

Reduce fluctuations in capacity to improve the accessibility of radiotherapy treatment cost-effectively

P. E. Joustra · R. Kolfin · N. M. van Dijk ·
C. C. E. Koning · P. J. M. Bakker

Published online: 22 October 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract This paper is motivated by a case study to reduce the throughput times for radiotherapy treatment. The goal is to find a cost-effective way to meet future throughput targets. A combination of queuing theory and computer simulation was used. First, computer simulation to detect the bottleneck(s) in a multi-step radiotherapy process. Despite, the investment in an additional linear accelerator, the main bottleneck turned out to be the outpatient department (OPD). Next, based on queuing theory, waiting times were improved by reducing the fluctuations in the OPD capacity. Computer simulation was used again to quantify the effect on the total throughput time of a radiotherapy patient. The results showed a reduction in both access times as well as waiting times prior to the consecutive steps: the preparation phase and actual treatment. The paper concludes with practical suggestions on how to reduce the fluctuations in capacity, and seems of interest for other radiotherapy departments or other multi-step situations in a hospital.

Keywords Discrete event simulation · Queuing theory · Radiotherapy department · Variability · Waiting lists

P. E. Joustra (✉) · R. Kolfin · P. J. M. Bakker

Department of Quality Assurance and Process Innovation, Academic Medical Center, Meibergdreef 9, Room D01-319, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands
e-mail: P.E.Joustra@AMC.UvA.nl

N. M. van Dijk

Department of Economics and Business, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands

C. C. E. Koning

Radiotherapy Department, Academic Medical Center, Meibergdreef 9, 1100 DD Amsterdam, The Netherlands

1 Introduction

1.1 Motivation

In the last few years, hospitals in The Netherlands have been moving towards more demand-driven than supply-driven institutions. For radiotherapy departments, this led to nationally accepted throughput time targets: for example, 80% of subacute patients have to start treatment within 7 calendar days of referral.

Throughput time is defined as the time between the initial referral and the actual start of radiotherapy treatment. This throughput time consists of several interdependent phases, such as waiting time for the first consultation, preparation for radiotherapy, and the actual start of the treatment.

To ensure that throughput time targets will be met in the future, cost-effective optimization strategies are preferable to costly investment in additional resources. Therefore, the challenge is to improve the efficiency of current resources.

Previous attempts to reduce throughput times for the radiotherapy department of the Academic Medical Center (AMC) in Amsterdam have not succeeded in reaching the national throughput time targets. These attempts have focused primarily on managing capacity at the presumed bottleneck: namely, the linear accelerators. However, extending the operational time of the linear accelerators did not lead to significant improvements in throughput time.

1.2 Radiotherapy literature

Until now, the literature concerning radiotherapy waiting lists and efficiency improvements has focused mainly on one single step in the entire radiotherapy process. Thomas et al. (2001) described a method for calculating the spare capacity needed in the outpatient department (OPD) in order to manage access times.¹ Munro and Potter (1994) analyzed the waiting time between the first consultation and the first fraction of radiation. Several studies have focused on the final step of the radiotherapy process: the actual treatment. Thomas (2003) analyzed the waiting time before a patient started radiotherapy after completing the preparation phase. An estimate for measuring the future workload of linear accelerators has been given by Delaney et al. (1997), and Williams et al. (2007) performed a strategic capacity analysis to quantify the effect of recommended access rates and optimal dose fractionation for radiotherapy activity. Ogulata et al. (2009) developed a scheduling approach to optimize the use of the bottleneck treatment device and tested this scheduling approach in different circumstances with simulation.

Few articles described and assessed the entire process. Dickof et al. (2001) analyzed all steps of the radiotherapy process (the first consultation at the OPD, the preparation, and treatment on the linear accelerators) separately using three different models to forecast waiting times before each individual step and not the total

¹ In this article we use the term access time for the number of days a patient has to wait until the first appointment at the OPD.

throughput time. Kapamara et al. (2007) modeled the entire teletherapy process and focused especially on the preparation phase and the actual treatment.

In order to decide whether a specific patient has met the throughput time target, however, it is necessary to analyze all steps simultaneously. For example, a high access time can be compensated in the subsequent steps of the process, so that the patient can still meet the performance target.

1.3 Variability

As outlined, the challenge is to improve efficiency rather than to invest in additional resources. As is well-known from queuing theory, waiting times in queues generally increase by higher the variability in inter-arrival and process times (see “Appendix 1”). Basic queuing results that express the effect of these types of variability date back to the Pollaczek–Khintchine formula (Pollaczek 1957) and to Kingman (1962).

Litvak and Long (2000) described a variability-based methodology and identified *artificial* variability that unnecessarily increases the inefficiency. The basic concepts of arrival and service variability are frequently applied to health services.

Litvak et al. (2005) reduced the variability in the demand for nursing staff by smoothing the elective admissions. Lynk (1995) described that merged hospitals would likely face lower variability in demand relative to the mean, while Baker et al. (2004) explored the relationship between demand variability and hospital costs.

