Noname manuscript No.
(will be inserted by the editor)

Chemotherapy Appointment Scheduling and Daily Outpatient–Nurse Assignment

Received: date / Accepted: date

Abstract Chemotherapy planning and patient–nurse assignment problems are complex multiobjective decision problems. Schedulers must make upstream decisions that affect daily operations. To improve productivity, we propose a two-stage procedure to schedule treatments for new patients, to plan nurse requirements, and to assign the daily patient mix to available nurses. We develop a mathematical formulation that uses a waiting list to take advantage of last-minute cancellations. In the first stage, we assign appointments to the new patients at the end of each day, we estimate the daily requirement for nurses, and we generate the waiting list. The second stage assigns patients to nurses while minimizing the number of nurses required. We test the procedure on realistically sized problems to demonstrate the impact on the cost effectiveness of the clinic.

Keywords Chemotherapy · Outpatients · Nurses · Appointments · Scheduling · Daily Assignment

1 Introduction

Assigning appointments to chemotherapy patients is a challenging task. Chemotherapy treatments are administered according to a regimen that maximizes the effect of the drugs on the cancer cells and minimizes the impact on healthy tissue. A regimen consists of a set of cycles that define chemotherapy appointments [1]. All the appointments, after the first one, depend on the regimen and resource availability. Usually rest days between appointments are fixed. Depending on the medication and the condition of the patient, appointments may be scheduled one or two days earlier or later than the recommended day; see, e.g., [14]. Only physicians can authorize such deviations [2], which are discouraged since they lead to patient dissatisfaction and may affect health outcomes.

The duration of each treatment, the acuity level of the patient, and the resources needed differ for each appointment. The duration is often based on the average of historical data for patients under the same regimen [2,15]. Most cancer treatment centers (CTC) use acuity levels to quantify the total workload. The acuity level defines the number of operations a nurse performs and the risk of complications [8]. Therefore, the level of nursing depends on the regimen. Nurses can monitor multiple patients when supervising chemotherapy appointments; this number depends on the medical condition of the group of patients. Each patient also requires at least one chair or bed.

The limited resources (chairs, lab technicians, pharmacists, oncologists, nurses) directly impact the delay between a patient's arrival in the system and the date of his/her first chemotherapy appointment. For example, Woodall et al. [18] show that nurse unavailability leads to inefficient patient flow; they demonstrated that only the limiting resources must be considered.

Chemotherapy appointment scheduling problems are complex multiobjective problems with conflicting criteria. The goals may include minimizing the number of days between the consult with the oncologist and the starting date of the treatment, reducing the in-clinic waiting time, balancing workloads among nurses, and minimizing overtime. Schedulers make decisions as patients arrive (without reconsidering previous decisions) which adds to the challenge. The decision makers try to balance care, health, and cost objectives [4], while, taking into consideration among others the mix of resources required to treat patients (time, nurse, pharmacists, etc).

The complexity of the problem usually leads to oversimplification because of the difficulty of schedulers to make multi-dimensional decisions. In most chemotherapy centers, schedulers only consider the maximum number of patients scheduled per day, the availability of oncologists or the availability of chairs. The most important variable is the schedulers' experience. Even thus, inefficiencies are inevitable because of last-minute changes: cancellations lead to an underutilization of resources, whereas adds-on and nurse absences lead to overtime.

Assigning appointments and resources to new patients makes the problem hard to solve with a single optimization model [9]. Turkcan et al. [15] were the first to propose a mathematical formulation that uses the solution to the planning problem (which assigns dates for the treatments for new patients without changing past decisions) as an input to the daily scheduling problem (which assigns a starting time, a chair, and a nurse). Furthermore, they introduced the concept of acuity level, limiting the total daily level based on the expected number of nurses and adding a constraint in the scheduling problem to limit the hourly workload for each nurse. They demonstrated that solving the problem at a tactical and operational level smooths the distribution of the daily workload; it decreases in-clinic waiting time, overtime, and treatment delays.

To respond to the requirements of CTCs, we present an online two-stage procedure to balance cost and health objectives. Our formulation is common to many CTCs although we were introduced to the problem by the oncology clinic of the Centre Hospitalier de l'Université de Montréal (CHUM) in Canada. The contribution of this paper is to incorporate the inherent uncertainty due to last-minute cancellations. As in [15], our first stage assigns, on daily basis, a date and time to treatments for new patients, while minimizing the number of nurses needed. We also introduce a waiting list for patients that might be scheduled due to last minutes cancellations. These patients are assigned to a "virtual" nurse. Our second stage solves the daily assignment problem for the final list of patients, again by minimizing the number of nurses. Moreover, we set a starting date and a recurring appointment time for all the appointments during the first stage; as this is comforting for the patients. We compare our approach to the first-come-first-served algorithm inspired by the current practice and the offline algorithm, which solves the problem once with the complete knowledge about all patients. To quantify the potential gains, we measure the productivity ratio, the nurse-power needed, the average daily workload assigned to the waiting list, and the number of patient starting late. We define the productivity ratio as the ratio between the number of hours of treatments provided and the number of available hours (i.e, worked by all nurses). The nurse-power represents the total labor force required to run the clinic. We evaluate the average daily workload by using the scale provided by the chemotherapy clinic. Finally, we measure lateness in respect to the prescribed delay to start treatment.

The paper is organized as follows. Section 2 presents a review of related studies, Section 3 gives the mathematical models for the planning and assignment problems, it also presents the simulation model used to generate last minute changes. Section 4 provides the different patient booking practice. Section 5 presents the evaluation of our models and algorithms, Section 6 discusses managerial insights, and Section 7 provides concluding remarks.

2 Related Work on Chemotherapy Appointment Scheduling

CTCs usually include several units: (1) Reception, (2) Vitals, (3) Laboratory, (4) Medical Consultation, (5) Pharmacy, and (6) Treatment Administration. Each unit requires a different combination of resources, and patients are processed by one or more units. Liang et al. [11] distinguish three patient types: medical consultation only, treatment administration only, and medical consultation with treatment administration. Lamé et al. [10] add the blood-testing step that typically occurs before each appointment, resulting in six patient types. All chemotherapy patients require a series of appointments characterized by precedence constraints and rest periods between treatments. Figure 1 illustrates the complex process of scheduling a patient.

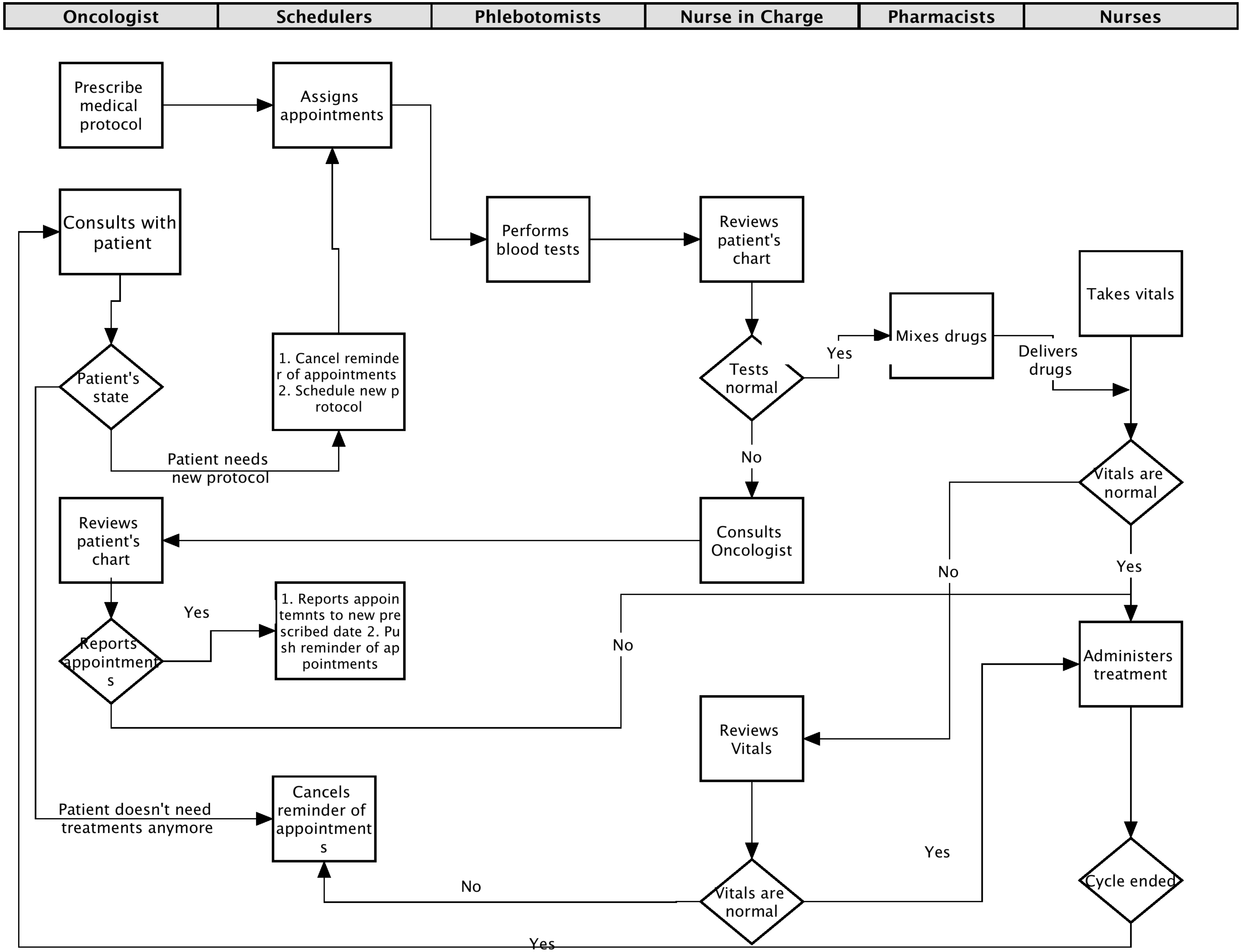

Fig. 1: Steps that lead to appointments: tasks completed and decisions made before administering chemotherapy treatments

