CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

Fine-Grained Simulation Optimization for the Design and Operations of a Multi-Activity Clinic

Philip Troy
Nadia Lahrichi
Dana Porubska
Lawrence Rosenberg

November 2018

CIRRELT-2018-47

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Fine-Grained Simulation Optimization for the Design and Operations of a Multi-Activity Clinic

Philip Troy[1], Nadia Lahrichi[2,*], Dana Porubska[3], Lawrence Rosenberg[1]

1. Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-Ouest-de-l'Ile-de-Montréal, 3755, chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 1E2
2. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Mathematics and Industrial Engineering, Polytechnique Montréal, P.O. Box 6079, Station Centre-Ville, Montréal, Canada H3C 3A7
3. Sir Mortimer B. Davis Jewish General Hospital, 3755, chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 1E2

Abstract. To ensure that patients are appropriately prepared for surgical procedures in a welcoming environment, the Sir Mortimer B. Davis Jewish General Hospital, a McGill University affiliated teaching hospital located in Montreal, is redesigning and relocating its existing presurgical screening clinic so that it provides additional services and is more patient friendly. Given the services being added, limited space, and the desire of senior management to minimize overtime costs, physician idle time, and excessive patient waiting times, we apply simulation optimization to the operations of the redesigned clinic. The simulation optimization is then used to evaluate the effect of possible design decisions to be made by senior management, to ensure that the resulting clinic meets their goals. In contrast to existing research which generally limits clinic optimization to just a few facets, we simultaneously optimize the clinic's multiple objectives at a fine-grained level with respect to individual decision variables for the start time of each physician, the appointment time of each patient, and the start, break, and lunch times of each staff member. To perform the optimization, we apply a simple heuristic to a simulation model of the clinic. We show, with this simple heuristic, that simultaneously optimizing the clinic's multiple objectives by adjusting decision variables at this more granular level can significantly reduce physician idle time, staff overtime, and excessive patient waiting. This in turn makes it possible to evaluate design decisions in context of optimized operations. These results suggest the usefulness of this approach to other multi-activity clinics such as cancer treatment clinics.

Keywords: Patient flow optimization, simulation, healthcare application.

---

* Corresponding author: Nadia.Lahrichi@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec

Bibliothèque et Archives Canada, 2018

1 Introduction

The Quebec Ministry of Health and Social Services' guide to "Planning surgical activities" [1] recommends a particular process for patients requiring elective surgery. The process begins with physicians' decisions to perform surgical procedures. It is followed, at scheduled times, by patient visits to the presurgical screening clinic, which is responsible for ensuring that patients are medically prepared for the procedure, and for providing education needed to ensure that patients are otherwise prepared. This visit is followed, at scheduled times, by the procedures, which are in turn followed by recovery in a post-anesthesia care unit and possibly a surgical ward.

The presurgical screening clinic plays a central role in this process. The literature [1] shows that full-service presurgical screening clinics reduce risk and postoperative hospital stays by ensuring that patients are properly prepared and sufficiently healthy for the procedures to be performed. These clinics also reduce the cancellations that occur when patients are found to be not ready. Therefore, senior management of the surgical services department at Montreal's Sir Mortimer B. Davis Jewish General Hospital decided to redesign and relocate its presurgical screening clinic so that it includes medication reconciliation and individualized nurse education for those patients determined to need these services. This would be in addition to the services provided by the existing clinic, which include registration, electrocardiograms (ECGs), physician exams, insurance information collection, and blood testing. Senior management also decided to create a more patient friendly environment.

Despite the importance of the existing clinic, little attention had been given to understanding and optimizing its processes. When the processes were examined, it was found that the time required for each part of the process is stochastic in nature, and whereas almost all patients and physicians arrive on time, some do arrive a little late. In addition, there are often long delays, during which patients wait in surgical gowns in public hallways for exams. After being examined by the clinic's physicians and getting dressed, patients also have to wait for other tasks. As a result, the average cumulative wait time in the clinic varies between one hour and one hour and a half, and the 50^{th} percentile and 75^{th} percentile are one hour and 33 minutes and one hour 54 minutes respectively. Senior management expects that this situation will be exacerbated by the expansion of the clinic's activities and the number of patients to be seen. These waiting times, and their expected increase, motivated interest in minimizing excessive patient waiting time, which we define to equal the total waiting between activities in a visit less a threshold of sixty minutes, so as to improve the patient experience.

To ensure that the redesigned clinic meets senior managements' goals of minimizing staff overtime costs, physician idle time, and excessive patient waiting, we apply simulation optimization to the operations of the redesigned clinic. In particular, we optimize the clinic's multiple objectives at a fine-grained level with respect to individual decision variables for the start time of each physician, the appointment time of each patient, and the start, break, and lunch times of each staff member, for each of several scenarios. These scenarios were selected to evaluate the effect of certain characteristics of the redesigned clinic such as the number of exam rooms, and whether or not the collection of insurance information should be grouped together with patient registration. To perform the optimization, we apply a simple heuristic to a simulation model of the clinic. In contrast to existing research where simulation is primarily used to analyze the flow in a clinic, or its scheduling policies, or to determine the best values of a few variables, the primary contribution of this research is that we show, with this simple heuristic, that simultaneously optimizing the clinics's multiple objectives with respect to decision variables at this more granular level can significantly reduce physician idle time, staff overtime, and excessive patient waiting. This suggests the usefulness of applying this approach to other multi-activity clinics such as cancer treatment clinics or high risk pregnancy clinics.

The paper is organized as follows. In Section 2 we describe the existing clinic, the proposed clinic, and the challenges expected in implementing the proposed clinic, and in Section 3 we review the related literature. Next, in Section 4 we first present the scenarios to be analyzed, so that we can subsequently describe, in the rest of the section, the simulation model, the approach used to verify and validate the model, the approach used to compute the performance of the processes evaluated with the model, and the local search heuristic developed to optimize operations. Finally, in Section 5 we present and analyze our results and in Section 6 we present our conclusions.

2 The processes

2.1 Patient journey through the existing clinic

In the existing clinic, clerks and nurses are scheduled to arrive when the clinic opens for the day, while physicians have a more flexible schedule, i.e. the days and length of the clinics they wish to provide. We note that in Quebec physicians are not staff and work on a fee-for-service basis (where the fee is paid for by the government, and not the hospital). Two clerks, three nurses and three technicians (one for the ECG, one for blood taking and the pharmacy technician) are available during the clinic. Two to three physicians are usually available.

Patients are booked at fifteen-minute intervals starting at 7:30 and are then told when they should arrive at the clinic. We intentionally do not consider no-shows and cancellations in our analysis after investigating the data and discussions with management who indicated that these were not an issue. After seeing the clerk to register their presence and to provide insurance information for their hospital stays, if they had not previously done so, patients meet with a pharmacy technician to reconcile their medications. They are then asked to change into a surgical gown in one of two changing rooms; since no bags or lockers are provided because of space constraints, patients carry their belongings with them when moving between locations in the clinic.

Once in a gown, patients requiring an ECG wait in a public waiting area to have it taken by a technician in a dedicated ECG exam room. They then wait again, in the same public waiting area, until the physician to which they have been assigned becomes available; an administrator ensures that each physician is assigned their fair share of the number of patients so that they can each bill the government for approximately the same amount. Each physician uses one exam room. After being seen by a physician, patients change back into their street clothes. If so instructed by the physician, they have blood samples taken by a technician, and they collect urine samples in the bathroom. Except when waiting times for physicians are excessive, only patients undergoing outpatient surgery are given nurse education, usually at the end of their visit to the clinic.

A flow chart of the current process is presented in Figure 1. This figure shows the flow of patients from the time they arrive to the clinic to the time they finish their visit, in context of their interactions with clinic resources. Usually, one break for staff (administrative, technicians and nurses) is scheduled to start between 2 to 3 hours after arrival, and lunch to start after 2 to 2.5 hours after the morning break. One should note that physicians are not scheduled to take breaks as they typically only see patients for around five hours per day. Physicians leave when they have finished their activities, and the remaining staff leave at the end of their shift, or later if they haven't finished their activities for the day by that time.