Elkhuizen et al. (2007) showed that a variability reduction in the duration of CT scans would lead to lower waiting times and a higher utilization. Strum et al. (2003) applied econometric methods to accurately predict surgery times to reduce the variability between scheduled and actual surgery times.

A third variability factor for waiting times to appear can be the fluctuations in capacity. Capacity variability is essentially different from the variability in process times; in OPDs fluctuations in process times lead to higher waiting times in the waiting room while fluctuations in capacity lead to higher access times.

For a single step process, queuing theory can be used to explain why less variability in capacity leads to reduced waiting/access times. But even for a single step process with variability in capacity only some limited expressions are available (e.g. Jaiswal 1968; Doshi 1986). In addition, Tian and Zang (2006) presented a more recent review about of priority queues. For a multi-step process as in the present case no analytic results at all seem to be available to express the effect of variability in capacity. Preferably, we would experiment with the variability in capacity directly to show the effect of more variability in capacity. However, because no queuing formula is available to support this type of experimentation, we used an indirect way to generate variability in capacity by applying a priority queuing model. Applying these formula to our case study shows that the average waiting time decreases with reduced fluctuations in capacity (for details see “Appendix 2”). However, to experiment with the variability in capacity in a multi-step process we developed a simulation model. With this simulation model we were able to quantify the effect of a reduction in variability in capacity for both access times as well as

waiting times prior to the consecutive steps: the preparation phase and actual treatment.

Surprisingly, we have not been able to find articles about the effect of reducing fluctuations in capacity in health care literature. In particular, the effect of a more stable capacity of the first step in a multi-step process on the total throughput times does not seem to be addressed in health care literature.

1.4 Objectives

The main objectives of the present study therefore were:

1. To determine which aspects of the process are primarily responsible for preventing the AMC radiotherapy department from meeting the throughput time targets.
2. To search for cost-effective solutions to reduce the throughput times and to provide quantitative support to compare alternative solutions.
3. To show that capacity variability reduction in the multi-step process can be highly effective.

2 Problem description

2.1 The AMC radiotherapy department

At the start of this study in May 2006, the AMC radiotherapy department was staffed by 9 radiation oncologists, 5 clinical physicians, and 30 technicians. The department had one simulator and one CT simulator for the preparation and three linear accelerators for the actual treatment. A fourth accelerator was scheduled for 2007. In 2006, 2,350 new patients had a first consultation and 7,200 patients were seen in follow-up. During the entire year, a total of around 32,500 radiation fractions were administered.

Table 1 contains the performance targets for the different patient groups set by the Dutch Society of Radiotherapy and Oncology. We did not include acute patients in our study, because these patients were seen outside regular consultation periods. Therefore, we focused primarily on subacute patients and those with other indications. In the first quarter of 2006, only 50% of the subacute patients started treatment within 7 calendar days of referral though the target is at least 80% (see Table 1). The performance for patients with other indications was even worse: 36% of the patients started treatment within 21 calendar days of referral.

2.2 The multi-step radiotherapy process

The study began with the process mapping of the entire radiotherapy treatment process. We observed the different steps in the process, held interviews with experts from the department, and developed a flowchart, which was validated by the experts.

Table 1 Performance targets for different types of patients

| Patient group | Patients within throughput time target (%) | Throughput time target (days) | Maximum waiting time (days) |
|-------------------|--|-------------------------------|-----------------------------|
| Acute | 100 | 1 | – |
| Subacute | 80 | 7 | 10 |
| Other indications | 80 | 21 | 28 |

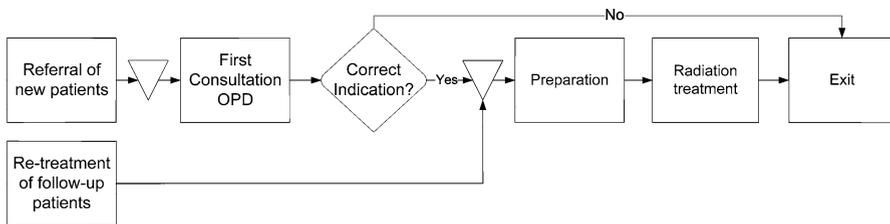
**Fig. 1** Sequential steps in the radiotherapy process

Figure 1 shows the primary flow within the radiotherapy department of the AMC. The sequential steps are: (1) the first consultation at the OPD performed by a radiation oncologist, (2) the preparation for the treatment, and (3) the actual treatment by use of a linear accelerator and performed by a technician.

All new patients require a first consultation and confirmation that radiotherapy is indicated before preparation for treatment and the actual treatment can start. Both types of patients use the same capacity for new patients at the OPD and are scheduled with the FCFS principle in 1-h-slots.

