected data are presented: country, date, type of data (generated, real, generator) and number of instances. The country is associated with either the location of the hospital if applicable, or the location where the research group is located. The date refers to the acquisition date of the data if it is real, or when it was generated otherwise. Regarding the type of data, it can be real when the set is original and provided by hospitals, it can be generated based on information provided by hospitals or only on projections made by researchers and it can be a generator of data that, as the name implies, is made up of parameters and distributions that allow the generation of new data sets depending on the available parameters.

The first observation that can be taken from observing the characteristics presented in the table is that the majority of collected data are old. This fact can bring some disadvantage in the development of instances, since more recent instances are, more they resemble the current state of surgical activity, but still due to the limited number of data sets available, these were the ones that could be collected. Regarding the type of data, 5 data generators, 2 generated data sets and 9 real data sets were collected. Although an effort was made to collect as much real data as possible, this task was also hampered by the small number

Table 1: List of articles and related references of the articles that provided the 14 data sets.

	Article	Reference
A	Optimizing a multiple objective surgical case sequencing problem	[8]
B	Robust surgery loading	[9]
C	Application of a patient flow model to a surgery department	[10]
D	A decision support system for cyclic master surgery scheduling with multiple objectives	[11]
E	Visualizing the Demand for Various Resources as a Function of the Master Surgery Schedule: A Case Study	[12]
F	Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources	[13]
G	Bi-Criteria Scheduling of Surgical Services for an Outpatient Procedure Center	[14]
H	Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning	[15]
I	*Anónimo*	-
J	Process Modeling of ICU Patient Flow: Effect of Daily Load Leveling of Elective Surgeries on ICU Diversion	[16]
K	Integrating Data Mining and Optimization Techniques on Surgery Scheduling	[17]
L	Planning and scheduling of semi-urgent surgeries	[18]
M	A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP)	[19]
N	Master surgery scheduling with consideration of multiple downstream units	[20]

of articles that use them and by the confidentiality agreements signed between researchers and hospitals. Even so, the number of real data is higher than those generated, which for the barriers found is positive. The generated data sets are also valuable, but compared to the real ones, have some disadvantages already mentioned. The data generators, despite not being instances in themselves, allow the generation of new sets and therefore new instances.

It was possible to gather a total of 262 instances, of which 37 are constituted by real data and 225 by generated data. This discrepancy is reasonable, since it is possible to produce a large number of sets in the data generation, like for example set A that is constituted with 224 instance that were generated to test its methodology. It is important to bear in mind that the actual data sets are directly extracted from documents provided by hospitals. For all the obstacles founded in the process, it is possible to say that the number of possible developed instances is remarkable.

In essence, it was concluded that the collection of data with the contacts made was successful due to the number of sets obtained and the variability of their origin. The number of possible instances resulting from this collection is high and sufficient for the creation of a consistent platform that will enable the use of the developed instances to compare existing methodologies in the field.

Table 2: Characteristics of the collected data (country, date, type of data and number of instances).

Data set	Country	Date	Type of data	Nº instances
A	Belgium	2005	Generated	224
B	Netherlands	2006	Generated	1
C	Italy	2008	Real	1
D	Belgium	2004	Real	1
E	Belgium	2004	Real	1
F	Netherlands	2011	Generator	-
G	USA	2006	Generator	-
H	Netherlands	2006	Generator	-
I	USA	2014	Generator	-
J	United Kingdom	2009	Real	1
K	Portugal	2011	Real	10
L	Netherlands	2007	Real	1
M	Italy	2010	Real	1
N	Netherlands	2008	Generator	-
O	Portugal	2008-2019	Real	14
P	Portugal	2015-2019	Real	7

3.2. Development of benchmark instances

3.2.1 Analysis of collected data

It was possible to collect a total of 16 sets from various researchers and from the two previously mentioned Portuguese hospitals. Each set is constituted by one or more instances and each one of these by several types of parameters. This way, it was necessary to previously analyze the type of data that were collected, in order to conclude whether they address the two OR planning problems (MSSP and EPSP), but also to examine their content and the type of data processing that it would be necessary to developed the various instances. This analysis is then presented for each of the data sets obtained.

