



CASE STUDY

Organizing multidisciplinary care for children with neuromuscular diseases at the Academic Medical Center, Amsterdam

Nikky Kortbeek^{1,2,3,4},
M. F. van der Velde^{1,2} and
N. Litvak^{1,2}

¹Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands; ²Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands; ³Department of Quality and Process Innovation, Academic Medical Center Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; ⁴Rhythm, Houtsingel 5, 2719 EA Zoetermeer, The Netherlands

Correspondence: Nikky Kortbeek,
Correspondence: M. F. van der Velde,
Correspondence: N. Litvak, Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands

Abstract

The Academic Medical Center (AMC) in Amsterdam, The Netherlands, recently opened the 'Children's Muscle Center Amsterdam' (CMCA). The CMCA diagnoses and treats children with neuromuscular diseases. The patients with such diseases require care from a variety of clinicians. Through the establishment of the CMCA, children and their parents will generally visit the hospital only once a year, while previously they used to visit on average six times a year. This is a major improvement, because the hospital visits are both physically and psychologically demanding for the patients. This paper describes how quantitative modelling supports the design and operations of the CMCA. First, an integer linear program is presented that selects which patients are to be invited for a treatment day and schedules the required combination of consultations, examinations and treatments on one day. Second, the integer linear program is used as input to a simulation study to estimate the capacity of the CMCA, expressed in terms of the distribution of the number patients that can be seen on one diagnosis day. Finally, a queueing model is formulated to predict the access time distributions based upon the simulation outcomes under various demand scenarios. Its contribution on the case under study is twofold. First, we design highly constrained appointment schedules for multiple patients that require service from multiple disciplines' resources. Second, we study the effect of the trade-offs between scheduling constraints and access times. As such, the contribution of this case study paper is that it illustrates the value of applying Operations Research techniques in complex healthcare settings, by designing context-specific combinations of mathematical models, thereby improving delivery of the highly-constrained multidisciplinary care.

Health Systems (2017). doi:10.1057/s41306-016-0020-5

Keywords: healthcare management; patient flow; appointment scheduling; queueing systems; integer linear programming

Introduction

The Academic Medical Center (AMC) Amsterdam, The Netherlands, recently opened a center for children with neuromuscular diseases. Neuromuscular diseases is the generic term for a broad set of disorders which impair the functioning of the muscles via muscle or nerve pathology. Most of the diseases are progressive in time, sometimes leading to an early death of the patient (European Neuromuscular Centre, 2012).

Communicated by Prof P Harper.

Received: 27 March 2015

Revised: 15 November 2016

Accepted: 30 November 2016

Most neuromuscular diseases have no cure, so the goal of the treatment is to reduce symptoms, and increase both the mobility and life expectancy (Medline Plus, 2012). Examples of neuromuscular diseases are the diseases of Duchenne, Becker and Charcot Marie Tooth (Vereniging Spierziekten Nederland, 2011). In this paper, we present the quantitative modelling that supports the AMC in the design and operations of the CMCA.

Children with neuromuscular diseases typically need care from various physicians and therapists: a rehabilitation physician, a neurologist, a clinical geneticist, a cardiologist and a pneumonologist. In addition, psychologists, dieticians and even cardiac surgeons may be required. Traditionally, these patients (and their parents) are forced to make separate visits to these various physicians, possibly working in different outpatient clinics, which imposes several risks and challenges. A lack of coordination between the disciplines may result in under- or overtreatment, treatments may be performed in a non-optimal order, or certain aspects of the disease may be overlooked. However, the necessary accurate coordination is a challenging task because different disciplines are accommodated at different locations within the hospital, and often have different routines and conflicting schedules. This article illustrates the potential of applying operations research techniques in supporting the realization of efficient and patient-centred care in multidisciplinary-focused care facilities.

In order to meet the challenge and improve the care quality, the AMC has decided to cluster the expertise of the involved care providers in the ‘Children’s Muscle Center Amsterdam’ (CMCA). Regularly, a treatment day is organized on which the required disciplines come together to see multiple patients. Physicians discuss the condition of the different patients, so that diagnoses are settled earlier and treatments are better customized. With the opening of the CMCA, children and their parents will generally visit the hospital only once a year, while previously they used to visit on average six times a year. This is a major improvement, because the hospital visits are both physically and psychologically demanding for the patients. Also, the CMCA will take away the great responsibility and burden from the parents: to gather all relevant information from the different hospital visits and to schedule the right appointments at the right time. Consequently, the CMCA will simultaneously increase quality of care and patient-centredness.

We address the following challenges encountered at the AMC in making the centralized care at CMCA a reality. (1) All physicians should cooperate and reserve time for the treatment days in their already busy schedules. (2) For each treatment, day patients have to be selected and scheduled in an optimal way. Since the schedules are heavily constrained, construction by hand is very time consuming and does not guarantee the best solution. (3) Due to the small size of the patient group, the treatment days are not often organized, and due to the many constraints, only a few patients can be

scheduled in one treatment day. Consequently, long access times may arise (i.e. the time between the day of an appointment request and the actual appointment date).

We contribute to the logistic questions on two levels:

- *Day scheduling method* First, we present a method that solves the task of scheduling the required consultations, diagnostics and treatments on one day. To this end, we develop and solve an Integer Linear Program (ILP) that simultaneously selects which patients are to be invited for a particular treatment day, and generates an optimal day schedule that complies with all relevant restrictions and preferences.
- *Access time evaluation* Second, we evaluate the access times for new patients, using a queueing model. As input, this model uses probability distributions of the number of patients that can be invited to a treatment day. This input is obtained by simulations: we repeatedly generate sets of patients using historical data, and then apply our developed day scheduling algorithm.

The research reported in this article has been conducted in close cooperation with the neuromuscular disease care experts. Since the CMCA has started very recently, demand predictions are uncertain: estimations run from 20 to 50 new patients per year. Therefore, we considered various demand scenarios. We have analysed how the constraints proposed by the hospital (such as the availability of the specialists) affect the access times, detected the bottlenecks, and proposed the improvements. The outcomes of this study are used to advise the AMC on how frequently the treatment days should be organized, and how to resolve the bottlenecks in the day schedules. The approach can be applied in a general setting of multidisciplinary care, faced with highly constraint scheduling and service level guarantees for access times.

This paper addresses three prominent open challenges in the healthcare appointment scheduling literature, as stated, for example, in Gupta & Denton (2008) and Hulshof *et al* (2012): (1) planning coordinated packages of care for patients needing treatment from several health services; (2) scheduling in highly constrained situations, and (3) formulating models that link day scheduling with access time calculations. Our case study shows that in complex multidisciplinary settings, a context-specific approach is essential, and that it is important to acknowledge the interrelation between decisions made with respect to day schedules and their influence on access times.

This paper is organized as follows. The “[Literature and Contribution](#)” section provides an overview of the related literature. The “[Case study](#)” section describes the characteristics of the setting of the case study. The “[Appointment scheduling](#)” section presents the ILP model for the planning of a treatment day. In this section, planning algorithm is applied to data of the targeted patient group, and the results for these patients are presented. Based on our results of this planning algorithm, an access time

model is derived in “Access time analysis” section, and the numerical results are given for the AMC case. The paper ends with a detailed discussion on the merits and limitations of the proposed ILP-based methods, and the implications and suggestions for providing the high-quality health care of the patients in the concluding “Discussion” section.

Literature and contribution

There is a vast literature on designing appointment systems in healthcare. Two comprehensive surveys are provided in Cayirli & Veral (2003) and Gupta & Denton (2008). Appointment systems can be regarded as a combination of two distinct queueing systems. The first queueing system concerns customers making an appointment and waiting until the day the appointment takes place (the access process). The second queueing system concerns the process of a service session during a particular day (the day process).

The literature on appointment scheduling has mainly focused on the day process; specifically, on scheduling *single appointments* (i.e. each patient needs just one consultation) on a particular day for an *individual service provider* (Cayirli & Veral, 2003). The objectives that are typically considered are minimizing customer waiting time, maximizing resource utilization, and minimizing resource idle time. The influence of the designed schedules on patient access times is typically neglected.

Scheduling *multiple appointments* at once for a *single resource* for a planning horizon of one day or one week is addressed in Chien *et al* (2008, 2009) and Podgorelec & Kokol (1997). These references study the scheduling of a particular day for a given set of patients with given sets of physical therapy treatments. Chien *et al* (2008, 2009) formulate this problem as a hybrid shop scheduling problem, and solve it by developing a genetic algorithm (Chien *et al*, 2008), combined with data mining techniques in later work (Chien *et al*, 2009). Podgorelec & Kokol (1997) develop a scheduling algorithm based on genetic algorithms and machine learning.

ILP approaches for highly constrained *monodisciplinary* treatment planning consisting of *multiple appointments* are employed in Conforti *et al* (2008, 2011), Ogulata *et al* (2008) and Turkcan *et al* (2012) for radiotherapy and chemotherapy treatments. For these patients treatments have to be scheduled during a given number of weeks, strictly taking into account the required rest periods.