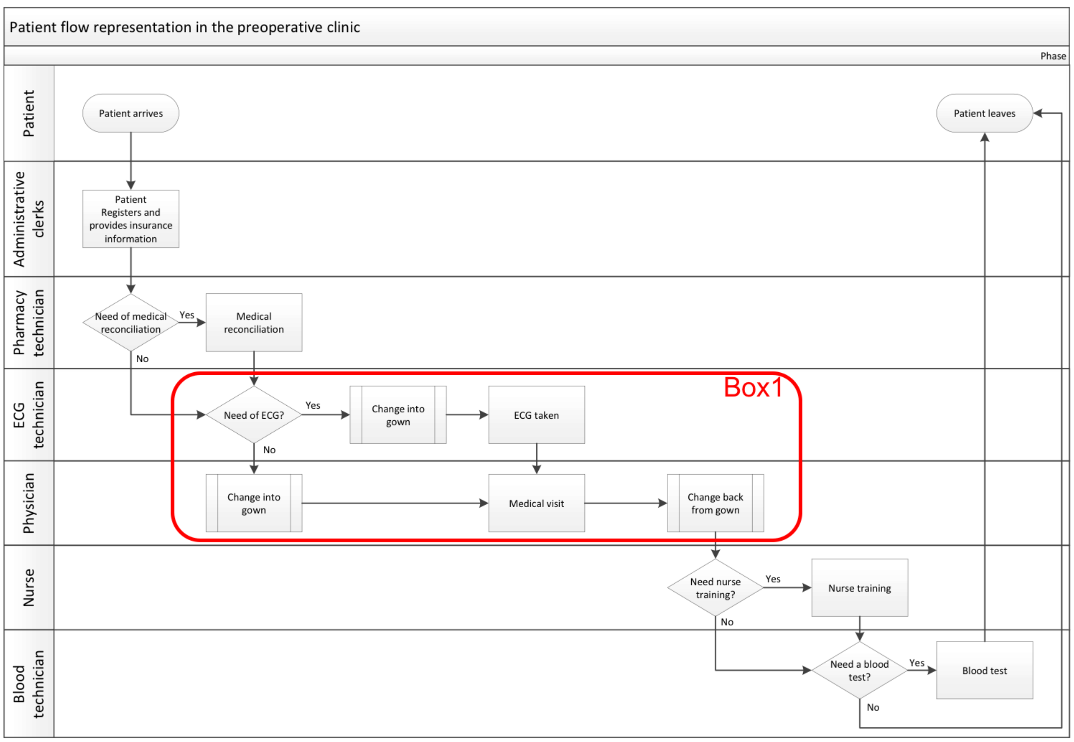

Figure 1: Illustration of the patient flow in the preoperative clinic

2.2 Patient journey through the proposed clinic

A different and more patient-centered process is envisioned for the proposed clinic. In particular, while patients will still be told when to arrive, those appointment times will be set so as to simultaneously optimize all of the operating goals of the clinic. Likewise, physician start times, and administrative staff, nurse, blood taker, pharmacy technician, and ECG technician start, break, and lunch times will be set as part of the same optimization. With the expanding role of the clinic, patients will be triaged in advance in order to assign them to a general physician or an internist in the case of complex diseases.

In this proposed process, after registering their presence at the clinic with administrative staff, patients who haven't already done so will meet with the pharmacy technician to have their medications reconciled. They will then wait for an exam room to become available, unless administrative staff become available in which case they will be asked to provide those staff their insurance information. Upon entering an exam room, patients will change into a gown, wait for the ECG technician if they need to have an ECG, and then wait for a physician to give them an exam. They will then change back into street clothes and leave the room, without having to wait in a gown in a public area. After leaving the room, patients will watch a DVD created to educate them about how to prepare for their procedure, after which they will meet individually with a nurse who will ensure that they understand how to prepare for their procedure. They will also provide their insurance information to administrative staff if they have not already done so, and provide blood samples and a urine sample if so directed by the physician during the exam.

Figure 1 illustrates the process. The red box shows the major change in the flow; instead of changing into a gown (and changing back) in the changing room, all the activities included in the box now will occur in the exam room.

From a patient perspective, there are many advantages to the proposed process. In particular, the scheduling changes are designed to minimize long waits, patients do not have to wait in hallways in a gown, and all patients always receive individualized nurse education. We note that the proposed process reflects a paradigm shift from patients lining up to see physicians in their "office" to a paradigm where patients are given the privacy of the exam room and the physician instead visits patients in the exam rooms in which they are waiting to be seen.

These proposed changes lead to a number of challenges. These include determining the number of exam rooms, how the exam rooms are assigned to physicians, and whether the proposed process is feasible. These proposed changes also lead to conflicting objectives that need to be handled simultaneously in a balanced manner: how to minimize physician idle time, staff overtime, and excessive patient waiting, all in context of the variability inherent in the processing times.

3 Related literature

To ensure that the redesigned clinic meets their goals, senior management needs to decide on the number of exam rooms, whether general practioners and the internist should work in the clinic in the same time, and whether both administrative tasks should be performed at the same time, preferably in context of optimized operation of the clinic. To optimize clinic operations, we need to simultaneously minimize physician idle time, staff overtime, and excessive patient waiting.

A very large amount of published research addresses individual facets of this problem [23]. For example Morrice et al. [20] shows how effective triaging of patients can increase the number of patients seen in a specific clinic, and Pinedo et al [22] provide an overview of the scheduling problems that appear in health care (timetabling, reservations and appointments). They discuss the similarities and the differences between the problem formulations and solution techniques used in the various different industries.

Our literature review focuses on research that addresses multiple facets simultaneously. We note several common elements in that literature. One of those [7, 15] is that patients seem to tolerate short waiting periods, i.e., up to thirty to forty minutes (for example in 75% of the cases) for visits consisting of a single activity such as a physician exam. A second common element is the use of simulation [11, 18] to identify the importance of particular factors in a healthcare operation [10], to analyze problematic processes and to experiment with process changes to improve those processes [24], and to compare and optimize process approaches. A third common element is that the set of outcomes that are measured includes a measure of physician and staff idle time or utilization, the length of the clinic or a measure of additional costs due to clinics running too long, patient waiting time, and patient throughput. A fourth common element is the set of factors used to control the process in context of the outcomes just mentioned; these factors include staffing levels, the number and allocation of exam rooms, the number of patients that are booked, the ordering of patients, and patient appointment times.

Several studies simultaneously include two or more of these common elements. For example, Hashimoto and Bell [14], in context of a multi-physician clinic with sequential providers, found that having too many physicians in the clinic increased waiting times by overloading other staff members. To address this they reduced the number of physicians and then adjusted the number of other staff members. Swisher and Jacobson [28], in context of planning a clinic, address profitability, patient satisfaction with respect to waiting times, and staff satisfaction issues. They started by developing a clinic effectiveness measure that included patient waiting times of more than 15 minutes or 30% of total visit times, and medical staff dissatisfaction caused by reduced break times. They then developed a discrete event simulation model to compute the clinic effectiveness for different combinations of the number of physician assistants or nurse practitioners, nurses, medical assistants, check-in rooms, exam rooms, and specialty rooms. Huschka et al. [17] used simulation to evaluate the ability of a new outpatient procedure center to provide a needed throughput. When they found that it could not, they used simulation to experimentally evaluate different operational approaches, nurse staffing levels, recovery space allocations, and patient appointment times to simultaneously minimize patient waiting and obtain the required throughput.

Other studies investigate improving patient throughput using booking strategies and space allocation. Groothuis et al. [13] looked at the problem of improving patient throughput in context of catherization rooms, where procedure start times were often delayed by the late arrival of patients. In particular, they applied simulation to evaluate the effect of two scheduling policies on the number of procedures that could be performed per day and the proportion of days the procedures were performed without overtime. The first strategy was to perform only procedures that could be started before 4 p.m., and the second strategy was to fix the number of patient procedures per day. They also determined the increase in procedures that could be performed by 1) precluding the need to wait for patients, 2) adjusting rooms, and 3) prepping patients outside the catherization room. Berg et al. [4] simultaneously addressed determining the number of patients to book for particular days, how patients of different types should be sequenced within the day, and the time between appointments, while considering the variability of procedure times and high patient no-show rates. In particular, they developed a two-stage stochastic mixed-integer program to maximize the profit of an outpatient procedure center with respect to these factors. Their work highlights the benefits of stochastic optimization and double booking. They also show that it is better to allocate slots later in the day for procedures with a larger duration variance and patients with higher no-show probabilities. In [3], Berg et al. tested the effect of different numbers of physicians, rooms, arrival times, and room turn-around times on the number of patients treated, the resource utilization, and the patient waiting times. Huang [16] investigated how the existing patient types could be grouped into appointment or scheduling groups to minimize the total cost of clinic flow and scheduling flexibility in a Primary Care Facility using simulation based optimization.

Fewer studies include the idle time of the resources (generally physicians) and staff overtime/clinic operating time. White et al. [33] used simulation to evaluate the effect of different patient sequencing approaches and different exam room allocations on physician idle time, patient waiting time, and total clinic operating time. Another study [26] added resident and student involvement. That study used simulation to study the joint effects of on-time clinic starts, patient sequencing, patient appointment times, and exam room allocation on patient waiting time, clinic end time, physician idle time, and waiting and exam room occupancy. Taheri et al. [29] used simulation to evaluate the effect of having nurses change their activities based on current needs and different patient sequencing. They also studied the effect of the timing of patient appointments on nursing requirements, overtime, and unbalanced nursing loads. In Oh et al. [21], guidelines for scheduling (by minimizing a weighted measure of provider idle time and patient wait time) in primary care are provided. They show that classifying patients and keeping empty slots can be used to compensate for variability in service times.

Finally, Berg and Denton [5] reviewed many of the issues discussed above in context of an outpatient procedure center. They included multiple facets of the problem but not overtime or idle time.