A certain percentage of follow-up patients receive re-treatment. Although this group is not relevant from an OPD perspective (because they use time slots dedicated to follow-up patients which are different from the time slots dedicated to new patients), they are important because they use the same capacity for the preparation and treatment process.

The OPD is an appointment-based system with a first-fit algorithm for patients. The earliest moment a patient can be seen is the day after referral.

Because the actual treatment has to start within a limited number of days after the preparation is finished, the start of the preparation process depends on the available capacity of the linear accelerators. Another point is that due to the treatment offered, all treatment sessions have to be scheduled on consecutive days. The only interruption allowed in the treatment is during weekends when the whole department is closed. This also implies that treatment cannot start on Friday, because then the gap between the first and the second fraction is considered to be too long.

2.3 Capacity analysis

A capacity analysis was performed to find the bottleneck(s) in different steps of the process. We analyzed the data concerning referrals, the capacity of the OPD, the

Table 2 Overview of data

| Description of the data | Value |
|---|------------------------|
| Average arrival rate of new patients | 32.5 per week |
| Number of new consultations | 25–42 per week |
| Average waiting time between the preparation phase and the actual treatment for subacute patients | 3 days |
| Average waiting time between the preparation phase and the actual treatment for patients with other indications | 12 days |
| Preparation time of subacute patients | 2 days |
| Preparation time of patients with other indications | Lognormal (10, 3) days |
| Availability per linear accelerator | 40 h per week |
| Average number of patients requiring treatment | 29.5 per week |
| Required number of fractions per patient | 1–35 |
| Duration of sessions | 10, 20 or 30 min |

duration of the preparation phase, the number of fractions needed, the duration of a treatment session, and the capacity of the linear accelerators (see Table 2 for an overview of the data). We used this information to calculate the utilization of each separate step.

The average access time for subacute patients and patients with other indications was 6 and 7 calendar days, respectively. The utilization of the OPD was high: an average of 32.5 new patients arrived per week and an average of 33.5 new patients could be seen. There were substantial fluctuations in demand and capacity: the number of referrals per week fitted to a poisson distribution and the weekly capacity for new patients in the OPD (as based upon real life data from the whole of 2006) fitted to a discrete uniform distribution with range 25–42 consultations. Subsequently, the weekly number of 1-h-slots is randomly selected from this uniform distribution. With a utilization of 97% and fluctuations in both demand and capacity, high access times were likely.

To verify that the preparation phase was not a bottleneck in the treatment process, we checked whether the capacity of the simulator and CT simulator were exceeded during the simulation. Other preparation phase activities—for example, determining the exact area to be treated and the correct dose fractionation—are performed by radiation oncologists and/or technicians and were not considered to be a bottleneck.

The total duration of the preparation of patients with other indications fitted a lognormal distribution (with a mean of 10 calendar days and 3 days standard deviation). For subacute patients, preparation was always finished within 2 days. The reason the preparation phase was short for subacute patients is that all these patients receive palliative treatment (for example, patients with bone metastasis, who are in severe pain) requiring only one or two radiation fractions, and the treatment planning is relatively simple. Also, the tumor is often superficial, so determining the exact area to be treated is less time consuming.

The three linear accelerators were available as scheduled (on average, 120 h per week). The data showed that breakdowns rarely occur and regular maintenance of the machines (which are interchangeable) did not influence availability, so we assumed that both aspects have only a negligible effect on waiting time.

To determine the utilization for the linear accelerators, we divided the total demand by the scheduled availability of the linear accelerators. Including the follow-up patients who needed re-treatment, the average number of patients requiring starting treatment is 29.5 per week. Depending on the protocol, a patient needs 1–35 fractions of 10, 20 or 30 min per session; on average, a patient needs 224 min of radiotherapy treatment. This results in a total weekly demand of approximately 110 h, which leads to a utilization of 92%. The measured average waiting time between the preparation phase and the actual treatment was 3 days for subacute patients and almost 12 days for patients with other indications.

3 Modeling

As will be outlined in more detail below, a combination of methods was used:

- Queuing theory to obtain insights into the underlying reasons for waiting times to arise.
- Computer simulation to (1) study these insights, (2) detect potential bottlenecks, and (3) quantify the effect of adjustments.

3.1 Queuing theory

Queuing theory is an established mathematical discipline to explain and compute waiting times in service processes (for example, within health care see Green et al. 2006), as based upon mathematical expressions (for example, “Appendices 1 and 2”). For the strongly simplified situation of a single server queue, this will briefly be clarified in the appendices. Nevertheless, for the complicated multi-step radiotherapy process, the same insights are still applicable. However, as exact formulas for a multi-step situation are no longer available, this has to be investigated by computer simulation.

3.2 Computer simulation

According to Shannon (1975), computer simulation is the process of designing a model of a real system and conducting experiments with this model on a computer to gain insight into the system or to evaluate different strategies by means of numeric results and animation. Multiple types of computer simulation are used to analyze health care systems (Brailsford et al. 2009). More specifically, to reduce waiting times for radiotherapy several articles have described the use of Monte Carlo simulation (Thomas et al. 2001; Munro and Potter 1994; Thomas 2003).