Set A

Set A consists of 224 instances of generated data based on hospital information. Each one of the instances represents a list of planned surgeries, but the number of surgeries varies between instances despite the parameters being the same. For each surgery, the following parameters are presented: **type of surgery, expected surgery duration, Post-Anesthesia Care Unit (PACU) length of stay, bottleneck instruments and other information related to either the patient or the surgery**. Each instance was received in a *txt* file with only the values associated with each parameter and, in parallel, the correspondent author sent the code used to generate the instances, that allow to read them.

Set B

This was obtained from the author of the only existing open access *website* constituted with generated instances. In addition to his platform, he also sent a larger instance also constituted with generated data. This is representative of a list of surgeries with the following parameters: **surgery date, allocated OR, surgical specialty, planned surgery duration, standard deviation of surgery duration and surgery duration**. Additionally, a data generator with a distribution was collected to create new lists of performed surgeries with information about the surgical specialty and the mean duration and standard deviation of the duration. It was also made available the block availability per day and the probability of urgent surgeries arrival per specialty, which allows the generation of lists related to planned surgeries.

Set C

Composed only by an instance of real data provided by a hospital center. This instance is a list of 100 patients that focuses on the **surgery priority** and the **waiting times** at various stages of the entire surgical process (surgery scheduling, hospital admission, waiting for surgery and waiting for discharge) and also the sum of these times, which in some cases is invalid due to the lack of information about some waiting times. However, the number of this occurrences is not significant to discard the present instance.

Sets D and E

Sets D and set E were provided by the same author but correspond to two different data sets used in different articles. Set D is composed by an instance with several planned surgeries and associated information such as the **surgery date and day of the week, allocated ward and surgeon and length of stay in the hospital**. Set E is the link between the **list of available surgeons and the resources consumption by each one of them**, where the number of time blocks and shifts in which they are applied is shown.

Set F

Set F does not have a list of individual patients, but groups of patients that are divided according to their age, the level of surgery complexity and the Intensive Care Unit (ICU) and OR length of stay. For each group, the **arrival rate per day, planned surgery duration, average length of stay before surgery and ICU and Medium Care Unit (MCU) length of stay distributions** are presented. The **capacity of multiple resources** by day of the week is also available. This data set is presented in the literature [13] and was developed using real data provided by a hospital, but it can also be used to generate new data for individual patient lists.

Set G

Set G is also a data generator based on real hospital parameters, which is available in the respective article [14]. For each surgical specialty, the following parameters were provided: **number of surgeries, mean surgery duration, standard deviation of duration**, as well as the **distribution** associated with each stage of the surgical process. Distributions and associated parameters related to **patient transfer time between units** and the **turnover time between surgeries** are also presented. The existing parameters allow the creation of new lists for planning surgeries.

Set H

Set H is a data generator that despite having been submitted by a different author and referring to a different article, is quite similar to data set F, but with less information than this one. It was later understood that the authors who sent the two generators carried out the two articles together and these are based on the same hospital data. Thus, as set F is more complete than set H, this one will be excluded and will not be used for the development of instances.

Set I

It consists of a list of surgical specialties that for each of them presents **number of surgeries, average number of surgeries per week, average surgery duration, average turnover time between surgeries and standard deviation of the number of surgeries, duration and turnover time**. This data set was partially provided by a hospital and adapted by the investigators who used it for their article. Like set F it can be used in its original format or to create new data sets.

Set J

Set J consists of an instance composed by real data provided by a hospital. This contains the **patient inflow and outflow of 4 ICU** for one month. Despite being an instance that contains only one parameter, it was presented because it can be used to, for example, analyze the maximum number of surgeries that can be scheduled depending on the number of ICU available beds. Therefore giving the possibility of its use for this type of studies, increase the parameters variability and the methodologies that can be reached with the formulated instances.

Set K

Set K has 10 instances, where each one corresponds to a waiting list for each existing surgical specialty in a Portuguese hospital. Each list has several parameters such as **priority, surgical specialty, allocated surgeon, waiting time for surgery and planned duration of surgery**. Furthermore, the **block availability** is also referred to for each shift and each surgical specialty.

