Health Care Management Science (2021) 24:26–40
https://doi.org/10.1007/s10729-020-09528-y

A robust optimization model for tactical capacity planning in an outpatient setting

Nazanin Aslani[1] · Onur Kuzgunkaya[1] · Navneet Vidyarthi[2] · Daria Terekhov[1]

Received: 15 September 2019 / Accepted: 26 October 2020 / Published online: 20 November 2020

Abstract
Tactical capacity planning is a key element of planning and control decisions in healthcare settings, focusing on the medium-term allocation of a clinic's resources to appointments of different types. One of the most scarce resources in healthcare is physician time. Due to uncertainty in demand for appointments, it is difficult to provide an exact match between the planned physician availability and appointment requests. Our study uses cardinality-constrained robust optimization to develop tactical capacity plans which are robust against uncertainty, providing a feasible allocation of capacity for all realizations of demand to the extent allowed by the budget of uncertainty. The outpatient setting we consider sees first-visit patients and re-visit patients, and both patient types have access time targets. We experimentally evaluate our robust model and its practical implications under different levels of conservatism. We show that we can guarantee 100% feasibility of the robust tactical capacity plan while not being fully conservative, which will lead to the clinic saving money while being able to meet demand despite uncertainty. We also show how the robust model helps us to identify the critical time periods leading to worst case physician peak load, which could be valuable to decision-makers. Throughout the experiments, we find that the step of translating available data into an uncertainty set can influence the true conservatism of a solution.

Keywords Tactical capacity planning · Outpatient clinic · Operations research · Cardinality-constrained robust optimization · Demand uncertainty · Access time

Highlights

- Our robust optimization model provides a tactical capacity plan for an outpatient setting with first-visit and re-visit patients dependent on a decision-maker's risk attitude toward demand uncertainty.
- We show that the step of translating available data into optimization model parameters (specifically, uncertainty set) can influence the true conservatism of a robust solution and requires careful consideration in practice.
- Our analysis of robust solutions shows when worst-case realizations have the most impact, which can be useful at the tactical planning level to arrange alternative means to accommodate surges in demand.

✉ Daria Terekhov
daria.terekhov@concordia.ca

Nazanin Aslani
na_aslan@encs.concordia.ca

Onur Kuzgunkaya
onur.kuzgunkaya@concordia.ca

Navneet Vidyarthi
N.Vidyarthi@concordia.ca

1 Department of Mechanical, Industrial and Aerospace Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, 1455 de Maisonneuve Blvd. W., Montréal, Québec, H3G 1M8, Canada

2 Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, 1450 Guy St, Montréal, Québec, H3H 0A1, Canada

1 Introduction

One of the essential steps toward addressing physician scarcity and long access times in healthcare is tactical capacity planning (TCP). TCP involves making tactical-level decisions, i.e., medium-term planning decisions for a group of patients instead of individual ones [2]. TCP requires determining the resource capacity (e.g., total number of available physician hours) and the allocation of that capacity to different patient classes to meet certain performance targets [13, 16].

In outpatient settings, where care delivery happens without overnight hospitalization, a typical performance target is based on access time, which is the time between the arrival of an appointment request and the scheduled appointment time.

We study TCP for an outpatient setting with the following characteristics: two appointment types, corresponding to appointments for new patients and follow-up visits; dependence between appointment types, i.e., the number of follow-ups depends on the number of new patients; and access time targets for each appointment type. This setting was first described and studied by Nguyen et al. [16], who developed a deterministic model for finding the total required physician time and its allocation to each patient type to meet access time targets in an outpatient urology clinic.

However, planning using a deterministic approach may lead to the clinic facing periods of high congestion and inability to meet access targets when demand deviates from the particular scenario used to construct the plan. To overcome this issue, we use robust optimization to create plans that can meet access targets in spite of demand uncertainty. We make a general assumption that there is uncertainty in new patient demand, without knowledge of the exact probability distribution.

Using the cardinality-constrained method of Bertsimas and Sim [4], we develop a robust TCP model for the outpatient setting described by Nguyen et al. [16]. We conduct an extensive set of experiments to determine the level of robustness based on cost and infeasibility probability of a robust solution. In addition to model development and experimentation, we explicitly consider the steps required in practice before and after a mathematical model is run. First, through numerical experiments we show that the step of translating available data into an uncertainty set can influence the true conservatism of a robust solution. Second, we develop recommendations for planners through interpreting the output of the robust TCP model. In particular, information about when worst-case realizations have the most impact can be useful at the tactical planning level to arrange alternative means to accommodate surge in demand.

The paper is organized as follows. Section 2 describes recent work in robust optimization in healthcare planning and scheduling. Section 3 describes the problem. Section 4 focuses on the development of a cardinality-constrained robust optimization model to address demand uncertainty. In Section 5, we present experimental results and in Section 6, we discuss the three key observations from our study. Section 7 concludes the paper.

2 Robust optimization in healthcare planning and scheduling

One of the major challenges in the development of planning and scheduling models for healthcare environments is the availability of data and data uncertainty. If data uncertainty is ignored during the model development process, and the realization of data is different from the nominal values, several constraints may be violated, and an optimal solution found using the nominal data may in practice not be optimal or even feasible. In some cases, where enough information regarding the input parameters of the model is available for finding a probability distribution that fits well, one can use stochastic programming techniques to minimize an expected objective function. However, a lack of reliable data may mean that fitting a representative distribution is not possible. In this case, an alternative approach is robust optimization (RO) [1], which does not require knowledge of the probability distribution of uncertain parameters [8, 12]; instead, RO assumes that the value of an uncertain parameter belongs to an uncertainty set.

The quality of a robust approach is evaluated based on two criteria: remaining feasible despite changing parameter values and the cost of doing so. The cost of a robust solution is attributed to potential over-conservatism and is measured by evaluating a trade-off between the robustness and the optimal objective value.

Another reason for using RO is computational tractability [1] in comparison to scenario-based approaches. Nguyen et al. [17] extended their previous paper [16] to address demand uncertainty by formulating a stochastic linear optimization model with chance constraints, assuming either full or partial knowledge of the uncertainty in each period of the planning horizon. In contrast, in this paper, we do not assume any knowledge of the probability distribution and develop a robust tactical capacity planning model based on a particular RO approach known as cardinality-constrained RO.

Drawbacks of the other RO approaches are the issue of overconservatism and computational intractability [4, 5]. In cardinality-constrained RO, the overconservatism is handled through a polyhedral uncertainty set in which the level of conservatism is controlled by defining the *budget of uncertainty* (Γ). If the budget of uncertainty is zero, the decision-maker is not conservative at all. However, if the budget of uncertainty is equal to the number of constraints in which the uncertain parameter exists, the decision maker has the highest level of conservatism; the decision-maker can control the conservatism by defining a budget of uncertainty between zero and highest level of conservatism where the cardinality of the parameters permitted to change is constrained [4, 5]. This approach guarantees that the solution is feasible if less than Γ uncertain input parameters change. Furthermore, it provides a probabilistic guarantee that even if more than Γ change, then the robust solution will be feasible with high probability. Unlike other RO approaches, this approach has the advantages of providing computationally tractable models (as robust counter parts are linear optimization problems) as well as simplicity that is appealing to decision-makers; therefore, it has been successfully applied in many areas, including logistics and production systems [11], tactical planning in supply chains [3] and healthcare management problems [1].