In context of this published research, the research presented in this paper is at the forefront of current work in this field, for both its application, and for the granularity of its decision variables in context of a simulation model. In particular, this work contributes to the literature by simultaneously addressing in a coherent manner several different logistical problems related to operating a presurgical screening clinic, including almost all of the problems faced by outpatient clinics. Furthermore, it does so in the context of optimizing a simulation model in which uncertainty in processing and arrival times is considered, and the start, break, and lunch time of each staff member, the start time of each physician, and the time of each patient appointment are individually scheduled, providing a maximum degree of control over the clinic's operations; this is in contrast to the studies referenced above which limit their control factors or decision variables. And while the results of the research presented in this paper do suggest approaches practitioners can use to improve patient scheduling, more importantly, they also show practitioners that the use of booking strategies is not always the best strategy. Instead, for our problem, and very possibly for other multiple activity clinics, individually optimizing each patient appointment times, physician start times, and staff members' start, break and lunch times yields the best results in context of the multiple objectives defined for the clinic.

4 Simulation scenarios, model and optimization

In this section, we specify the scenarios that guided the development of the simulation model. We then specify the data collected for the model, the unusual aspects of the design of the model, how the model was verified and validated, the operations optimization problem, and the local search algorithm used to optimize the model.

4.1 Scenario analysis

In order to provide senior management with insights as to the effects of possible design decisions on the optimized operations of the clinic, we evaluated the five scenarios listed immediately below for five, six and seven exam rooms. Except where specified otherwise, these scenarios assume that individualized nurse education is provided to each patient, that registration and insurance information collection are treated as two separate activities, that 35 patients are booked for each day, that the threshold for excessive waiting, $\mathcal{W}$, is 60 minutes, that the one internist sees patients much later in the morning than do the two general practitioners, that there are two admissions staff, that there are three technicians (one ECG technician, one technician for taking blood, one pharmacy technician), that there are three nurses, and that the weighting scheme for the objective function uses the values specified at the start of Section 5.

- (Scenario 1) Excessive wait time thresholds: This scenario makes it possible for senior management to see the effect of different excessive wait time thresholds on optimized clinic operations.
- (Scenario 2) Increased demand: This scenario makes it possible for senior management to see the effect of an increased daily patient count on optimized clinic operations.
- (Scenario 3) Physician scheduling: This scenario makes it possible for senior management to determine the feasibility of reducing the number of hours the clinic operates each day by having all of the physicians work at the same time.
- (Scenario 4) Merging administrative activities: This scenario makes it possible for senior management to see whether there is any substantial benefit in merging the two administrative activities.
- (Scenario 5) Booking rules: This scenario makes it possible for senior management to see the effect of optimization, versus the use of un-optimized booking rules, on clinic operations.

4.2 Model data

Two types of data were needed for the model: patient profiles (list of activities), and probability distributions of the processing time for each activity. Patient profiles were specified in a matrix with a row for each patient, a column for each type of activity, and flags in cells indicating the activities each patient needed to undergo while in the clinic. A summary of the patient profile data is provided in Table 1. The list of activities in that table reflects the usual sequence in the clinic (refer to Figure 1).

To determine the probability distribution of the processing time for each activity, we developed a form that was approved by the quality department. The simulation modelling

Table 1: Summary of the patient profile data

	Profile					
	1	2	3	4	5	6
Registration	√	√	√	√	√	√
Insurance information	√	√	√	√	√	√
DVD education	√	√	√	√	√	√
Pharmacist visit				√		√
ECG			√	√	√	√
GP exam		√	√	√		
Internist exam					√	√
RN Individual education	√	√	√	√	√	√
Blood taking	√	√	√	√	√	√
Proportion	14%	14%	50%	8%	7%	7%

did not require ethics approval. We collected data during one week at the clinic during which we asked all 98 patients visiting the clinic to fill out the form (see Appendix A). With the help and support of the clinic staff, 90 patients filled out the form. Eight patients refused to participate in the study which was performed on a voluntary basis. Patients were asked to write down the start and end times of each activity as they progressed through their visit to the clinic. In addition to making it possible for us to collect task-duration information, this exercise helped us to better understand the problems facing the clinic, to better understand patient needs, and to prioritize the scenarios to test.

Table 2 summarizes the data collected. We define n to be the number of patients booked on each day and $\bar{n}$ to be the number of patients not wishing to participate in the study. For each task, μ represents the average duration and σ the standard deviation. The minimal and maximal durations are also presented. The last three columns show the total duration of a visit to the clinic, the time spent interacting with staff and the time spent waiting for staff (usually between tasks). The clinic usually runs in the morning (from 7:30 to 14:00), two to three days a week. Note that we removed extreme outliers from the observed data, including the time for an ECG that was recorded for one patient as taking 16 min, but included time during which the patient waited for the ECG technician to come to the room. We also removed the time it took to take blood for one patient who fainted in the middle of the process.

For the simulation, we used triangular distributions for the service times. We did so because actual service-time do not fit to any known distribution, and because triangu-

Table 2: Summary of the data

				Activity Duration Time (minutes)							Total Time (hours)		
Day	n	$\bar{n}$		Adm	In gown	ECG	Doc	Out of gown	Train	Blood	Visit	Interaction with staff	Wait for staff
1	28	2	μ	11	2	3	12	2	22	7	02:30	00:51	01:38
			σ	4	1	2	5	3	14	6	00:32	00:19	00:31
			Min	5	0	2	5	1	1	2	01:30	00:21	00:20
			Mode	10	2	2	10	2	-	5	-	-	-
			Max	25	5	16	22	20	45	28	03:27	01:42	02:27
2	34	4	μ	11	2	4	10	2	31	6	01:52	00:47	01:05
			σ	7	1	1	5	1	17	5	00:45	00:23	00:37
			Min	2	0	3	4	1	3	2	00:37	00:19	00:04
			Mode	10	4	3	10	2	-	5	-	-	-
			Max	40	5	10	25	6	60	30	03:45	01:52	02:19
3	28	2	μ	10	2	3	12	3	27	9	02:31	00:52	01:38
			σ	4	2	1	6	2	16	19	00:52	00:23	00:43
			Min	2	0	2	3	0	4	1	01:12	00:24	00:16
			Mode	8	2	3	10	2	30	5	-	-	-
			Max	21	12	8	30	9	48	100	04:13	02:05	03:01
Simulation	35		Min	1+10	2	3	8	2	15	2			
model	or		Mode	2+15	3	4	12	3	20	4			
	40		Max	3+20	5	6	25	5	45	7			

lar distributions have the advantage of being centrally located. We also used triangular distributions because they tend to increase the variance over other distributions with the same bounds that are more centrally located, such as the normal distribution. This would thus exacerbate, rather than possibly minimize, queuing issues due to variability. The last three lines in the tables reflect the values used in the simulation model. We used the observed data and discussions with management to generate the minimum duration (Min), the mode (Mode) and the maximum duration (Max). Since the new activities in the clinic also include a visit with the internist for patients with particular needs, a visit with the pharmacy technician and DVD education, in collaboration with management and staff we estimated the Min, Mode and Max times for them to be: (30,45,60) for the internist, (15,20,40) for the pharmacy technician and (30,33,45) for watching the DVD.

Given that patients will undergo the same tasks in the redesigned clinic, the estimated distributions are used to optimize and test the performance of that clinic.

4.3 Model design

To facilitate patient flow in the redesigned clinic, we wanted patients waiting primarily for a particular task to be able to perform other tasks as the resources needed for them to do so became available. For example, we wanted patients primarily waiting to enter an exam room to be able to provide their insurance information should they still need to do so, should one of the admissions staff become available before an exam room became available. From a conceptual modelling perspective, this requires individual patients to be in multiple queues.

We identified two approaches for implementing this modelling need. In the first approach, as suggested by Tocher [31], individual patients are added to multiple queue objects when those patients are available for more than one task. This requires removing a patient from all the queues the patient is in when that patient's turn for being served occurs in any of the queues. The computational impact of doing so would be significant with the simulation tool we used (Simul8).

The second approach is to implement separate queue objects for each combination of resources, e.g. to have a queue for a patient just waiting for an exam room, a queue for a patient waiting just to provide insurance information, and a third queue for patients simultaneously waiting for an exam room and to provide insurance information. Given our expectation that a large number of simulation replications would be needed to optimize the model, that this approach would reduce the time it would take to perform the simulation and the optimization, and that it would also make it easier to verify and obtain face validity of the simulation model, we chose this approach. The reader may refer to [32] for more details.

To facilitate visualization for verification and vaidation purposes, we decided to animate the model so that the state of each activity, resource, and individual queue is represented. With such a representation, organized like a swim-lane diagram so that different entities are in different rows, it is straightforward to see the transitions that occur at each moment in time. It is also straightforward to modify the quantities of each type of entity, e.g., exam rooms, since such a change does not affect the visual layout of the model. Since this project started with a management decision to move to a larger space with more exam rooms, the relative ease with which the model makes it possible to evaluate changes was critical to winning buy-in from the stakeholders.