There has been little research on the scheduling of multiple appointments over a large horizon; the focus has been on daily scheduling problems, which are reviewed in [10]. For larger horizons, there are two types of studies:

- **Next-Day Chemotherapy:** Blood tests and medical consultations are scheduled at least one day prior to the chemotherapy appointment. This helps to decrease the in-clinic waiting time and to improve the efficiency of pharmacists and nurses [6].
- **Same-day Chemotherapy:** Patients undergo all the procedures on the same day. This allows for an integrated process and reduces drug waste [16].

2.1 Next-Day Chemotherapy

Multi-stage methods are typically used to solve next-day chemotherapy problems. The first stage sets the date and/or time of the first appointment; the second stage solves the patient–resource assignment problem. Turckan et al. [15] were the first to formulate a two-stage model; they assign patients to a day by considering the medical protocol, acuity level, and resource requirements. They demonstrate that only limiting resources must be considered. Alvarado et al. [3] use a two-stage mean-risk stochastic programming model to set the date and time while considering the material capacity and the uncertainty related to acuity level, appointment duration, and availability of nurses. The first objective minimizes deviations from the recommended starting date, and the second minimizes scheduling conflicts. Condotta and Shakhlevich [5] develop a multi-stage method. The first stage assigns starting dates using multiday templates; the second stage assigns patients to nurses using intraday templates; and the third stage reassigns patients to reduce schedule conflicts. The templates are generated by solving an integer programming model for a set of artificial patients.

Only two studies solve the problem in a single stage. Gocgun and Puterman [7] use a Markov decision process, minimizing the cost of deviating from the specified date. They propose an approximate dynamic programming approach and a heuristic to test different scheduling policies such as allocating patients to the earliest date. Finally, Wong [17] allows deviation from the medical protocol but tries to maximize adherence to it by minimizing the completion date of all the appointments and by minimizing the number of times a patient is not assigned to his/her primary nurse.

2.2 Same-Day Chemotherapy

CTCs offer more than one service (medical consultation, blood testing, etc.), but mathematical formulations that take into account multiple CTC services are uncommon. Sadki et al. published two papers that consider joint medical and chemotherapy appointments: [12] sets the appointment period and the doctor's schedule with the goal of balancing the workload among beds, and [13] allocates medical and chemotherapy appointments with the aim of controlling the workload over the planning horizon. Finally, Liang et al. [11] solve a similar problem and generates templates for doctor and chemotherapy appointments that can be used to allocate appointments to incoming patients.
In conclusion, we note the lack of focus on the resolution of the appointment assignment problem as well as the daily operational problems, except for [3, 5, 15]. We also note that very few studies consider last minute changes. We focus our study on a Next-Day setting. This type of setting is the most common, since medical visits and blood tests are usually operated by different departments of the clinic. Dividing the problem also allows for more flexibility, and permits adjustments in regard to last minute changes.

3 Methodology and Mathematical Formulation

In our model, we consider the chemotherapy appointment scheduling problem and the daily operational schedule of the clinic. We consider the set of new patients that arrive every day. We classify patients in respect to the regimen prescribed: this dictates the hourly workload, the daily workload, the complexity, the time requirements and the appointments patterns. We also consider nurses' ability to absorb the hourly and daily workloads: each has a limited hourly and daily capacity. Figure 2 provides an overview of this procedure.
The first stage solves the **planning problem** at the tactical level. This stage is subdivided into two problems: (1) Appointment scheduling problem (assign a date, and a time to incoming patients) (P1-Problem) and (2) Staff planning problem (evaluate the daily nurse-power) (P2-Problem). P1-problem schedules patients to fit within the overall capacity of the clinic. P2-problem evaluates the number of required nurses, and assigns the patients to nurses or the waiting list. For example, in a clinic with only 2 nurses, P1-problem would schedule patients with respect to the maximum number a nurse can supervise per day. P2-problem would assign patients to the nurses with respect of the allowed mix of patients following the guideline established by the clinic. For example, a maximum of 1 patient of high

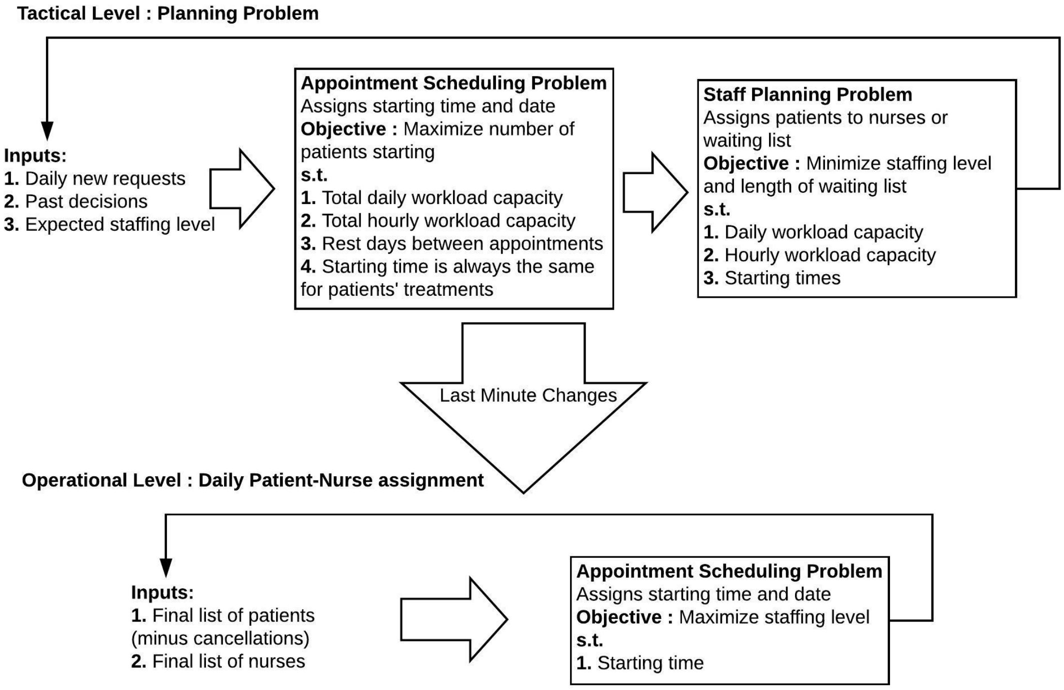

Fig. 2: Overview of the two-stage procedure.

complexity can be assigned to each nurse. Once the capacity of each nurse is reached, patients are assigned to the waiting list.

In the second stage, we solve the daily Patient-Nurse Assignment Problem at the **operational level**. Between two treatments, patients complete a blood test to determine if the next appointment can go ahead. After each cycle, patients typically consult with an oncologist to follow on the progress of the treatment. Four outcomes are possible: treatment continues as arranged; the next appointment is canceled and reported; a new prescription is prepared; or treatment is ceased. Last-minute changes are a result of this distinctive characteristic of chemotherapy. We use a simulation model at this step to include patient cancellations and nurse absences.

The second stage assigns the updated list of patients to the updated set of nurses scheduled to work on that day. At this stage, we relax the capacity constraint. This allows us to assign all patients, including those in the waiting list. This is common practice to avoid canceling appointments.

This procedure is launched at the end of each day for the set of new patients who arrived. The planning horizon of the problem is defined as the maximum recommended delay between the day of diagnosis and the first treatment. This delay varies from a clinic to another. However, it is usually 10 working days [2]. For each request, we define the set of treatments, the length, the acuity level, the hourly workload, the daily workload and, the list of characteristics. We also define the starting interval, and the priority level. The CTC has a limited human and material resources. One of the bottlenecks is the number of nurses, which varies during the day because of break periods. The number of chairs/beds is also limiting, and pharmacists are required to meet with all new patients to go over their medications.

Next sections present the formulations for each one of these three problems.

3.1 Appointment Scheduling Problem (P1-Problem)

We solve the problem at the end of every day for the set of new patients without changing past decisions. The first step considers the overall availability of nurses, chairs and pharmacists. We use this information to limit the workload assigned to each day with the objective to maximize the number of patients starting their treatment within the rolling horizon. The starting day and time will both be provided.