In discrete event simulation, the operation of a system is represented as a chronological sequence of events. Each event occurs at a particular moment and

marks a change of state in the system. This means that time evolution is incorporated in a natural way. In contrast to discrete event simulation, the passage of time does not play a substantive role with Monte Carlo simulation. Therefore, Monte Carlo simulation is better suited to a waiting time analysis of the individual steps in the radiotherapy process than to analyzing the entire multi-step radiotherapy process. Monte Carlo has its limitations if quantitative analysis of the whole process is necessary. This is illustrated by Dickof et al. (2001), who used three separate models to perform a quantitative analysis for all steps individually; however, all conclusions drawn for the entire process are of a qualitative nature. In order to analyze the total throughput time for a patient, we had to incorporate the complete process into a single model. Therefore, discrete event simulation is best suited to model this throughput time problem.

The simulation model was developed using MedModel simulation software Version 7 by ProModel Corporation. This simulation model had to be validated to ensure that it would deliver accurate waiting time predictions for future scenarios. Validation is the process of determining whether a model is an accurate representation of the system for the particular objectives of the study (Fishman and Kiviat 1968). The simulation model was validated successfully for both types of patients by comparing the results of the model with the actual data from current practice. The differences with respect to the average and the 80th percentile of the waiting times were less than 10%.

After the input of the simulation model has been determined, the output gives an average percentage and confidence interval of patients with a waiting time less than or equal to the critical number of days (e.g., 7 days for subacute patients). To decide whether the throughput time target has been met, the lower bound of the 95% confidence interval must exceed the required percentage (in our case, 80% for both types).

4 Results

The simulation model was developed to obtain quantitative support and to compare different scenarios.

As shown in Fig. 1, the radiotherapy process has three sequential steps: (1) the first consultation at the OPD, (2) the preparation, and (3) the actual treatment by use of a linear accelerator. For the different scenarios, OPD capacity and the number of linear accelerators could be altered.

4.1 Extra linear accelerator

The radiotherapy department had already planned an extra linear accelerator to remove the presumed bottleneck in the process and to make sure performance targets would be met in the future. Therefore, we started the experimentation by adding a fourth linear accelerator without making any other adjustments to the validated model of the current situation. With four linear accelerators, fewer than 60% of the subacute patients started actual treatment within 7 days (see Table 3).

Table 3 Results of the various scenarios

| Scenario | Average weekly number of consultations and the distribution | Number of linear accelerators | Variability in OPD capacity | Subacute patients within 7 days (%) | Patients with other indications within 21 days (%) |
|----------|---|-------------------------------|-----------------------------|-------------------------------------|--|
| 1 | Current situation | 3 | Current | 44 | 39 |
| 2 | Extra linear accelerator | 4 | Current | 58 | 73 |
| 3 | Extra OPD capacity | 3 | Current | 84 | 61 |
| 4 | Extra OPD capacity | 4 | Current | 79 | 78 |
| 5 | Extra OPD capacity | 4 | Current | 85 | 92 |
| 6 | Variability reduction in OPD capacity | 4 | Half of current range | 86 | 90 |
| 7 | Variability reduction in OPD capacity | 4 | 50% chance of 33 or 34 | 80 | 90 |

Although performance for patients with other indications was better (73% start treatment within 21 days), the targets were not being met.

With the additional linear accelerator, the capacity in the last process step seemed to be sufficient so we did not try to optimize the planning of the linear accelerators. Instead, we focused on the first and second step in the radiotherapy process.

For patients with other indications the preparation time was relatively long (10 days on average) compared with the throughput time target of 21 days. However, because the preparation of subacute patients took less than 2 days, a reduction in preparation time was only feasible for patients with other indications. Nevertheless, a possible reduction for these patients will not lead to less pressure on OPD capacity, because the subacute patients are leading with regard to OPD capacity to reach its more stringent performance target. Only when both types of patients no longer share the same capacity in the OPD will reducing preparation time for patients with other indications lead to less pressure on OPD capacity.

Therefore, we focused our attention on the OPD first.

4.2 Extra OPD capacity

A possible alternative solution to meet the performance targets was to increase OPD capacity. In the original situation with three linear accelerators, 36 consultations per week were required to meet the throughput time target for the subacute patients. Unfortunately, the performance for patients with other indications was still not sufficient. Therefore, either the preparation time for patients with other indications should be reduced or the number of linear accelerators should be increased to four. Although a significant reduction of the preparation time for patients with other indications might be possible, we only explored the expansion of the number of linear accelerators because the department had already planned an extra linear accelerator.

Therefore, the scenarios following were simulated with four linear accelerators. A weekly capacity of 34 consultations was not sufficient to meet the performance targets. The throughput time targets for both types of patients were met with 35 consultations per week.