Finally, the simulation model of the clinic requires a representation of the patients, staff members, physicians, DVD players, and exam rooms. In many simulation modelling environments the norm is to model the patients as work items (first-class simulation objects) and the remaining items as resources, i.e., second-class simulation objects lacking the full flexibility of first-class objects.

However, we wanted all representations to be modeled as first-class simulation objects, to make it possible to individually set the arrival times for patients and physicians, and the start times, break times, and lunch times for staff. This also made it possible to develop simulation logic, activated by programmable event handlers provided by the simulation development environment, to assign staff members, physicians, exam rooms, and DVD players to individual patients, and vice versa, based on rules for the flow of patients through the clinic, and availability of the appropriate objects.

4.4 Model verification and validation

Carson [6] defines verification as “the processes and techniques that the model developer uses to assure that his or her model is correct and matches any agreed-upon specifications and assumptions.” He further defines validation as “the processes and techniques that the model developer, model customer and decision makers jointly use to assure that the model represents the real system (or proposed real system) to a sufficient level of accuracy.”

Carson goes on to say that ”it should also be noted that no model is ever 100% verified or validated” and that ”any model is a representation of a system, and the model's behavior is at best an approximation to the system's behavior. When we (loosely) say that a model has been verified or validated, we mean that we have explicitly carried out a series of tasks to verify and validate our model to the degree necessary for our purposes.”

Robinson [25] presents verification and validation (V&V) from a slightly different perspective. He reminds us that ”that there is no such thing as absolute validity” and that ”the aim of V&V is not to prove that a model is correct, since this is not possible. Instead, the aim is to try and prove that a model is in fact incorrect. If it cannot be proved that a model is incorrect, then V&V has served to increase confidence in the model and its results.”

Unfortunately, we note that we cannot apply an empirical approach to validating the model, due to the fact that since the redesigned clinic has not yet been put into operation, we cannot yet obtain data about its operations. In context of that limitation, the two perspectives mentioned above, and the purpose of our modeling, and instead of trying to prove that the model is valid as suggested by Carson, Sargent [27] and Kleijnen [19], we instead apply Robinson's approach to try to prove that the model is incorrect. Then if we are not able to do so, we can conclude that our efforts have "served to increase confidence in the model and its results."

Applying this approach with tests from all of these authors, the approaches we used to try to prove that the model is incorrect include:

- *Checking the code:* Because the code used to specify simulation logic was heavily commented and highly modularized, we were able to and did carefully check the code for errors.

- *Visual checks with the animation:* Taking advantage of the fact that each run lasted less than 24 hours (of simulated time), that only 35 patients are served in the clinic each day, and that each patient only needs to go through a small list of activities, we traced through the simulation for several (simulated) days, an event at a time, to try to find:

  - Patients that were not pulled from the appropriate queue to an appropriate activity when a physician or staff member became available
  - Idle physicians or staff members that were not pulled from the correct queue to the appropriate activity when a patient became available.
  - Task times that were not plausible according to the distributions for the model.
  - Events not being properly added to the simulation's future-event list.
  - Patients that did not leave the clinic before the end of the day.
  - Staff not leaving before the end of the day.

- *Degeneracy tests:* As part of our scenario analyses, we tried to find the following situations that did not meet our expectation of the behavior of the redesigned clinic:

  - Decreasing the threshold for excessive waiting times decreasing expected total costs.

  - Decreasing the threshold for excessive waiting times decreasing physician idle time.
  - Increasing the number of patients served by the clinic each day decreasing expected total costs.
  - Increasing the weight of excessive patient waiting, relative to the other weights, increasing excessive patient waiting.
  - Increasing the weight of excessive patient waiting, relative to the other weights, decreasing physician idle time.
  - Adding exam rooms, when there were only a few exam rooms, increasing the expected total costs.
  - Combining registration and the collection of insurance information decreasing expected total costs.

- *Face Validity*:Presenting the underlying assumptions of the model, a high level perspective of the logic of the model, animation of the model as described above, and the results of the optimized model to the head nurse of the clinic and to the chief of surgical services, and either of them having any reason to doubt the validity of the model for the purposes of redesigning the clinic, or of the value of optimizing operations.

As discussed above, since none of the above approaches to invalidate the model were successful, we and the head nurse and the chief of servical services are confident that the model can be used to redesign and build the clinic.

4.5 Variables and objective function

Having verified and validated the model, our next step was to adapt it to measure the operating cost of different configurations, particularly with respect to three factors. The first factor was physician idle time. In Quebec hospitals, physicians are compensated for each service provided to each patient, and generally wish to minimize idle time between patients.

The second of these factors was the cost of staffing overtime, which the hospital wished to minimize. Staff may not leave for the day until all of their activities are finished.

The third factor was excessive patient waiting. We define waiting time to be the time after the patient arrives at the clinic and before the patient finishes all of their activities in the clinic during which a patient is not busy with a clinic activity, for example, the time during which a patient is sitting in an exam room after having their ECG taken but before a physician arrives. The existing presurgical screening clinic had a reputation, as do many of the hospital's clinics, for making patients wait a very long time. Thus, a decision was made to try to minimize excessive waiting time, which was defined to include the total waiting between activities in a visit, minus a threshold of sixty minutes. The threshold of sixty minutes was selected so as to approximately extrapolate the thirty to forty minutes discussed in the literature [7, 15] as acceptable waiting for a physician visit, to allow for the multiple activities that patients undergo during visits to our clinic.

Given these three factors we defined the following decision variables for each category of player:

- x_i = start time of person $i \in I$, where I is the union of D the set of physicians (GPs and internists), S the set of administrative staff, T the set of technicians (blood and ECG technicians) and H the set of nurses and pharmacy technicians;
- b_i = break time of $i \in J$, where $J = S \cup T \cup H$. We note that doctors typically do not take breaks;
- l_i = lunch time of $i \in I$; and,
- y_p = appointment time of patient $p \in P$ where P is the set of patients seen per day.

Discussions with management and documentation such as collective agreements, helped to determine the allowed range of values for these decision variables. From the start time of the clinic, the latest allowable arrival time for everyone is eight hours. For staff, the minimum time from arrival until their break is two hours and the longest allowable period of time is three hours, except for the ECG technician for whom it is four hours; the maximum time for the ECG technician was set to a large value because there is only one ECG technician and because ECGs are needed by many of the patients. For staff, the shortest allowable period from the end of a break to the start of lunch is two hours and the largest amount of time is three hours. One should note that these values are parameters which can readily be adapted for other settings.

The objective function is defined as the expectation of a weighted sum of idle time, overtime, excessive wait time and the number of patients who do not finish all of their clinic activities by the end of the day. The weights for the staff overtime, β, were based on estimates of their actual overtime costs. Section 5 presents a discussion of the weights associated with the physician idle time α and the weights for the excessive patient waiting γ. Different contexts lead to different values for these weights, for example selecting α to be larger than β increases the weight of the physician idle time in the objective. The penalty for patients not completing all of their clinic activities before the end of the day is set to a large value, \$1000 per patient, so as to encourage the optimization to prevent such occurrences.

We also defined the following variables:

- I_i =total idle time for physician $i \in D$;
- O_j = total overtime for $j \in J$;
- W_p = waiting time for patient $p \in P$;
- $Z_p = 0$ indicates that patient $p \in P$ has finished the process (and equals 1 otherwise).

All times are expressed in minutes. Given this notation, the objective function is:

Minimize

$$\mathbb{E}\left(\alpha \sum_{i \in D} I_i + \sum_{j \in J} \beta_j O_j + \gamma \sum_{p \in P} \max\{0, W_p - \mathcal{W}\} + 1000 \sum_{p \in P} Z_p\right)$$

where $\mathcal{W}$ is defined to be a reasonable patient waiting time.

4.6 Local search

Given specific space, staffing levels and parameters for the scenarios, the objective function was minimized with respect to the start times of staff and physicians and appointment times of patients, as well as the break and lunch times of staff. Though we plan on finding or developing a more global optimization procedure in future work [2, 34], for our initial analysis, to simplify the coding of the search function, we used the search heuristic discussed in (Algorithm 1), which uses a simple movement, i.e. a change of the booking appointment of a patient, that has proved efficient in many applications of combinatorial problems, such as vehicle routing [8], without embedding it in a more sophisticated approach [12, 30]. The first option available for moving a booking appointment was to do so one patient at a time; a second option was to do so two patients at a time, either by switching their appointment times or by changing their appointments times one a time. Because of its simplicity we selected the first approach, and alternatively used different neighbourhoods: each patients' appointment could be changed by 120, 60, 30, 15 or 5 minutes during the first, second, third, fourth and fifth set of 250 iterations (see parameter δ in Algorithm 1). This approach precludes appointment times from being arbitrary (e.g. 8:12), and also precludes the need for adding an anti-cycling rule such as that used in tabu search.

Algorithm 1 provides the specifics of how we change from one neighbourhood to the other.