Most of the literature on healthcare scheduling, including the aforementioned articles, does not take access times into account. Conforti *et al* (2011) make a step in this direction by developing an ILP for radiotherapy treatment planning, that has the objective to maximize the number of patients from the waiting list that receive an appointment. As, such they indirectly minimize access times. Our contribution compared to Conforti *et al* (2011) is that we provide an explicit evaluation of the access times, using the ILP results as an input to the

queueing system that models the access process. With this approach, we are able to quantify consequences of decisions made on the level of the day schedules, on the access time in the long run.

Our problem under study shows similarities to the resource-constrained project scheduling problem [see De-meulemeester & Herroelen (2006)], as jobs (patients), precedence constraints, process times and resource limitations are involved. Examples of applications of resource-constrained project scheduling to healthcare settings can be found in Naber & Kolisch (2014) and Roland *et al* (2010). We will model the problem under study with an ILP, by which the planning problem can be modelled appropriately, and multiple objectives can be weighted rationally. Applying resource-constrained project scheduling problem to our problem under study is not straightforward, but it might have the potential to provide a more efficient technique than we present in this paper. This is an interesting direction for further research.

On the other end of the spectrum, models that do evaluate access times for outpatient settings typically do not consider day schedules. For example, continuous-time queueing models for access times have been studied in Green & Soares (2007) and Green *et al* (2006), under various assumptions on the access process, including time-dependent demand (Green *et al* 2006). The length of hospital waiting lists were studied in Worthington (1987), and further, for example, in Goddard & Tavakoli (2008). In order to preserve the discrete-time nature, Kortbeek *et al* (2014) presents a slotted queueing model in discrete time that is solved by a generating function approach based upon (Bruneel & Wuyts, 1994).

In the current paper, we evaluate access times using discrete time queueing model with batch service. Queueing systems with batch service were first considered by Bailey (1954), motivated, as in this paper, by evaluation of access times for out-patients in hospitals. Other applications mentioned in the literature are in transport, control of traffic flows, and manufacturing. There is a vast literature on the analysis and numerical evaluation of queues with batch service, see e.g. Chaudhry *et al* (1987), Gold & Tran-Gia (1993) and Neuts (1967, 1979). In this paper, we approximate the queueing process with a finite Markov chain and use the renewal theory to derive stationary waiting times. In the literature, we could not find results on waiting times in the type of queueing system that we consider, therefore, we also provide the analysis. We believe that the proposed model is natural and useful for the integral analysis of day schedules and access times.

Summarizing, our contribution is twofold. First, we design appointment schedules for *multiple patients* that require service from *multiple disciplines/resources*. The day schedules are highly constrained and we integrate the patient selection decision given a certain waiting list. Second, we study the effect of produced schedules on the access times of the patients. The trade-offs between

scheduling constraints and access times, to the best of our knowledge, have not been addressed before. Our quantitative evaluations of this trade-off resulted in recommendations that have been adopted at the AMC.

Case study

At the CMCA children are treated up to eighteen years old, who either already have been diagnosed with a neuromuscular disease (follow-up patients), or are suspected of having one (new patients). The CMCA does not invite follow-up and new patients on the same day, because a different team of physicians is required. Therefore, 'diagnosis days' are organized for patients suspected of having a neuromuscular disease, and 'follow-up days' for patients who have already been diagnosed. A summary of the patient flow is displayed in Figure 1.

Preconsultation When a physician suspects a neuromuscular disease, the patient and the parents are first asked to fill out a questionnaire at home, which is then assessed by the CMCA. If the questionnaire does not support the suspicion of a neuromuscular disease, the patient will not be invited to the CMCA.

Diagnosis day If a patient is eligible for a diagnosis day, a set of required consultations and examinations is determined during a meeting between the 'core members' of the multidisciplinary treatment team. The core members are the physicians who are together responsible for the patient's treatment. They will all see the patient during the diagnosis day. The team is completed by a nurse practitioner who provides administrative support to both physicians and patients. For diagnosis days the core of the team is formed by a paediatric neurologist, a clinical geneticist, and the nurse practitioner.

Next, the patient is scheduled to come to the AMC for a diagnosis day. On this day, there will first be an intake meeting between the patient and the nurse practitioner. Then, the prescribed consultations and examinations will

take place. During the afternoon, a Multidisciplinary Team Meeting (MTM) is scheduled in which the outcomes of the consultations and examinations are discussed. In the MTM, the diagnosis is settled, and a care plan is designed. After the MTM, the neurologist shares the conclusions with the patient. During the day the nurse practitioner is present to act as a host for the patient and to guide the patient through the different examinations. The day finalizes with a meeting between the patient, parents, and the nurse practitioner, to answer remaining questions and to explain the further care pathway. If it is not possible to settle the diagnosis on the same day (which is exceptional), additional examinations are scheduled for a new visit and the care plan is formulated outside the scope of a diagnosis day of the CMCA.

Follow-up programme. If the conclusion of the diagnosis day is that the patient has a neuromuscular disease, he/she will continue to the follow-up programme. During this programme, the physicians monitor the health of the patient and give advice on how to reduce and handle the symptoms. Most of the check-ups have to be performed annually, therefore the patient will visit a follow-up day once a year. Here, the core of the team consists of a paediatric neurologist, a paediatrician, a rehabilitation physician and the nurse practitioner. Also, the set of required appointments is different than for the diagnosis day and depends on the type and severity of the disease from which the patient suffers. The set-up of a follow-up day is as follows: an intake with the nurse practitioner, examinations, a first MTM, consultations, a second MTM, a feedback consultation by a rehabilitation physician, and a final consultation with a nurse practitioner. When the patient turns 18, he will proceed to the adults track.

Day schedule

A month prior to a diagnosis day, the patients are selected who are preferably invited for the diagnosis day of the next month. If there are at least two candidate patients, a diagnosis day is scheduled. Otherwise, the hospital considers it to be inefficient. A feasible day schedule has to be composed to assess how many patients can actually be invited. The schedule of a treatment day is highly constrained: for example some physicians are only available on specific times of the day, for some consultations several physicians have to be present, appointment precedence constraints have to be satisfied, and all results of diagnostic tests have to be available before the MTM can start.

Based upon the predominant suspected disease, patients are assigned to one of four patient types. For each type, the percentage of patients that require a certain appointment is listed in Table 1. These numbers are based on data from the patients who have been treated before outside the CMCA augmented with estimations of the involved physicians.

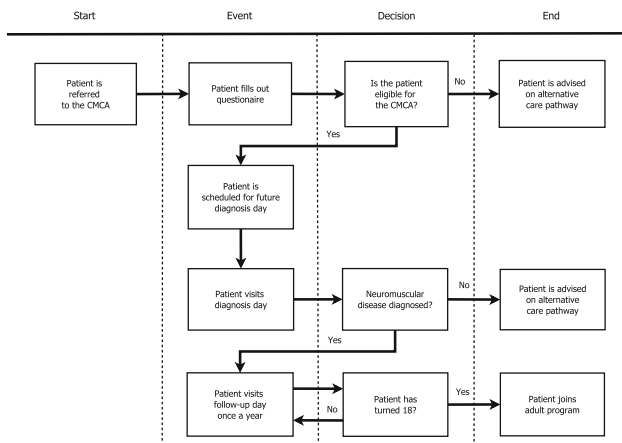


Figure 1 Patient flow diagram.

Table 1 The required appointments per patient type (necessary appointments: preconditional for a patient to be invited; desirable: preferably added if the day schedule allows)

| Patient type | Appointment (min) | | | | | | | | | | |
|-------------------------------|--------------------------------|--------------------------|------------------|----------|--------------------|-------------------------------|--------------------------------|------------------------|--------------------------|-------------------------------|----------|
| | Necessary | | | | | Desirable | | | | | |
| | Intake and evaluation (3 x 15) | Clinical geneticist (45) | Neurologist (45) | MTM (15) | Paediatrician (30) | Rehabilitation physician (30) | Physiotherapist (45) | Blood examination (15) | Clinical photograph (15) | Cardiac ultrasound / ECG (60) | EMG (75) |
| | % | | | | | | Required by x% of the patients | | | | |
| Myopathy (MP) | 39 | | | | 50 | 50 | 50 | 100 | 100 | 50 | 15 |
| Neuropathy (NP) | 39 | | | | 50 | 50 | - | 100 | 100 | - | 50 |
| Spinal muscular atrophy (SMA) | 20 | | | | 50 | 50 | 50 | 100 | 100 | - | 15 |
| Neuromotor disease (NMD) | 2 | | | | 50 | 50 | - | 50 | 100 | - | - |

| Patient type | Appointment (min) | | | |
|-------------------------------|--------------------------------|----------|------------|----------------------------|
| | Desirable | | | |
| | Muscle ultrasound (45) | MRI (60) | X-ray (30) | Needle biopsy muscle (120) |
| | Required by x% of the patients | | | |
| Myopathy (MP) | 100 | 15 | 20 | 15 |
| Neuropathy (NP) | - | - | - | - |
| Spinal muscular atrophy (SMA) | 50 | - | 10 | - |
| Neuromotor disease (NMD) | - | - | - | - |

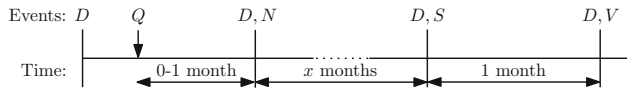


Figure 2 Timeline of patient access time to diagnosis day (Legend: *D* diagnosis day, *Q* questionnaire received, *N* not yet scheduled because of the waiting list, *S* scheduled for next treatment day, *V* visit takes place).