Parameters
J: Rolling horizon (days)
H: Number of time slots per day
L: Set of material resources (chairs and beds)
R: Set of new requests
S: Set of already-scheduled requests
P: Set of all patients, $P = \{S \cup \mathrm{R}\}$
A: List of patient characteristics
V_l: Number of available chairs and beds per day
I^p: Set of tasks for patient $p \in P$
$a^p_{(i+1),i}$: Rest days between two consecutive appointments $i \in I^p$
D^p_i: Duration of appointment $i \in I^p$
C^p_i: Acuity level of appointment $i \in I^p$
L^p_{li}: Number of chairs/beds needed for appointment $i \in I^p$
G^p_i: Hourly workload of appointment $i \in I^p$
W^p_i: Daily workload of appointment $i \in I^p$
S^s_i: Starting time of appointment $i \in I^s$
B^s_i: Scheduled date of appointment $i \in I^s$
e^r: Priority level of patient $r \in R$
$[a^r; b^r]$: Prescribed starting interval (days) for patient $r \in R$
P^p_{ai}: 1 if task $i \in I^p$ has characteristic a; 0 otherwise
M_a: Daily upper bound on the number of treatments with characteristic $a \in A$
T_j: Expected number of nurses per day
Q_{jh}: Expected number of nurses per day and time slot
N: Maximum hourly workload per nurse per time slot
W: Maximum daily workload per nurse per day

Variables
x^p_{ij}: 1 if appointment $i \in I^p$ is scheduled on day j; 0 otherwise
y^p_i: 1 if appointment $i \in I^p$ is accomplished during planning horizon; 0 otherwise
u^p_{ijh}: 1 if appointment $i \in I^p$ is scheduled to start in time slot h on day j; 0 otherwise

Model

$$max \sum_{r\in R}\sum_{i\in I^r}\left(e^r y^r_1 + \sum_{j\in J}\frac{1}{j}x^r_{ij}\right) \tag{1a}$$

s.t.

$$x^s_{ij} = 1 \quad \forall i \in I^s, \forall s \in S, j = B^s_i \tag{1b}$$

$$u^s_{ijh} = 1 \quad \forall i \in I^s, \forall s \in S, j = B^s_i, h = S^s_i \tag{1c}$$

$$\sum_{p\in P}\sum_{i\in I^p}\sum_{h'=max(1,h+1-D^p_i)}^{min(h,H+1-D^p_i)} L^p_{il}u^p_{ijh'} \le V_l \quad \forall l \in L, \forall j \in J, \forall h \in H \tag{1d}$$

$$\sum_{p\in P}\sum_{i\in I^p}\sum_{h'=max(1,h+1-D^p_i)}^{min(h,H+1-D^p_i)} G^p_i u^p_{ijh'} \le NQ_{jh} \quad \forall j \in J, \forall h \in H \tag{1e}$$

$$\sum_{p\in P}\sum_{i\in I^p} u^p_{ijh} \le Q_{jh} \quad \forall j \in J, \forall h \in H \tag{1f}$$

$$\sum_{p\in P}\sum_{i\in I^p} P^p_{ai}x^p_{ij} \le M_a \quad \forall a \in A, \forall j \in J, \forall h \in H \tag{1g}$$

$$\sum_{p\in P}\sum_{i\in I^p} W^p_i x^p_{ij} \le WT_j \quad \forall j \in J \tag{1h}$$

$$\sum_{j\in J}\sum_{h\in H} hu^p_{ijh} + D_i \le |H| \quad \forall i \in I^p, \forall p \in P \tag{1i}$$

$$y^r_1 a^r \le \sum_j jx^r_{1j} \le y^r_1 b^r \quad \forall r \in R \tag{1j}$$

$$y^p_i \le y^p_{(i-1)} \quad \forall i \in \{I^p | i > 1\}, \forall p \in P \tag{1k}$$

$$y^p_i\left(\sum_{j\in J} jx^p_{ij} - \sum_{j\in J} jx^p_{(i-1)j}\right) = y^p_i a^p_{(i-1)i} \quad \forall i \in \{I^p | i > 1\}, \forall p \in P \tag{1l}$$

$$y^p_i\left(\sum_{h\in H} h\sum_{j\in J} u^p_{ijh} - \sum_{h\in H} h\sum_{j\in J} u^p_{(i-1)jh}\right) = 0 \quad \forall i \in \{I^p | i > 1\}, \forall p \in P \tag{1m}$$

$$\sum_{j\in J} x^p_{ij} = y^p_i \quad \forall i \in I^p, \forall p \in P \tag{1n}$$

$$\sum_{j\in J}\sum_{h\in H} ju^p_{ijh} = \sum_{j\in J} jx^p_{ij} \quad \forall i \in I^p, \forall p \in P \tag{1o}$$

$$x^p_{ij} \in \{0,1\} \quad \forall i \in I^p, \forall j \in J \tag{1p}$$

$$y^p_i \in \{0,1\} \quad \forall i \in I^p \tag{1q}$$

$$u^p_{ijh} \in \{0,1\} \quad \forall i \in I^p, \forall j \in J, \forall h \in H \tag{1r}$$

The objective maximizes the total number of scheduled appointments. The first term e^r is the priority level of new requests. The second term sums the number of scheduled appointments for each patient $r \in R$. The term $\frac{1}{j}$ discriminates between requests with the same importance, and the same number of treatments. It also ensures that appointments are assigned as soon as possible.

The model allocates appointments to new patients without changing past decisions (1b), (1c). Constraint (1d) limits the maximum number of chairs and beds occupied in each time slot. Constraint (1e) limits the appointments per time slot by summing over the nursing care (G_i^P) required. Nurses are permitted to set up one patient per time slot (1f), so the number of appointments started in a time slot is bounded by the expected number of nurses Q_{jh}. Appointments must be finished before closing time (1i). We limit the number of treatments depending on their characteristics (1g). Finally, Constraint (1h) limits the daily workload. The acuity level is a good representation of the workload, but it does not consider the type of treatment or the fact that two regimens can have the same acuity level but different daily workloads. We developed a weight table based on the CTC scheduling rules to better estimate the workload (see Appendix A). The requirements of the prescribed regimen are guaranteed by Constraints (1j) to (1m). Constraint (1j) ensures that the first appointment for all new requests $r \in R$ starts in the recommended interval $[a^r; b^r]$. Constraint (1l) ensures that precedence rules and rest days between consecutive tasks are satisfied, and Constraint (1m) ensures that a patient's appointments occur at the same time. The linearization of (1l) and (1m) is presented in Appendix B. Finally, the links between x_{ij}^p and y_i^p and between x_{ij}^p and u_{ijh}^p are imposed by constraints (1n) and (1o).

3.2 Nurse Planning Problem (P2-Problem)

The objective of this step is to evaluate the level of nurses necessary everyday. Usually, all appointments are scheduled at least one week in advance, except for last minute add-ons and cancellations. We use the following formulation to assign appointments to nurses, minimizing the number of nurses without exceeding the nurse capacities. The maximum number of nurses corresponds to the budgeted staff. For example, the CHUM schedules in average between 10 and 14 nurses each day. In this problem, we generate a schedule for at most 14 nurses. The unassigned patients are added to the waiting list.

Parameters
H: Number of time slots
F: Set of available nurses on day j, $f = \{1...virtual\}$, and $F^- = F \setminus \{virtual\}$
K: Set of appointments to be completed on day j
S^k: Starting time of appointment k
C^k: Acuity level of appointment k
D^k: Duration of appointment k
G^k: Hourly capacity needed to complete appointment k
W^k: Daily capacity needed to complete appointment k
N: Maximum level of workload per nurse per time slot
W: Maximum level of workload per nurse per day
P_a^k: 1 if appointment k has characteristic a; 0 otherwise
M_{af}: Maximum number of appointments with characteristic a nurse $f \in F^-$ can handle.

Variables
z_{fh}^k: 1 if appointment k is assigned to nurse f at time h; 0 otherwise
v_f: 1 if nurse f is assigned; 0 otherwise
w_{fh} 1 if nurse f is on break in time slot h; 0 otherwise
τ_f: Integer variable: number of tasks handled, $\{0...E\}$
σ_f: Integer variable: first overflow level, $\{0...B\}$
ι_f: Integer variable: second overflow level, $\{0...C\}$

The objective function minimizes the number of nurses needed, the number of extra patients assigned to each nurse which correspond to the different level of overflow (E,B,C), and the number of patients assigned to the virtual nurse ($\alpha < \beta < \gamma < \eta < \epsilon$). The figure bellow illustrates the structure of the objective function.