Set L

Set L was acquired through an article that studies the planning of semi-urgent surgeries. From this, only one table was picked with the **number of elective, urgent, semi-urgent, overtime or canceled surgeries** and with **OR capacity** per week. Since this was the only information available for sharing and this is quite limiting, is not enough to form an instance in itself and therefore the data set L will also be excluded.

Set M

It comprises an instance with approximately 400 planned surgeries and the respective parameters associated with each surgery: **surgery date, priority, surgical specialty, surgery duration and expected length of stay**.

Set N

Like some of the previous data generators, this one is available in the respective article from which it was taken and presents several parameters for a total of 7 surgical specialties. The parameters are: **expected number of operated patients, probability of ICU patient need, expected ICU length of stay and expected ward length of stay**. The present set is related and can be used for planning the Master Surgery Schedule taking into account the hospital units associated with the OR.

Set O

Set O was provided by one of the Portuguese hospitals, CHLN. It contains 14 instances with real surgical data that can be summarized in three different types: **waiting lists, planned surgeries lists and lists with *timestamps***. Waiting lists have the **surgery request date, patient's date of birth and gender, priority, type of surgery, surgical specialty and allocated surgeon**. Regarding the planned surgeries, it is available the **expected date for surgery and request date, allocated OR, priority, type of surgery, surgical specialty and allocated medical team**. Finally, the last surgery lists have several ***timestamps*** of the entire surgical process, from the time the patient enters the OR until the recovery. By analyzing the various existing documents, it was possible to observe that there were several data from surgeries that were repeated and, for this reason, it was necessary to exclude some instances of this set, resulting in only 10 instances in total.

Set P

Set P refers to data provided by the other Portuguese hospital (HESE), in which 7 instances were presented. Among these, the parameters provided are the same as those referred in the previous data set (O) and are also divided into three types of data: **waiting lists, planned surgeries lists and lists with *timestamps***. For the same reason mentioned above, some instances were excluded due to the existence of duplicate surgeries in them, leaving a total of 5 instances.

After analyzing the content of the 16 collected data sets, two of these (H and L) were excluded, leaving a total of 14 data sets to continue the development of instances. After this step, it was clear that some data processing would have to be carried out for the formation of the various instances, in order to standardize them. The following section describes this process, as well as the title and definition associated with each parameter included in the instances.

3.2.2 Standardization and development of instances

To develop the compilation of instances thru the 14 data sets, it is essential to create a uniform template to convert the various data sets into this model. For this, it was necessary to define a range of characteristics to be included in the format of each instance. It was defined that in all instances each column corresponded to a parameter and the lines to individual surgeries, surgical specialties or groups of patients, depending on the case. Any additional hospital information that was referent to all instances would not be included in each one of them but would be stated through the *website* or a separate document. Since the instances were converted into a CSV file, it was important that their organization was as clear and simple as possible. During the development of the instances in *Excel*, each document cannot be composed of more than one sheet, as this

is not compatible with the CSV format, and cannot exist joined cells due to the same reason.

Through the analysis carried out to the 14 collected data sets, the parameters that would be included in each instance were selected and defined, to improve consistency between the various developed instances. The following parameters were established:

- **Surgery duration:** corresponds to the time elapsed from the patient's OR entry until the exit, and is represented in minutes.
- **Planned surgery duration:** corresponds to the expected duration of the described procedure from the surgery duration, and is also represented in minutes.
- **Length of stay:** refers to the time the patient is at the hospital in days. Within each instance, it can also be specified the OR time, the ICU, MCU, PACU time or preoperative time.
- **Resources utilization:** by resources it can be intended material resources such as beds, instruments, devices and human resources such as surgeons, nurses, assistants. It can be displayed in number of resource or utilization hours.
- **Resources capacity:** which can refer to the same resources listed above and is presented in the same units.
- **Patient characteristics:** such as age, date of birth and gender.
- **Surgery characteristics:** such as type of surgery, date, day of the week and allocated OR, ward and surgeon.
- **Surgical specialty:** corresponds to the specialty with which the surgery is associated, such as general, cardiovascular or plastic surgery.
- **Priority:** each surgery has a defined priority, depending on the urgency associated with it. It usually ranges from 0 to 4, with 4 being the most priority surgeries.
- **Standard deviation of surgery duration:** degree of variation in the surgery duration, and is presented in minutes.
- **Mean surgery duration:** based on historical records, it is the average duration associated with each type of surgery or with surgeries in a specific specialty. Displayed in minutes.
- **Expected number of urgent surgeries:** it is based on historical records and represents the expected number of urgent surgeries to be received at the hospital, per week or day.
- **Block availability:** displays the number of hours each OR is available to perform surgeries.
- **Waiting time:** it can refer to the waiting time between the request and the day of surgery, but also the waiting time at the hospital, between entering the hospital and the surgery itself, or between surgery and hospital discharge. Displayed in days or minutes.
- **Arrival rate of patients:** number of patients entering the hospital per hour, day or week.
- **Transfer time between unit:** it is presented in minutes and represents the time to transfer a patient between the OR and its associated hospital units (PACU, ICU, for example).
- **Turnover time:** represents the time between the end of one surgery and the beginning of another, and is presented in minutes.
- **Mean number of performed surgeries:** based on historical records and represents the number of surgeries performed per day or week or per OR.
- **Expected number of performed surgeries:** based on historical records and represents the expected number of surgeries performed per day or week or per OR.
- **Standard deviation of number of performed surgeries:** indicates the degree of variation in the number of surgeries performed per day, week or OR.
- **Number of patients at ICU**
- **ICU probability:**probability that a patient will need to stay at ICU.
- ***Timestamps*:** temporal marks of various events in the surgical process, such as entering the OR or the beginning of anesthesia.

As was observed above, 23 parameters are extracted from the collected data sets. Although a general definition has been developed for each parameter, within each instance and associated supporting document, the parameters mentioned in each instance are specified, especially in the case of patient and surgery characteristics. Each parameter unit is also displayed in the supporting document.

Moreover is important to uniform the instances in terms of structure. Although some sets were already in the defined standardize format, in others it was necessary to make some changes. For example, in data set A, it was necessary to analyze the collected instances through the code used to generate them. Additionally, each instance was based on the type of surgery instead of being constituted by a list of surgeries, with changes being made, the parameters are now presented for each surgery rather than for each type of surgery. Data set K consisted of 10 instances, each one of them was a waiting list corresponding to only one surgical specialty and it was possible to gather the various waiting lists in only two instances. In other sets the data was received in a *txt* file, in which each line corresponded to a parameter, since it was previously defined the opposite, the change was made to correspond each column to a parameter. In addition, in sets O and P, as they were directly provided by hospitals, they were not organized like the remaining data, existing a high number of duplicates within the same instance and even between several instances. It was also common to find empty records, that didn't have enough information about the correspondent surgery. Therefore, even in these two sets, several instances were excluded for the reasons previously mentioned. In general, structural changes were carried out in all instances and some redundant information and parameters were excluded if they didn't fit into the study in question.

Regarding the data generators, they weren't considered as instances but are quite valuable as they allow the generation of new data and can also be adapted to a certain distribution. The parameters presented in each generator may not be enough to generate a complete instance, but it can be useful for some studies developed that need a specific type of data. Although most of them are already available in the literature, from now on it will be possible to find a compilation of these.

3.2.3 Analysis of developed instances

After analyzing and standardizing the aforementioned data sets, a total of 247 instances and 5 sets of data generators were created and compiled to be published. In table 3, the number of instances and data generators that resulted from each set is presented.

Table 3: Number of instances and data generators (represented by G) that were extracted from each data set.

	A	B	C	D	E	F	G	I	J	K	M	N	O	P
Nº Instances	224	1 + G	1	1	1	G	G	G	1	2	1	G	10	5

Analyzing the instances, 225 are made up of generated data and 22 are constituted with real data. In general, it was possible to gather instances consisting on lists about planned surgeries and others about performed surgeries. Within the lists of performed surgeries, several types were also identified with information like waiting times, patient flow or resources consumption for each surgery. Several instances were also built consisting on waiting lists of patients for surgeries and others composed on timestamps related to the surgical process.

Within the several types of developed instances, the parameters presented in each one of them were listed, to enable their analysis. Table 4 shows the correspondence between the various parameters and the data sets were they are located, as well as the percentage of utilization of each parameter in the created instances.

Figure 2 presents the percentages (in descending order) of utilization for the parameters in the created instances, to facilitate their analysis and comparison with the parameters mentioned from the literature review.