The core team members subdivide for each patient the list of required consultations and examinations in ‘necessary’ and ‘desirable’ appointments. A patient is invited to a diagnosis day if all necessary appointment can be scheduled. Further, the trade-off has to be addressed between skipping some of the desirable (but not necessary) appointments in order to invite more patients, or keeping all appointments and inviting less patients. Clearly, the latter option will result in longer access times. In “[Appointment scheduling](#)” section, an integer linear program is presented that simultaneously addresses: (1) rational patient selection in conjunction with the appointments to be executed, and (2) the creation of a day schedule. The resulting access times are analysed in “[Access time analysis](#)” section.

Access times

The next concern at the CMCA is the access times for diagnostic patients. We note that for follow-up patients the access times are no major issue because they need to revisit the CMCA between 12 and 15 months after their last treatment day. Then, follow-up days are well predictable and can be always scheduled within the required time frame.

The access time of a diagnostic patient is counted as the number of days between the date when a completed questionnaire is received at the hospital, and the date of the patient’s visit. The timeline between these two dates is illustrated in Figure 2. The AMC strives for a maximum access time of seven weeks for diagnosis days, which is quite ambitious because initially the time for a diagnosis day was reserved only once per month. Due to the variety in patient types, the complexity of the set of scheduling constraints, and diverging availability of the different care providers, the number of patients that can be invited for each diagnosis day cannot be easily predicted.

In “[Access time analysis](#)” section, a queueing model is presented and the access time distribution for diagnostic patients is derived, based upon the probability distribution of the number of patients that can be seen on a particular diagnosis day. We provide recommendations on the desired frequency of the diagnosis days so that the access times do not exceed the prescribed norms.

Appointment scheduling

In this section, a mathematical model is formulated that has the purpose to decide which patients are invited to visit the center for the next treatment day (both

diagnosis and follow-up) and to compute an optimal schedule for this day. We first give an overview of the properties of the model. Next, we describe our experimental setup, and finally, present the results for diagnosis days. For clarity of presentation, the detailed mathematical formulation of the model is presented in the appendix. The model was developed in close cooperation with the CMCA healthcare professionals. Several versions of the model were designed and tested. Each time, the formulation and the inclusion or exclusion of specific constraints and objectives were discussed based upon the outcomes on various test problems.

Model formulation

We formulate the day scheduling problem as an Integer Linear Program (ILP). In our model, a treatment day is divided in time slots of equal length. The decisions to be made are twofold: which patient gets which appointment at what time slot, and with which types of resource(s) (the resource(s) can be staff members and/or equipment). Thus, the decision variables are as follows:

$$z_{j,p,s,t} = \begin{cases} 1 & \text{if patient } j \text{ has an appointment with resource } s \\ & \text{for procedure } p \text{ starting at time slot } t, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

We say that a patient has a *complete visit* if all his/her appointments are scheduled (both the necessary and desirable appointments, recall from “[Case study](#)” section). If some of the desirable appointments are omitted, we say that a patient has a *partial visit*. We are interested in aspects such as which patients have a complete visit, which have a partial visit, and at what time they have which type of appointment. The formal description of these variables can be found in Table 5, see Appendix A. The constraints and objectives will be described in the next subsections, while the mathematical formulation is given in Appendix A.

Constraints

We distinguish several types of constraints:

Patient selection A patient has most of his/her appointments, all of them, or none. The visit of such a patient is thus a complete visit, a partial visit, or the patient is not scheduled. The amount of appointments that are allowed to be omitted is patient specific. At least two patients must be scheduled on a treatment day; otherwise, the treatment day is cancelled. For reasons of fairness, patients are scheduled according to the first-come-first-served (FCFS) discipline. This implies that the decision on *which* patients are invited is equivalent to the decision on *how many* patients are invited from the beginning of the queue.

Basic constraints A patient gets each treatment at most once, a treatment is carried out by a resource that has

the necessary qualifications, and a resource can only be scheduled at one place at a time and should be available.

Precedence constraints Some treatments have to be performed before others; there is a minimum amount of time between the starting times of some combinations of treatments.

CMCA-specific appointment constraints Some of the appointments of a patient can take place simultaneously. For example, an orthopaedist and physiotherapist can treat one patient at the same time. A patient needs time to rest; therefore, at least half an hour per three hours must be free from appointments. Some appointments are obligatory. If these are not scheduled, the patient cannot be scheduled.

Patient-specific constraints A patient cannot get more appointments than he/she can physically handle in a day. Therefore, sometimes appointments of a patient have to be spread over several days. This number of days is given per patient. However, an appropriate fraction of total appointment time should be scheduled on the first treatment day in order to avoid scheduling problems in future.

Multidisciplinary Team Meeting constraints The MTM is scheduled as a series of appointments, one per patient, in consecutive time slots, to make it just one meeting. All core team members should be present at the MTM. The starting time of the MTM is flexible, sometimes physicians prefer that the MTM starts at a fixed time.

Defining constraints This last set of constraints is required for obtaining a correct mathematical formulation. The defining constraints determine the starting and the end times of the patients and staff members, and produce the indicators of whether or not a patient has an appointment at a certain time slot.

Objectives

The following objectives have been formulated, in descending order of priority:

1. Maximize the number of patients that have a complete visit
2. Maximize the number of patients that have a partial visit
3. Maximize the treatment time of all scheduled patients
4. Minimize the idle time in the schedules of the staff
5. Minimize the idle time in the schedules of the patients

The objective of the ILP is to maximize the sum of the weighted rewards on these objectives. In the objective function (see Eq. (33) in Appendix A), objectives 1, 2, 3, 4 and 5 are, respectively, rewarded by weight factors $\alpha, \beta, \gamma, \delta,$ and ϵ . One may observe that the objective function contains multiple goals that are possibly in conflict. By varying the weight factors in the objective function, the relative importance of the various goals can be specified.

The calibration of weight factors is done in close cooperation with the care professionals. Consider, for example, the first and second objectives. Assume that one patient with a full visit is as important as two patients with a partial visit. As the visit of a patient is a binary variable, this leads to $\alpha = 2\beta$. The other objectives are measured in the number of time slots: they have significantly smaller values. This is compensated by appropriate scaling of the weight factors.

Experimental setup

The ILP is implemented using the program AIMMS. The solver employed is Cplex 12.2, using the branch and bound technique. All input parameters of the ILP are set

Table 2 Resource availabilities on a diagnosis day

| Resource | 8:00 | 9:00 | 10:00 | 11:00 | 12:00 | 13:00 | 14:00 | 15:00 | 16:00 | 17:00 |
|--------------------------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| Neurologist | | | | | | | | | | |
| Clinical geneticist | | | | | | | | | | |
| Nurse practitioner | | | | | | | | | | |
| Rehabilitation physician | | | | | | | | | | |
| Physiotherapist | | | | | | | | | | |
| Blood examination | | | | | | | | | | |
| Clinical photograph | | | | | | | | | | |
| Cardiac ultrasound / ECG | | | | | | | | | | |
| EMG | | | | | | | | | | |
| Muscle ultrasound | | | | | | | | | | |
| MRI | | | | | | | | | | |
| X-ray | | | | | | | | | | |

Table 3 The weight factors for the objective function

| Objective | Weight factor | Importance | Value |
|-----------|---------------|------------|-------|
| 1 | α | 10 | 100 |
| 2 | β | 8 | 50 |
| 3 | γ | 10 | 2 |
| 4 | δ | 6 | 20 |
| 5 | ϵ | 5 | 2 |

according to the AMC data. Table 2 displays the availability of all resources during a diagnosis day. For brevity, we do not include other data such as appointments precedence and qualifications of the staff to carry out the procedures. Table 3 lists the values used for weight factors. To determine these values, we let the CMCA clinicians score the relative importance of objective on a 0–10 scale. As the objectives are not measured on the same scale, we applied normalization to each weight factor. The normalization factors, multiplied by the relative importance, resulted, after several calibration runs, in the listed weight factor values. In addition, a test set of 5900 patients was constructed at random, based on the data given in Table 1.

Under the given constraints and data, the maximum number of patients that fits in a diagnosis day is five. This can be seen as follows. All patients require 45-min appointment with the clinical geneticist, 15 min in the MTM and 15 min in evaluation (see Table 1). The MTM and the evaluation take place after all other appointments. As all patients need these appointments and only one staff member is available, this yields 75 min per patient. Now consider the availability of the clinical geneticist in Table 2. These appointments can only start at 10:30 and the last appointments should finish the latest at 17:30. Thus, at most $\lfloor 7 \cdot 60 / 75 \rfloor = \lfloor 5.6 \rfloor = 5$ patients can visit a diagnosis day.