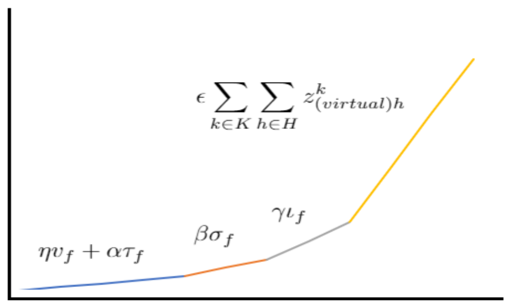

Fig. 3: Cost Structure of the Objective Function

Constraints (2b) and (2c) ensure that decisions made by the scheduling problem are unchanged, that each task is assigned, and that nurses set up at most one patient per time slot. Constraints (2e), and (2f) ensure that the patient mix is satisfied. Constraints (2g) and (2h) control the maximum hourly workload and the daily workload.
A nurse is entitled to a one-hour lunch break starting at 11 a.m., 12 p.m., or 1 p.m. The working hours are divided into 30-minute time slots, so 11 a.m. corresponds to the seventh period, 12 p.m. to the eighth, and so on. The number of nurses on break at the same time is one third of the total number of nurses present.

When nurses are on break, their patients are attended to by other medical staff. Constraints (2m) to (2r) ensure that these restrictions are satisfied. Moreover, a nurse on break is unable to set up a patient (Constraint (2s)). Finally, to break the symmetry of the model due to the equivalence between nurses, we add Constraint (2i).

Model

$$min \sum_{f \in F^-} (\eta v_f + \alpha \tau_f + \beta \sigma_f + \gamma \iota_f) + \epsilon \sum_{k \in K} \sum_{h \in H} z^k_{(virtual)h} \quad (2a)$$

s.t.

$$\sum_{f \in F} z^k_{fh} = 1 \quad \forall k \in K, h = H^k \quad (2b)$$

$$\sum_{k \in K} z^k_{fh} \leq 1 \quad \forall f \in F^-, \forall h \in H \quad (2c)$$

$$\sum_{h \in H} z^k_{fh} \leq v_f \quad \forall k \in K, \forall f \in F \quad (2d)$$

$$\sum_{k \in K} \sum_{h \in H} z^k_{fh} = \tau_f + \sigma_f + \iota_f \quad \forall f \in F^- \quad (2e)$$

$$\sum_{k \in K} \sum_{h \in H} P^k_a z^k_{fh} \leq M_{af} \quad \forall a \in A, \forall f \in F^- \quad (2f)$$

$$\sum_{k \in K} \sum_{h'=max(1,h+1-D^k)}^{min(h,H+1-D^k)} G^k z^k_{fh'} \leq N \quad \forall f \in F^-, \forall h \in h, \forall j \in J \quad (2g)$$

$$\sum_{k} \sum_{h} W^k z^k_{fh} \leq W \quad \forall f \in F^- \quad (2h)$$

$$v_{f+1} \leq v_f \quad \forall f \in F^- \quad (2i)$$

$$v_f \in \{0,1\} \quad \forall f \in F \quad (2j)$$

$$z^k_{fh}, w_{fh} \in \{0,1\} \quad \forall f \in F, \forall h \in H \quad (2k)$$

$$\tau_f, \sigma_f, \iota_f \in \mathbb{N} \quad \forall f \in F \quad (2l)$$

Distribution of Lunch Breaks

$$w_{fh} + w_{f(h+2)} + w_{f(h+4)} = v_f \quad h = 7, \forall f \in F^- \quad (2m)$$

$$w_{fh} + w_{f(h+2)} + w_{f(h+4)} = v_f \quad h = 8, \forall f \in F^- \quad (2n)$$

$$hw_{fh} + (h+2)w_{f(h+2)} + (h+4)w_{f(h+4)} \leq (h+1)w_{f(h+1)} + (h+3)w_{f(h+3)} + (h+5)w_{f(h+5)} \quad h = 7, \forall f \in F^- \quad (2o)$$

$$0 \leq w_{f(h-1)} - w_{fh} + w_{f(h+1)} \leq 1 \quad \forall h \in H, \forall f \in F^- \quad (2p)$$

$$0 \leq hw_{fh} + (h+1)\, w_{f(h+1)} \leq 1 \quad \forall h \in H, \forall f \in F^- \quad (2q)$$

$$\sum_{f \in F} w_{fh} \leq \frac{1}{3} \sum_{f \in F^-} v_f \quad \forall h = \{7,9,11\} \quad (2r)$$

$$w_{fh} + z^k_{fh} \leq 1 \quad \forall f \in F^-, \forall k \in K, \forall h = \{7, ...,12\} \quad (2s)$$

3.3 Simulation Model of Cancellations and Absences

Between the tactical and the operational level, one needs to adjust to consider sources of uncertainty. As explained earlier, two components need to be considered: patient cancellations and nurse absences. We build a simulation model to generate these events and the output of the simulation is used as input for the PNA–problem (3.4). Figures 6 and 7 in Appendix illustrate the procedure used to generate cancellations and absences.

Cancellations Patients arrive in the system on the date and time allocated by model (3.1). A decision module then determines if their appointments will be canceled; the probability distribution is a TRIA(4, 13, 20)%. This distribution is based on the CTC data. The outcome of canceled patients is decided by the second decision module:

- Outcome 1: All subsequent treatments are delayed one week (probability 90%);
- Outcome 2: All subsequent treatments are canceled.

Absences Nurses are created as a function of their assigned schedules. A decision module then determines if they are available or absent. The probability of absence is set to 1%.

3.4 PNA-Problem: Daily Patient–Nurse Assignment Problem

The model solves the daily–assignment problem for the final set of patients and nurses. It uses as input the solution given by the first phase as described in Figure 2. The mathematical model is similar to the planning problem, but we modify the objective function (3a) and the set of nurses. First, we delete the virtual nurse. Second, we use the value of v_f as an input to select the set of working nurses, F. Finally, we modify constraints (2g) and (2h). We increased the terms Ψ and χ iteratively until a feasible solution was found. In practice, planners prioritize a well distributed workload. They also "try to never" exceed more than 4 patients per hour or 8 per day. Therefore, after discussion with the personnel of the clinic, we relax the workload constraints to accommodate all patients. The terms Ψ and χ respectively permits us to relax the constraints related to the hourly workload, and daily workload. Since the total clinics workload is controlled in (3.1), we ensure that nurses' have a reasonable hourly and daily workload.

Model

$$min \quad \sum_{f \in F} v_f \tag{3a}$$

s.t.

$$\sum_{k \in K} \sum_{h'=max(1,h+1-D^k)}^{min(h,H+1-D^k)} G^k z^k_{fh'} \leq \Psi N \quad \forall f \in F, \forall h \in h, \forall j \in J \tag{3b}$$

$$\sum_k \sum_h W_k z^k_{fh} \leq \chi W \quad \forall f \in F \tag{3c}$$

$$z^k_{fh}, v_f, w_{fh} \in \{0,1\} \tag{3d}$$

$$\tau_f, \sigma_f, \iota_f \in \mathbb{N} \tag{3e}$$

4 Algorithms

In this section we introduce an offline algorithm that will allow us to evaluate the quality of our booking strategy. We also present a first-come-first-served (FCFS) algorithm inspired by the current practice, and the proposed online algorithm that allows to schedule the set of patients which arrive daily at the CTC. The algorithms have a similar structure, first we assign patients' treatments to a nurse, a date, and a time, second we evaluate the level of staff required, third we simulate cancellations, finally we built the daily schedule. We collect results after each step. We run all algorithm for the same time horizon to be able to compare our results (see Section 5 for more details on the design of experiments).

4.1 Offline

The offline algorithm (see Algorithm 1) generates a solution for the offline problem at hand with all the data. The problem is formulated as an mixed integer problem, it assigns appointments and a nurse to each patient. We input the complete set of patients to solve the scheduling and solve it for the time horizon. The objective function minimizes the number of nurses and the number of patients assigned to the virtual nurse (waiting list) (see appendix C). The set constraints are as for the P1-problem (3.1) and the P2-problem (3.2); the mathematical formulation varies slightly. The second step of the algorithm simulates cancellations; the last step solves the PNA-Problem which generate daily schedules. We use the simulation model in this experiment to simulate last minute cancellations and to represent as closely as possible what would happen in practice. Algorithm 1 generates the solution for the offline problem with all information known in advance, the simulation introduces last minute cancellations. The last phase solves the daily assignment problem with the goal of finding a solution that uses less resources than the one generated in phase 1.

Algorithm 1 Offline Algorithm

1: P = Generate patient mix
2: Solve offline problem (C)
3: Simulate cancellations and absences (3.3)
4: Classify appointment $i \in I^p$ by date $j \in PP$
5: **for** $j \in PP$ **do**
6: Solve PNA-problem (3.4)
7: **end for**

4.2 First Come First Served

The FCFS algorithm is inspired by the current situation at the CTC, where schedulers assign appointments to patients as they arrive. The patient is assigned to the first available slot, and all his/her appointments are at always the same time. To decide if a time slot is available, the schedulers consider the following:

1. **Chair/Bed Capacity:** the number of chairs and beds occupied in each time slot must not exceed the number available.
2. **Number of Patients:** this is limited by the average specified by the CTC.
3. **Pharmacist Capacity:** the number of new patients must not exceed the daily limit of the pharmacists.