Algorithm 1 Outline of NeighborhoodSearchHeuristic

```
Set z' to a very large value;
Generate initial values for all V decision variables;
Set s to generated values;
Set j = 1;
Set δ = 120 being the variation in appointment time to apply to the variables;
Set trial i = 1;
while i <= 2500 do
  TestNeighborhoodForImprovedSolution(j, δ); (see Algorithm 2)
  If j > X: j = 1, otherwise j = j + 1;
  Set i = i + 1.
  if MODULUS(i, 250) = 1 then
    UpdateDeltaMinutes: While (δ > 15minutes), δ = δ/2 (i.e, decrease by half); other-
    wise: if (δ = 15minutes), δ = δ/3; otherwise: δ = 120.
  end if
end while
```

Algorithm 2 explains how the moves are evaluated and accepted. After each feasible move in Algorithm 1, we run the simulation and compute the objective function. We run two local search around x_j by increasing this appointment time by δ minutes and decreasing it by 2δ minutes.

Algorithm 2 Outline of *TestNeighborhoodForImprovedSolution(j, δ)*

```
Set Improvement = 0;
for x_j = x_j + δ and x_j = x_j − 2δ do
   if x_j is feasible then
      Run a trial of the simulation;
      Compute the objective function value z with results from trial;
      if z < z' then
         Set z' = z;
         Set Improvement = 1;
      end if
   end if
end for
if (Improvement == 0) then
   Set x_j = x_j + δ;
else if (Improvement == 1) then
   Set x_j = x_j + 2δ;
end if
```

Figure 2 presents a high level view of our approach and how the the optimization is applied to the simulation model.

Note that trials consists of 100 simulation runs, where each run is one day long. Pseudo random service times are generated individually for each activity for each of the 100 runs in the trial; these service times are reused for all trials so as to facilitate comparison of trial results. It takes approximately two seconds to simulate 100 days for one scenario with a specific combination of parameters, and arrival, break and lunch times. For each scenario with a specific set of parameters, the optimization heuristic typically evaluates up to 2000 combinations of arrival, break and lunch times. It typically takes between 75 and 90 minutes to perform the optimization of one scenario with a specific set of parameters. These times were obtained on a computer using an i7-4790 CPU running at 3.6 GHz. The simulation software used was Simul8 2013 running on top of Windows 10 which was running in a QEMU/KVM virtual machine on top of Lubuntu 16.04.

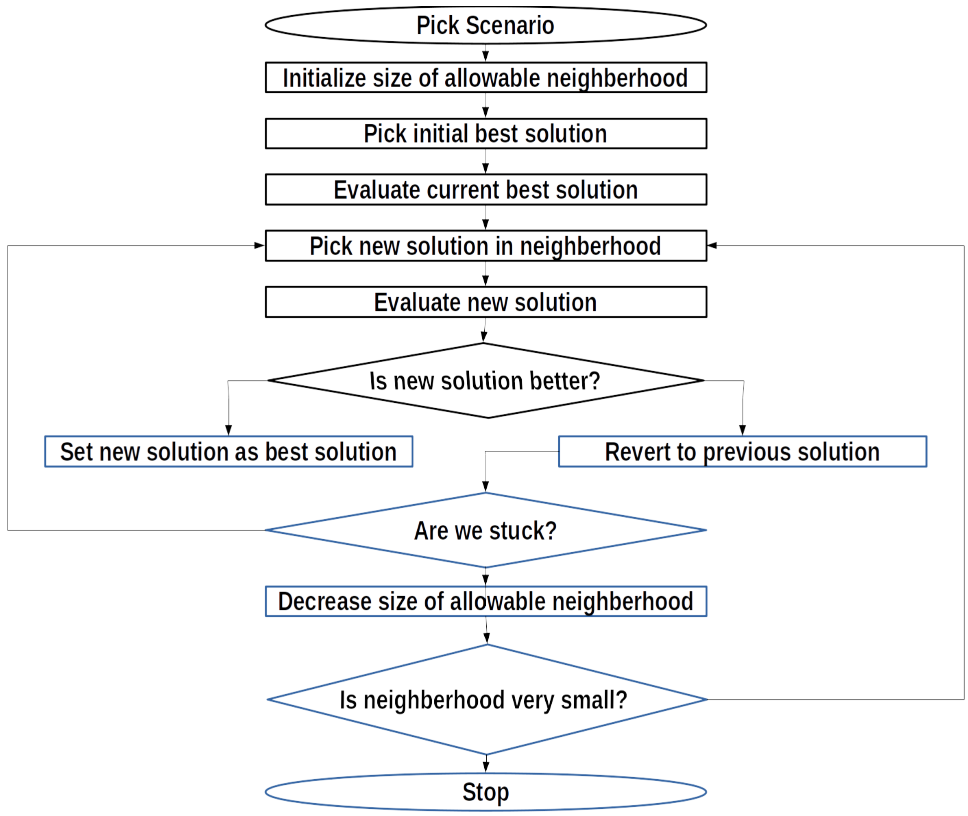

Figure 2: High level view of our approach

5 Experiments and results

In this section we present the results for each of the scenarios described in Section 4.1. As mentioned there, the purpose of these scenarios is to provide senior management with insights as to the effects of several parameters on the optimized operations of the clinic. As a baseline for comparison, we assume that individualized nurse education is provided to each patient, that registration and insurance information collection are treated as two separate activities, that 35 patients are booked for each day, that the threshold for excessive waiting, $\mathcal{W}$, is 60 minutes, that the one internist sees patients much later in the morning than do the two general practitioners, that there are two admissions staff, that there are three technicians (one ECG technician, one technician for taking blood, one pharmacy technician), and that there are three nurses. Also, as part of the baseline, the objective function weight for physician idle hours (α) is 100, the weight for administrative staff overtime hours (β_1) is 30, the weight for technician overtime hours (β_2) is 45, the weight for the nurse overtime hours (β_3) is 80, and the weight for excessive patient weighting hours (γ) is 20.

Except where specified otherwise, all the scenario analyses assume these baseline characteristics.

5.1 (Scenario 1) Excessive wait time thresholds

The purpose of this scenario is to demonstrate to senior management the effect of different excessive wait time thresholds, in context of different exam room counts, on optimized clinic operations, so they can select the threshold that they feel is most appropriate for the clinic. To demonstrate these effects, we evaluate the optimized total expected cost operating cost of using excessive wait time thresholds of 60, 45, and 0 minutes with 5, 6, and 7 exam rooms.

Table 3 displays the results of this scenario. The first set of columns displays the results with $\mathcal{W} = 60$, the second set of columns displays the results for $\mathcal{W} = 45$, and the third set of columns displays the results for $\mathcal{W} = 0$.

Table 3: Comparative results when the threshold of excessive wait time (in minutes) varies

			Baseline ($\mathcal{W} = 60$)			$\mathcal{W} = 45$			$\mathcal{W} = 0$		
	Exam Rooms		5	6	7	5	6	7	5	6	7
Idle Minutes	GP	μ	6	8	1	13	6	10	24	19	38
		p	[0;15]	[1;17]	[0;10]	[3;32]	[2;17]	[0;28]	[9;39]	[5;33]	[25;56]
	In	μ	1	3	1	2	3	5	8	5	12
		p	[0;10]	[0;19]	[0;10]	[0;10]	[0;16]	[0;22]	[0;29]	[0;19]	[0;34]
Overtime Minutes	AS	μ	0	6	7	3	12	1	2	1	23
		p	[0;4]	[0;20]	[0;18]	[0;37]	[0;37]	[0;7]	[0;19]	[0;18]	[0;49]
	BT	μ	0	0	0	0	0	12	6	3	6
		p	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;56]	[0;28]	[0;21]	[0;25]
	ET	μ	0	0	0	0	0	0	0	0	0
		p	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]
	RN	μ	0	1	2	3	1	1	4	3	1
		p	[0;3]	[0;5]	[0;11]	[0;13]	[0;6]	[0;8]	[0;12]	[0;9]	[0;8]
Excessive Wait Minutes	PA	μ	1	2	4	4	6	6	27	28	35
		p	[0;4]	[0;8]	[1;9]	[1;9]	[2;13]	[2;11]	[22;36]	[22;36]	[30;43]
Total Cost	*Opt.*	μ	36	63	70	111	117	125	436	411	591
		σ	19	34	30	*41*	*34*	*40*	*56*	*47*	*53*
	Not	μ	1424	1172	1136	1544	1292	1258	1971	1720	1689

The results show that while excessive waiting time is almost independent of the number of exam rooms (between 1 and 4 minutes for $\mathcal{W} = 60$, 4 and 6 for $\mathcal{W} = 45$ and 27 to 35 for $\mathcal{W} = 0$), there appears to be a small dependency for GP idle time and staff overtime. When optimization is not used, the solutions improve with five or more exam rooms, but in all cases, the benefit of optimization is extremely clear. We observe that the results are better with five rooms and very similar with six and seven rooms when $\mathcal{W} = 60$. We also observe that reducing $\mathcal{W} = 60$ to $\mathcal{W} = 45$, only decreases the average excessive wait by 10 minutes, while the GP idle time increases by 7 to 9 minutes with 5 rooms and 7 rooms and stays similar with 6 rooms. Finally, when minimizing directly patient wait time, average physician idle time increases considerably.