A run is stopped as soon as the gap between the LP-bound and the best solution so far is less than 1%, i.e. when a nearly optimal solution has been found. In all cases, the first four objectives have found their optimal values by then, the fifth not necessarily, though. The ILP when run with five patients has 9000–10000 constraints and 3500–4000 variables. The solution time in most cases ranges from seconds to minutes. In some exceptional cases, this can increase to hours. Analysis of these cases showed that the patient set was selected within a few minutes, and the solver keeps searching for the optimal solution that minimizes the idle time of staff and patients.

Computed schedules

As an illustration of a resulting day schedule, the result for an exemplary schedule of a diagnosis day is displayed in Figure 3. Overall, several bottlenecks have been identified in the scheduling of diagnosis days. The following issues restrict the capacity of the CMCA, and need consideration when the CMCA desires to expand:

- Each patient has to visit the clinical geneticist for 45 min. However, this physician is available only from 10:30 h. These consultations have to take place before the examinations. Since a fifth patient can visit the clinical geneticist at the earliest at 13:30 h, just a short time is left for the examinations.
- The results of the blood examination have to be known before the MTM starts. However, obtaining these results takes two hours, and the blood examination cannot be done before the consultation with the clinical geneticist. Therefore, at the most, three patients only can have a blood examination, regardless of their other appointments.
- Each patient has two appointments after the MTM, one with the neurologist, and one with the nurse practitioner. In combination with the growing length of the MTM when there are more patients, this results in less time for consultations and examinations before the MTM.
- Half of the patients with a neurological disease need to have an EMG examination. The examination takes more than an hour, and the outpatient clinic is closed during the lunch time (12:00–13:00 h). Therefore, at the most, two patients can have this examination on a single day. When there are five patients on one diagnosis day, just one patient can take the EMG examination, regardless of the other appointments needed.
- As not all patients have the physical ability to spend a full day at the hospital, it should be considered to divide the required services for the patient between multiple days. This is part of the model given in Appendix A. For each patient, a maximum of appointments or appointment time can then be given to balance the load for a patient. The hospital, however, did not adopt this suggestion, and therefore, it is left out of the results.

Access time analysis

In this section, we study the access times of diagnostic patients. Recall that the access time is defined as the time span from returning the questionnaire until being present at a diagnosis day. First, we analyse the capacity of a diagnosis day of the CMCA by simulation. Then, a queueing model is formulated to evaluate access time distribution. Finally, we present numerical results on various demand scenarios.

Number of scheduled patients per diagnosis day

The number of patients that can be scheduled in one diagnosis day is a defining determinant for access times. However, due to the complexity of the scheduling problem, this number cannot be directly modelled or predicted; therefore, a simulation study has been performed to determine its distribution. For consecutive diagnosis days, the first patients on the generated test list of 5900 patients were selected, and an optimal schedule was constructed. When a patient was scheduled, he/she was deleted from the list.

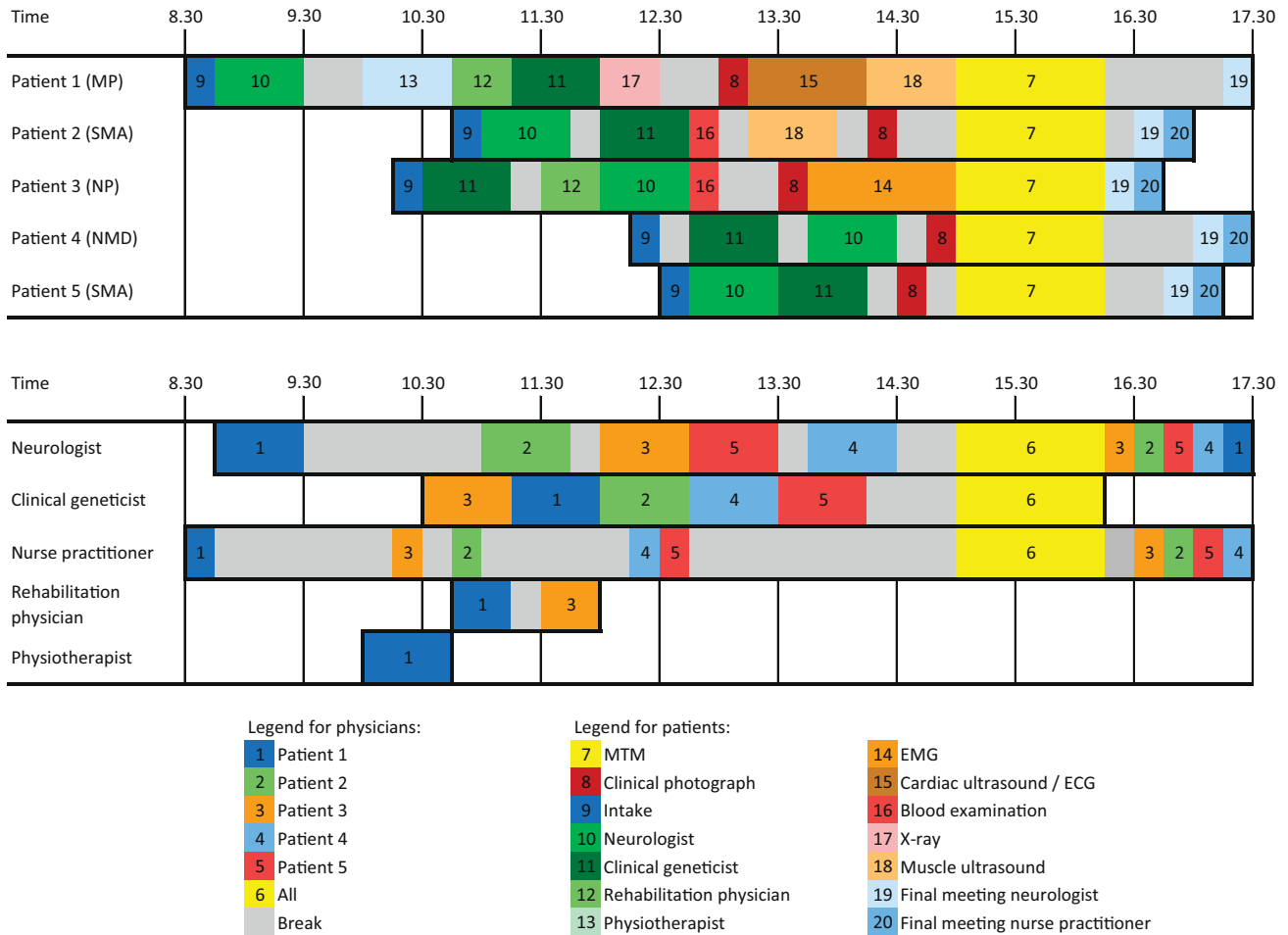


Figure 3 An exemplary day schedule for a diagnosis day.

Table 4 shows the distribution of the number of patient visits scheduled, given the waiting list size. Observe that, at the most, five patients are invited, and when the waiting list consists of two or three patients, all patients are invited. If two patients are scheduled, both have a complete visit. If three or more patients are scheduled, then in almost all cases (99%), three patients have a complete visit and the others have a partial visit; otherwise, two patients have a complete visit and the others have a partial visit.

The simulation results have shown that mainly the needs of the first four scheduled patients, and not the needs of the fifth patient, determine whether the fifth patient can be scheduled or not. Thus, we can assume that the number of patients scheduled on a diagnosis day depends only on the size of the waiting list, and is independent of how many and which patients were scheduled for other diagnosis days.

Queueing model description

We model the arrivals of new patients as a Poisson process, of which the arrival rate, λ (patients per year), is

known. Note that since the CMCA recently started, no historical data on patient arrivals is available. Since patient arrivals originate from a large population of independent (possible) patients, the Poisson assumption is a reasonable assumption. The service discipline is FCFS. A year is split into m time periods of equal length, with one diagnosis day per time period. At the opening of the CMCA, the value for m proposed by the AMC is twelve.

Recall the procedure given in Figure 1. An access time of a patient consists of three parts. (1) The time until the end of the time period. This time is stochastic and has a uniform distribution. (2) The number of full time periods the patient has to wait until being scheduled. This is stochastic, and has a discrete distribution W that has to be determined. (3) The time between being scheduled and the actual visit to the hospital. This time is deterministic.

Let A_n be the random number of arrivals in time period n . Denote by Q_n the number of waiting patients at the end of time period n . Out of Q_n waiting patients, a random number B_n of patients are scheduled for diagnosis day n . The distribution of B_n depends on the value of Q_n . This

Table 4 The distribution of the capacity of diagnosis days

| Length waiting list | Number of patients scheduled | | | | | |
|---------------------|------------------------------|---|------|------|-------|-------|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | 100% | – | – | – | – | – |
| 1 | 100% | – | – | – | – | – |
| 2 | – | – | 100% | – | – | – |
| 3 | – | – | – | 100% | – | – |
| 4 | – | – | – | 0.3% | 99.7% | – |
| 5 | – | – | – | 0.3% | 13.7% | 86.0% |

dependence is described by the conditional probabilities: $P(B_n = b|Q_n = q)$, $q \geq 0$, and $0 \leq b \leq q$. For the case study, these are the probabilities obtained from simulations as discussed in “Number of scheduled patients per diagnosis day” section (see Table 4).