4.3 Variability reduction in OPD capacity

Increasing OPD capacity in combination with the four linear accelerators will ensure the performance targets are met in the future. Unfortunately, this is not a cost-effective solution. An alternative is to apply the variability reduction principle to OPD capacity. Because in practice there were substantial fluctuations in OPD capacity (between 25 and 42 consultations per week), which can be regarded as artificial variability (Litvak and Long 2000), we also experimented with a more stable capacity so that demand before preparation and actual treatment would be more evenly distributed (see Fig. 2).

Simulation showed that by reducing the variance of OPD capacity for new patients to half of the current range, the throughput time targets will be met with only 34 consultations per week. Thus, the spare capacity needed to reach the

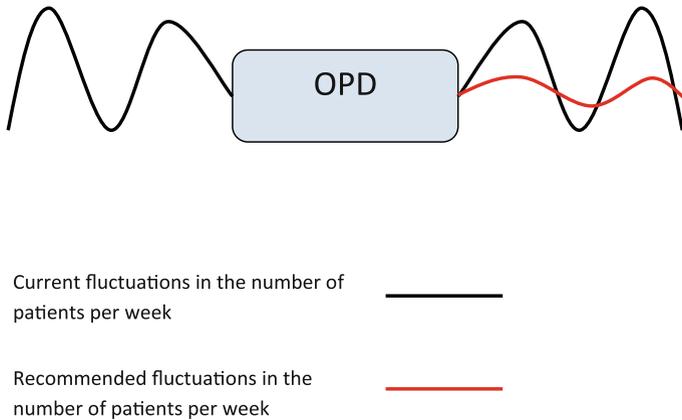


Fig. 2 Level of fluctuations in the number of patients per week

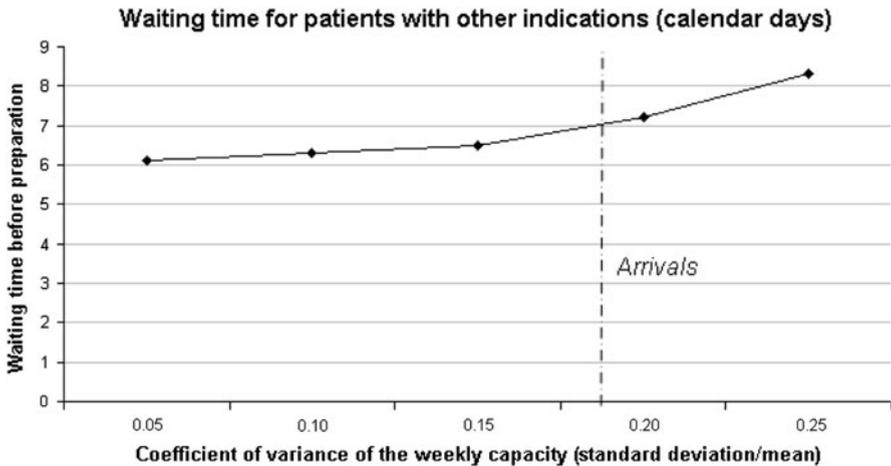


Fig. 3 Effect of fluctuations in OPD capacity on the waiting time before preparation

throughput time targets with a relatively stable number of first consultations per week is 40% less (1.5 consultations in stead of 2.5 consultations per week) than with the current, highly fluctuating capacity in the OPD.

If the capacity was even more stable, an average weekly capacity of 33.5 consultations would be sufficient to meet the throughput time targets. In this situation, each week would have an equal chance of 33 or 34 consultations.

With the current fluctuations in capacity, the variance in the number of patients per week was even greater after the OPD than before. Figure 3 shows that large fluctuations in the weekly number of consultations could add as much as 2 days to the average waiting time before preparation.

4.4 Remark

We applied the variability reduction principle to OPD capacity and did not discuss the application of the principle to the demand and duration of a first consultation. The weekly number of new patients was considered to be an external factor that could not be influenced. Additionally, fluctuations in the actual duration of a consultation were small and no sub-activities were found that could be reallocated.

5 Discussion

The first objective of our study was to analyze which aspects of the process were responsible for preventing the department from meeting the throughput time targets. With the scheduled extra linear accelerator, the radiotherapy department expected that the throughput time targets would be met in the future, but this was not the case. The simulation model indicated that OPD capacity was an important additional reason that the targets were not met.

In the original situation with three linear accelerators, 36 consultations per week were required to meet the throughput time target for the subacute patients. Unfortunately, the performance for patients with other indications was not sufficient. Therefore, either the preparation time for patients with other indications should be reduced or the number of linear accelerators should be increased to four. Although a significant reduction of the preparation time for patients with other indications might be possible, we only explored the expansion of the number of linear accelerators because the department had already planned an extra linear accelerator. With four linear accelerators and 35 consultations per week, the throughput time targets were met for both types of patients.