At this stage, the schedulers do not consider the patient mix because of a lack of information. The exact number of nurses available is not yet known. Moreover, it is not efficient to assign patients to nurses without knowing the final mix (cancellations and absences). The schedulers assign the patients to nurses at least 24 hours before the appointment day. They make last-minute adjustments in the morning in an attempt to minimize the number of nurses required.

Algorithm 2 FCFS Algorithm

```
P = Generate a patient
Assign the first date j and time h for patient p subject to:
    1. Chair/Bed Capacity
    2. Number of Patients
    3. Pharmacist Capacity
Classify i ∈ I^p by date j
for j ∈ PP do
    Simulate cancellations and absences (3.3)
    K.list = Sort patients according to starting time h
end for
while K.list.!empty() do
    k.list = K.list.dequeue()
    Assign appointment k to nurse f subject to:
        1. Number of patients per nurse is not exceeded
        2. Mix of patients is respected
end while
```

Algorithm 3 Online Algorithm

```
P = Generate daily patient mix
Initialize J, R = patients with starting date a^r = j, S= already-scheduled patients
Solve the scheduling problem (3.4)
Classify appointment i ∈ I^p by date j ∈ PP
for j ∈ PP do
    Solve planning problem (3.2)
    Simulate cancellations and absences (3.3)
    Classify appointment i^p by date j ∈ PP
    Solve daily patient–nurse assignment problem (3.4)
end for
```

4.3 Proposed Booking Approach

We use the online algorithm to measure how our approach performs. We test two options:

1. **Flexible starting delay:** Given a fixed number of nurses and patients, we relax constraint (1j) and allow patients to have their first appointment after the last day of the prescribed interval (b^r).
2. **Flexible number of nurses:** Given a fixed set of patients, we ensure that all patients have their first appointment within the prescribed interval [a^r, b^r]. If necessary, we increase the number of nurses (T_j) until a feasible solution is reached.

For each of these options, we consider the following:

1. **No Adjustment:** The appointment time is fixed by the scheduling problem (3.1) and remains the same for all subsequent appointments.
2. **Partial Adjustment:** The appointment time is fixed by the scheduling problem. We allow the model to change this time for up to 10 patients within a two-hour time window (Constraint (2b) is relaxed for these patients) when we solve the planning problem (3.2).
3. **Complete Adjustment:** The appointment time is fixed by the planning problem; Constraint (2b) is relaxed.

Algorithm 3 solves (3.1) followed by (3.2) and then simulates cancellations and absences. The final step reassigns patients using (3.4).

4.4 Numerical Example

In this section, we illustrate the steps of the FCFS, in this example with and without a waitlist. The second example illustrates how the online algorithms works with and without a waitlist. We consider a clinic with 1 nurse and 4 chairs. The CTC is open from 8am to 12pm (8 time slots of 30 minutes each). The nurse is allowed a 30 minutes break between 10am and 12pm (not mandatory). She can handle a maximum of 4 patients per day, a total hourly workload of 4, and a total daily workload of 8. Moreover, the nurse can treat a maximum of one patient with an acuity level of 4, and a maximum of 2 patients with acuity level of 3, without ever exceeding the mix of one patient of acuity level of 3 and 4 per day. Finally, the nurse can only setup one patient per time slot. We solve the problem for a rolling horizon of 2 days.

Table 1: Sample of new patients arrived on Day 0

ID	Regimen	Hourly workload	Daily workload	Starting interval	Priority	Treatment Duration
834	1 treatment every 7 days, 6 cycles	3	3	2	10	2
685	1 treatment every 28 days, 2 cycles	4	4	1	10	4
105	1 treatment every 14 days, 4 cycles	2	2	4	7	3
472	1 treatment break of 10 days, 5 cycles	3	3	10	1	1
265	1 treatment every 10 days followed by 3 consecutive treatments, 5 cycles	1	2	6	5	2
650	2 consecutive treatments 18 days break, 6 cycles	1	1	7	4	0.5

The FCFS algorithm would schedule patients one by one, as they arrive to the first available time slot. Therefore, patient 834 would be scheduled to start on day 1 at 8am. Patient 685 would only start his treatment on day 2 at 8am. Patient 685 cannot start treatment on day 1, because the maximum hourly workload a nurse can handle is 4. Therefore, patient 685 is unable to start and finish his treatment before closing time without exceeding the nurses' hourly workload. Since patient 685 needed to start treatments within 24h, the delay is equal to 1. This approach easily generates lateness. We continue scheduling patients as they arrive. On the day of treatment based on the last list of patients, the algorithm distributes workload among nurses as fairly as possible. The main goal is always to treat all patients. Therefore, when it is not possible to respect all constraints, hourly workload and daily workload constraints are relaxed. Figure 4 (a) illustrates the schedule of nurse 1 for days 1 and 2.

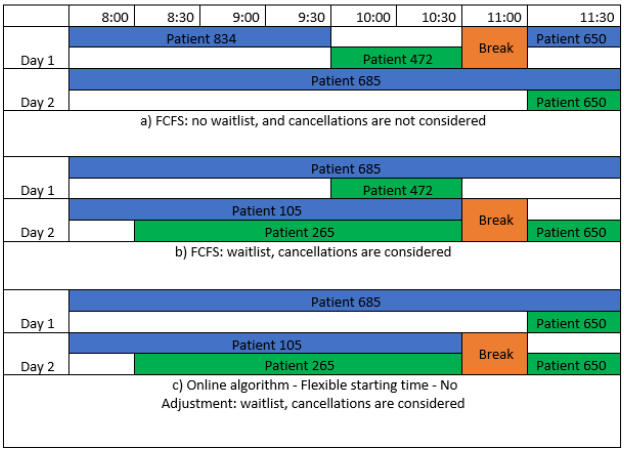

Fig. 4: Schedules generated by the FCFS and the online algorithm

We add the simulation of the cancellations. After the simulation patient 834 is canceled. This leaves only patients 472, 650 and 685 in the schedule.

4.5 FCFS : Waitlist

We now solve the same example using a waiting list. The waiting list allows us to schedule patients as long as the overall capacity is not exceeded. This means that we can schedule up to 4 patients per day. Therefore, we a) add patient 685 to the waiting list of day 1, b) patient 105 is scheduled to day 2, and c) patients 265 and 650 are added to the waiting list of day 2. We do take a risk by adding the patient to the waiting list, however, we use this as a buffer. In case of cancellations, these patients will be shifted to an available nurse. The final schedule after cancellations is illustrated in Figure 4 (b). This schedule allows the clinic to start patients' treatments sooner, but it has the risk of overbooking nurses.

4.6 Online

The online algorithm proposed would book all patients at the end of the day, when all the set of new patients is known. We solve the problem for the online algorithm-Flexible starting time- No Adjustment. Therefore, patients 685 and 834 would be prioritized by model P1-Problem (in regards to their priority level). Moreover, patients 834 and 650 would be assigned to the waiting list of day one. Patients 105 and 265 would be scheduled to start on day 2. In this scenario, the objective function is maximized. The final schedule after cancellations (patient 834) is illustrated in figure 4 (c).

The solution generated by the online algorithm assigns as many patients as the solution generated by FCFS. However when we analyze the quality of the solution, we notice that the solution of the online algorithm books patients according to their priority level, and generates a more balanced schedule in which the nurse is not overloaded. We also observe that the usage of the wailing list drastically improve the solution of the FCFS algorithm.

5 Planning and Daily Scheduling Optimization Results

Our goal is to gain insight into how the inherent uncertainty resulting from last-minute cancellations affects the cost effectiveness of the clinic. To achieve this, we compare our online algorithm to the offline algorithm and the FCFS. To measure the quality of the solutions generated by the online algorithm, we compare the productivity ratio, the nurse-power, the average daily workload assigned to the waiting list, and the delay to start treatment . The productivity ratio, and the nurse-power respectively allows us to measure the cost efficiency of the clinic. We define the productivity ratio as the ratio between the number of hours of treatments provided and the number of available hours (i.e, worked by all nurses). The higher the ratio the better is the productivity. The nurse-power illustrates the labor force required to provide all the services in the clinic; the lower this number the better. We also measure the daily workload since it affects the quality of care provided by the nurse. These three indicators measure the cost efficiency of the chemotherapy clinic. The last indicator, lateness, measures how the center is handling health objectives. We calculate the number of patients scheduled to their first appointment after the prescribed delay, because this directly affects the chance of success of the treatment.

We experiment with 10 instances inspired by a real clinic. An average of 15 new requests arrives per day which results in a total number of patient that ranges from 700 to 900 per instance. Patients arrive with a prescription and a starting interval. The four most common patterns are: (1) every 21 days (38%), (2) every 7 days (19%), (3) every 14 days (13%), and (4) every 28 days (6%). In total, there are 44 patterns. The treatments of each pattern are characterized by a duration and an acuity level, e.g., a 21-day pattern has 18 combinations of duration/acuity level: (1,1);(1,2);(2,1);(2,2);(3,1);(3,2);(3;3);...;(8,4). In theory, patients are booked for their first appointment within 10 working days. Figure 5 illustrates the distribution of patients as a function of their starting intervals. The rate of cancellations follows a triangular distribution TRIA(4,13,20)[%]. The average cancellation rate is 13%. The distribution is based on the data collected at the chemotherapy center. See section 3.3 for more details.