The main performance characteristic of interest is the access time of the patients. Denote by W_n the access time of a patient that arrived at time period n . The distribution of W_n depends on the queue length at the *beginning* of time period n , denoted by Y_n . The following equations hold:

$$Q_n = Y_n + A_n, \text{ and} \quad (2)$$

$$Y_{n+1} = Q_n - B_n, \quad n = 1, 2, \dots \quad (3)$$

Here A_n is independent of the other random variables, and B_n depends on Q_n . Assuming that the arrival rate is not too high, it will often happen that all waiting patients in the queue are scheduled. In that case, the stochastic process W_n will soon reach stationarity; therefore, we choose to obtain its stationary distribution W . To this end, we first determine the stationary distribution Y of Y_n and then obtain the distribution of W using the renewal theory argument. The details of the derivation are provided in Appendix B.

Numerical results

The distribution of W is evaluated numerically, by approximating Y_n with a finite Markov chain. This is justified by the fact that in practice we can find a large enough N (say, $N = 100$) such that Y_n never reaches or exceeds N . The details are provided in Appendix B. The initial frequency of diagnosis days that the CMCA will apply is (at most) twelve such days per year, and therefore, $m = 12$. Next, as stated in “Case study” section, the arrival rate is estimated to be between 20 and 50 patients per year. The time between being scheduled for a diagnosis day and the actual visit to CMCA is precisely one month. For these input parameters, we obtain the total average access times, from the time the questionnaire is received till the hospital visit. The results are presented in Figure 4.

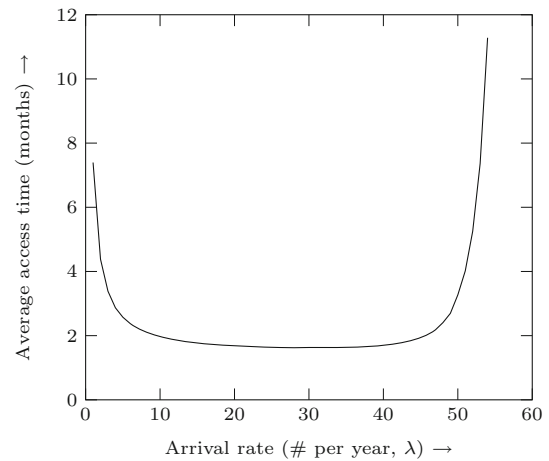


Figure 4 Average access times in months ($m = 12$).

We see that excessively large access times are observed in two extreme cases. When the arrival rate is small, less than ten patients per year, large access times arise because at least two patients have to be scheduled on a single day, and thus arriving patients often have to wait for another arrival. When there are more than fifty patients a year, large average access times arise since the maximal capacity is almost met. In between, the average access time is stable at a value just below two months.

The shape of the distribution of the access times heavily depends on the arrival rate λ . This can be seen in Figure 5. When the arrival rate is low, the moment in a month when a patient arrives does not have any influence on the distribution of the access times. However, as the value of λ increases, a heavier dependence shows, because the patients that arrive at the beginning of a month have a considerably higher chance to be scheduled earlier.

We emphasize that the access times are heavily dependent on the constraints of the scheduling problem. When the bottleneck constraints, as mentioned in “Computed schedules” section, are relaxed, then larger groups of patients can be scheduled together. For example, if the clinical geneticist would be available all day, it will in

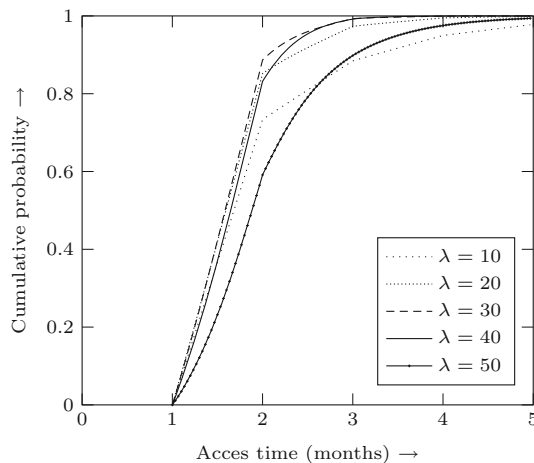


Figure 5 Distributions of access times ($m = 12$).

some cases be possible to schedule one more patient per day. This will increase the maximum capacity and result in smaller access times.

Concluding, the number of patient arrivals is uncertain, which is currently around 20 patients a year, and is expected to increase. Figure 4 shows that expected access times are very similar for arrival rates between 10 and 45 patients per year. Therefore, the proposed setup of the CMCA is robust to this data uncertainty.

Discussion

We showed how a combination of operations research techniques, including integer linear programming, simulation and queuing theory helps the AMC in organizing care for children with neuromuscular diseases. The treatment center embodies a transformation from supply-driven to demand-driven patient care. Customized diagnostics and treatment can be offered in a combined visit. To realize this, all practical constraints and preferences were collected and incorporated in an ILP by which feasible day schedules for multiple patient visits can be constructed. Simulations give insight into the capacity of the CMCA, given the availability of staff and equipment and estimates on patient demand in the number of arrivals and required appointments. Finally, a queuing model predicts the access time distributions for diagnostic patients based upon the simulation outcomes.

Formulating the day scheduling model was an iterative process, intensively involving the clinicians. Results on initial formulations predicted very long patient access times. Presenting these results to the clinicians, yielded that some highly restricting constraints were loosened or deleted. For example, there was a constraint that the physicians were to have at least two appointments per day. Thus, only patients with the same needs could be scheduled together, resulting in long access times. Another example, where the AMC still struggles, is the choice whether the MTM should start at a fixed time or

not. Although a fixed time is preferred by the clinicians, from a patient's point of view, based on the experimentation outcomes, we strongly recommend it to be flexible. We believe that the benefit of quantitative analysis in such a 'negotiation' process is that it rationalizes the process of realizing a good trade-off between interests of clinicians and patients.

The value of applying Operations Research to this healthcare delivery problem has been expressed in both its process and its outcomes. The process of modelling leads to better understanding and recognition of a problem. The outcomes of mathematical models make it possible to prospectively assess the consequences of various alternative interventions, without actually changing the system. Modelling is highly suitable in healthcare settings—since experimenting in practice may induce risks for patients and field experimenting takes more time, is more costly, and offers less statistical reliability. Moreover, since healthcare environments are generally politically charged, by quantifying the impact of potential solutions, fact-based rather than feeling-based decision making can be realized.

A limitation of our study is the availability of accurate data. Since the center just opened its doors, no historical data was available besides data on realizations of how the treatment was previously delivered by different outpatient clinics. In addition, patient-type mix and required appointments could only be estimated from physician's expert opinions and data on the former patient population. However, in "Computed schedules" section, it is observed that the bottlenecks in the schedules are mainly in the availability of resources and not in the exact evaluation of the patient-type mix and appointment requirements. Furthermore, having a focused care center may increase the attractiveness for patients to come to the AMC, which makes predictions on the number of patient arrivals highly uncertain. This is accounted for by studying several demand scenarios in Figure 4. It can be seen that the average access time is similar for the realistic demand scenarios. Concluding, the model is robust to the data inaccuracy. As data inaccuracy is a consequence of the lack of historical data, we recommend the center to constantly monitor its operations. We furthermore recommend to regularly repeat the analysis so to reconsider both the frequency of carrying out treatment days and the staff/equipment availability during a treatment day.

In this paper, we have considered the FCFS discipline for patients to be admitted from the waiting list. There might be a discipline which gives a better performance, if such a discipline increases the number of patients that can be seen on one day. Investigating the existence of such a policy is an interesting direction for future research. However, when changing the service discipline, the justification for the assumption of independence between the batches of patients scheduled on consecutive treatment days has to be reconsidered. Even more prominent, the issue of practical acceptance needs to be

addressed, since it is questionable how clinically acceptable it would be to not schedule the patient that has been on the waiting list the longest.

The first patients have visited the CMCA in a pilot phase of the entire treatment concept. During this pilot phase, the nurse practitioner enters the needs of the patients in an Excel sheet. Given a set of patients with prescribed consultations and examinations, and the availabilities of the staff and equipment, the optimal schedule is determined using AIMMS. This is not the desired ultimate state, since it still requires copying the resulting appointments in the electronic agenda system by hand. In addition, the AMC strongly opposes the implementation of different software tools in different parts of the hospital, to prevent

the maintenance and support task of ICT department to become inefficient if not impossible. Therefore, the scheduling algorithm is intended to be incorporated in the new hospital-wide electronic agenda system that is currently under construction. For such a system, it will be required to be able to communicate with an ILP solver, which will be a main challenge for the ICT design. Modern ICT systems for hospital organization increasingly often embrace operations research solutions, in particular, in capacity evaluation and appointment scheduling. In the near future, advanced quantitative schemes, as the one described in this article, should become a standard part of hospitals' integral ICT support, for transparent and efficient planning of high-quality care.