Cardinality-constrained RO in healthcare planning has mainly been applied in inpatient settings. For example, Denton et al. [9] apply cardinality-constrained RO for allocating surgery blocks to operating rooms to deal with uncertainty in surgery duration. A surgery block is defined as one or more consecutive surgeries which are performed by a specific surgeon in the same operating room during an eight-hour period. The objective of their model is to allocate surgery blocks to operating rooms to minimize the worst possible over-time cost for all realizations of surgery block duration within the defined range. Their use of cardinality-constrained RO is attributed to the uncertainty due to the limitation in data availability for surgery durations as well as the capability of decision makers to provide reasonable estimates for the lower and upper bounds for surgery durations. Tang and Wang [20] develop an RO model for allocating operating room time to different sub-specialty of surgeries to deal with uncertainty in demand for surgeries, considering both elective and emergency cases. The developed RO model is based on implementor/adversary algorithm of Bienstock [7], which decomposes the original problem into a master problem (implementor) and a sub-problem (adversary). The sub-problem chooses a value for uncertain demand which deviates from its nominal value and controls the maximum number of uncertain demands based on the cardinality - constrained RO method. The master problem minimizes the revenue loss for the shortage of operating rooms as well as a penalty cost for idleness of operating rooms for the generated demand from the sub-problem. Geranmayeh [10] develops a cardinality-constrained RO model to allocate blocks of operating rooms to each surgeon to deal with uncertainty in the number of referrals to the inpatient ward. The number of referrals cannot be estimated via an average because doing so would force the surgeon to have same number of inpatients per surgery block. However, the decision for referral can be made only after the surgery is done. The choice of cardinality-constrained RO is due to the fact that the probability distribution for the number of referrals is not known, but the range can be estimated. The objective of the model is protecting the hospital against congestion in the inpatient ward due to the highest possible number of referrals to the inpatient unit.

To the best of our knowledge, the literature on healthcare planning models in outpatient settings that has handled uncertainty using RO is scarce. Pour [18] develops a cardinality-constrained RO model to schedule patients from a radiotherapy waiting list to deal with uncertainty in treatment duration and treatment time. The goal of their model is maximizing the number of scheduled patients from the waiting list for the worst possible scenario of treatment durations. Shalamzari Mirahmadi [19] applies cardinality-constrained RO method to develop an admission planning model for multi-priority independent patients for magnetic resonance imaging (MRI) exam in the presence of patient arrival uncertainty. The objective of their model is to minimize weighted patient waiting time for the highest possible demand for MRI exams. Furthermore, among the studies described above, [19] is the only one that focuses on uncertainty that arises in the right-hand side of a constraint. In our study, similar to [19], we consider demand uncertainty that arises in the right-hand side of two constraints. However, it is worthwhile noting that in our study, patient types are *dependent* and we explicitly consider demand uncertainty only in one type of appointment (first-visit patients) which implicitly impacts the demand for the other type of appointments (re-visit patients). Furthermore, we control the level of conservatism for both the total demand in the planning horizon as well as the demand in a single period.

3 Problem description

This study focuses on the outpatient clinic setting described by Nguyen et al. [16] and we adopt and, where necessary, extend their notation (see Section 4.1 and Appendix A). This clinic follows a re-entry appointment system based on which patient appointments are classified into the two general categories of first-visit (FV) and re-visit (RV) patients. The two patient types are dependent: FV patients may need to revisit the outpatient clinic for completing their care, at which point they become RV patients; otherwise they are discharged. The clinic requires a tactical capacity planning tool that will determine the minimum physician capacity required to ensure that access ttargets of both patient classes are met.

A key characteristic of the problem is uncertainty in patient demand. We assume that the first-visit patients are considered as external demand and the main source of uncertainty. In addition, patients that are required to do a subsequent visit are following the recommendations and do so to complete their treatments. Therefore, uncertainty in follow-up visits is also implicitly considered, since a proportion of FV patients who are not discharged become RV patients.

FV demand is assumed to arrive during the *arrival horizon*, denoted $\mathcal{S} = \{1, \ldots, S\}$ and measured in weeks. Appointments for both FV and RV appointments should be scheduled in the planning horizon, $\mathcal{T} = \{1, \ldots, T\}$, which is defined as an extension of the arrival horizon and should be long enough to cover the maximum access time allowed for the last arriving FV and RV requests from the arrival horizon.

The tactical capacity tool should operate in a rolling horizon fashion as shown in Fig. 1, with a plan of length T being constructed every S time periods. FV demand arriving in $\{1, \ldots, S\}$ can be scheduled anywhere in $\{1, \ldots, T\}$. FV demand that arrives after time S is considered in the planning problem of the subsequent period. FV patients arriving in $\{1, \ldots, S\}$ that are assigned an appointment in $\{S + 1, \ldots, T\}$ are considered fixed when a new tactical capacity plan is developed at time $S + 1$; similarly, the appointments of RV patients scheduled in $\{S + 1, \ldots, T\}$ remain fixed for the planning problem solved at $S+1$; these patients are called *pre-assigned*.

RV patients who cannot be scheduled in the current planning horizon will be postponed to the next planning horizon, which is modeled by scheduling them in period $T + 1$. Therefore, we define index set $\mathcal{T}' = \{1, \ldots, T + 1\}$. We denote the total number of postponed RV patients ("scheduled" at $T + 1$) as Ψ and assume that these become pre-assigned RV patients for the first week of the next planning horizon. Unlike in previous work [16], we treat Ψ as a parameter that controls the trade-off between the number of patients seen in the current planning horizon (for which the access targets are enforced) and the number of patients postponed until the next planning horizon.

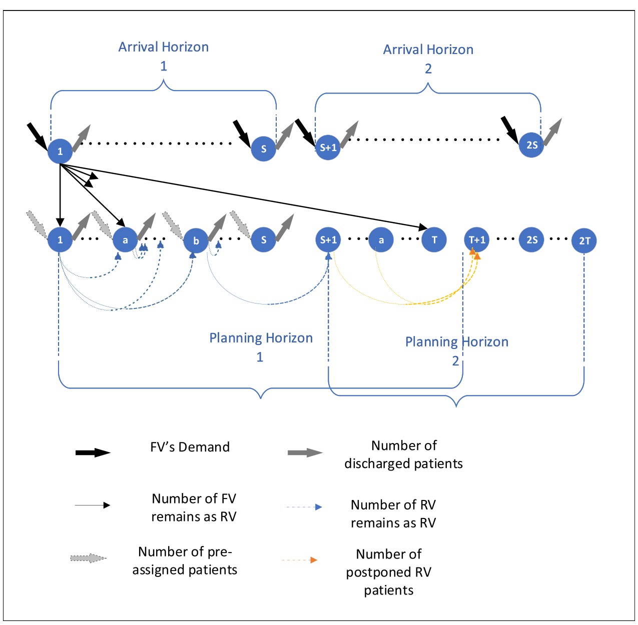

Fig. 1 Rolling horizon schematic for the tactical capacity planning tool, adapted from Figure 3 of the paper by Nguyen et al. [16]

4 Robust tactical capacity planning model

We develop a robust tactical capacity planning (RTCP) model to ensure care accessibility in the presence of uncertainty in demand.

Among the operations research methods, we choose robust optimization (RO) since it will allow us to identify the smallest required physician time that will be feasible regardless of what realization of demand uncertainty occurs in the uncertainty set (the set of possible uncertain parameter values).

While Nguyen et al. [16]'s deterministic tactical capacity planning (DTCP) model (see Appendix A) minimizes the maximum required physician time for a particular demand scenario, our proposed RTCP model minimizes the maximum required physician time for the worst case realization of demand in an uncertainty set. This is equivalent to minimizing the highest potential peak load for physicians under demand uncertainty.

4.1 Notation

Here, we define the notation required to develop the RTCP model. Additional notation related to Nguyen et al. [16]'s DTCP model is provided in Appendix A.

Decision Variables

- $z_{i,j}$ Number of FV patients who make a request in period i and have their appointment scheduled in period j.
- q Maximum required capacity for the worst case realization of demand per period (minutes).

Parameters

- f_i Number of FV appointment requests (demand) in time period i.