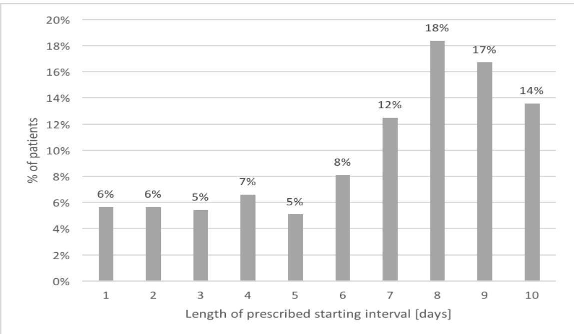

Fig. 5: Distribution of prescribed starting intervals.

Each algorithm runs for a time horizon (PP) of 54 days, and the length of the rolling horizon $|J|$ was set to 10 working days to match the maximum delay. At the beginning of the booking process, the calendar is empty. The CTC is open Monday to Friday from 8 a.m. to 4 p.m. Each day has 16 time slots of 30 minutes. We compare the online, FCFS, and offline algorithms. The algorithms were developed in Julia/0.5.2. The callable libraries of Cplex 12.6.2 were used to solve the models (3.1), (3.2), and (3.4) and the model in Appendix C. The simulation used Rockwell's Arena 15 simulation software. See 3.3 for more details on the simulation model.

5.1 Tradeoff between waiting list and number of nurses

We discuss finding a compromise between the number of nurses and the use of a waiting list. Table 2 illustrates the nurse-power required. It shows that if the patients must start within their prescribed intervals (Flexible # of nurses), the CTC needs to increase the average number of nurses scheduled per week. Column 3 illustrates the average number of nurses that the clinic would need to schedule if the waiting list is not used, all patients are assigned to a nurse regardless of the probability of cancellation. Column 4 gives the average number of nurses required when the waiting list is used (Cancellation distribution TRIA(4,13,20)[%]). We observe that by using a waiting list, we can reduce the nurse-power required by an average of 1 nurse across all scenarios. This result shows the importance of taking into consideration cancellations when measuring the clinic requirements in terms of nurses.

Table 2: Impact of fixed starting interval on number of nurses

		Average number of nurse per week (No Waitlist)	Average number of nurse per week (Waitlist)
Offline	-	13.1 ± 1.0	10.5 ± 0.8
FCFS	-	15.0 ± 0.6	13.4 ± 0.6
Flexible starting delay	No Adjustment	13.4 ± 0.5	12.4 ± 0.5
	Partial Adjustment	13.4 ± 0.7	12.4 ± 0.2
	Adjustment	11.6 ± 0.9	11.4 ± 0.3
Flexible # nurses	No Adjustment	14.0 ± 0.3	13.8 ± 0.6
	Partial Adjustment	14.0 ± 0.8	13.3 ± 0.3
	Adjustment	13.5 ± 0.2	13.3 ± 0.5

5.2 Impact of flexible # of nurses on lateness

Table 3 shows that the number of patients starting their treatment late regardless of the usage of a waiting list. For each of the 4 scenarios, the number of patients starting late stays unchanged, however the number of nurses required increases without the usage of a waiting list as demonstrated in table 2. We observe that the average number decreases by 47% when we use the online algorithm with a flexible starting delay. As noted, the online algorithm with a flexible number of nurses schedules all requests within the interval $[a^r, b^r]$, but more nurses are needed.

5.3 Tradeoff of a flexible starting time on the productivity

Table 2 demonstrates the advantage of considering a waitlist. Since, we become aware of the absentees only 24 hours before their appointment, we assign patients to

Table 3: Number of days late and number of patients with late first appointment

	Range in # days late	Average # patients	# of patients on time
Offline	0	0	837 ± 71
FCFS	[1,8]	9.9 ± 3.8	826 ± 58
Flexible starting delay	[1,5]	5.5 ± 3.0	831 ± 69
Online Flexible # nurses	0	0	837 ± 71

a temporary nurse, the overflow is assigned to the waitlist. Once we confirm the final list of patients, we reassign patients to nurses. In the current practice, the time of appointment stay unchanged. In this experiment, we discuss the impact on the waiting list of allowing the online algorithm to modify the time of the appointments when solving the planning problem. We evaluate the impact on the total daily workload assigned to the waiting list and on the nurses' productivity ratio.
Table 4 shows the total daily workload in the waiting list (assigned to the virtual nurse) for the planing problem. As expected, the workload is null when applying the offline algorithm; likewise, no tasks are assigned to the virtual nurse when complete adjustment is permitted; This shows that all patients are assigned to available nurses and that the waiting list is not used. This observation allows to demonstrate that fixing the starting time severely impact the effectiveness of the clinic. Fixing the time appointments is comforting for the patients, it is also a common practice in the field. However, we observe a 60% improvement over FCFS when we use the online algorithm with no adjustment or partial adjustment. We observe that "Online Algorithm, No Adjustment" and "Online Algorithm, Partial Adjustment" give similar averages. However, the standard variation is different which suggests that the waiting list varies depending on the patients selected when we solve 3.2.

Table 4: Total average daily workload assigned to the waiting list

	Daily Workload
Offline Algorithm	0 ± 0
FCFS	196 ± 31
Online Algorithm, No Adjustment	76 ± 21
Online Algorithm, Partial Adjustment	76 ± 44
Online Algorithm, Complete Adjustment	0 ± 0

The average workload assigned to the waiting list is measure using the scale presented in table 5. This table represents how much of the nurse's capacity is needed to treat a patient. Therefore, a workload of 76 can represent 10 patients with a workload of 7, one patient with a workload of 4, and one patient of workload 2. It could also represent 19 patients with a workload of 4. Table 5 contains the distribution of each type of patient.

Table 5: Daily workload associated with each type of task, W_i^p

Daily Workload	% of patients	Definition
7	0.4%	Requires most of the nurse's daily capacity. Can be matched to one patient of daily workload of 1 or 2
4	0.7%	Patients of acuity level 4 or 3 with some special characteristics
3	7.0%	Patients of acuity level 4 or 3 without special characteristics
2	8.2%	Patients of acuity level 2 with some special characteristics
1	83.7%	Patients of acuity level 1 or 2 without special characteristics

Table 6 shows the average productivity ratio of each working hour. The online algorithm with no adjustment has the lowest productivity ratio after FCFS. Constraining the model to book appointments in the same time slots decreases the average productivity ratio by 24% compared to that for the theoretically optimal offline solution. Allowing the online algorithm to reassign the starting time within a two-hour window for up to 10 patients improves the ratio by 8%. Complete adjustment results in an increase of 23%.

Table 6: Online scheduling compared to the theoretically offline solution and the current scheduling practice.

	Flexible starting delay	Daily average workload
Offline	-	2.64 ± 0.70
FCFS	-	1.95 ± 0.52
Online Algorithm, No Adjustment	$\surd$	2.00 ± 0.64
Online Algorithm, Partial Adjustment	$\surd$	2.11 ± 0.63
Online Algorithm, Complete Adjustment	$\surd$	2.41 ± 0.62

Notes: -Ratios indicate the average treatment hours for one hour of nursing time
-Higher ratios indicate higher productivity.

5.4 Computational Time

In this section, we summarize the computational time for the two-stage approach.
The scheduling problem has $\sum_{p\in P} |I_p|(|J|(|H|+1)+1)$ binary variables and $|J|(|H|+(|A|-1))+\sum_{p\in P} 6|I_p|+2|I_p||J|+2|R|$ constraints. The planning problem has $|H|(|K||F|+(|F|-1))$ binary variables, $3(|F|-1)$ integer variables, and up to $3|K|+|F|(|A|+|H||J|)$ constraints.

Table 7 illustrates the range of the parameters of our formulation.

Table 7: Value of parameters used for experiments

Parameter	$\lvert R\rvert$	$\lvert I^p\rvert$	$\lvert J\rvert$	$\lvert H\rvert$	$\lvert F\rvert$	$\lvert K\rvert$
Interval	11–22	3–16	10	16	$\{12, 14\}$	35–102

Table 8 summarizes the average computational time at each iteration of the 10 instances tested the online algorithm. An iteration refers to the problem we solve at the end of each day, which correspond to scheduling an average of 15 new requests for a rolling horizon of 10 days. The table presents the time required to solve the offline algorithm as well.

Table 8: Computational Time

	Average Computational Time (s)	Maximum Value Observed (s)
Offline	39514	65423
Online, Scheduling Problem (3.2)	172	198
Online, Planning & Assignment Problems (3.4)	1.65	17.52

6 Insights and Recommendations

The experiments suggest that the modelisation of a waiting list, and the usage of an optimization tool has great potential. First, the waiting list helps decrease the nurses' requirement. Second, an optimization tool improves productivity ratio, and lateness. Finally, since the computational time is reasonable, we were able to consider several constraints.