References

- BAILEY N (1954) On queueing processes with bulk service. *Journal of the Royal Statistical Society Series B (Methodological)* **18**(1), 80–87.
- BRUNEEL H and WUYTS I (1994) Analysis of discrete-time multiserver queueing models with constant service times. *Operations Research Letters* **15**(5), 231–236.
- CAYIRLI T and VERAL E (2003) Outpatient scheduling in health care: a review of literature. *Production and Operations Management* **12**(4), 519–549.
- CHAUDHRY M, MADILL B and BRIERE G (1987) Computational analysis of steady-state probabilities of $M|G^{a,b}|1$ and related nonbulk queues. *Queueing systems* **2**(2), 93–114.
- CHIEN C, HUANG Y, and HU C (2009) A hybrid approach of data mining and genetic algorithms for rehabilitation scheduling. *International Journal of Manufacturing Technology and Management* **16**(1–2), 76–100.
- CHIEN C, TSENG F and CHEN C (2008) An evolutionary approach to rehabilitation patient scheduling: a case study. *European Journal of Operational Research* **189**(3), 1234–1253.
- CONFORTI D, GUERRIERO F and GUIDO R (2008) Optimization models for radiotherapy patient scheduling. *4OR: A Quarterly Journal of Operations Research*, **6**(3), 263–278.
- CONFORTI D, GUERRIERO F, GUIDO R and VELTRI M (2011) An optimal decision-making approach for the management of radiotherapy patients. *OR Spectrum* **33**(1), 123–148.
- DEMEULEMEESTER EL and HERROELEN WS (2006) Project scheduling: a research handbook, vol. 49. Springer, New York (2006)
- EUROPEAN NEUROMUSCULAR CENTRE (2012) <http://www.enmc.org> (accessed 29 February 2012).
- GODDARD J. and TAVAKOLI M. (2008) Efficiency and welfare implications of managed public sector hospital waiting lists. *European Journal of Operational Research* **184**(2), 778–792.
- GOLD H and TRAN-GIA P (1993) Performance analysis of a batch service queue arising out of manufacturing system modelling. *Queueing Systems* **14**(3) 413–426.
- GREEN L and SOARES J (2007) Computing time-dependent waiting time probabilities in $M(t)|M|s(t)$ queueing systems. *Manufacturing & Service Operations Management* **9**(1), 54–61.
- GREEN L, SOARES J, GIGLIO J and GREEN R (2006) Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine* **13**(1), 61–68.
- GUPTA D and DENTON B (2008) Appointment scheduling in health care: challenges and opportunities. *IIE Transactions* **40**(9), 800–819.
- HULSHOF P, KORTBEEK N, BOUCHERIE R, HANS E and BAKKER P (2012) Taxonomic classification of planning decisions in health care: a review of the state of the art in OR/MS. *Health Systems* **1**(2), 129–175.
- KORTBEEK N, ZONDERLAND ME, BRAAKSMA A, VLIEN IMH, BOUCHERIE RJ, LITVAK N, ET AL. (2014) Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals. *Performance Evaluation* **80**(10), 5–26.
- MEDLINE PLUS (2012) <http://www.nlm.nih.gov/medlineplus/neuromusculardisorders.html> (accessed 29 February 2012).
- NABER A. and KOLISCH R (2014) MIP models for resource-constrained project scheduling with flexible resource profiles. *European Journal of Operational Research* **239**(2), 335–348.
- NEUTS M (1967) A general class of bulk queues with poisson input. *The Annals of Mathematical Statistics* **38**(3), 759–770.
- NEUTS M (1979) Queues solvable without Rouché's theorem. *Operations Research* **27**(4), 767–781.
- OGULATA S, KOYUNCU M and KARAKAS E (2008) Personnel and patient scheduling in the high demanded hospital services: a case study in the physiotherapy service. *Journal of Medical Systems* **32**(3), 221–228.
- PODGORELEC V and KOKOL P (1997) Genetic algorithm based system for patient scheduling in highly constrained situations. *Journal of medical systems* **21**(6), 417–427.
- ROLAND, B., DI MARTINELLI, C., RIANE, F., AND POCHE, Y. (2010). Scheduling an operating theatre under human resource constraints. *Computers & Industrial Engineering*, **58**(2), 212–220.
- TURKCAN A, ZENG B and LAWLEY M (2012) Chemotherapy operations planning and scheduling. *IIE Transactions on Healthcare Systems Engineering* **2**(1), 31–49.
- VERENIGING SPIERZIEKTEN NEDERLAND (In Dutch) (2012) <http://www.vsn.nl/spierziekten> (accessed 5 August 2011).
- WINSTON W (2003) Operations Research: Applications and Algorithms. Duxbury Press, Pacific Grove
- WORTHINGTON D (1987) Queueing models for hospital waiting lists. *Journal of the Operational Research Society* **38**(5), 413–422.

Appendix A: Mathematical formulation appointment scheduling

This appendix contains the mathematical formulation of the ILP described in “[Model formulation](#)” section. This model is applicable to both types of treatment days: diagnosis and follow-up days. The constraints are used for both days unless indicated otherwise.

Variables and parameters

Recall from (1) that the decision variables are denoted by $z_{j,p,s,t} \in \{0, 1\}$, which equals one if patient j has an appointment with resource s for procedure p starting at time slot t . Other variables and parameters will be introduced when they are used for the first time. A complete list of sets, indices, variables, parameters and their properties can be found in Tables 5 and 6.

Constraints

In the ILP, several types of constraints are considered. We distinguish: constraints on the selection of patients, basic planning constraints, precedence constraints, appointment constraints, MTM constraints, and defining constraints. Below each constraint is presented in detail.

Selection of patients. A patient can either have a complete visit (if all appointments are scheduled), or a partial visit (if most appointments are scheduled), or not be scheduled at all. We denote by g_j and e_j the binary variables that indicate whether a patient has, respectively, a complete visit ($g_j = 1$) or a partial visit ($e_j = 1$). A patient cannot have a partial and complete visit at the same time, and therefore we have

$$g_j + e_j \leq 1, \quad \text{for all } j. \quad (4)$$

Patient j receives the number o_j of desirable but not necessary appointments. The binary parameter $N_{j,p}$ denotes whether patient j needs procedure p or not. Denote by the binary variable $x_{j,p,t}$ whether patient j has an appointment for procedure p starting at time slot t or not. Constraint (5) below ensures that a patient with a partial visit receives at least as many appointments as necessary. By using a big- M constraint (see Winston Winston (2003)), where $M_1 = |P|$, we make sure that this constraint is always satisfied for patients with a complete visit or no visit:

$$M_1 \cdot (1 - e_j) + o_j + \sum_{p,t} x_{j,p,t} \geq \sum_p N_{j,p}, \quad \text{for all } j. \quad (5)$$

At least two patients should be scheduled on a treatment day; otherwise it is cancelled:

Table 5 Sets and indices ILP

| Set | Description | Index |
|---------------------------------------|-------------|--------------|
| J | Patients | j, j', j'' |
| P | Procedures | p, p' |
| S | Resources | s |
| $T = \{1, 2, \dots, t_{\text{end}}\}$ | Time slots | t, t' |

$$\sum_j (g_j + e_j) \geq 2. \quad (6)$$

Patients are scheduled according to the FCFS discipline:

$$g_j + e_j \geq g_{j'} + e_{j'}, \quad \text{for all } j, j' \in J \text{ such that } j < j'. \quad (7)$$

Basic planning constraints Denote by $L_{j,p}$ the number of time slots patient j needs to undergo procedure p . If $L_{j,p} = 0$, then the patient does not need the procedure. Constraint (8) states that a patient gets each treatment at the most once:

$$\sum_t x_{j,p,t} \leq 1, \quad \text{for all } j, p \text{ such that } L_{j,p} > 0. \quad (8)$$

Denote by the binary $Q_{s,p}$ whether resource s is qualified to perform procedure p . Each scheduled procedure should have a qualified resource performing it at the intended time. This constraint, as given in (9), holds for all procedures except the MTM, for which we have a separate set of constraints. Furthermore, (9) in combination with (8) ensures that no dummy appointments are scheduled to reduce the idle time of staff members:

$$\sum_s z_{j,p,s,t} \cdot Q_{s,p} = x_{j,p,t}, \quad \text{for all } i, j, p \quad (9)$$

such that $p \neq p_{\text{MTM}}$ and $L_{j,p} > 0$.