5.2 (Scenario 2) Increased demand

The purpose of this scenario is to demonstrate to senior management the effect of increased demand on optimized clinic operations. To determine that effect, we evaluate the optimized total expected cost operating cost of 35 and 40 patient visits per day with 5, 6, and 7 exam rooms. We see that when increasing the number of patient visits per day to 40, admission staff overtime increases, suggesting that the admission staff would be very near to their capacity with the increased patient visits. We also see that there is a small but nonetheless expensive increase in per nurse overtime and per general practitioner idle time. All three of these amounts are somewhat smaller with six exam rooms than with five exam rooms. These results suggest if the clinic were to expand, it should be possible to run it efficiently, though it might be desirable to add some additional admission staffing and nursing each day.

5.3 (Scenario 3) Physician scheduling

The purpose of this scenario is to determine the feasibility of reducing the number of hours the clinic operates each day by having all of the physicians work at the same time. Keeping in mind that two rooms are reserved for the sole use of the internist, the results in Table 5 show that that having all the physicians working at the same time very significantly increases physician idle time.

Table 4: Expected time and cost as a function of the number of patients served each day

	Exam Rooms			Baseline - 35 patients 5	6	7	40 patients 5	6	7
Average duration (in mins)	Idle	GP	μ	6	8	1	8	5	6
			p	[0;15]	[1;17]	[0;10]	[0;24]	[0;20]	[0;21]
		In	μ	1	3	1	3	1	0
			p	[0;10]	[0;19]	[0;10]	[0;19]	[0;11]	[0;8]
	Over	AS	μ	0	6	7	26	24	32
			p	[0;4]	[0;20]	[0;18]	[6;43]	[0;48]	[10;49]
		BT	μ	0	0	0	1	0	0
			p	[0;0]	[0;0]	[0;0]	[0;9]	[0;0]	[0;0]
		ET	μ	0	0	0	0	0	0
			p	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]
		RN	μ	0	1	2	11	5	3
			p	[0;3]	[0;5]	[0;11]	[1;20]	[0;12]	[0;12]
	Excessive wait	PA	μ	1	2	4	4	3	4
			p	[0;4]	[0;8]	[1;9]	[1;12]	[1;10]	[2;13]
	Total cost	*Opt.*	μ	36	63	70	162	118	156
			σ	*19*	*34*	*30*	*56*	*54*	*54*
		Uopt.	μ	1424	1172	1136	2022	1683	1624

5.4 (Scenario 4) Merging administrative activities

The purpose of this scenario is to determine whether it is better to merge administrative activities or to perform them separately. To do so, we used the baseline settings with six exam rooms and replicated it with the administrative activities merged. The rationale for performing this test is the possibility that having patients wait to provide information to administrative staff twice might result in increasing the total cost. However, the results show that total costs increase significantly (see Table 6) when patients have to perform all of their administrative activities at registration because patients end up waiting a lot more time to go through the merged administrative activities, which in turn results in very significantly increase in general practitioner idle time while they wait for patients.

5.5 (Scenario 5) Booking rules

In practice, when booking patients and scheduling staff, decision makers often look for "easy to implement" rules they can use to perform these activities. Thus in this section, in context of our thesis that operations should be optimized, we compare several "easy to implement" rules with optimization in context of the baseline settings which senior management agreed to.

Table 5: Expected time and cost as a function of physician scheduling

				Baseline			All physicians in morning		
	Exam Rooms			5	6	7	5	6	7
Average duration (in mins)	Idle	GP	μ	6	8	1	52	9	3
			p	[0;15]	[1;17]	[0;10]	[36;69]	[1;23]	[0;13]
		In	μ	1	3	1	131	85	117
			p	[0;10]	[0;19]	[0;10]	[111;154]	[74;98]	[108;135]
	Over	AS	μ	0	6	7	5	5	3
			p	[0;4]	[0;20]	[0;18]	[0;39]	[0;35]	[0;19]
		BT	μ	0	0	0	4	0	0
			p	[0;0]	[0;0]	[0;0]	[0;26]	[0;0]	[0;0]
		ET	μ	0	0	0	0	0	0
			p	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]	[0;0]
		RN	μ	0	1	2	1	0	2
			p	[0;3]	[0;5]	[0;11]	[0;6]	[0;5]	[0;9]
	Excessive wait	PA	μ	1	2	4	2	2	2
			p	[0;4]	[0;8]	[1;9]	[0;5]	[0;6]	[0;7]
	Total cost	*Opt.*	μ	36	63	70	421	205	247
			σ	*19*	*34*	*30*	*50*	*37*	*37*
		Uopt.	μ	1424	1172	1136	1405	5711	2710

The rules are the following:

- Rule 1: one patient is booked to each 10 minute time slot randomly (i.e, without using any information on their profile). This is the rule used currently in the clinic;
- Rule 2: up to two patients are booked to each 10 minute time slot randomly;
- Rule 3: heaviest patients are booked first. This requires that the head nurse evaluates all files to evaluate the profile of each patient; and,
- Rule 4: lightest patients are booked first.

We note that the research literature often suggests processing patients with the largest variability in resource usage last; this in turn suggests that Rule 4 ought to do the best, as heavier patients tends to have higher variability in their total usage of resources.

The results can be found in Table 7. For each rule we show the average (μ) for staff overtime, physician idle time, and excessive patient waiting for the clinic using five, six or seven exam rooms. We also display the utilization rate (ρ) for staff and physicians, so as to provide another perspective on the relative performance of the four rules. While these strategies are easy to implement, the results clearly show that these strategies do

Table 6: Scenario 4 - Comparative results of merging administrative activities with 6 exam rooms

				Baseline	Merged Administrative Activities
Average duration (in mins)	Idle	GP	μ	8	42
			p	[1;17]	[29;58]
		In	μ	3	2
			p	[0;19]	[0;13]
	Over	AS	μ	6	0
			p	[0;20]	[0;0]
		BT	μ	0	4
			p	[0;0]	[0;28]
		ET	μ	0	1
			p	[0;0]	[0;11]
		RN	μ	1	3
			p	[0;5]	[0;11]
	Excessive wait	PA	μ	2	14
			p	[0;8]	[9;21]
	Total cost	*Opt.*	μ	63	324
			σ	*34*	*54*
		Uopt.	μ	1172	4367

not perform nearly as well as optimization, with respect to the criterion usually used in practice: idle time, overtime and wait time. In particular, we observe that the random booking rules (Rule 1 and 2) perform poorly, which is very interesting since Rule 2 is essentially the approach currently used in the clinic, since no questions are asked of the patients in order to determine if they should be booked at a different time. Also with respect to Rules 1 and 2, we see that while there is no clear difference in internist idle time when using five, six or seven exam rooms, GP idle time and excessive patient waiting time decrease as we increase the number of rooms.

When comparing Rules 3 and 4, we observe that Rule 4, i.e. booking "lighter" patients first (and heaviest patients last), significantly reduces average excessive waiting time for patients from 38 to 6 minutes when five exam rooms are used and from 32 to 6 minutes when six or seven exam rooms are used. No notable change is observed for internist idle time but GP idle time also decreases as we increase the number of rooms.

Finally, when comparing the average of the resulting idle times, overtime, excessive patient waiting and utilization obtained using Rule 4 to the results obtained by optimization, we see that the optimization yields much better results than does Rule 4; in particular, average idle time is much reduced for physicians (from 212 minutes to 5 minutes), while

Table 7: Testing known strategies for booking

		Resource	Number of exam rooms 5 μ	5 ρ	6 μ	6 ρ	7 μ	7 ρ
Rule 1: random booking	Idle Time	GP	317	40%	232	45%	206	48%
		In	221	50%	220	51%	219	51%
	Overtime	AS	60	75%	30	77%	16	80%
		BT	97	53%	24	50%	14	51%
		ET	60	51%	0	55%	0	64%
		PH	0	72%	0	72%	0	72%
		RN	71	75%	34	79%	25	80%
	Excessive wait time	PA	46		11		6	
Rule 2: double booking	Idle Time	GP	146	57%	121	61%	84	69%
		In	76	75%	76	75%	76	75%
	Overtime	AS	7	93%	0	92%	0	95%
		BT	9	51%	0	51%	0	51%
		ET	3	73%	0	73%	0	79%
		PH	0	88%	0	88%	0	88%
		RN	12	82%	7	82%	5	83%
	Excessive wait time	PA	49		46		42	
Rule 3: Heavy patients first	Idle Time	GP	126	60%	89	68%	80	70%
		In	33	87%	33	87%	33	87%
	Overtime	AS	0	88%	0	87%	1	86%
		BT	0	51%	0	51%	0	51%
		ET	0	69%	0	76%	0	77%
		PH	0	98%	0	98%	0	98%
		RN	57	77%	54	77%	54	77%
	Excessive wait time	PA	38		32		32	
Rule 4: Lighter patients first	Idle Time	GP	212	47%	186	50%	186	50%
		In	19	92%	20	92%	20	92%
	Overtime	AS	60	78%	69	79%	71	79%
		BT	143	59%	145	59%	145	59%
		ET	2	43%	3	43%	3	43%
		PH	0	56%	0	56%	0	56%
		RN	48	80%	47	80%	47	80%
	Excessive wait time	PA	6		6		6	
Fully optimized	Idle Time	GP	5	97%	3	99%	3	99%
		In	0	100%	0	100%	3	99%
	Overtime	AS	1	83%	1	83%	1	86%
		BT	2	51%	2	51%	2	51%
		ET	0	58%	0	57%	1	46%
		PH	0	55%	0	78%	0	64%
		RN	3	88%	2	87%	2	88%
	Excessive wait time	PA	2		2		2	

Note: All durations are in minutes

average staff overtime and average excessive patient waiting are both also reduced.