4.2 Uncertainty set

We assume that uncertainty in f_i happens in the arrival horizon without the knowledge of the specific periods of the arrival horizon that are affected. The demand in each period is independent across the planning horizon. One of the benefits of robust optimization is that one does not need the full distribution information. The knowledge of the mean and standard deviation is sufficient [6]. This assumption matches reality since we cannot know in advance in which week the uncertainty will be manifested. We assume that demand at each period is modeled as a non-negative, bounded random variable $\tilde{f}_i, i \in S$, which is independent of $\tilde{f}_j$, $j \in S$, $j \neq i$. To incorporate the uncertainty, we focus on the case when the number of FV patient requests in the arrival horizon at each period exceeds its nominal value, since our model does not have any penalties for idleness. Specifically, we assume that each uncertain parameter, $\tilde{f}_i$, takes values in the box uncertainty set of $[0, \bar{f}_i + \hat{f}_i]$, where $\bar{f}_i$ is known as the *nominal* value and $\hat{f}_i$ is the maximum deviation of the uncertain parameter from its nominal value. The uncertain parameter $\tilde{f}_i$ appears in constraints:

$$\sum_{z_{i,j} \in L^{100}} z_{i,j} = \sum_{i \in S} \tilde{f}_i, \tag{1}$$

$$\sum_{j=i}^{T} z_{i,j} = \tilde{f}_i \ \forall i \in S, \tag{2}$$

where Eq. 1 ensures that the 100th percentile access time target is met for all FV patient (uncertain) demand, and Eq. 2 ensures that FV patient demand at every period i of the arrival horizon is scheduled (at i or later).

4.3 Budget of uncertainty

We define a budget of uncertainty based on the assumption that uncertainty happens in the arrival horizon without the knowledge of the specific period(s) affected. Specifically, the budget of uncertainty Γ takes values in $[0, s]$, where s (number of periods in the arrival horizon) is equal to the maximum number of periods which can incorporate uncertainty over the arrival horizon. In other words, the budget of uncertainty (Γ) is the degree of freedom that the decision-maker can consider regarding number of periods in the arrival horizon s in which the number of FV patient requests (f_i) deviates from its nominal values ($\Gamma = s$ will be the worst case). The proposed robust formulation provides a solution capable of handling extreme demand realizations to the extent allowed by the budget of uncertainty.

4.4 Robust formulation of constraint (1)

Protection function aims to guarantee feasibility of the constraint with uncertain parameters through adding the following function:

$$\eta(\Gamma) = \max_{\{P \cup \{t\} | P \subseteq S, |P| = \lfloor \Gamma \rfloor, t \in S \setminus P\}} \{\sum_{i \in P} \hat{f}_i + (\Gamma - \lfloor \Gamma \rfloor) \hat{f}_t\}. \tag{3}$$

Let κ_i be the proportion of deviation of uncertain parameter $\tilde{f}_i$ from the the nominal value $\bar{f}_i$ towards the maximum weekly demand of $\bar{f}_i + \hat{f}_i$, with $0 \le \kappa_i \le 1, \forall i \in S$. Before deriving the robust equivalent, we rewrite constraint (1), as follows:

$$\sum_{z_{i,j} \in L^{100}} z_{i,j} = \sum_{i \in S} \bar{f}_i + \sum_{i \in S} \kappa_i \hat{f}_i \tag{4}$$

Below, the linear equivalent for the non-linear protection function η is provided. The objective function of the following problem is maximizing the proportion of demand uncertainty in each period of arrival horizon:

$$\max \ \sum_{i \in S} (\kappa_i \hat{f}_i) \tag{5}$$

$$\text{s.t.} \ \sum_{i \in S} \kappa_i \le \Gamma \tag{6}$$

$$0 \le \kappa_i \le 1 \ \ \forall i \in S. \tag{7}$$

Due to the fact that the presented equivalent model is feasible and bounded, by strong duality theorem, its dual model is also feasible and bounded. The dual model of the above linear program is derived by introducing the dual variables λ (constraint (6)) and μ_i (constraint (7)):

$$\min \ \lambda \Gamma + \sum_{i \in S} \mu_i \tag{8}$$

$$\text{s.t.} \ \lambda + \mu_i \ge \hat{f}_i \qquad \forall i \in S \tag{9}$$

$$\lambda \ge 0, \mu_i \ge 0 \quad \forall i \in S. \tag{10}$$

Therefore, the linear robust formulation of constraint (1) is as follows:

$$\sum_{z_{i,j} \in L^{100}} z_{i,j} = (\sum_{i \in S} \bar{f}_i) + \lambda \Gamma + \sum_{i \in S} \mu_i, \tag{11}$$

$$\lambda + \mu_i \ge \hat{f}_i \ \forall i \in S, \tag{12}$$

$$\lambda \ge 0, \mu_i \ge 0, \forall i \in S. \tag{13}$$

4.5 Robust reformulation of constraint (2)

Constraint (2) will be protected against uncertainty if we can obtain the value for κ_i (proportion of deviation of $\tilde{f}_i$ from its nominal value in each period of arrival horizon). This can be realized by including dual feasibility and strong duality conditions as follows:

$$\sum_{j=i}^{T} z_{i,j} = \bar{f}_i + \kappa_i \hat{f}_i \ \forall i \in S, \tag{14}$$

$$\sum_{i \in S} \kappa_i \hat{f}_i = \lambda \Gamma + \sum_{i \in S} \mu_i, \tag{15}$$

$$\sum_{i \in S} \kappa_i \le \Gamma, \tag{16}$$

$$0 \le \kappa_i \le 1 \ \forall i \in S, \tag{17}$$

$$\lambda + \mu_i \ge \hat{f}_i \ \forall i \in S, \tag{18}$$

$$\lambda \geq 0, \mu_i \geq 0, \forall i \in S. \tag{19}$$

4.6 Complete RTCP model

The complete RTCP model based on considering a fixed number of postponed RV patients is given below:

$$\min\ q, \tag{20}$$

$$\text{s.t.} \quad (29)$$

$$\sum_{j=i}^{T} z_{i,j} = \bar{f}_i + \kappa_i \hat{f}_i, \quad \forall i \in S, \tag{21}$$

$$(32) - (42),$$

$$\sum_{z_{i,j} \in L^{100}} z_{i,j} = (\sum_{i \in S} \bar{f}_i) + \lambda\Gamma + \sum_{i \in S} \mu_i, \tag{22}$$

$$\lambda + \mu_i \geq \hat{f}_i, \ \forall i \in S, \tag{23}$$

$$\sum_{i \in S} \kappa_i \hat{f}_i = \lambda\Gamma + \sum_{i \in S} \mu_i, \tag{24}$$

$$\sum_{i \in S} \kappa_i \leq \Gamma, \tag{25}$$

$$0 \leq \kappa_i \leq 1, \ \forall i \in S, \tag{26}$$

$$\lambda \geq 0, \quad \mu_i \geq 0, \forall i \in S, \tag{27}$$

$$(39) - (50).$$

5 Experimental results

The RO approach of Bertsimas and Sim [4] hedges against uncertainty such that the proposed solution is feasible for a given budget of uncertainty. The issue is that setting this budget of uncertainty has to be done prior to the realization of uncertainty. Thus, we need to answer the following two questions: 1) How to ensure the feasibility of RTCP under all demand realizations? 2) How to find a balance between the level of protection of robust TCP and additional cost of robust TCP? In order to answer these questions, we will provide some numerical results for the implementation of both the deterministic and the robust model in IBM ILOG CPLEX Optimization Studio V12.6.2.0 on Lenovo ThinkPad X260 Corei7 2.60 GHz.

5.1 Translation of Demand Data into Uncertainty Sets

In order to answer the two questions, we set up experiments in which $\bar{f}_i$ is based on the data from [15], which includes the weekly FV demand over a 52-week arrival horizon which should be scheduled in 82 weeks of the planning horizon. The data set includes available physician time in minutes in each week of the planning horizon. For our experiments, we take the values for the pre-assigned FV and RV patients as well as the nominal demand values from Appendix A2.2 of [15]; these nominal demand values are the actual realizations of demand from year 2009 and 2010 in an outpatient (urology) clinic from Tan Tock Seng Hospital in Singapore. The values for the other input parameters used in the experiments are listed in Table 3 in the appendix. In order to incorporate uncertainty, we should define an appropriate interval as an uncertainty set for FV demand in each week of the planning horizon. Usually the minimum and the maximum observed data are used as the lower and upper bounds of an uncertainty set [14]. Since the uncertain parameter is in the right-hand side of the constraint, we consider only the maximum observed demand to find the maximum deviation from nominal demand value.