Denote by the binary $A_{s,t}$ whether resource s is available for the CMCA at time t . Recall that $z_{j,p,s,t} = 1$ if procedure p starts at time t . The following constraint (10) ensures that resource s is available and is scheduled only for one procedure with one patient at a time. The latter is done by checking whether or not an appointment has started in the past that has not yet finished:

$$\sum_{j,p} \sum_{t'=\max\{1, t-L_{j,p}+1\}}^t z_{j,p,s,t'} \leq A_{s,t}, \quad \text{for all } s, t. \quad (10)$$

Precedence constraints Some treatments have to be performed in a given order. For example, an intake appointment of a patient should be scheduled before all other appointments of the day. Denote with the binary $H_{p,p'}$ whether procedure p should be performed before procedure p' in case a patient needs both procedures. Since the constraint only holds for appointments p, p' that are both scheduled, we introduce the binary variable $c_{j,p,p'}$ which is one if both p and p' are scheduled for patient j . This variable is only relevant when both procedures are needed by the patient and there is a precedence constraint. The following constraint (11) ensures $c_{j,p,p'} = 1$ when both procedures p and p' are needed:

$$\sum_t (x_{j,p,t} + x_{j,p',t}) - 1 \leq c_{j,p,p'},$$

for all j, p, p' such that $H_{p,p'} = 1, L_{j,p} > 0, L_{j,p'} > 0$. (11)

Table 6 Parameters and variables ILP

| Notation | Description |
|----------------------------------|--|
| <i>Binary parameters</i> | |
| $Q_{s,p}$ | 1 if resource s is qualified to perform procedure p |
| $A_{s,t}$ | 1 if resource s is available in time slot t |
| $C_{p,p'}$ | 1 if procedures p and p' can be performed simultaneously |
| $H_{p,p'}$ | 1 if procedure p has to be performed before procedure p' |
| $N_{j,p}$ | 1 if patient j needs to undergo procedure p |
| $E_{j,p}$ | 1 if procedure p is necessary for patient j |
| <i>Integer parameters</i> | |
| $F_{p,p'}$ | Minimal number of time slots before start of procedure p' after start of p |
| t_{MTM} | Starting time slot of the MTM |
| p_{MTM} | Procedure number of the MTM |
| $L_{j,p}$ | Number of time slots that procedure p takes for patient j . $L_{j,p} = 0$ indicates that procedure p is not required for patient j . |
| m_j | Maximum number of appointment time slots patient j can handle on a day |
| d_j | Number of treatment days over which the appointments of patient j may be spread |
| o_j | Maximum number of desirable appointments that patient j is allowed to skip in a partial visit |
| <i>Real parameters</i> | |
| u_s | Relative weight of idle time of staff member s |
| k_j | Relative weight of patient j |
| <i>Binary variables</i> | |
| $Z_{j,p,s,t}$ | 1 if patient j has an appointment with resource s for procedure p starting at time slot t |
| $x_{j,p,t}$ | 1 if patient j has an appointment for procedure p starting at time slot t |
| g_j | 1 if patient j has a complete visit |
| e_j | 1 if patient j has a partial visit |
| $b_{j,t}$ | 1 if patient j has an appointment at time slot t |
| $c_{j,p,p'}$ | 1 if patient j has both appointments p and p' scheduled |
| <i>General integer variables</i> | |
| y_j^{\min} | First time slot at which patient j has an appointment |
| y_j^{\max} | Last time slot at which patient j has an appointment |
| y_s^{\min} | First time slot at which staff member s has an appointment |
| y_s^{\max} | Last time slot at which staff member s has an appointment |

Now, we can formulate the precedence constraint, in combination with the constraint on a minimum amount of time between the starting times of some combinations of treatments. This minimum amount of time slots is denoted by $F_{p,p'}$. The big- M formulation ensures that the constraint is always satisfied when $c_{j,p,p'} = 0$. In this constraint, a value of $M_2 = 3 \cdot |T|$ suffices:

$$\sum_t t \cdot (x_{j,p',t} - x_{j,p,t}) - L_{j,p} - F_{p,p'} + (1 - c_{j,p,p'}) \cdot M_2 \geq 0, \quad \text{for all } j, p, p' \text{ such that } H_{p,p'} = 1, L_{j,p} > 0, L_{j,p'} > 0. \quad (12)$$

CMCA appointment constraints A patient can get some of his treatments simultaneously. Denote by the binary $C_{p,p'}$ whether the procedures p and p' can be carried out for the same patient at the same time. The next constraint checks, for each time slot and each combination of appointments, whether they are being performed or not. This is only relevant if the two procedures cannot be performed concurrently, and a qualified resource should be available. This is not relevant for the MTM, since patients are not present there.

$$\sum_{t'=\max\{1,t-L_{j,p}+1\}}^t x_{j,p,t'} + \sum_{t'=\max\{1,t-L_{j,p'}+1\}}^t x_{j,p',t'} \leq 1, \quad \text{for all } j, p, p', t \text{ such that } C_{p,p'} = 0, p > p', \sum_s A_{t,k} \cdot (Q_{p,s} + Q_{p',s}) > 0, p \neq p_{MTM}, p' \neq p_{MTM}. \quad (13)$$

A patient needs a time to rest. Therefore, in the span of three hours, there is at least half an hour free from appointments. These breaks should have the length of at least one quarter of an hour. Denote by the binary variable $b_{j,t}$ whether patient j has an appointment at time t or not. Again, the MTM is not considered as an appointment because the patient is not present at the MTM. If a time slot is 15 min, then we deduce from the above that 2 or every 12 slots must be empty. This gives the following constraint:

$$\sum_{t'=t}^{t+11} b_{j,t} \leq 10, \quad \text{for all } j, t. \quad (14)$$

Note that when time slots have a different length, constraint (14) can be easily adjusted to ensure the patient

has enough time to rest. However, an additional constraint will be necessary so that each break is at least 15 min long.

Some appointments are obligatory: the ‘necessary’ appointments. If these are not scheduled, the patient cannot visit the CMCA. Denote by the binary $E_{j,p}$ whether an appointment is necessary or not. Thus, for all appointments that are necessary, we require

$$g_j + e_j \leq \sum_t x_{j,p,t}, \quad \text{for all } j, t \text{ such that } E_{j,p} = 1. \quad (15)$$

Patient-specific constraints Denote by m_j the maximum number of time slots of appointments that patient j can have on one treatment day. Then we have the following constraint (16) that says that a patient gets no more treatment time than he/she can handle on a day:

$$\sum_t b_{j,t} \leq m_j, \quad \text{for all } j. \quad (16)$$

Sometimes, appointments of a patients have to be spread over several days because of the requirements of the patient. The number of treatment days patient j has left until all appointments are scheduled, is given by d_j . The right-hand side of (17) below is the total appointment length of patient j in one treatment day. If $d_j > 1$, then only complete visit is allowed. Constraint (17) ensures that at least fraction $1/d_j$ of the treatment is accomplished in one day for patient j who has a complete visit:

$$g_j \cdot \frac{1}{d_j} \cdot \sum_{p \neq p_{MTM}} L_{j,p} \leq \sum_t \sum_{p \neq p_{MTM}} x_{j,p,t} \cdot L_{j,p}, \quad \text{for all } j. \quad (17)$$

Multidisciplinary Team Meeting constraints We schedule the MTM as one appointment for all patients. Using precedence constraints below, we will ensure the length of the MTM is correct. The core team members are formally assigned to the first patient:

$$\sum_s z_{j,p,s,t} = x_{j,p,t} \cdot \sum_s Q_{p,s} \quad \text{for all } t, p = p_{MTM}, j = 1. \quad (18)$$

Now we define the precedence constraints for the MTM. Some of the appointments must be finished before the MTM. Thus, we define a constraint similar to (12)

$$\begin{aligned} \sum_t t \cdot (x_{j',p',t} - x_{j,p,t}) - L_{j,p} - F_{p,p'} + (1 - c_{j,p,p'}) \cdot M_2 &\geq 0, \\ \text{for all } j, j', p, p' \text{ such that } H_{p,p'} &= 1, L_{j,p} > 0, p' = p_{MTM}. \end{aligned} \quad (19)$$

Some appointments can start only after the MTM. The length of the MTM for scheduled patient j is $(g_j + e_j) \cdot L_{j,p_{MTM}}$. This yields a constraint similar to (19)

$$\begin{aligned} \sum_t t \cdot (x_{j,p,t} - x_{j',p',t}) - \sum_{j''} L_{j'',p'} \cdot (g_{j''} + e_{j''}) - F_{p,p'} \\ + (1 - \sum_t x_{j',p',t}) \cdot M_2 + (1 - \sum_t x_{j,p,t}) \cdot M_2 &\geq 0, \\ \text{for all } j, j', p, p' \text{ such that } p' = p_{MTM}, L_{j,p'} > 0, H_{p',p} &= 1. \end{aligned} \quad (20)$$

Sometimes it is desirable to always start the MTM at a fixed time. Denote by t_{MTM} the time slot in which the MTM should start. Then we obtain a constraint for the starting time of each MTM appointment:

$$\sum_t t x_{j,p,t} = t_{MTM}, \quad \text{for all } j, p = p_{MTM}. \quad (21)$$

Defining constraints This group of constraints determines the starting and the end times of the patients and staff members. Denote by y_j^{min} the first time slot when patient j has an appointment. We define y_j^{min} through a big- M constraint (22) that holds for all t smaller than the starting time of the first appointment. Here $M_3 = |T|$ is sufficient:

$$y_j^{min} \leq M_3 + (t - M_3) \cdot x_{j,p,t}, \quad \text{for all } j, p, t \text{ such that } p \neq p_{MTM}. \quad (22)$$

The last time slot when patient j has an appointment, y_j^{max} , is determined by the following constraint:

$$y_j^{max} \geq (t + L_{j,p}) \cdot x_{j,p,t}, \quad \text{for all } j, p, t \text{ such that } p \neq p_{MTM}. \quad (23)$$

Note that when patient j is not planned, y_j^{min} and y_j^{max} can take any integer value in the interval $[0, \dots, |T|]$.