5.6 Discussion and insights for managers

While the analysis of the different scenarios provides information needed by senior management needing to make decisions on the redesigned clinic, the most important result of this analysis is how it highlights the extent of the decrease in total expected costs provided by optimizing operations, even over the rules discussed in the last subsection.

To understand why the optimization performs so much better than those rules, and in particular Rule 4, we first compare the results of Rule 4 to those obtained with full optimization. Looking at Figure 3, we see that following Rule 4, all staff arrive at 7:30 and physicians arrive at 8:00 and 8:30. In contrast, when applying the optimization, admistrative staff start times, break times, and lunch times are staggered to better match demand of the clinic. In particular, one administrative staff arrives at the opening at the clinic while the second one arrives between 2 and 3 hours later. Nurses clearly do not need to be at the clinic when it opens so they are scheduled to arrive later in the day. Likewise, the start time of the second GP is also delayed to 2 hours after the clinic opens. And while some of these times could have been guessed without using optimization, it is unlikely that the result would have been as good, particularly with respect to the scheduling of break and lunch times.

The second insight as to why this occurs can be seen by looking at the optimized patient appointment times in Table 8, which shows the booking times for patients starting at 8:00am. The "Rule 4" column shows that one patient at a time is booked every 10 minutes. The number in the column shows the category of patients (informal scaling used internally): for example, one patient of category 1 (lightest patients) is scheduled at 8:00am, another one at 8:10am and so on. The last appointment of the day is at 1:40pm with a patient of the heaviest category. The remaining columns refer to the booking appointments obtained from the optimization for 5, 6 or 7 exam rooms respectively. ".+." in the table reflects that two patients are booked at the same time. In these three columns we observe that: 1) patients are not booked at a regular interval; 2) double booking is used at strategic moments in the day; 3) patients from different categories are mixed during the day. The key takeaway from this is that while some benefit can be achieved by applying "easy to

Table 8: Illustration of booking appointment times obtained when fully optimizing the problem

First half of the day					Second half of the day				
Time	Rule 4	5 rooms	6 rooms	7 rooms	Time	Rule 4	5 rooms	6 rooms	7 rooms
8:00	1			1	11:30	3		4	
8:05					11:35			3	
8:10	1	1	1		11:40	3	3+4		3
8:15			1		11:45		3		
8:20	1	1	1		11:50	3	4	3	
8:25				1	11:55			4	
8:30	1	1	1	1	12:00	3	3		2
8:35		1			12:05		2		2
8:40	1		1		12:10	3	3		3
8:45					12:15			3	
8:50	2			1	12:20	3		2+3	
8:55					12:25				3
9:00	2	1			12:30	4		3	3
9:05					12:35			5	2
9:10	2				12:40	4	2+5	3	5
9:15					12:45		2		
9:20	2		1	2	12:50	4			
9:25		2		3	12:55				6
9:30	2	4		3+4	13:00	5		5	
9:35					13:05				5
9:40	3	3	2+3	1+2	13:10	5	5		
9:45			2	3	13:15				
9:50	3	3+3	3+3		13:20	5			
9:55				4	13:25				
10:00	3	3	2+3		13:30	6		5	5
10:05				3	13:35		5		
10:10	3		3		13:40	6			
10:15			3	3	13:45				
10:20	3	3	3	3	13:50			6	
10:25					13:55		6		6
10:30	3	3+3	3		14:00				
10:35					14:05				
10:40	3	3	3+4		14:10				
10:45					14:15				
10:50	3	3		3+3	14:20				
10:55				4	14:25				
11:00	3	2+3	3	3+3	14:30				
11:05					14:35				6
11:10	3	3	3	3	14:40		6		
11:15					14:45				
11:20	3	3+3	2	3	14:50			5	
11:25					14:55				

implement" rules to scheduling patients and staff, optimizing the model yields much better results.

6 Conclusion

In this paper, we present a research project in which we address the multiple challenges of designing and operating a complex clinic. In contrast to existing research, we optimize the operations of a simulation model of clinic designs at a fine-grained level by including individual decision variables for the start time of each physician, the appointment time of each patient and the start, break, and lunch times of each staff member, with the goal of minimizing the costs of staff overtime, physician idle time, and excessive patient waiting time. This optimization capability is then used to determine the tradeoffs between the number of exam rooms and to evaluate the tradeoffs inherent to other design decisions. We show that simultaneously optimizing the decision variables at this more granular level can significantly reduce physician idle times, staff overtime, and excessive patient waiting all at the same time, even though we only used a very simple heuristic to perform the optimization. These results also suggest the applicability of this approach to other multi-activity clinics such as cancer treatment clinics.

Given the richness of this problem and the potential to apply the results of this work to different types of clinics, further analysis is expected. In particular, we plan on attempting to refine our analysis of the activity times of individual patients, to try to identify whether some of those activity times can be predicted by the patient's profile, and if so, to adapt the model to take advantage of that information. We also plan to redevelop the model in a compiled programming language with a significantly improved heuristic to facilitate more dynamic use of the model.

References

[1] Guide de gestion, planification des activités chirurgicales, mécanisme central de gestion de l'accès aux services spécialisés et surspécialisés. Ministère de la Santé et des Services sociaux du Québec, www.msss.gouv.qc.ca, September 2010 (2010)

[2] Amaran, S., Sahinidis, N., Sharda, B., Bury, S.J., Simulation optimization: A review of algorithms and applications. *Ann Oper Res*, 240: 351. https://doi.org/10.1007/s10479-015-2019-x

[3] Berg, B.P., Denton, B.T., Nelson H., Balasubramanian, H., Rahman, A., Bailey, A., Lindor, K. (2010). A discrete event simulation model to evaluate operational performance of a colonoscopy suite. *Medical Decision Making*, 30(3): 380-7

[4] Berg, B.P., Denton, B.T., Erdogan, S.A., Rohleder, T., Huschka T. (2014). Optimal booking and scheduling in outpatient procedure centers. *Computers & Operations Research*, 50: 24-37

[5] Berg, B.P., Denton, B.T. (2012). Appointment planning and scheduling in outpatient procedure centers. *In* Handbook of Healthcare System Scheduling. International Series in Operations Research & Management Science, 168

[6] Carson, J. (2002). Model verification and validation. *In* Proceedings of the 2002 Winter Simulation Conference. http://informs-sim.org/wsc02papers/008.pdf

[7] Cartwright, A., Windsor, J. (1992). Outpatients and their doctors: A study of patients, potential patients, general practitioners and hospital doctors. HMSO London 1992

[8] Cordeau, J.-F., Laporte, G., Mercier, A. (2001). A unified tabu search heuristic for vehicle routing problems with time windows. *Journal of the Operational Research Society*, 52: 928-936

[9] Chiandussia, G., Codegone, Ferrero, S., Varesio, F.E., Comparison of multi-objective optimization methodologies for engineering applications (2012) *Computers & Mathematics with Applications*, 63:5, 912-942

[10] Côté, M.J. (1999). Patient flow and resource utilization in an outpatient clinic. *Socio-economic Planning Sciences*, 33(3): 231-245

[11] Fetter, R.B., Thompson, J.D. (1966). Patients' Waiting Time and Doctors' Idle Time in the Outpatient Setting, Health Services Research Summer 1966

[12] Glover, F., Kelly, J.P., Laguna, M. (1996). New advances and applications of combining simulation and optimization. *In* Proceedings of the 1996 Winter Simulation Conference, J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain (Eds)

[13] Groothuis, S., van Merode, G., Hasman, A. (2001). Simulation as decision tool for capacity planning. *Computer Methods and Programs in Biomedicine*, 66(2-3): 139-151

[14] Hashimoto, F., Bell, S.J. (1996). Improving outpatient clinic staffing and scheduling with computer simulation. *Journal of General Internal Medicine*, 11(3): 182-4

[15] Huang, X.M. (1994). Patient waiting in an outpatient clinic and its applications. *Health Service Management Research*, 7: 2-8

[16] Huang, Y.L. (2016). The Development of Patient Scheduling Groups for an Effective Appointment System. *Appl Clin Inform*, 7(1): 43?58

[17] Huschka, T., Denton, B.T., Narr, B., Thompson, A. (2008). Using simulation in the implementation of an outpatient procedure center. *In* Proceedings of the 2008 Winter Simulation Conference