The waiting list would enable CTCs to take advantage of cancellations. By overbooking patients, the clinic would be able to better use its resources. In average, the required number of nurses decreases by 1, even in the case of FCFS. This implies that modeling a waiting list is beneficial regardless of the implementation of an optimization tool. Only by taking advantage of the inherent uncertainty resulting from last minute cancellations, a CTC can make an economy of one complete salary.

Combining a waiting list to a flexible starting time also improves the productivity ratio. We acknowledge that changing times is a sensitive matter, and the clinic must carefully consider the viability of this approach. It would adversely affect patients with fixed transportation arrangements or other appointments. Changing the time of appointment is undesirable. To intentionally incorporate this practice, one could add a parameter to the model to prevent adjustments for the patients with no flexibility. The Implementation of this practice would require further studies to verify that enough patients are willing to consider this option. Our experiment also demonstrates the benefice of partially adjusting starting time which would affect only a small amount of patients, a maximum of 10 in this study. We evaluate the gain in productivity of using partial adjustment to 5.5% when compared to the solution generated by the No Adjustment scenario, and to 8% when compared to FCFS.

Relaxing the starting interval constraints (1j) and choosing a large planning horizon reduces the risk of infeasible solutions to the planning problem. The resulting formulation is limited only by the capacity constraints. FCFS algorithm generates twice as much lateness than the proposed approach. When studying similar problems, one can find the optimal level of resources to balance lateness and nurses' requirement. Moreover, to discourage lateness, we formulate the objective function so as to break the symmetry between equivalent requests, and to favor solutions where the appointments are scheduled as soon as possible.

The modelisation of the waiting list has advantages. It acts as a buffer, absorbing the excess workload of the scheduling problem. The number of patients assigned to any given time slot is restricted by the available material resources in stage one, the maximum daily workload, the maximum hourly workload, and the quota of the type of patients. Therefore, the workload assigned in stage one almost always exceeds the nurse capacity. To ensure a feasible solution, the procedure needs a cushion to absorb the excess workload.

Finally, we note that the flexibility (to start treatment and on the number of nurses) copes with the issues of uncertainty in most cases. The results of the offline algorithm show that if we were able to control the uncertainty and the variability, we would be able to schedule all patients on time while optimizing the efficiency of the clinic.

7 Conclusions

We have developed a mathematical formulation for the planning and daily patient–nurse assignment problems. We have used the model to evaluate the impact of adjusting appointment times, using a waiting list, and relaxing the starting day restriction. Analysis demonstrated that allowing algorithms to change starting time has an impact on nurses' productivity ratio and on the total workload assigned to the waiting list. Moreover, the results showed that scheduling all the patients in their prescribed intervals increases the number of nurses required.

The daily patient–nurse assignment demonstrated the advantages of considering cancellations and absences.

It also enabled us to illustrate the potential gain of allowing changes to the starting times. Furthermore, the computational time is reasonable. The model solves the scheduling problem on average within 180 s (3 min); it solves the assignment problem within seconds. In conclusion, an optimization tool would decrease the workload of the schedulers and the nurse-in-charge.

References

1. Le cancer. `https://www.societederecherchesurlecancer.ca/page.aspx?pid=2093`
2. Interview with members of the managerial team at the chemotherapy clinic of Notre-Dame hospital. Personal communication (2016)
3. Alvarado, M., Ntaimo, L.: Chemotherapy appointment scheduling under uncertainty using mean-risk stochastic integer programming. Health Care Management Science (2016). DOI 10.1007/s10729-016-9380-4
4. Berwick, D.M., Nolan, T.W., Whittington, J.: The triple aim: Care, health, and cost. Health Affairs **27**(3), 759–769 (2008)
5. Condotta, A., Shakhlevich, N.: Scheduling patient appointments via multilevel template: A case study in chemotherapy. Operations Research for Health Care **3**(3), 129–144 (2014)
6. Dobish, R.: Next-day chemotherapy scheduling: A multidisciplinary approach to solving workload issues in a tertiary oncology center. Journal of Oncology Pharmacy Practice **9**(1), 37–42 (2003)
7. Gocgun, Y., Puterman, M.L.: Dynamic scheduling with due dates and time windows: an application to chemotherapy patient appointment booking. Health care management science **17**(1), 60–76 (2014)
8. Green, E., Preyra, C., Stewart, J., McLennan, C., Bland, R., Dus, T., Langhorn, M., Beattie, K., Cheung, A., Hertz, S., Sechter, H., Burns, J., Angus, H., Sawka, C.: Determining resource intensity weights in ambulatory chemotherapy related to nursing workload **22**, 114–28 (2012)
9. Hahn-Goldberg, S., Carter, M.W., Beck, J.C.: Dynamic template scheduling to address uncertainty in complex scheduling problems: a case study on chemotherapy outpatient scheduling. In: Society for Health Systems Conference, Las Vegas, NV (2012)
10. Lamé, G., Jouini, O., Stal-Le Cardinal, J.: Outpatient chemotherapy planning: A literature review with insights from a case study. IIE Transactions on Healthcare Systems Engineering **6**(3), 127–139 (2016)
11. Liang, B., Turkcan, A., Ceyhan, M.E., Stuart, K.: Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic. International Journal of Production Research **53**(24), 7177–7190 (2015)
12. Sadki, A., Xie, X., Chauvin, F.: Patients assignment for an oncology outpatient unit. In: Automation Science and Engineering (CASE), 2010 IEEE Conference on, pp. 891–896. IEEE (2010)
13. Sadki, A., Xie, X., Chauvin, F.: Appointment scheduling of oncology outpatients. In: Automation Science and Engineering (CASE), 2011 IEEE Conference on, pp. 513–518. IEEE (2011)
14. Santibáñez, P., Aristizabal, R., Puterman, M.L., Chow, V.S., Huang, W., Kollmannsberger, C., Nordin, T., Runzer, N., Tyldesley, S.: Operations research methods improve chemotherapy patient appointment scheduling. The Joint Commission Journal on Quality and Patient Safety **38**(12), 541–553 (2012)
15. Turkcan, A., Zeng, B., Lawley, M.: Chemotherapy operations planning and scheduling. IIE Transactions on Healthcare Systems Engineering **2**(1), 31–49 (2012)
16. Van Merode, G.G., Groothuis, S., Schoenmakers, M., Boersma, H.H.: Simulation studies and the alignment of interests. Health Care Management Science **5**(2), 97–102 (2002)
17. Wong Kee Yan, N.: Chemotherapy outpatient scheduling at the Segal Cancer Center using mixed integer programming models. Ph.D. thesis, Concordia University (2017)
18. Woodall, J.C., Gosselin, T., Boswell, A., Murr, M., Denton, B.T.: Improving patient access to chemotherapy treatment at Duke Cancer Institute. Interfaces **43**(5), 449–461 (2013)

Appendices

A Measure of Hourly Workload (G_i^p)

The CTC prepared a list of rules related to workload balancing. We use it to develop two parameters to facilitate the estimation of workload. The first, G_i^p (Table 9), represents the hourly workload that the patient requires.

Table 9: Hourly workload associated with each type of task, G_i^p

Hourly Workload	Definition
4	Requires full attention
3	Can be matched to patients requiring little attention. Represents patients with acuity level 4 or 3 without special characteristics
2	Can be matched to patients requiring some or little attention. Represents patients with acuity level 1 or 2 with research protocol
1	Requires little attention. Represents patients with acuity level 1 or 2

B Linearization of the Formulation

Constraints (1l) and (1m) are both nonlinear. We use a big-M approach to linearize them: Constraints (4a) and (4b) replace (1l), and Constraints (4c) and (4d) replace (1m). The parameter Θ_{i^p} plays the role of the big M; it is tightened as much as possible in each case.

$$\left(\sum_{j\in J} jx^p_{(i-1)j} + a^p_{(i-1),i}\right) - \sum_{j\in J} jx^p_{ij} \le (1-y^p_i)\,\Theta^p_i \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{4a}$$

$$\sum_{j\in J} jx^p_{ij} \le \sum_{j\in J} jx^p_{(i-1)j} + a^p_{(i-1),i} \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{4b}$$

$$\sum_{h\in H} h \sum_{j\in J} u^p_{(i-1)jh} - \sum_{h\in H} h \sum_{j\in J} u^p_{ijh} \le (1-y^p_i)\,|H| \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{4c}$$

$$\sum_{h\in H} h \sum_{j\in J} u^p_{ijh} \le \sum_{h\in H} h \sum_{j\in J} u^p_{(i-1)jh} \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{4d}$$

C Formulation of Offline Model

Variables

x^p_{ifjh}: 1 if treatment i^p of patient p is assigned to nurse f on day j in time slot h; 0 otherwise
y^p_i: 1 if treatment i^p of patient p is assigned; 0 otherwise
z_{fj}: 1 if nurse f is assigned; 0 otherwise
v_{fjh}: 1 if nurse f is on break in time slot h; 0 otherwise
τ_{fj}: Integer variable: number of tasks handled, $\{0...E\}$
σ_{fj}: Integer variable: first overflow level, $\{0...B\}$
ι_{fj}: Integer variable: second overflow level, $\{0...C\}$