In a similar fashion, we can derive the minimum and maximum values for staff members:

$$y_s^{min} \leq M_3 + (t - M_3) \cdot z_{j,p,s,t}, \quad \text{for all } j, p, s, t, \text{ and} \quad (24)$$

$$y_s^{max} \geq (t + L_{j,p}) \cdot z_{j,p,s,t}, \quad \text{for all } j, p, s, t. \quad (25)$$

The next constraint determines whether a patient has an appointment at a certain time or not; recall that this is denoted by the binary variable $b_{j,t}$. The following constraint forces $b_{j,t} = 1$ when a patient has an appointment. The big- M , is needed because a patient can have multiple appointments at one time slot. Here we take with $M_4 = |P|$:

$$\sum_{p \neq p_{MTM}} \sum_{t' = \max\{1, t - L_{j,p} + 1\}}^t x_{j,p,t'} \leq M_4 \cdot b_{j,t}, \quad \text{for all } j, t. \quad (26)$$

The following constraint (27) ensures $b_{j,t} = 0$ whenever patient j has no appointment at time t

$$b_{j,t} \leq \sum_{p \neq p_{MTM}} \sum_{t' = \max\{1, t - L_{j,p} + 1\}}^t x_{j,p,t'}, \quad \text{for all } j, t. \quad (27)$$

Objective function

The objective function is a linear combination of the five objective functions as described below.

Maximize the number of patients that have a complete visit Denote by k_j the relative weight of patient j . Then we want to maximize the following expression:

$$\sum_j g_j \cdot k_j. \quad (28)$$

Maximize the number of patients that have a partial visit This expression is similar to (28)

$$\sum_j e_j \cdot k_j. \quad (29)$$

Maximize the treatment time of all scheduled patients This objective maximizes the utilization of the hospital facilities. Note that constraint (8) ensures that no dummy appointments are being scheduled. If two appointments with lengths $L_{j,p}$ and $L_{j,p'}$ are scheduled at the same time, then we need to add $L_{j,p} + L_{j,p'}$ to the total treatment time. Thus, we want to maximize

$$\sum_{j,p,t} x_{j,p,t} \cdot L_{j,p}. \quad (30)$$

Minimize the idle time in the schedules of the staff We have already defined the starting and end times of a staff member. Since the idle time of some staff members (or resources) might be more important than that of others, we assign a relative weight u_s to the idle time of a staff member (or resource) s . Thus, we wish to minimize the following expression:

$$\sum_s u_s \cdot \left(y_s^{\max} - y_s^{\min} - \sum_{j,p,t} z_{j,p,s,t} \cdot L_{j,p} \right). \quad (31)$$

Minimize the idle time in the schedules of the patients It is assumed that the idle time of each patient is equally important. Note that constraint (14) ensures that each patient has enough time to rest. Then, the total idle time of the patients equals to

$$\sum_j \left(y_j^{\max} - y_j^{\min} - \sum_t b_{j,t} \right). \quad (32)$$

The expressions (28)–(32) contribute to the objective functions, each having its own relative importance. The coefficients determining the relative importance are given by α , β , γ , δ , and ϵ , which are calibrated as discussed in “Appointment scheduling” section. Thus, we obtain the following objective function:

$$\max \quad \alpha \cdot (28) + \beta \cdot (29) + \gamma \cdot (30) - \delta \cdot (31) - \epsilon \cdot (32). \quad (33)$$

Appendix B: Derivation of the access time distribution

In this appendix, we derive the stationary distribution of the stationary waiting time W as defined in “Queueing model description” section.

First, we write the transition probabilities for Y_n . From (3), by conditioning on $[Q_n = q]$ and noting that $P(B_n = q - i | Q_n = q) = 0$ whenever $q - i < 0$, we obtain

$$P(Y_{n+1} = i | Y_n = j) = \sum_{q=\max(i,j)}^{\infty} P(B_n = q - i | Q_n = q) \cdot P(Q_n = q).$$

Next, using (2) we get

$$P(Y_{n+1} = i | Y_n = j) = \sum_{q=\max(i,j)}^{\infty} P(B_n = q - i | Q_n = q) \cdot P(A_n = q - j), \quad i \geq q. \quad (34)$$

From the transition probabilities in (34), we determine the stationary distribution Y of Y_n . In the case study, we obtain an approximation for the stationary distribution. To this end, we bound the maximal value of Y_n with some large number N so that $P[Y \geq N]$ is sufficiently close to zero. Then, the stationary distribution for the bounded chain is computed by numerically solving the balance equations. Finally, we approximate Y with the stationary distribution of the bounded Markov chain, and use $P[Y_n = k] = 0$ when $k \geq N$.

Now our goal is to derive the stationary waiting times. To this end, denote by $P(W \in \mathcal{A} | Y = i)$ the stationary probability that waiting time of an arriving patient is the number in a set $\mathcal{A} \subset \{0, 1, \dots\}$, provided that there were $i = 0, 1, \dots$ waiting patients at the beginning of the time slot of the arrival. Consider the sequence of time periods n such that $[Y_n = i]$. The distribution of the waiting times of the A_n patients arriving in such time period is completely defined by Y_n . Thus, given Y_n , these waiting times are independent of the waiting times of the patients arriving in the other time periods. Denote by $A_n^{(k)}$ the number of patients that have arrived in time period n and have to wait k time slots before being scheduled. Then using the renewal reward argument, we write

$$\begin{aligned} P(W = k | Y = i) &= \frac{\mathbb{E}[A_n^{(k)} | Y_n = i]}{\mathbb{E}[A_n | Y_n = i]} = \frac{\mathbb{E}[A_n^{(k)} | Y_n = i]}{\mathbb{E}[A_n]} \\ &= \frac{\mathbb{E}[A_n^{(k)} | Y_n = i]}{\lambda/m}. \end{aligned} \quad (35)$$

Let us now define the probability $P(W \geq 1 | Y = i)$. From (35,) it follows that

$$P(W \geq 1 | Y = i) = \frac{\mathbb{E}[A_n - A_n^{(0)} | Y_n = i]}{\lambda/m}, \quad (36)$$

where for the numerator, we write

$$\begin{aligned} \mathbb{E}[A_n - A_n^{(0)} | Y_n = i, A_n = k_n] &= \sum_{b_n=0}^{i+k_n} P(B_n = b_n | Q_n = i + k_n) \\ &\times \mathbb{E}[A_n - A_n^{(0)} | Y_n = i, A_n = k_n, B_n = b_n], \end{aligned} \quad (37)$$

and for the last component, above holds:

$$\mathbb{E}[A_n - A_n^{(0)} | Y_n = i, A_n = k_n, B_n = b_n] = \min \{k_n, i + k_n - b_n\}. \quad (38)$$

Using (36)–(38), the probability $P(W \geq 1 | Y = i)$ can be directly computed.

Similarly, we can write the expression for $P(W \geq 2 | Y = i)$. Note that sometimes a patient has to wait longer because there are not enough patients on the list to form a batch of a minimal size. Thus, the waiting times of the patients arriving in time slot n depend also on the arrivals in time slot $n + 1$. Specifically, we derive the following:

$$\begin{aligned} P(W \geq 2 | Y = i) &= \frac{m}{\lambda} \sum_{k_n=0}^{\infty} P(A_n = k_n) \sum_{b_n=0}^{i+k_n} P(B_n = b_n | Q_n = i + k_n) \\ &\times \mathbb{E}[A_n - A_n^{(0)} - A_n^{(1)} | Y = i, A_n = k_n, B_n = b_n], \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}[A_n - A_n^{(0)} - A_n^{(1)} | Y = i, A_n = k_n, B_n = b_n] &= \\ &\sum_{k_{n+1}=0}^{\infty} P(A_{n+1} = k_{n+1}) \\ &\times \sum_{b_{n+1}=0}^{i+k_n+k_{n+1}+b_n} P(B_{n+1} = b_{n+1} | Q_{n+1} = i + k_n + k_{n+1} + b_n) \\ &\times \mathbb{E}[A - A^{(0)} - A^{(1)} | Y = i, A_n = k_n, B_n = b_n, A_{n+1} \\ &= k_{n+1}, B_{n+1} = b_{n+1}], \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[A_n - A_n^{(0)} - A_n^{(1)} | Y = i, A_n = k_n, B_n = b_n, A_{n+1} = k_{n+1}, B_{n+1} \\ = b_{n+1}] &= \max \{0, \min \{k_n, i + k_n - b_n - b_{n+1}\}\}. \end{aligned}$$

In a similar fashion, we derive $P(W \geq k | Y = i)$ for $k = 3, 4, \dots$, and finally, $P(W = 0 | Y = i) = 1 - P(W \geq 1 | Y = i)$.

To complete the derivation, we assume that the system functions in a stationary regime, and we use the full probability formula, where the exchange of the limit and the summation is justified by the dominating convergence theorem:

$$P(W \geq k) = \sum_{i=0}^{\infty} P(Y = i) P(W \geq k | Y = i), \quad k = 0, 1, \dots$$