Resource Optimisation for Cancer Pathways with Aggregate Diagnostic Demand: a Perishable Inventory Approach

Edilson F. Arruda*[1,2], Paul Harper[1], Tracey England[1], Daniel Gartner[1], Emma Aspland[1], Fabrício O. Ourique[3], and Tom Crosby[4]

[1]School of Mathematics, Cardiff University, Senghennydd Rd, Cardiff CF24 4AG, UK.
[2]Alberto Luiz Coimbra Institute- Graduate School and Research in Engineering, Federal University of Rio de Janeiro, Rio de Janeiro RJ, Brazil.
[3]Federal University of Santa Catarina, Araranguá SC, Brazil.
[4]Velindre Cancer Centre, Cardiff CF14 2TL, UK.

May 21, 2020

Abstract

This work proposes a novel framework for planning the capacity of diagnostic tests in cancer pathways that considers the aggregate demand of referrals from multiple cancer specialties (sites). The framework includes an analytic tool that recursively assesses the overall daily demand for each diagnostic test and considers general distributions for both the incoming cancer referrals and the number of required specific tests for any given patient. By disaggregating the problem with respect to each diagnostic test, we are able to model the system as a perishable inventory problem that can be solved by means of generalised G/D/C queuing models, where the capacity C is allowed to vary and can be seen as a random variable that is adjusted according to prescribed performance measures. The approach aims to provide public health and cancer services with recommendations to align capacity and demand for cancer diagnostic tests effectively and efficiently. Our case study illustrates the applicability of our methods on lung cancer referrals from UK's National Health Service. Healthcare Modelling, Capacity Planning, Inventory Control, Queuing Systems

*Corresponding Author: ArrudaEF@cardiff.ac.uk, efarruda@po.coppe.ufrj.br

1 Introduction

Increasing demand on healthcare services caused by an ageing population paired with frequent budget constraints leads to a constant need to optimise healthcare resources (Capan et al., 2017). In particular, the demand for cancer care services in the United Kingdom (UK) has been constantly rising over the last 20 years (Saville et al., 2019). That, in turn, gave rise to capacity issues that have been identified by multiple government bodies (e.g., Welsh Government, 2019; South Yorkshire, Bassetlaw & North Derbyshire Cancer Alliance, 2019; Scottish Government, 2017; Meskarian et al., 2017).

The literature presents a number of alternatives for improving the management of demand and capacity within healthcare services. For example, a proper management of demand when a number of competing services are made available can be used to promote a better usage of the existing capacity, while also improving the quality of the service (e.g., Scottish Government, 2017; Arruda et al., 2019). In contrast, scheduling and capacity allocation models can also be implemented to promote an optimised use of the available capacity within a prescribed planning horizon (e.g., Culpan et al., 2019; Woznitza et al., 2018). Reviews in (Capan et al., 2017; Marynissen and Demeulemeester, 2019; Hulshof et al., 2012) provide an overview of analytic and simulation tools for decision making in healthcare systems in general. For more in-depth analysis of the literature regarding cancer care and chemotherapy management, see Saville et al. (2019) and Shi et al. (2014). In this paper we study a capacity planning problem for cancer diagnostic services considering the steady state demand for diagnostic tests incoming from all available cancer specialties, which are also referred to as cancer sites.

The complex nature of healthcare services gives rise to a complex supply chain that can be viewed as a network of interacting services and supplies (Martins et al., 2019). Martins et al. (2019) discuss the importance of considering the nature of these interactions in healthcare models, whilst also arguing that such models often lack a "networks perspective". In order to capture the interaction of services and supplies, some authors opt to model the flow of services that are part of the scope of their studies. This gives rise to complex analytic tools that often resort to unrealistic simplifying assumptions, or to simulation tools that are only able to compare a small number of alternatives (Saville et al., 2019; Alagoz et al., 2011). The related literature includes a model of network community services in Canada (Bidhandi et al., 2019) that uses classical M/M/1 queuing models to estimate the probability of a delayed service. For more details of queueing models and a definition of what is meant by an M/M/1 queue, see (Shortle et al., 2018). A similar approach is applied by Wu et al. (2019) for multi-stage bed allocation in hospitals. In contrast, Nguyen et al. (2018) propose a deterministic approximation for a capacity allocation problem applied to an outpatient clinic, whereas Xiao et al. (2018) develop a deterministic model for scheduling rehabilitation services. Deterministic methods for scheduling and capacity allocation in healthcare also appear in the works of Gartner et al. (2018), Nguyen et al. (2015) and Hulshof et al. (2013).

Related works also include material logistics applications in healthcare, which were recently reviewed by Ahmadi et al. (2019). These applications encompass inventory models with restrictive assumptions on the demand distribution (Ahmadi et al., 2019; Rosales et al., 2015), which is often assumed to follow a Poisson process (Rosales et al., 2015). Another possible drawback of such models is that they generally assume exogenous demand for supplies, an assumption that is violated in practice when the demand is indirect, such as the demand for surgical supplies or diagnostic tests. While patient arrival is generally exogenous, the demand for surgical supplies or diagnostic tests is endogenous and is generated by patients that were previously admitted in the system. This typically generates a large variety of patient pathways. In such a context, Garg et al. (2008) applied Markov models to search for frequent and/or costly pathways in healthcare data. Finally, a recent study by Richers et al. (2019) accounts for indirect demand in surgical supplies inventory.