Model

$$min \sum_{f\in F\setminus\{virtual\}} \sum_{j\in J} \eta z_{fj} + \alpha\tau_{fj} + \beta\sigma_{fj} + \gamma\iota_{fj} + \epsilon \sum_{p\in P}\sum_{i^p\in i^p}\sum_{j\in J}\sum_{h\in H} x^p_{i(virtual)jh} \tag{5a}$$

s.t.

$$\sum_{p\in P}\sum_{i\in I^p}\sum_{f\in F}\sum_{h'=max(1,h+1-D^p_i)}^{min(h,H+1-D^p_i)} L^p_{li} x^p_{ifjh'} \le V_l \quad \forall j \in PP, \forall h \in H, \forall l \in L \tag{5b}$$

$$\sum_{f\in F}\sum_{j\in PP}\sum_{h\in H} hx^p_{ifjh} + D^p_i \le |H| \quad \forall i \in I^p, \forall p \in P \tag{5c}$$

$$\sum_{p\in P}\sum_{i\in i^p}\sum_{h'=max(1,h+1-D^p_i)}^{min(h,H+1-D^p_i)} G^p_i x^p_{ijfh'} \le N \quad \forall f \in F\setminus\{virtual\}, \forall j \in PP, \forall h \in H \tag{5d}$$

$$\sum_{p\in P}\sum_{i\in I^p}\sum_{h\in H} W^p_i x^p_{ifjh'} \le W \quad \forall f \in F, \forall j \in PP, \forall h \in H \tag{5e}$$

$$\sum_{p\in P}\sum_{i\in I^p} x^p_{ifjh} \le 1 \quad \forall f \in F\setminus\{virtual\}, \forall j \in PP, \forall h \in H \tag{5f}$$

$$\sum_{p\in P}\sum_{i\in i^p}\sum_{h\in H} P^p_{ai} x^p_{ifjh} \le M_a \quad \forall f \in F\setminus\{virtual\}, \forall j \in PP, \forall a \in A \tag{5g}$$

$$a^p \le \sum_{j\in PP}\sum_{h\in H} jx^p_{1fjh} \le b^p \quad \forall p \in P \tag{5h}$$

$$y^p_i \le y^p_{(i-1)} \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{5i}$$

$$y^p_i\left(\sum_{j\in PP} jx^p_{ifjh} - \sum_{j\in PP} jx^p_{(i-1)fjh}\right) = y^p_i a^p_{(i-1),i} \quad \forall i \in \{I^p|i>1\}, \forall p \in P \tag{5j}$$

$$v_{fjh} + v_{fj(h+2)} + v_{fj(h+4)} = z_{fj} \quad h=7, \forall f \in F\setminus\{virtual\}, \forall j \in PP \tag{5k}$$

$$v_{fjh} + v_{fj(h+2)} + v_{fj(h+4)} = z_{fj} \quad h=8, \forall f \in F\setminus\{virtual\}, \forall j \in PP \tag{5l}$$

$$hv_{fjh} + (h+2)v_{fj(h+2)} + (h+4)v_{fj(h+4)} \le (h+1)v_{fj(h+1)} + (h+3)v_{fj(h+3)} + (h+5)v_{fj(h+5)} \quad h=7, \forall f \in F\setminus\{virtual\}, \forall j \in PP \tag{5m}$$

$$0 \le v_{fj(h-1)} - v_{fjh} + v_{fj(h+1)} \le 1 \quad \forall f \in F\setminus\{virtual\}, \forall j \in PP, \forall h \in H \tag{5n}$$

$$0 \leq hv_{fjh} + (h+1)\, v_{fj(h+1)} \leq 1 \quad \forall f \in F \setminus \{virtual\}, \forall j \in PP, \forall h \in H \tag{5o}$$

$$\sum_{f \in F} v_{fjh} \leq \frac{1}{3} \sum_{f \in F \setminus \{virtual\}} z_{fj} \quad \forall h = \{7, 9, 11\} \tag{5p}$$

$$v_{fjh} + x^p_{ifjh} \leq 1 \quad \forall p \in P, \forall \in I^p, \forall f \in F \setminus \{virtual\}, \forall j \in PP, \forall h = \{7, 8, 9, 10, 11, 12\} \tag{5q}$$

$$\sum_{f \in F} \sum_{j \in PP} \sum_{h \in H} x^p_{ifjh} = y^p_i \quad \forall i \in I^p, \forall p \in P \tag{5r}$$

$$\sum_{p \in P} \sum_{i \in I^p} \sum_{h \in H} x^p_{ifjh} = \tau_{fj} + \sigma_{fj} + \iota_{fj} \quad \forall f \in F \setminus \{virtual\}, \forall \in PP \tag{5s}$$

$$x^p_{ifjh}, y^p_i, z_{fjh}, v_{fjh} \in \{0, 1\} \tag{5t}$$

$$\tau_{fj}, \sigma_{fj}, \iota_{fj} \in \mathbb{N} \tag{5u}$$

D Outline of the procedure to generate cancellations and absences

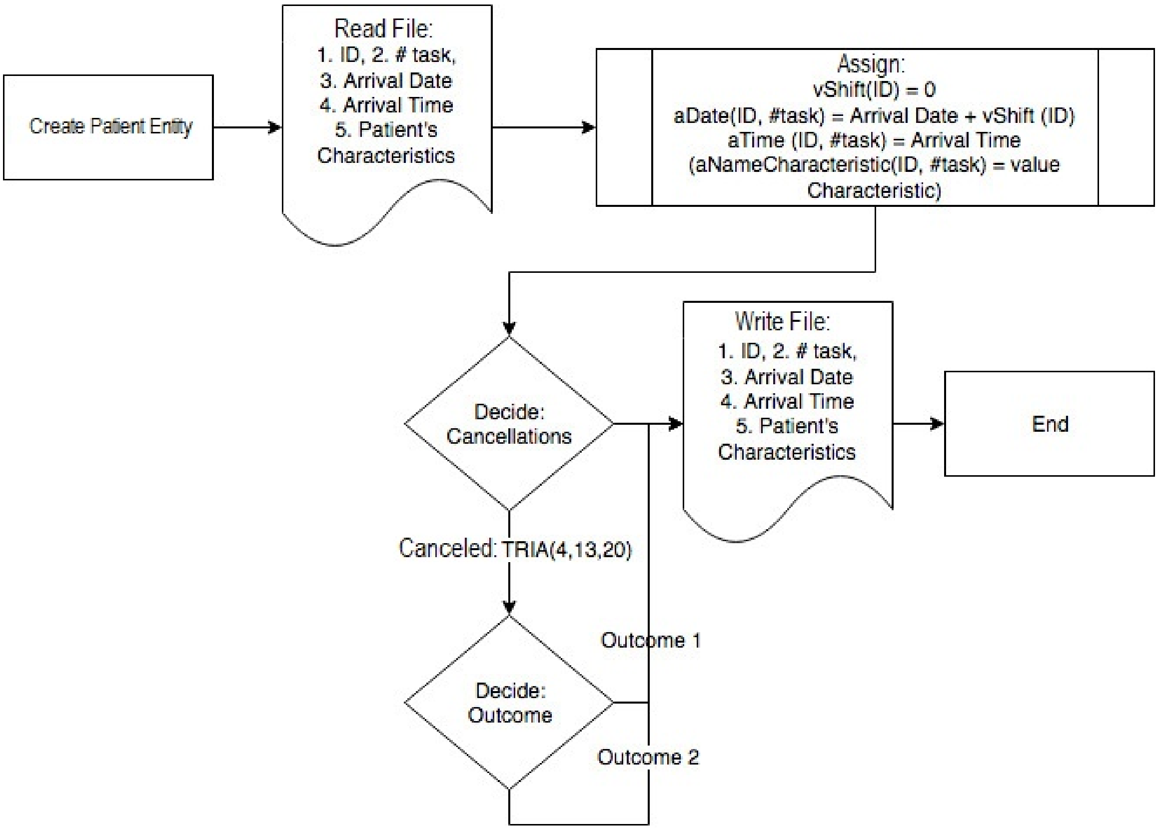

Fig. 6: Simulation of patient cancellations

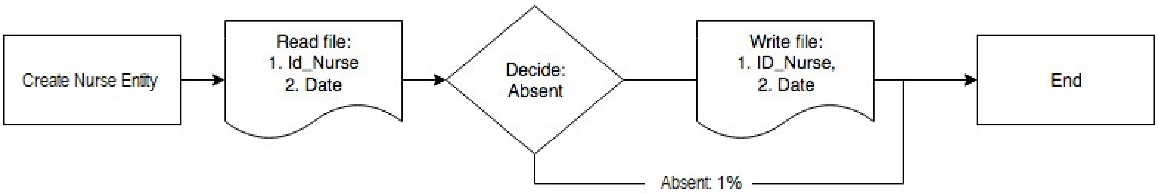

Fig. 7: Simulation of nurse absences