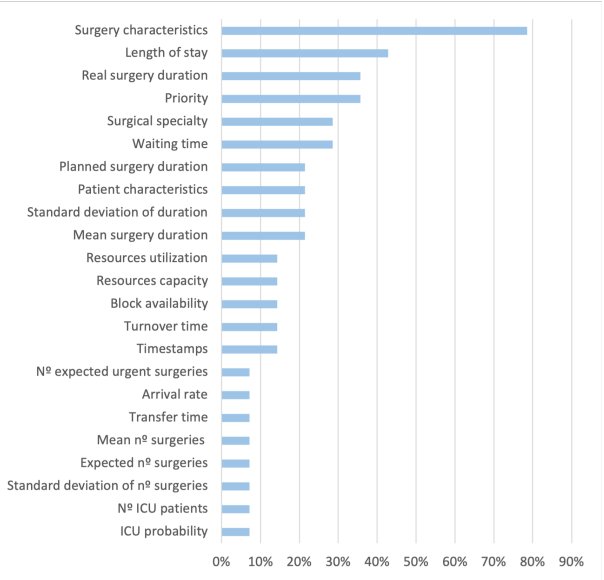

Figure 2: Parameters used in the developed instances and correspondent percentage of utilisation.

Various analyzes can be performed on the obtained results. Although the general objective of the dissertation is the compilation of several instances, within it there are other goals that are intended to be achieved. As previously mentioned, the compilation of as many instances as possible, incorporated with a high level of parameters variability, but also with the parameters most used in the literature, is useful for the study of MSSP and EPSP, related to the optimization of OR planning.

Analyzing the parameters presented in each set, it was possible to build several types of instances or generators with several types of data (table 4). Some are related to lists of surgeries already performed, such as set B, in which it is common to have the planned and real surgery duration, the length of stay and, more specifically, in sets O and P, which are made up of timestamps including the exact time an event occurred like the OR entry and exit or entry and exit from the hospital, as well as the beginning or end of anesthesia. This type of parameters is only accessible when surgeries have already taken place but are difficult to obtain through hospitals and therefore, even if it is a list of performed surgeries, it is more usual to know the expected surgery duration, a parameter that is defined in advance, than its real duration. Regarding the lists of planned (to be performed) surgeries, as in sets D and M, it is usual to refer to the average and expected surgery duration and other general patient and surgery characteristics. The third type of instances is the waiting lists, present in sets O and P, which include the waiting time for surgery, patient characteristics and surgery priority. Within the collected instances there are more parameters that represent the preoperative than the postoperative. Even so, several stages of the surgical process were addressed, as well as several key features of the surgery planning that could be applied to different studies.

The compilation contains a total of 23 parameters, being the most used: **surgery characteristics, length of stay, surgery duration, priority, surgical specialty and waiting time.** Among all parameters, there is a big difference mainly in the utilization percentage of the surgery characteristics, due to the fact that this parameter encompasses several information regarding the surgery like date and hour. The rest of the utilization percentages vary between 43% and 7%. Comparing with the analysis carried out in the literature, it was possible to gather in the instances the 15 most used parameters mentioned in the reviewed articles and were collected even more parameters that allowed to increase the level of versatility of the compilation. On the other hand, it was intended to collect a higher number of waiting lists, as they are commonly wanted by researchers to test their developed methodologies on scheduling waiting surgeries, but that was impossible to achieve.

3.3. Publication of benchmark instances

3.3.1 Sharing confirmation

After the completion of the instances, it was important to confirm again with the corresponding authors that their sharing could be carried out and some clarifications regarding the data set and some feedback was also asked. Thus, an email was sent to the 14 authors, with the developed instances attached, of which 12 confirmation were obtained. For the remaining two the email was still forwarded but still unanswered. As future work its is advised to try again to contact the two corresponding authors, in order to collect their confirmation, since the instances are already prompt to be shared. As a confirmation of these two has not been received so far, only the 12 data sets proceeded to the publication of the benchmark instances.

3.3.2 Format of instances

With the 12 data sets, the process of sharing and publishing the respective benchmark instances continued. For each set, a folder was created that includes all the instances associated in CSV files, that can be converted into all other existing formats.