This paper addresses the problem of capacity planning for diagnostic tests in the cancer pathway, considering the incoming demand for various cancer specialties. At the operational level, a patient typically requires multiple services and appointments since being referred to a cancer pathway (Sauré et al., 2012; Aspland et al., 2019). While multi-appointment scheduling problems are reviewed by Marynissen and Demeulemeester (2019), more specifically, Romero et al. (2013) developed a simulation model for the treatment phase of skin cancer. A spreadsheet simulation tool for a more general diagnostic and treatment unit was introduced by Bowers et al. (2005), whereas Bikker et al. (2015) chose to optimise the allocation of consultant doctor's activities to accelerate the access to radiotherapy treatments. Refer to Aspland et al. (2019) for a comprehensive survey of clinical pathway modelling. Other applications of operational research to cancer care are reported by Saville et al. (2019). They identified a gap in the literature regarding optimisation methods for cancer diagnosis and staging. In the present paper we seek to contribute to bridging this gap by addressing the planning stage of the diagnostic and staging phases of a cancer pathway.

This study proposes an innovative analytic tool for capacity planning that makes no assumption on the distribution of the cancer referrals, nor on the distribution of the diagnostic and staging tests required by a referred patient. In order to keep this level of generality, we avoid a direct modelling of the pathway, which has previously given rise to involved models that are very difficult to solve (e.g., Sauré et al., 2012; Castro and Petrovic, 2012). Instead, we exploit the problem's structure to simplify the model by disaggregating it with respect to each diagnostic test. The rationale is somewhat similar to the demand disaggregation in (Suárez-Vega et al., 2017). It also bears some similarities to agent-based approaches in which the agents examine the state of the system and take decisions (e.g., Fuller et al., 2019).

By taking into account the aggregated demand for diagnostic tests generated by each cancer specialty, we are able to produce a simple and easy to use analytic tool that captures the essential characteristics of the problem whilst also enabling the user to optimise performance. For each diagnostic test, we start by assessing the probability distribution of the demand produced by referrals

of each cancer specialty. We propose a recursive procedure to obtain such a distribution. Then, the overall demand for the referred diagnostic test can be obtained as the convolution of the demands of all cancer specialties, which is also obtained by means of a recursive procedure.

We consider the capacity planning problem for each diagnostic test as an inventory problem with perishable items. At each day, the system possesses a given capacity, which is viewed as an inventory of tests that can be used on either incoming or queued patients. The inventory contains all the appointments that are available for a given test on a specific day. These appointments, however, expire at the end of the day and will be no longer available in the capacity inventory of the following day. Hence, unused capacity can be treated analogously to expired inventory that needs to be discarded. Once we are in possession of the probability distribution of the aggregate demand for tests, the problem then becomes an inventory problem in which the decision maker seeks a balance between unused capacity and the stationary distribution of the number of queued tests. Once the capacity is defined, the problem can analogously be seen as a G/D/C queue that can be solved analytically, by means of $Z-$transform (Chaudhry and Kim, 2003) or by means of embedded Markov models (Shortle et al., 2018). Hence, the inventory model can be solved by determining an adequate number of daily slots C in the equivalent queuing model that provides the desirable balance between unused capacity and waiting times.

The research and modelling approach is motivated by working with oncologists and managers across Wales as part of a wider Cancer Research UK (CRUK) funded project. The Welsh Government has recently set out its approach to improving cancer services and outcomes with a significant focus on the earlier detection of cancers and greater understanding and efficiency of the pathways patients take. Subsequently, the Cabinet Secretary announced the introduction of a single cancer pathway (SCP) and its implementation must be properly tested and evaluated to understand the impact on patient care, treatment outcomes and the wider health system. Our research is therefore aimed at supporting the Welsh Government in making recommendations to align capacity to best match demand in an effective and efficient manner, and to ultimately improve patient care and outcomes.

The remainder of this paper is organised as follows. Section 2 introduces and motivates the problem. Section 3 features the mathematical formulation, which is divided in two subsections. Subsection 3.1 presents a recursive algorithm for the evaluation of the aggregate demand per test for a single cancer specialty, whereas Subsection 3.2 introduces a recursive evaluation of the overall demand for tests considering all cancer specialties. Section 4 discusses capacity planning and the evaluation of the long-term behaviour of the system for prescribed capacities. A case study to illustrate the approach based on data from the U.K.'s National Health Service (NHS) is presented in Section 5. Finally, Section 6 concludes the paper.

2 Problem Description

This paper studies the indirect demand for supplies or resources in healthcare systems, and is concerned in particular with the demand for diagnostic tests in cancer pathways. That notwithstanding, it is worth mentioning that the proposed model is general enough to encompass any other system with similar characteristics.