A cost-effective alternative for increasing OPD capacity is to reduce the substantial fluctuations in this capacity, for example reducing the fluctuations in capacity by 50% will save one consultation per week. Our results demonstrate that a reduction of this artificial variability (Litvak and Long 2000) will lead to lower throughput times, because in addition to lower access times, the waiting times before the subsequent steps in the process, namely the preparation and actual treatment, will decrease as well. An explanation for this phenomenon is that if the utilization rate of the OPD is high, a more stable OPD capacity will decrease the fluctuations in the weekly number of patients arriving before the preparation and actual treatment. As by Kingman (1962) this will lead to lower waiting times before the preparation and actual treatment (see “Appendix 1” for the formulas).

5.1 How to reduce variability in OPD capacity in practice?

The reason for such substantial fluctuations in OPD capacity was the department’s reactive strategy. In practice, the number of first consultations for new patients in the OPD was only increased when the waiting time was far above the target level for a considerable period. To reduce the increased access times, capacity was raised substantially for a short period. This resulted in a high variability in OPD capacity.

To be able to act proactively instead of reactively, an overview of the total number of consultations for new patients per week for the coming quarter has been introduced. All national holidays, physicians' vacations, and otherwise cancelled consultations are incorporated into the overview so that management can see in advance when OPD capacity will be too low. Because this information is available a few months before the week in question, management has enough time to compensate for the anticipated temporary insufficient capacity.

Some suggestions are: a national holiday should be compensated directly during the other days of that specific week. To spread the workload better, the follow-up patients can be scheduled in the week before and after a week with a national holiday, so that doctors have more time to see new patients. Management can also ensure doctors spread their own holidays more evenly to prevent sharp declines in capacity. By introducing a backup consultation per doctor, the capacity can be increased easily if the available, future capacity was expected to drop below a certain threshold value.

5.2 Conclusion

A combination of queuing theory and computer simulation was used. Computer simulation to find the bottlenecks in the multi-step radiotherapy process and to quantitatively compare the alternative solutions to reduce the throughput times efficiently. Queuing theory to provide the insight of variability and some supportive formula.

Despite, the investment in an additional linear accelerator to remove the presumed bottleneck, the throughput time targets were still not met. The simulation model indicated that the capacity of the OPD had to be increased. A combination of increased OPD capacity and a reduction of the preparation time for patients with other indications, would probably be sufficient to meet the performance targets for both types of patients. Although this suggests that the additional linear accelerator might not have been necessary after all, we continued our experimentation with an fourth linear accelerators as it was already purchased.

Furthermore, to include the dynamic interaction between the different steps of the process, all three steps in the radiotherapy process were incorporated into one discrete event simulation model. This enabled us to quantify the effect of the substantial fluctuations in OPD capacity on both access times in the OPD and waiting times before preparation and actual treatment. Our radiotherapy case study has shown that a lower variability in the weekly number of consultations in the OPD did result in a considerable reduction in throughput times. Variability reduction in OPD capacity thus appeared more cost-effective than just adding more capacity to ensure the performance targets would be met in the future.

This new insight seems to be of interest for other radiotherapy departments as well as to other situations with dynamic interaction between the consecutive steps.

Acknowledgments The authors are grateful to Mr. B. Marijt, Mr. M. Van der Ouderaa, and Dr. M. Hulshof of the radiotherapy department at the AMC for their collaboration during the project. The authors would also like to thank Ms. D. Constant of the Team Innovation and Process Management at the AMC for her contribution to the study. Finally, the authors are grateful to the anonymous reviewers for their constructive comments which have been beneficial for this version of the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1: Queuing formula on variability in demand and process times

In this appendix the queuing formula to explain the relationship between waiting times and variability in demand (and process times) is described.

Step 1: Variability in service times

The Pollaczek–Khintchine formula (Pollaczek 1957) expresses that the average waiting time of a system with fluctuations in the service times (general distribution) is a function of the variation coefficient of the service times (the standard deviation divided by the mean) and the average waiting time with fixed service times. This formula is scientifically correct for a single capacity unit.

$$W_{M|G|1} = (1 + c_s^2) \times W_{M|D|1} \quad (1)$$

with c_s the variation coefficient of the service time distribution.

This formula proves that a lower variance in the service time distribution leads to a reduction in the average waiting time. The average waiting time of a system with exponential service times ($c_s = 1$) is twice the average waiting time with fixed service times ($c_s = 0$).

Step 2: Variability in arrivals

Kingman (1962) expanded the Pollaczek–Khintchine formula as an approximation of the average waiting time for a system with generally distributed service times and generally distributed interarrival times. He assumes that the randomness of the interarrival process has the same effect on the mean waiting time as does the randomness in the service times.

$$W_{G|G|1} \approx (c_a^2 + c_s^2) \times W_{M|D|1} \quad (2)$$

with c_a the variation coefficient of the interarrival time.

Again, a reduction in the fluctuations of the number of arrivals and/or the service times will lead to a decrease in waiting times.