We calculate $\bar{f}_i$ and $\hat{f}_i$ in three ways, based on yearly, seasonal, and monthly maximum values:

- **Yearly:** We use the maximum and minimum realized FV requests over the arrival horizon (52 weeks):
  - Let $f_{max}^{(52)} = \max_{i \in \{1,\ldots,52\}} f_i$ and $f_{min}^{(52)} = \min_{i \in \{1,\ldots,52\}} f_i$.
  - To calculate the nominal value, we set $\bar{f}_i^{(52)} = \frac{f_{max}^{(52)} + f_{min}^{(52)}}{2}$ for all $i = 1, \ldots, 52$.
  - To calculate the maximum deviation from the nominal, we find $\hat{f}_i^{(52)} = f_{max}^{(52)} - \bar{f}_i$ for all $i = 1, \ldots, 52$.
- **Seasonal:** We use the maximum and minimum realized FV requests over each 13-week period:
  - Let $h_{max,j}^{(13)} = \max_{i \in \{1+13j,\ldots,13+13j\}} f_i$ and $h_{min,j}^{(13)} = \min_{i \in \{1+13j,\ldots,13+13j\}} f_i$ for each season $j = 0, \ldots, 3$.
  - To calculate the nominal value, we find $\bar{h}_j^{(13)} = \frac{h_{max,j}^{(13)} + h_{min,j}^{(13)}}{2}$ for $j = 0, \ldots, 3$.
  - To calculate the maximum deviation from the nominal, we find $\hat{h}_j^{(13)} = h_{max,j}^{(13)} - \bar{h}_j^{(13)}$ for all $j = 0, \ldots, 3$.
  - We set $\bar{f}_i^{(13)} = \bar{h}_j^{(13)}$ and $\hat{f}_i^{(13)} = \hat{h}_j^{(13)}$ whenever $i \in \{1 + 13j, \ldots, 13 + 13j\}$ for $j = 0, \ldots, 3$.
- **Monthly:** We use the maximum and minimum realized FV requests over each 4-week period:
  - Let $h_{max,j}^{(4)} = \max_{i \in \{1+4j,\ldots,4+4j\}} f_i$ for $j = 0, \ldots, 12$ and $h_{min,j}^{(4)} = \min_{i \in \{1+4j,\ldots,4+4j\}} f_i$ for $j = 0, \ldots, 12$.
  - To calculate the nominal value, we find $\bar{h}_j^{(4)} = \frac{h_{max,j}^{(4)} + h_{min,j}^{(4)}}{2}$ for $j = 0, \ldots, 12$.
  - To calculate the maximum deviation from the nominal, we find $\hat{h}_j^{(4)} = h_{max,j}^{(4)} - \bar{h}_j^{(4)}$ for all $j = 0, \ldots, 12$.
  - We set $\bar{f}_i^{(4)} = \bar{h}_j^{(4)}$ and $\hat{f}_i^{(4)} = \hat{h}_j^{(4)}$ whenever $i \in \{1 + 4j, \ldots, 4 + 4j\}$ for $j = 0, \ldots, 12$.

Table 1 Values of annual, seasonal and monthly based uncertainty sets

Weeks			Uncertainty set		
Year	Season	4weeks	Bounds/year	Bounds/season	Bounds/4 weeks
w1-w52	w1-w13	w1-w4	[181 ± 67]	[162 ± 47]	[157 ± 43]
		w5-w8			[186 ± 23]
		w9-w12			[186 ± 22]
	w14-w26	w13-w16			[171 ± 17]
		w17-w20		[189 ± 38]	[185 ± 34]
		w21-w24			[194 ± 13]
	w27-w39	w25-w28			[180 ± 47]
		w29-w32		[190 ± 58]	[181 ± 34]
		w33-w36			[218 ± 30]
	w40-w52	w37-w40			[198 ± 47]
		w41-w44		[200 ± 45]	[213 ± 32]
		w45-w48			[189 ± 17]
		w49-w52			[186 ± 31]

Table 1 presents the bound values for the three types of uncertainty sets defined above. We employ the uncertainty set bounds to randomly generate demand scenarios. For example, to generate one instance of a problem based on uncertainty set $[\bar{f}_i^{(52)}, \bar{f}_i^{(52)} + \hat{f}_i^{(52)}]$, we draw a random sample from Uniform$[\bar{f}_i^{(52)}, \bar{f}_i^{(52)} + \hat{f}_i^{(52)}]$ for each i. We note that the above three methods of calculating uncertainty set bounds induce different levels of variability, as the interval $[\bar{f}_i^{(52)}, \bar{f}_i^{(52)} + \hat{f}_i^{(52)}]$ is wider than $[\bar{f}_i^{(13)}, \bar{f}_i^{(13)} + \hat{f}_i^{(13)}]$, which is in turn wider than $[\bar{f}_i^{(4)}, \bar{f}_i^{(4)} + \hat{f}_i^{(4)}]$, for each $i = 1, \ldots, 52$.

All other parameter values are set the same way as in Nguyen et al.'s [16]'s work.

5.2 Results

We provide the results of the experiment for the RTCP model given in Section 4.6 in two stages. In the first stage, we run the DTCP model with the demand scenario used by [16] to find the value for Ψ (number of RV patients postponed to $T + 1$). We find this number to be 12929 patients; we set $\Psi = 12929$ for all subsequent experiments. In practice, this number should be determined in consultation with the clinic; an alternative approach is discussed in Section 6.

Following the cardinality-constrained RO literature, we evaluate the probability of infeasibility of the robust optimal solution to control over-conservatism and the price of robustness, defined as the trade-off between cost and feasibility of the robust optimal solution. We also compare the robust solution with the worst-case deterministic solution. We subsequently analyze the results from our robust model to identify the most critical time periods in the planning horizon. All experiments are performed under the three uncertainty set definitions provided in Section 5.1.

Trade-off Between Infeasibility Probability of Robust Optimal Solution and Level of Conservatism: To calculate the empirical infeasibility probability of RO solution, we evaluate the feasibility of robust optimal solution for different level of robustness (conservatism) in the presence of randomly generated demand. To do so, we apply Monte-Carlo simulation. Therefore, for all the values of budget of uncertainty from [0, 52], we randomly generate 200 scenarios for uncertain demand from the uniform distributions of the three defined uncertainty sets $[\bar{f}_i^{(52)}, \bar{f}_i^{(52)} + \hat{f}_i^{(52)}]$, $[\bar{f}_i^{(13)}, \bar{f}_i^{(13)} + \hat{f}_i^{(13)}]$, and $[\bar{f}_i^{(4)}, \bar{f}_i^{(4)} + \hat{f}_i^{(4)}]$, for each i. The generated scenarios are the simulated values of demand. The proposed RTCP model becomes infeasible if the available physician time is not sufficient to meet generated random demand. In other words, for each demand scenario the optimal solution corresponding to physician capacity from the RTCP model is inserted as a parameter into the DTCP model where f_i is replaced by generated demand scenarios. Thereafter, we calculate the empirical probability of RO solution infeasibility by dividing the number of infeasible instances by 200 (the total number of simulated demand scenarios for each uncertainty set type). Therefore, we solve the deterministic model $200 \times 52 \times 3 = 31200$ times to find the trade-off between infeasibility probability and level of conservatism (robustness).