A *txt* file was also added to each folder, in order to help the reading of the various instances but also with additional information, such as characteristics collected about the hospital in question and the reference to the article or hospital that provided the data set, when so is permitted. The purpose of this file is to facilitate the use of the various instances for entities that need them. They are all standardized, at the beginning it shows the number of instances that the data set has and its origin, if it was obtained through contact with authors or if it was from one of the mentioned Portuguese hospitals. Afterwards, some notes are left that can complement the reading of the instances to which the document refers. Furthermore, if there is additional information about the hospital from which the data comes, it is presented in this document as it is general to all of them and may be essential for their use. The various types of parameters are: **number of surgical specialties**, **number of ORs**, **available block time in hours**, **number of surgeons**, **number of**

Table 4: Correspondence between the selected parameters and the data sets where they are presented. The last column shows the percentage of utilization of each parameter in the 14 data sets.

	A	B	C	D	E	F	G	I	J	K	M	N	O	P	%
Surgery duration		X									X		X	X	**36%**
Planned surgery duration	X	X				X				X					**21%**
Length of stay	X			X		X					X	X		X	**43%**
Resources utilization	X				X										**14%**
Resources capacity					X	X									**14%**
Patient characteristics	X												X	X	**21%**
Surgery characteristics	X	X		X	X		X	X	X	X	X		X	X	**79%**
Surgical specialty		X								X	X	X			**29%**
Priority			X							X	X		X	X	**36%**
Standard deviation of surgery duration		X					X	X							**21%**
Mean surgery duration		X					X	X							**21%**
Expected nº of urgent surgeries		X													**7%**
Block availability		X								X					**14%**
Waiting time			X							X			X	X	**29%**
Arrival rate						X									**7%**
Transfer time							X								**7%**
Turnover time							X	X							**14%**
Mean nº surgeries								X							**7%**
Expected nº surgeries												X			**7%**
Standard deviation of nº surgeries								X							**7%**
Nº ICU patients									X						**7%**
ICU probability												X			**7%**
Timestamps													X	X	**14%**

ICUs, **number of wards** and **number of beds**. Finally, instructions are presented to reading each instance by defining each parameter that each column contains. If exists similar instances in each data set, their instructions are gathered and if not, they are described separately.

The standardization and the attempt to facilitate the reading and use of the instances is extremely important. It is relevant to bear in mind that the interpretation of each instance is more complicated for someone who does not build them but wants to use it, so the maximum effort was made to clarify all the parameters referred to, as well as all the information that was obtained through contact with the authors. After organizing the instances, a website was developed for the last stage of the project.

3.3.3 Development of website

To achieve the objectives proposed for the project, a website was developed with *WordPress*, to enable the publication of the various instances created. To the website was given the name *"Operating room planning and scheduling benchmark set"*.

The website created has four main pages: Home, Real data, Generated data and data generators. On the main page (Home) the developed project is presented, the main objectives to be achieved with the creation of the platform are enumerated and the available content and how it was developed is described. In each of the remaining pages (Real data, Generated data and Data generators) an explanation is made about their content, the number of available data sets and a brief explanation on how they were built. Within each of the pages there are several tabs and each tab corresponds to a set of instances. Within each separator, the number and origin of each set is mentioned, as well as a list of the parameters that contains. To have access to each set of instances, it is necessary to download the folder corresponding to each one of them. Finally, at the end of the website, the scope in which the platform was developed is also presented and an email address was also provided for contact in case there are questions regarding the platform or the instances, or to allow the sharing of more data sets.

After developing the website, it was necessary to define the domain to host it. *WordPress* itself provides a free domain but with some restrictions, that hinder the platform's intended operation. This prevents the users from downloading folders, making it impossible to share the developed folders for each set. Another disadvantage is the storage space, which is quite limited. Since *Instituto Superior Técnico* provides a domain for each student, giving the possibility to host a website, this option was taken. Using the domain provided by the university it is possible to overcome the two major restrictions mentioned above. However, it was not yet possible to do the transition between the domains and, for this reason, the website is not yet available.