Note that with exponentially distributed interarrival times, Kingman's formula decreases to the Pollaczek–Khintchine ($c_a = 1$).

Appendix 2: Queuing formula on variability in capacity

In this appendix we use queuing theory to quantitatively explain why the access times decrease if the fluctuations in OPD capacity are reduced. Note that capacity variability is essentially different from the variability in process times; in hospitals fluctuations in process times lead to higher waiting times in the waiting room while fluctuations in capacity lead to higher access times.

Preferably, we would experiment with the distribution of the capacity directly to show the effect of more variability in capacity. However, because no queuing

formula is available to support this type of experimentation, we used an indirect way to generate variability in capacity by a priority queuing example.

Priority queuing is a specific category of queuing models (e.g. Jaiswal 1968; Doshi 1986). A priority queuing system has multiple queues that contain jobs of a certain priority and all types of jobs are served by the same capacity units. Jobs with higher priority are served first, which makes priority queuing suitable for modeling resource vacations and breakdowns. The priority model used by D’Antone (2000) is suitable for modeling fixed process times and fluctuating capacity simultaneously, as is the case in our radiotherapy OPD. The vacations, breakdowns, or cancellations are modeled as jobs with high priority and the patients as jobs with low priority. Both types of jobs are handled by the same single server.

In the practical situation, the maximum weekly capacity was 42 consultations. Simulation shows that on average 35 consultations are necessary to meet the throughput time targets. This indicates that on average 7 consultations are cancelled each week, which can also be seen as vacations or breakdowns. The average waiting time for patients (W_2) can be calculated by formulas:

$$W_i = \frac{W_0}{(1 - \sigma_i)(1 - \sigma_{i-1})} \tag{3}$$

with

$$W_0 = 1/2 \sum_{i=1}^N \lambda_i s_i^2 \tag{4}$$

In these equations, N is the number of priority classes, λ_i represents the arrival rates of jobs with priority i , s_i represents the mean service time $1/\mu_i$ of jobs with priority i , and $\sigma_i = \sum_{k=1}^i \rho_k$ is the utilization of the server by customers of priority $1 - i$.

We will apply these formulas to evaluate three scenarios with an equal effective utilization for patients but with varying level of fluctuations in OPD capacity. A higher number of cancelled consultations (λ_1) results in more fluctuations in the effective number of consultations for patients. Note that ρ_2 is not the effective utilization for patients, because the effective capacity for patients is 35 consultation in all scenarios (and thus the utilization is $32.5/35 = 0.93$ in all three scenarios).

- A. The original situation with 7 cancelled consultation per week results in an average waiting time of 0.23 weeks for patients (see Table 4).
- B. If the weekly number of cancelled consultations is halved, the average waiting time for patients is almost 9% lower than in the original situation.

Table 4 Average waiting time for patients with varying level of fluctuations in OPD capacity

| Scenario | N | λ_1 | λ_2 | ρ_1 | ρ_2 | $s_1 = s_2$ | W_2 |
|----------|-----|-------------|-------------|----------|-----------|-------------|-------|
| A | 2 | 7 | 32.5 | 7/42 | 32.5/42 | 1/42 | 0.23 |
| B | 2 | 3.5 | 32.5 | 3.5/38.5 | 32.5/38.5 | 1/38.5 | 0.21 |
| C | 2 | 0 | 32.5 | 0 | 32.5/35 | 1/35 | 0.19 |

- C. For the situation without any cancellations, D'Antone's formula can be applied with $\lambda_1 = 0$, but the standard queuing formula for an M/D/1 system is also suitable. Compared to the original situation, the reduction in average waiting time is 18%.

These queuing results show that reducing fluctuations in OPD capacity also reduces the access times.