Figure 2 represents the impact of the budget of uncertainty on the magnitude of infeasibility for the robust solution for the 200 generated demand scenarios for the three types of uncertainty sets. Figure 2 shows that the budget of uncertainty required to ensure feasibility is

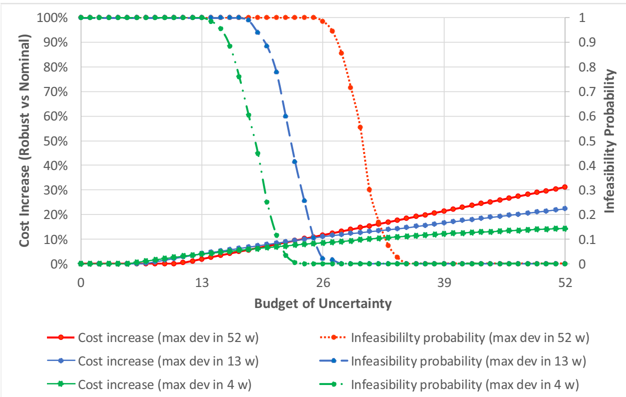

Fig. 2 Cost of robustness and infeasibility probability for RO solution

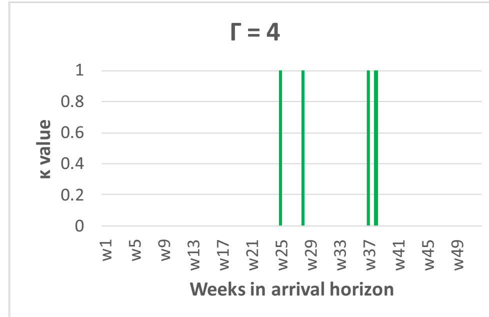

Fig. 3 κ Values for 4-Week-Based Uncertainty Set & $\Gamma = 4$

35, 28, and 24 for the 52-week, 13-week and 4-week uncertainty sets. In addition to the budget of uncertainty, another element that could control over-conservatism in practice is the subset of data (in our case this corresponds to some period of time) used to determine the bounds of the uncertainty set. We considered three periods over which the bounds were estimated from the data, i.e., 52 weeks, 13 weeks and 4 weeks. Since $\hat{f_i}^{(52)} \geq \hat{f_i}^{(13)}$ and $\hat{f_i}^{(52)} \geq \hat{f_i}^{(4)}$ for every i, we see that a higher budget of uncertainty is required to ensure feasibility in the case when the uncertainty set is estimated based on all 52 weeks. If the variation of demand is, for instance, highly seasonal (e.g., substantially lower demand values in the summer as opposed to winter), then the results based on $\hat{f_i}^{(52)}$ will be unnecessarily conservative.

Price of Robustness: The price of robustness presents the trade-off between the additional cost and the feasibility of the robust solution in terms of different budgets of uncertainty. The extra cost of robust solution is calculated as $\frac{q^R - q^N}{q^N}$, where q^R is the objective value of robust optimal solution and q^N is the objective value for nominal optimal solution. Figure 2 presents the price of robustness for different budgets of uncertainty.

We obtain from Fig. 2 the penalty of being fully confident about feasibility of robust solution. It can be seen that when demand variability is based on the uncertainty set from narrower interval (i.e., uncertainty sets based on monthly deviation), the penalty imposed by the robust model is at most 8.05%. However, when demand variability is based on the uncertainty set from wider interval, the system should pay a higher penalty to assure that all patient demand can be scheduled, i.e., 11.91% based on seasonal variation and 18.42% based on yearly variation. Furthermore, we can again interpret the above results as follows: if the variation of demand shows highly seasonal pattern (e.g., substantially lower demand values in the summer as opposed to winter) whereas the uncertainty set is estimated based on 52-week data (i.e., using $\bar{f_i}^{(52)}$ and $\hat{f_i}^{(52)}$), then an unnecessary cost increase of 6.51% is incurred.

Robust Solution vs. Worst-Case Solution: We compare the objective value of the robust problem (q^{RO}) for the budget of uncertainty for which the infeasibility probability becomes 0 with that of the worst case deterministic model (q^{wc}). Table 2 shows that q^{wc} - $q^{RO}_{avg} \geq 0$ for all the three types of uncertainty sets, implying that the robust solution protects fully against uncertainty at a lower cost than the worst-case solution.

Identification of the Critical Time Periods that Contribute to the Worst-Case Physician Peak Load: The objective of this analysis is finding the periods where realization of demand uncertainty will lead to the worst possible maximum physician peak load. The solution generated from the RO model provides a plan that performs well even when the realized demand $\tilde{f_i}$ is greater than its nominal value $\bar{f_i}$.

Table 2 Robust TCP vs. Worst-case TCP performance for various uncertainty sets

	Uncertainty set interval		
Objective value	4w	13w	52w
q^{wc}	130.262	139.290	149.330
q^{RO}	122.996	127.397	134.800

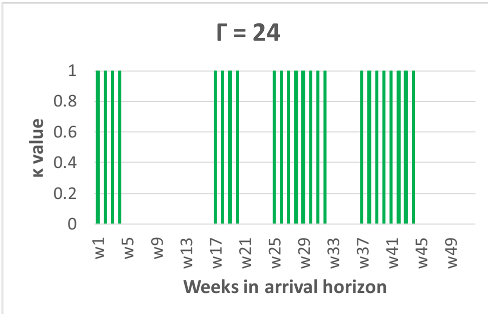

Fig. 4 κ Values for 4-Week-Based Uncertainty Set & $\Gamma = 24$

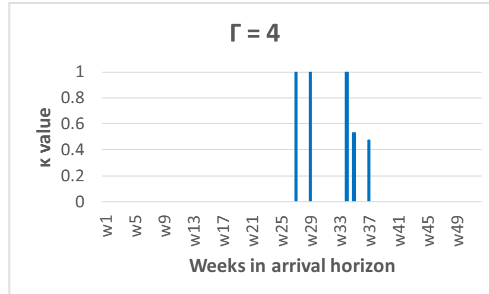

Fig. 6 κ values for yearly uncertainty set & $\Gamma = 4$

Having the knowledge of such critical time periods can help in tactical capacity planning decisions.

In particular, we can identify the critical weeks in the planning horizon that lead to the worst-case physician peak load by finding the value of κ_i. As defined in Section 4.6, κ_i is the proportion of deviation of $\tilde{f}_i$ from its nominal value $\bar{f}_i$ in each period of arrival horizon. Therefore, the periods with non-zero κ_i values are the critical ones.

Figures 3, 4 and 5 show the critical time periods for the 4-week-based uncertainty set. When the budget of uncertainty Γ is equal to 4, the selected critical weeks by the RTCP model are weeks 25, 28, 37 and 38. We think that this behaviour is due to the fact that the maximum $\hat{f}_i^{(4)}$ for all $i = 1, \ldots, 52$ (see Table 1) is 48, which is realized over weeks 25–28 and 37–40. For $\Gamma = 24$, six of the 4-week intervals with highest $\hat{f}_i^{(4)}$ are selected, which are weeks 1–4, 17–21, 25–28, 29–32, 37–40 and 41–44. For $\Gamma = 35$, the additional selected critical weeks compared to $\Gamma = 24$ are weeks 6–8, 33–36 and 49–52.

Figures 6, 7 and 8 show the critical time periods for the 13-week-based uncertainty set. When $\Gamma = 4$, the critical weeks selected by the RTCP model are 27, 29, 34, 35 and 37. The values of κ for weeks 35 and 37 are fractional. The reason for choosing these weeks as critical is that the maximum $\hat{f}_i^{(13)}$ for all $i = 1, \ldots, 52$ (see Table 1) is 58, which is realized over one season in weeks 27–39. For $\Gamma = 24$, the critical weeks are chosen from the two seasons with highest $\hat{f}_i^{(13)}$. Since 13 is not a factor of 24, κ values for some weeks are fractional. For $\Gamma = 35$, the critical weeks are from all the weeks in the two seasons with highest $\hat{f}_i^{(13)}$ and nine weeks of the season with third highest $\hat{f}_i^{(13)}$. As observed in Fig. 8, all the κ values are integer, although 13 is not a factor of 35. We conjecture that this behaviour is due to the fact that for seasonal-based uncertainty set, infeasibility probability of the robust solution becomes zero for $\Gamma = 28 < 35$, and that therefore the model does not need to be as "careful" in choosing the critical periods once the probability of infeasibility becomes 0; similarly, we suspect there are multiple equivalent solutions with different κ values after the budget of uncertainty is high enough to ensure no infeasibility. Figures 9, 10 and 11 depict the critical time periods for 52-week-based uncertainty set. Due to the fact $\hat{f}_i^{(52)}$ for all $i = 1, \ldots, 52$ is the same, the selected critical weeks are not based on the week with highest possible demand variation. As seen in Fig. 9, for $\Gamma = 4$ the weeks toward the end of arrival horizon are chosen due to the fewer remaining periods to allocate available physician time to demand as well as limited physician time availability in each period. Thereafter, for $\Gamma = 24$, the weeks from the beginning of arrival horizon are also chosen, in addition to the weeks at the end of arrival horizon.