Choosing a website for the publication of the developed compilation allowed to take the work carried out further, as it is an open access platform that is easily disseminated. This, due to the considerable number and the variability of the instances, is a starting point to fill the gap in the study of the optimization of OR planning regarding the lack of sharing surgical data and allowing the test and comparison of various methodologies and models developed in the field. Upon completion of the platform, it is essential that it be disclosed to the authors who provided their data sets, as well as other entities of interest.

4. Conclusions

The objective of this work is to compile data sets organized in benchmark instances, focused on parameters related to the OR planning and scheduling. This compilation intends to fill the small percentage of articles that test their methodologies after developing them, but also to allow a future comparison between the methodologies available in the literature. In these, preference was be given to real surgical data so that the test and comparison performed can be closer to real cases.

When performing a literature review in the scope of the OR planning, it is noticeable that it is rare to find available data sets to be used. So far, there is only one website that compiles several instances of generated data and a public drive also with a list of sets. Additionally, it was possible to know that exists a set of instances developed with real surgical data but which cannot be shared, being only used by a group of Dutch researchers. In the present work, two major barriers were found. The first concerns the privacy policies applied by hospitals, which hardly share data sets and when they do so for specific case studies, they also sign agreements that forbidden their disclosure, which

is one of the reasons for the unusual discovery of instances in the literature. The second barrier is associated with the parameters that function as inputs in the methodologies and, therefore, are the ones that compose the instances. For these, there is no standard model to be used in the studies, as well as a universal definition for each one of them.

With the aforementioned barriers and the current state of the art under consideration, several entities were contacted for data collection and the parameters of interest were defined. The data sets collected were standardized and transformed into benchmark instances that are intended to be published on an open access website that was developed.

In the first phase of the present work, 145 authors were contacted via email to provide their data sets, from which 54 responses were obtained and of those ones only 14 were positive. Two additional sets were added, one corresponding to each one of the Portuguese hospitals (CHLN and HESE), with whom a contact was made to collect their records. The collected data were analyzed in order to filter those that were within the scope of the two problems addressed (MSSP and EPSP), which led to 14 remaining sets that were used to formulate the instances.

With the 14 data sets used, it was possible to form 247 instances, of which 225 are constituted by generated data and 22 by real data. Additionally, 5 data generators were also collected, which despite not representing a complete instance, allow the generation of new data associated with the parameters that constituted them. The instances created are generally composed by lists of planned surgeries, lists of performed surgeries and waiting lists for surgery. Within those lists, was obtained 23 parameters and the most presents are: surgery characteristics, length of stay, surgery duration, priority, surgical specialty and waiting time. Additionally, the set of parameters used was defined and standardized to solve the previously mentioned barrier and tackle the uncertainty in the reading and use of instances.

Finally, after the development of the instances, they were edited to a uniform format and a document was developed for each set with additional information and to aid its reading. Later, a website was developed in *WordPress* for the publication of the created instances, which despite being finished, is not yet publicly accessible since it is not yet possible to host it in the domain provided by *Instituto Superior Técnico*.

In the end of carried out work, it is possible to consider that the objectives have been accomplished. Due to the scarce sharing of existing data sets in the literature, it was not expected to reach such a high number of data sets and consequent number of instances, and this aspect exceeded the expectations. The instances formed are composed by parameters of interest and commonly used in the study of the optimization of OR planning, and the versatile range that was possible to include is a bonus for the developed compilation.

The existence of benchmark instances does not contribute to the development of more optimization methodologies or models, but they allow their testing and comparison, raising the state of the art. As future work, regarding the developed platform, the hosting process will be finalized and will be disclosed to entities of interest. Additionally, more data sets may be collected to form more instances, mainly focused on real surgical data. In general, it will be beneficial to develop more platforms of this kind to promote cooperation between hospitals and researchers, allowing the progress of the literature as well as its application to the real world cases.

References

[1] D. A. Lungu, T. G. Ruggieri, and S. Nuti, "Decision making tools for managing waiting times and treatment rates in elective surgery," *BMC health services research*, vol. 19, no. 1, p. 369, 2019.

[2] Z. Abdelrasol, N. Harraz, and A. Eltawil, "Operating room scheduling problems: a survey and a proposed solution framework," in *Transactions on engineering technologies*, pp. 717–731, Springer, 2014.