[18] Jacobson, S.H., Hall, S.N., Swisher, J.R. (2006). Discrete-event simulation of health care systems. *In* Patient flow: Reducing Delay in Healthcare Delivery, International Series in Operations Research & Management Science, 91: 211-252

[19] Kleijnen, J., (1995). Verification and validation of simulation models, *European Journal of Operations Research*, 82: 145-162

[20] Morrice, D.J., Wang, D., Bard, J.F., Leykum, L.K. Noorily, S. and Veerapaneni, P. (2014). A Patient-Centered Surgical Home to Improve Outpatient Surgical Processes of Care and Outcomes." *IIE Transactions on Healthcare Systems Engineering*, 3: 119-134

[21] Oh, H., Muriel, A., Balasubramanian, H., Atkinson, K., Ptaszkiewicz, T. (2013). Guidelines for scheduling in primary care under different patient types and stochastic nurse and provider service times. *IIE Transactions on Healthcare Systems Engineering*, 3(4): 263-279

[22] Pinedo, M., Zacharias, C., Zhu, N. (2015). Scheduling in the service industries: An overview. *Journal of Systems Science and Systems Engineering*, 24: 1

[23] Rais, A., Viana, A. (2011). Operations research in healthcare: A survey. *International Transactions in Operational Research*, 18(1): 1-31

[24] Rohleder, T.R., Lewkonia, P., Bischak, D.P., Duffy, P., Hendijani, R. (2011). Using simulation modeling to improve patient flow at an outpatient orthopedic clinic. *Health Care Management Science*, 14(2): 135-45

[25] Robinson, S. (1997) Simulation model verification and validation: increasing the users' confidence. *In* Proceedings of the 1997 Winter Simulation Conference, Andradóttir, S., Healey, K.J., Withers, D.H., Nelson, B.L. (Eds)

[26] Santibáñez, P., Chow, V., French, J., Puterman, M., Tyldesley, S. (2009). Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency's ambulatory care unit through simulation. *Health Care Management Science*, 12(4): 392-407

[27] Sargent, R.B. (2011). Verification and validation of simulation models. *In* Proceedings of the 2011 Winter Simulation Conference, Jain, S., Creasey, R.R., Himmelspach, J., White, K.P. and Fu, M. (Eds)

[28] Swisher, J.R., Jacobson, S.H. (2002). Evaluating the design of a family practice healthcare clinic using discrete-event simulation. *Health Care Management Science*, 5: 77-88

[29] Taheri, J., Gellad, Z., Burchfield, D., Cooper, K. (2012). A simulation study to reduce nurse overtime and improve patient flow time at a hospital endoscopy unit. *In* Proceedings of the 2012 Winter Simulation Conference, C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher (Eds)

[30] Tekin, E., Sabuncuoglu, I. (2004). Simulation optimization: A comprehensive review on theory and applications. *IIE Transactions*, 36: 1067-1081

[31] Tocher, K.D. (1963). The Art of Simulation. English Universities Press Ltd., p. 184

[32] Troy, P., Lahrichi, N., Rosenberg, L. (2013). An alternative approach to modeling a pre-surgical screening clinic. *In* Proceedings of the 2013 Winter Simulation Conference, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl (Eds)

[33] White, D.L., Froehle, C.M., Klassen, K. (2011). The effect of integrated scheduling and capacity policies on clinical efficiency. *Production and Operations Management*, 20(3): 442-55

[34] Xu, J., Huang, E., Chen, C., Lee, L. (2015). Simulation Optimization: A Review and Exploration in the New Era of Cloud Computing and Big Data, *Asia-Pacific Journal of Operational Research* 32:3

Appendix A - Questionnaire

In order to improve pre-surgical screening services, the Surgery Program is collecting information on the time required for our patients to go through their various tests and interviews while at the pre-admission testing clinic. This data will then be used to improve the scheduling of our patients in order to minimize wait times.

You are invited to participate to this study through the completion of this questionnaire. If you have any questions or concerns regarding this survey, please feel free to ask the project leader or staff member present.

Your participation in this study is optional. The service you will be receiving at the clinic will not be affected by your decision to participate in this study. Your collaboration is important to us and all information obtained by you during this study will remain confidential and unidentifiable.

Gender	☐ Male	☐ Female			
Age	☐ less than 40	☐ 40-60	☐ 60-70	☐ 70-80	☐ more than 80

Do you have mobility difficulties that require you to take extra time to get dressed? ☐ Yes ☐ No

Do you require the aid of a person or an assistive device (cane, walker, wheelchair, etc.) to get out of a chair or walk short distances? ☐ Yes ☐ No

If yes, please check the appropriate box: ☐ Cane ☐ Walker ☐ Wheelchair ☐ Other

Did you bring a family member with you to the clinic? ☐ Yes ☐ No

For each of the tasks you go through, please fill up the blank space when relevant.

Appointment time: ________________ Leaving time: ________________

Thank you very much for participating in this study and helping us improve the quality of our services.

	Starting time	Ending time
Registration	:	:
Changing into gown	:	:
ECG	:	:
Meeting the doctor	:	:
Changing back from gown	:	:
Blood test	:	:
Urine test	:	:
Nurse training	:	:

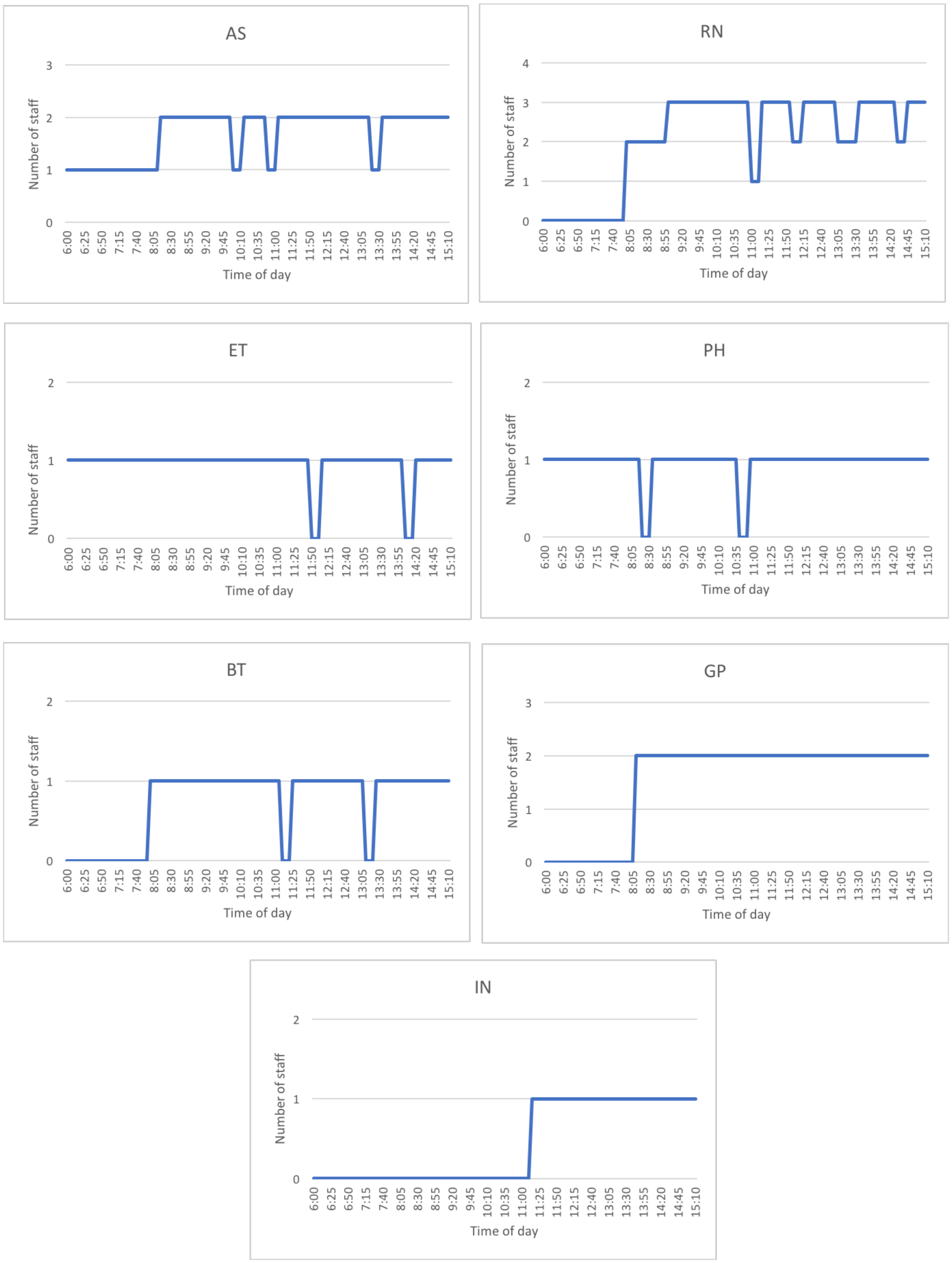

Figure 3: Schedules and break times for staff in the clinic with optimization for 6 exam rooms