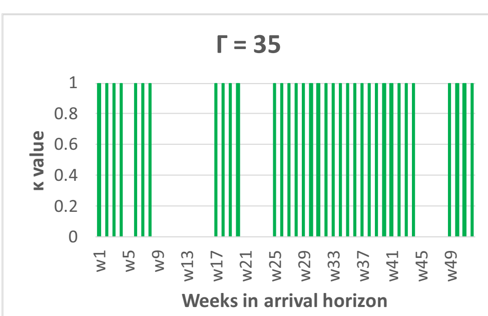

Fig. 5 κ Values for 4-Week-Based Uncertainty Set & $\Gamma = 35$

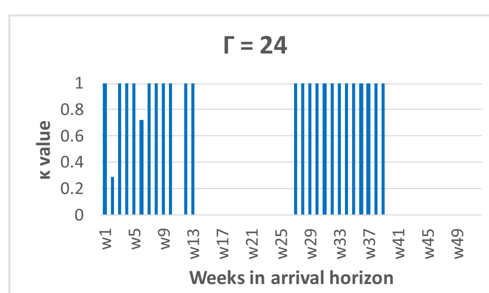

Fig. 7 κ values for yearly uncertainty set & $\Gamma = 24$

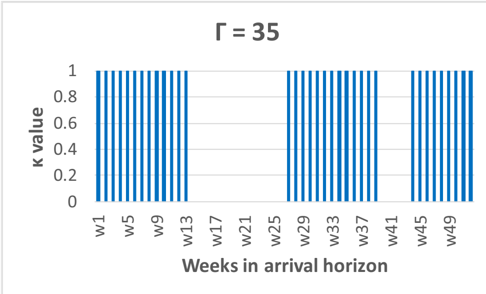

Fig. 8 κ values for yearly uncertainty set & $\Gamma = 35$

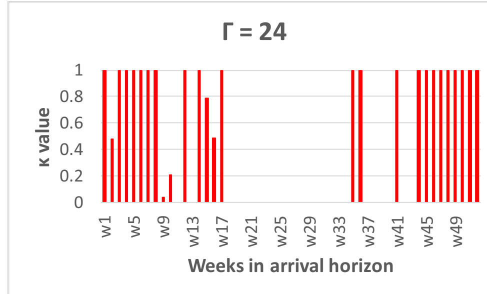

Fig. 10 κ values for yearly uncertainty set & $\Gamma = 24$

As observed from Figures 3–11, identifying the critical weeks that lead to worst-case physician workload is non-trivial and would be very challenging without solving a robust optimization model.

Figures 12, 13 and 14 depict the frequency of κ values over the weeks in the arrival horizon for the three defined uncertainty sets. The proposed robust tactical capacity planning model can help decision makers identify critical periods in a given planning horizon. As such, for a given demand uncertainty profile and budget of uncertainty, the proposed robust formulation highlights the periods where worst case realizations lead to capacity planning decisions. This information can be useful at the tactical planning level where decision makers can identify alternative measures to accommodate surges in demand. Examples of these measures include allocating overtime or having on-call physicians.

6 Discussion

Here, we further discuss three key observations from our study, namely the need to explicitly model the existence of a subsequent planning horizon even in a static model, the importance of considering the effect of uncertainty in particular time periods, and the necessity for careful translation of data into uncertainty sets.

Consideration of the Next Planning Horizon: First, in developing a static model for TCP, we need to explicitly consider the number of patients that may have to be postponed to the next planning horizon, since in reality the static model will be used in a dynamic setting. Above, we have shown that if a guideline for the maximum number of patients that could be postponed to the next planning horizon is either known or can be estimated from past data, then we recommend the use of our RTCP model.

Timing of Worst-Case Realizations: Second, we have shown, as expected, that if uncertainty in demand is not considered, then the calculated physician capacity is an underestimate of what is actually needed to achieve the required access time targets. We note that in our problem, the consideration of the worst-case realization of demand is done with respect to time; that is, our budget of uncertainty controls how many periods in the arrival horizon are subject to uncertainty, and the solution is based on identifying those periods in which achieving the highest potential demand would have the most impact on the maximum capacity required. Our results show that identifying such time periods without a robust model would have been difficult, as the determination of the timing of the worst-case realizations is dependent on the definition of the uncertainty sets and the budget of uncertainty, and since the allocation of appointments is a complex process that requires consideration of access time

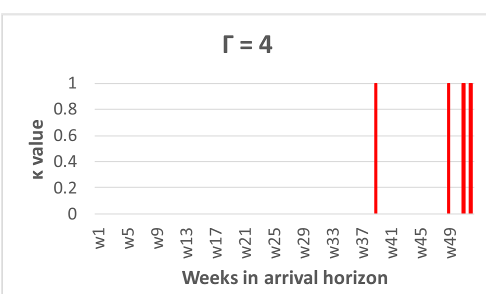

Fig. 9 κ values for yearly uncertainty set & $\Gamma = 4$

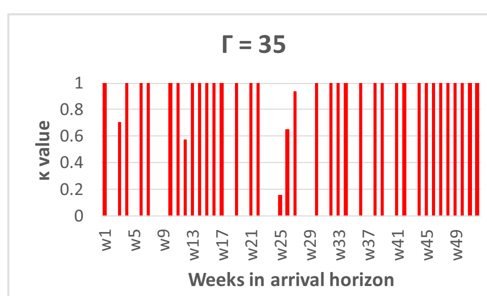

Fig. 11 κ values for yearly uncertainty set & $\Gamma = 35$

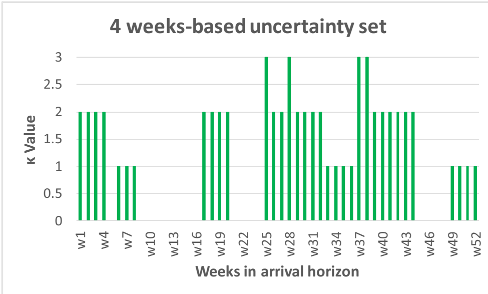

Fig. 12 Frequency of κ for monthly uncertainty set & $\Gamma = 4, 24, 35$

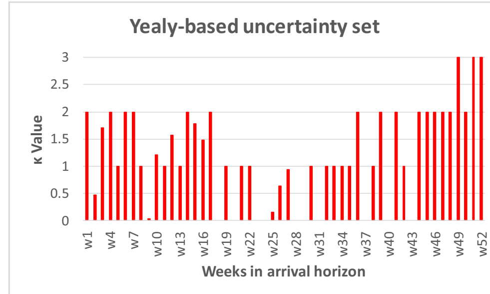

Fig. 14 Frequency of κ for yearly uncertainty set & $\Gamma = 4, 24, 35$

targets. For instance, the worst-case realizations for $\Gamma = 4$ with yearly-based uncertainty sets occurs at the end of the arrival horizon (see Fig. 9), since there is a small remaining number of periods in that planning horizon (and there is a constraint on how many patients can be allocated to the subsequent one); yet at other times, such as for $\Gamma = 24$ with yearly-based uncertainty sets (see Fig. 10), the worst-case realizations can in addition occur in the beginning of the arrival horizon since such realizations have an effect on the entire planning horizon.