References

- Baker LC, Phibbs CS, Guarino C, Supina D, Reynolds JL (2004) Within-year variation in hospital utilization and its implications for hospital costs. *J Health Econ* 23:191–211
- Brailsford SC, Harper PR, Patel B, Pitt M (2009) An analysis of the academic literature on simulation and modelling in health care. *J Simul* 3:130–140
- D'Antone I (2000) Pipeline length optimization in a digital asynchronous trigger controller. *Nucl Instrum Method Phys Res Sect A* 451:439–448
- Delaney GP, GebSKI V, Lunn AD, Lunn M, Rus M, Manderson C, Longlands AO (1997) Basis treatment equivalent (BTE): a new measure of linear accelerator workload. *Clin Oncol* 9:234–239
- Dickof P, Firth A, Foord C, Lusk V (2001) Managing radiation therapy queues. *Curr Oncol* 8:125–149
- Doshi BT (1986) Queuing system with vacation—a survey. *QUESTA* 1:29–66
- Elkhuizen SG, Van Sambeek JRC, Hans EW, Krabbendam JJ, Bakker PJM (2007) Applying the variety reduction principle to management of ancillary services. *Health Care Manag Rev* 32(1):37–45
- Fishman GS, Kiviat PJ (1968) The statistics of discrete-event simulation. *Simulation* 10:185–195
- Green LV, Soares J, Giglio JF, Green RA (2006) Using queuing theory to increase the effectiveness of physician staffing in the emergency department. *Acad Emerg Med* 13(1):61–68
- Jaiswal N (1968) Priority queues. Academic Press, New York
- Kapamara T, Sheibani K, Petrovic D, Haas O, Reeves C (2007) A simulation of a radiotherapy treatment system: a case study of a local cancer centre. ORP3 meeting
- Kingman JFC (1962) Some inequalities for the queue GIG/1. *Biometrika* 49:315–324
- Litvak E, Long MC (2000) Cost and quality under managed care: irreconcilable differences? *Am J Manag Care* 6(3):305–312
- Litvak E, Buerhaus PI, Davidoff F, Long MC, McManus ML, Berwick DM (2005) Managing unnecessary variability in patient demand to reduce nursing stress and improve patient safety. *Jt Comm J Qual Patient Saf* 31:330–338
- Lynk WJ (1995) The creation of economic efficiencies in hospital mergers. *J Health Econ* 14:507–530
- Munro AJ, Potter S (1994) Waiting times for radiotherapy treatment: not all that mysterious and certainly preventable. *Clin Oncol* 6:314–318
- Ogulata SN, Cetik MO, Koyuncu E, Koyuncu M (2009) A simulation approach for scheduling patients in the department of radiation oncology. *J Med Syst* 33:233–239
- Pollaczek F (1957) Problèmes stochastiques posés par le phénomène de formation d'une queue d'attente à un guichet et par des phénomènes apparentés. Gauthier Villars, Paris
- Shannon RE (1975) Systems simulation: the art and science. Prentice-Hall, Englewood Cliffs
- Strum DP, May JH, Sampson AR, Vargas LG, Spangler WE (2003) Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. *Anesthesiology* 98:232–240
- Thomas SJ (2003) Capacity and demand models for radiotherapy treatment machines. *Clin Oncol* 15:353–358
- Thomas SJ, Williams MV, Burnet NG, Baker CR (2001) How much surplus capacity is required to maintain low waiting times? *Clin Oncol* 13:23–28
- Tian N, Zhang ZG (2006) Vacation queuing models: theory and applications. Springer, Berlin
- Williams MV, Summers ET, Drinkwater K, Barrett A (2007) Radiotherapy dose fractionation, access and waiting times in the countries of the UK in 2005. *Clin Oncol* 19:273–286

Author Biographies

Paul Joustra (1975) is a project leader at the Academic Medical Centre in Amsterdam, The Netherlands. He is responsible for projects regarding patient logistics, including the radiotherapy project. Recently, he successfully defended his thesis “How to deal with fluctuations in hospital processes to improve accessibility” and became a PhD.

Rachid Kolfin (1981) studied Industrial Engineering and Management at the University of Twente in Enschede, The Netherlands. During his Master of Science he mainly focused on Operations Research. After finishing his master thesis Rachid became a logistic consultant for the Academic Medical Centre (AMC) in Amsterdam. At the AMC Rachid implemented discrete event simulation and other Operations Research methods to analyze and optimize the patient flows. After two years at the AMC Rachid went on to work at Talumis as a simulation consultant.

Prof. Nico M. van Dijk is responsible for the Operations Research and Management Program at the University of Amsterdam. He has a strong research interest in the area of stochastic OR, particularly Queuing. He also aims to popularize the potential of OR such as by articles in magazines or newspapers. He has been involved in a variety of practical projects, among which for the Dutch railways, the Dutch airport Schiphol, the Dutch Triple A, the Dutch Ministry of Health, the Dutch blood banks and various hospitals.

Prof. Dr. Carolina C. E. Koning (1946) had her Education in the Medical study University of Leiden from the period of 1965–1973. From 2002, She is the professor of the radiation Oncology Department AMC/UvA Amsterdam. She has served for several national guideline committees for melanoma, lung cancer, and cryopreservation of ovarian tissue and for several committees, She is/was a Member of several national Cancer Society Projects and committees (Dutch Cancer Society), Member of four committees of the Netherlands Health Council, Editorial board NTvG, NTvO and Oncologisch Vademecum.

Prof. Dr. Piet Bakker (1948) Medical Oncologist, Internist. Professor and Director of Quality of Care and Process Innovation PhD thesis 1992: “Development and application of methods for the measurement of cell kinetics and cell reproductive death in tumours”. Piet Bakker is a chaired professor in the Quality of Care Research and Process Innovation and head of the department KPI in the AMC and UvA. He was trained as an internist at the department of Internal Medicine at Binnengasthuis (University of Amsterdam). He later trained as a medical oncologist. From 1981 till 2010 he was working as a medical oncologist at the Academic Medical Centre (Amsterdam). As such he has a large experience in patient care, education, and research. He is a member of several oncology societies.