[3] B. Cardoen, E. Demeulemeester, and J. Beliën, "Operating room planning and scheduling: A literature review," *European journal of operational research*, vol. 201, no. 3, pp. 921–932, 2010.

[4] A. Fügener, S. Schiffels, and R. Kolisch, "Overutilization and underutilization of operating rooms-insights from behavioral health care operations management," *Health care management science*, vol. 20, no. 1, pp. 115–128, 2017.

[5] E. Demeulemeester, J. Beliën, B. Cardoen, and M. Samudra, "Operating room planning and scheduling," in *Handbook of healthcare operations management*, pp. 121–152, Springer, 2013.

[6] M. Samudra, C. Van Riet, E. Demeulemeester, B. Cardoen, N. Vansteenkiste, and F. E. Rademakers, "Scheduling operating rooms: achievements, challenges and pitfalls," *Journal of scheduling*, vol. 19, no. 5, pp. 493–525, 2016.

[7] M. Bauer, R. Wäschle, J. Rüggeberg, H. Meyer, C. Taube, M. Diemer, and M. Schuster, "The german perioperative procedural time glossary a concerted recommendation of the german societies of anaesthesiology, surgery and operating room management," *ANASTHESIOLOGIE & INTENSIVMEDIZIN*, vol. 57, pp. 669–683, 2016.

[8] B. Cardoen, E. Demeulemeester, and J. Beliën, "Optimizing a multiple objective surgical case sequencing problem," *International Journal of Production Economics*, vol. 119, no. 2, pp. 354–366, 2009.

[9] E. Hans, G. Wullink, M. Van Houdenhoven, and G. Kazemier, "Robust surgery loading," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1038–1050, 2008.

[10] D. Antonelli and T. Taurino, "Application of a patient flow model to a surgery department," in *2010 IEEE workshop on health care management (WHCM)*, pp. 1–6, IEEE, 2010.

[11] J. Beliën, E. Demeulemeester, and B. Cardoen, "A decision support system for cyclic master surgery scheduling with multiple objectives," *Journal of scheduling*, vol. 12, no. 2, pp. 147–161, 2009.

[12] J. Beliën, E. Demeulemeester, and B. Cardoen, "Visualizing the demand for various resources as a function of the master surgery schedule: A case study," *Journal of medical systems*, vol. 30, no. 5, pp. 343–350, 2006.

[13] I. Adan, J. Bekkers, N. Dellaert, J. Jeunet, and J. Vissers, "Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources," *European Journal of Operational Research*, vol. 213, no. 1, pp. 290–308, 2011.

[14] S. Gul, B. T. Denton, J. W. Fowler, and T. Huschka, "Bicriteria scheduling of surgical services for an outpatient procedure center," *Production and Operations management*, vol. 20, no. 3, pp. 406–417, 2011.

[15] I. Adan, J. Bekkers, N. Dellaert, J. Vissers, and X. Yu, "Patient mix optimisation and stochastic resource requirements: A case study in cardiothoracic surgery planning," *Health care management science*, vol. 12, no. 2, p. 129, 2009.

[16] A. Kolker, "Process modeling of icu patient flow: effect of daily load leveling of elective surgeries on icu diversion," *Journal of medical systems*, vol. 33, no. 1, pp. 27–40, 2009.

[17] C. Gomes, B. Almada-Lobo, J. Borges, and C. Soares, “Integrating data mining and optimization techniques on surgery scheduling,” in *International Conference on Advanced Data Mining and Applications*, pp. 589–602, Springer, 2012.

[18] M. E. Zonderland, R. J. Boucherie, N. Litvak, and C. L. Vleggeert-Lankamp, “Planning and scheduling of semi-urgent surgeries,” *Health Care Management Science*, vol. 13, no. 3, pp. 256–267, 2010.

[19] E. Tànfani and A. Testi, “A pre-assignment heuristic algorithm for the master surgical schedule problem (mssp),” *Annals of Operations Research*, vol. 178, no. 1, pp. 105–119, 2010.

[20] A. Fügener, E. W. Hans, R. Kolisch, N. Kortbeek, and P. T. Vanberkel, “Master surgery scheduling with consideration of multiple downstream units,” *European journal of operational research*, vol. 239, no. 1, pp. 227–236, 2014.