Translation of Data into Uncertainty Sets: The third observation is that the step of translating available data into an uncertainty set is important and should be carefully considered as it can lead to substantially different results and costs. In fact, the translation step can influence the true conservatism of a solution – an over-estimate or under-estimate in the uncertainty set definition will lead to an over-estimate or an under-estimate, respectively, of the resulting capacity and cost. For example, if the variation of demand is in reality highly seasonal (e.g., substantially lower demand values in the summer as opposed to winter) but the uncertainty set is estimated based on 52 weeks of data, then in some of our experiments an unnecessary cost of about 6.51% is incurred. Therefore, it is important to understand the given

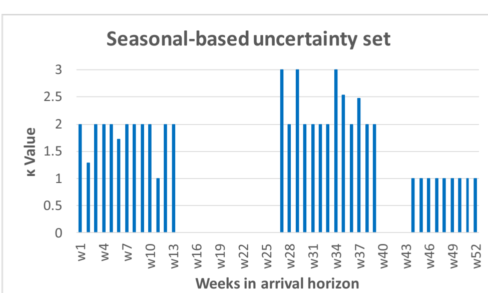

Fig. 13 Frequency of κ for seasonal uncertainty set & $\Gamma = 4, 24, 35$

data prior to the development of the uncertainty set as well as to validate the definitions of the uncertainty set with the stakeholders prior to, as well as following, the optimization process.

7 Conclusion

In this study, we developed a robust tactical capacity planning model via cardinality-constrained robust optimization which explicitly considers the number of patients who may need to be scheduled in a subsequent planning horizon. The robust tactical capacity planning model protects against uncertainty in the demand per time period. Due to the presence of the uncertain demand parameter in the right-hand side of two constraints, i.e., individual demand per period and aggregated demand over all the periods in the arrival horizon, formulating the robust tactical planning model required the use of both primal and dual constraints in the robust model. By employing cardinality-constrained robust optimization, we showed how to control over-conservatism through analyzing the trade-off between the budget of uncertainty and the feasibility of the robust plan based on different bounds of the uncertainty set. We also analyzed the price of robustness, i.e., the trade-off between feasibility and cost of a robust plan for different budgets of uncertainty and bounds of uncertainty set.

We showed that we can guarantee 100% feasibility of a robust tactical capacity plan while not being fully conservative, which will lead to cost savings for the clinic while being able to meet demand despite uncertainty. We also show how the robust model helps us to identify the critical time periods which contribute to worst-case physician peak load. The proposed robust tactical capacity planning model can help decision makers identify critical periods in a given planning horizon. As such, for a given demand uncertainty profile and budget of uncertainty, the proposed robust formulation highlights the periods where worst case realizations lead to capacity planning decisions. This information can be useful at the tactical planning level where decision-makers can identify alternative measures to accommodate surge in demand.

Future research includes extending the proposed RTCP model for the case when a guideline for the maximum number of postponed patients is not available or cannot be easily estimated from the data. In this case, we recommend considering a multi-objective RTCP model which can be calibrated via changing objective function weights to meet the clinic's goals. In particular, one can modify the objective function of the robust model to $\min h_1 q + h_2 \tau^r \sum_{i \in \mathcal{T}} y_{i,T+1}$, while removing the upper bound constraint (50). The two terms h_1 and h_2 would represent the penalty cost associated with each unit of extra required physician time and each extra RV patient postponed to period $T+1$, respectively. Thus, in this model Ψ is a variable and one can evaluate the trade-off between the maximum capacity for the current time period and Ψ. An additional direction for future work is the application of the proposed approach in outpatient settings with more than two appointment types.

Appendix A: Deterministic Tactical Capacity Planning

Nguyen et al. [16] presented a mixed-integer programming model for the deterministic version of the problem. The goal of their deterministic tactical capacity planning (DTCP) model is minimizing the maximum required physician time between weeks subject to ensuring weekly demand for appointments is satisfied and access time targets are met. Maximum required physician time can also be seen as physicians' peak load. The mathematical formulation of the DTCP model is presented in Eqs. (28)–(49). The model is based on a network flow representation with two sets of nodes, one for arrival periods of FV patient requests and the other for the scheduled periods for RV patient requests (see Fig. 1). There are four general sets of constraints: conservation of flow between FV nodes, conservation of flow between RV nodes, access time targets and finally required capacity for each patient type.

Decision Variables

- $z_{i,j}$ Number of FV patients who make a request in the i^{th} period and have their appointment scheduled in the j^{th} period.
- $x_{i,j}$ Number of FV patients who make a request in the i^{th} period and have appointment in the j^{th} period, and still remain in the system as RV patients after their appointment in the j^{th} period.
- $y_{i,j}$ Number of RV patients who have an appointment in the i^{th} period and have their next appointment in the j^{th} period, and still remain as RV patients after the j^{th} period.
- d_j^r The number of RV patients who are discharged after their appointment in period j.
- C_i^f Capacity in period j for FV patients (minutes).
- C_i^r Capacity in period j for RV patients (minutes).
- C_i Capacity in period j for both FV and RV patient types (minutes).
- q The maximum required capacity per period (minutes).

Parameters

- f_i Number of FV appointment requests (demand) in time period i.
- α, β Discharge rates for FV and RV patients, respectively ($0 < \alpha, \beta < 1$).
- r_j^f, r_j^r Number of pre-scheduled FV and RV patients with appointments in period j who still remain as RV patients after their appointments.
- u_m Appointment lead-time targets (number of time periods) for median of FV appointment requests (f_i).
- u_p Appointment lead-time targets (number of time periods) for p^{th} percentile ($0 < p < 1$) of FV appointment requests (f_i).
- u_{100} Appointment lead-time targets (number of time periods) for 100^{th} percentile of FV appointment requests (f_i).
- [a,b] Range of RV appointment access time target (number of time periods).
- $\bar{a}$ Mean RV appointment access time target (number of time periods).
- τ^f, τ^r Consultation times for FV and RV patients (minutes). These times are specified by the doctors.

Sets

- Z Set of all $z_{i,j} : j - i \geq 0$.
- L^m Set of all $z_{i,j} \in Z$ that have $j - i \leq u_m$.
- L^p Set of all $z_{i,j} \in Z$ that have $j - i \leq u_p$.
- L^{100} Set of all $z_{i,j} \in Z$ that have $j - i \leq u_{100}$.

The original DTCP formulation [16] is presented below:

$$\min \quad q \tag{28}$$

$$\text{s.t.} \quad q \geq C_j \; \forall j \in \mathcal{T} \tag{29}$$

$$\sum_{j=i}^{T} z_{i,j} = f_i \; \forall i \in \mathcal{S} \tag{30}$$

$$\sum_{j=i}^{T} z_{i,j} = 0 \; \forall i \in \mathcal{T} \backslash \mathcal{S} \tag{31}$$

$$x_{i,j} - (1-\alpha) z_{i,j} = 0 \; \forall i \in \mathcal{S}, \forall j \in \mathcal{T} \tag{32}$$

$$(r_j^f + r_j^r + \sum_{i=1}^{j} x_{i,j} + \sum_{i=1}^{j} y_{i,j}) - (d_j^r + \sum_{i=j}^{T} y_{j,i}) = 0 \quad \forall j \in \mathcal{S} \tag{33}$$

$$(r_j^f + r_j^r + \sum_{i=1}^{j} x_{i,j} + \sum_{i=1}^{j} y_{i,j}) - (d_j^r + \sum_{i=j}^{T+1} y_{j,i}) = 0 \quad \forall j \in \mathcal{T} \backslash \mathcal{S} \tag{34}$$

$$d_j^r - \beta(r_j^f + r_j^r + \sum_{i=1}^{j} y_{i,j} + \sum_{i=1}^{j} x_{i,j}) = 0 \; \forall j \in \mathcal{S} \tag{35}$$

$$y_{i,j} = 0 \forall j - i < a \; \forall i \in \mathcal{T}, \forall j \in \mathcal{T}^{'} \tag{36}$$

$$y_{i,j} = 0 \forall j - i > b \; \forall i \in \mathcal{T}, \forall, j \in \mathcal{T}^{'} \tag{37}$$

$$y_{i,j} = 0 \forall j \geq T + 1 \; \forall i \in \mathcal{S} \tag{38}$$

$$\sum_{j=1}^{T} \sum_{i=1}^{j} (j-i) y_{i,j} - \bar{a} \sum_{j=1}^{T} \sum_{i=1}^{j} y_{i,j} \leq 0 \tag{39}$$

$$\sum_{i=S+1}^{T} d_i^r = 0 \tag{40}$$

$$\sum_{z_{i,j} \in L^m} z_{i,j} \geq (1/2 \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} z_{i,j}) + 1 \tag{41}$$

$$\sum_{z_{i,j} \in L^p} z_{i,j} \geq p \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{T}} z_{i,j} \tag{42}$$

$$\sum_{z_{i,j} \in L^{100}} z_{i,j} = \sum_{i \in \mathcal{S}} f_i \tag{43}$$

$$z_{i,j} = 0 \;\; \forall i, j \in \mathcal{T} : j - i \geq u_{100} + 1 \tag{44}$$

$$z_{i,j} = 0 \;\; \forall i, j \in \mathcal{T} : j - i < 0 \tag{45}$$

$$C_j^f - (\tau^f r_j^f + \tau^f \sum_{i=1}^{j} z_{i,j}) = 0 \; \forall j \in \mathcal{T} \tag{46}$$

$$C_j^r - (\tau^r r_j^r + \tau^r \sum_{i=1}^{j} y_{i,j}) = 0 \; \forall j \in \mathcal{T} \tag{47}$$

$$C_j - (C_j^f + C_j^r) = 0 \; \forall j \in \mathcal{T} \tag{48}$$

$$z_{i,j}, x_{i,j}, y_{i,j^{'}}, C_j^f, C_j^r, d_i^f, d_j^r \geq 0 \; \forall i, j \in \mathcal{T}, \forall j^{'} \in \mathcal{T}^{'}. \tag{49}$$

The DTCP model minimizes the maximum required capacity (physician time). Constraint (29) defines a decision variable (q) for the maximum required capacity which should be greater than the total required physician time for both patient types. Conservation of flow at FV nodes is modeled via constraints (30)–(32). Constraint (30) assures FV patient demand over each period of arrival horizon is fully covered. Constraint Eq. 32 determines the proportion of scheduled FV patients who become RV patients.

Conservation of flow at RV nodes is modeled via constraints (33)–(35). Constraint (33) and (34) represent the balance between inflow and outflow for each RV node. The inflow into an RV node includes the total number of pre-scheduled RV and FV patients as well as total number of RV patients with an appointment at the RV node. The outflow from an RV node includes the total number of RV patients who will be discharged plus the number of RV patients who will revisit again after their appointment at the RV node. Constraint (35) shows the number of scheduled RV patients who will be discharged after their appointment at the RV node.

Control of FV and RV access times is achieved through (36)–(39). Constraints (36) and (37) ensure that the access time of the scheduled RV patients belongs to $[a, b]$, while (39) sets the RV mean access time target. Constraint (38) forces RV patients' appointments to be scheduled before the last date of planning horizon. Constraint (40) prohibits the discharge of RV patients if their appointment is made after the arrival horizon.

The access time targets for FV patients are defined through targets for median, p^{th} percentile and 100^{th} percentile of FV patient requests in Eqs. (41), (42) and (43), respectively. The capacity required for each patient type and total required capacity are defined in Eqs. (46), (47) and (48).

Appendix B: Modifications of Nguyen et al.'s Model

In this study, we modify the DTCP model [16] to control the number of scheduled RV patients in each planning horizon.

Doing so prevents congestion of RV patients assigned to time period $T + 1$, which we denote by Ψ. The reason for this congestion is the increase in Ψ as the level of uncertainty increases. We control Ψ to serve a strategically agreed number of patients at each planning horizon which is one of the main objectives of TCP [13]. In the original DTCP model, constraint (30) ensures that all FV patients are served.

Since RV patients are a consequence of FV patients, there is currently no constraint regarding the strategic number of RV patients that should be served. Therefore, when the demand for FV patients becomes uncertain, we would like to be sure the extra RV patients resulting from uncertain FV demand are also scheduled in the current planning horizon instead of being postponed to period $T + 1$. We therefore add constraint (50):

$$\sum_{i \in \mathcal{T}} y_{i,T+1} \leq \Psi. \tag{50}$$

In practice, the value of Ψ should be determined in consultation with the clinic. In this paper, we set Ψ by solving the DTCP model (28)–(49) for a particular demand scenario (the one used by [16]) and then insert the obtained value for Ψ in the right-hand side of (50).

Appendix C: Input Parameter Values Used in Experiments

References

1. Addis B, Carello G, Grosso A, Lanzarone E, Mattia S, Tànfani E (2015) Handling uncertainty in health care management using the cardinality-constrained approach: Advantages and remarks. Oper Res Health Care 4:1–4
2. Ahmadi-Javid A, Jalali Z, Klassen KJ (2017) Outpatient appointment systems in healthcare: a review of optimization studies. Eur J Oper Res 258(1):3–34
3. Bajgiran Sanei O, Kazemi Zanjani M, Nourelfath M (2017) Forest harvesting planning under uncertainty: a cardinality-constrained approach. Int J Prod Res 55(7):1914–1929
4. Bertsimas D, Sim M (2003) Robust discrete optimization and network flows. Math Program 98(1):49–71
5. Bertsimas D, Sim M (2004) The price of robustness. Oper Res 52(1):35–53
6. Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. Oper Res 54(1):150–168
7. Bienstock D (2007) Histogram models for robust portfolio optimization. J Comput Financ 11(1):1
8. Birge JR, Louveaux F (2011) Introduction to stochastic programming. Springer Science & Business Media
9. Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. Oper Res 58(4-part-1):802–816
10. Geranmayeh S (2015) Optimizing surgical scheduling through integer programming and robust optimization. PhD thesis, Université d'Ottawa/University of Ottawa
11. Hazır Ö, Dolgui A (2013) Assembly line balancing under uncertainty: Robust optimization models and exact solution method. Comput Ind Eng 65(2):261–267
12. Henrion R (2004) Introduction to chance-constrained programming. Tutorial paper for the Stochastic Programming Community home page
13. Hulshof PJ, Boucherie RJ, Hans EW, Hurink JL (2013) Tactical resource allocation and elective patient admission planning in care processes. Health Care Manag Sci 16(2):152–166

Table 3 Parameter values

Parameter description	Value	Time Unit	
FV appointment access time target	u_m	2	Weeks
	u_p	3	Weeks
	u_{100}	9	Weeks
RV appointment access time target	a	2	Weeks
	b	30	Weeks
	$\bar{a}$	30	Weeks
Consultation time	τ^f	15	Minutes
	τ^r	10	Minutes
Discharge rate	α	38	%
	β	2	%

14. Jalilvand-Nejad A, Shafaei R, Shahriari H (2016) Robust optimization under correlated polyhedral uncertainty set. Comput Ind Eng 92:82–94
15. Nguyen TBT (2014) Modelling, analysis and optimization in resource planning for outpatient clinics. PhD thesis, Nanyang Technological University
16. Nguyen TBT, Sivakumar AI, Graves SC (2015) A network flow approach for tactical resource planning in outpatient clinics. Health Care Manag Sci 18(2):124–136
17. Nguyen TBT, Sivakumar AI, Graves SC (2018) Capacity planning with demand uncertainty for outpatient clinics. Eur J Oper Res 267(1):338–348
18. Pour FH (2016) Robust radiotherapy appointment scheduling. Master's thesis, Concordia University Montreal, Quebec
19. Shalamzari Mirahmadi A (2018) Robust multi-class multi-period scheduling of MRI services with wait time targets. Master's thesis, University of Waterloo
20. Tang J, Wang Y (2015) An adjustable robust optimisation method for elective and emergency surgery capacity allocation with demand uncertainty. Int J Prod Res 53(24):7317–7328