The demand is generated by patients that arrive spontaneously to the system or are referred to some surgical procedure or to a specific pathway, such as a diagnostic pathway for a cancer specialty. The patient arrivals are stochastic and the decision maker has access to historical data on the probability distribution of arrivals for a given period of interest. However invaluable, the arrival distribution is not enough to characterise the demand for any given supply or resource. This is because the exact composition and quantity of resources and supplies needed for any given patient cannot be exactly foreseen a priori. Instead, such information is progressively revealed as the patient traverses the pathway. Nevertheless, the decision maker has access to historical data and hence can estimate the demand distribution of each incoming patient for any given supply or resource of interest. In Section 3, we introduce a procedure to evaluate the periodic demand of a given supply or resource making use of both the patient arrival distribution and the distribution of the individual demand per patient for all pathways that make use of this specific supply or resource.

Many systems conform to the proposed model. As an example, consider the demand for surgical supplies such as catheters, stents or specific prosthetics. While hospitals generally have an overall idea of the supplies that may be needed in any given surgery, the specific supplies that will be required, as well as the exact demand will only be known at the end of each surgery. Furthermore, hospitals generally hold data on the demand for specific surgeries, as well as information on the supplies demanded by similar surgeries in the past. In a very different domain, clients arrive stochastically at the supermarket, and each client requires a list of supplies that will only be known a posteriori. Hence, in order to predict future demand for any given product, the decision makers need not only forecasts of the number of arriving clients, but also probability distributions related to the shopping list of individual clients (e.g., Guidotti et al., 2019).

In spite of the level of generality, the proposed model was motivated by general cancer care pathways. For the sake of illustration, let us consider the diagnostic phase of lung cancer patients. Upon having access to the patient's medical history and undertaking an eventual physical examination, consultants may require a computerised tomography (CT) to search for possible lesions. The patient is then examined and waits for a radiological report that is issued a posteriori. Then, the consultant may prescribe a second CT in case the first report is inconclusive. Alternatively, the investigations may continue with other tests or a decision can be reached. Furthermore, another CT scan may or may not be required later in the pathway to search for metastases (Silvestri et al., 2013). In the same diagnostic pathway, multiple biopsies may be required as

the investigation progresses, and some may have to be repeated in case the collected sample is insufficient or considered inadequate, which will only be known a posteriori. Finally, since the investigation progresses according to the findings in the test results, neither the specific tests to be undertaken nor their quantities are known a-priori for any given patient entering the pathway.

3 Aggregate Demand Evaluation for a Diagnostic Test

Let us consider a set of $n \geq 1$ cancer specialties that are being monitored and treated in a given health service. To enter the pathway of a given cancer specialty, the patient must be referred to the service by a physician. For each cancer specialty, the pathway includes appointments with specialist consultants and medical teams, as well as a set of specialised diagnostic and staging tests. The former are designed to test for cancer, whereas the latter are tools to determine how advanced the cancer is. In this paper, we are interested in determining the number of daily slots that should be made available at any given day to satisfy the overall demand for each diagnostic or staging test, in such a way as to ensure that the patient does not have to wait excessively for an appointment should he or she be assigned the test.

3.1 Aggregate Demand for a Single Cancer Specialty

Firstly, let us consider the demand for a given diagnostic test, produced by incoming patients of a certain cancer specialty. Each incoming patient may or may not require this specific diagnostic test, but if they do require it, they may take the test multiple times. The studied problem involves two random variables, one that represents the number of incoming patients on a given day, and another to denote the number of times that an incoming patient will have to undertake the test under consideration. Let A_k be a random variable representing the number of patient referrals on any given day for cancer specialty $k \in \{1, \ldots, n\}$, which takes values from the set $\Omega_{A_k} = \{0, 1, , \ldots, N\}$. Let $p_{A_k}(m)$, $m \in \Omega_{A_k}$, denote the probability that exactly m patients are referred for cancer specialty k on a given day. In addition, define a random variable T_k to represent the overall demand for the test that is produced by cancer specialty k on a given day. It is clear that the number of tests performed is a function of the number of incoming referrals. Hence, by using the total probability theorem, we have:

$$P(T_k = j) = \sum_{m \in \Omega_{A_k}} P(T_k = j | A_k = m)\, p_{A_k}(m). \tag{3.1}$$

To simplify the notation, let V_m be the total number of tests given that we have exactly $m \in \Omega_{A_k}$ incoming referrals. This yields $P(V_m = j) = P(T_k = j | A_k = m)$, and hence Eq. (3.1) becomes:

$$P(T_k = j) = \sum_{m \in \Omega_{A_k}} P(V_m = j)\, p_{A_k}(m). \tag{3.2}$$

In order to fully assess Eq. (3.2), we need to find an expression for the first term in the right hand side. Let the random variable Y denote the number of tests required by a single referred patient, which takes values from the set of non-negative integers Ω_Y, and let $n_1 = \min(\Omega_Y)$ and $n_2 = \max(\Omega_Y)$ denote the minimum and maximum number of tests required by a single patient, respectively. Furthermore, let $P(Y = i)$, $i \in \Omega_Y$, denote the probability that a patient will require exactly i tests. From the definitions, it is clear that V_m is the sum of m independent and identically distributed (*iid*) random variables Y_l, $1 \leq l \leq m$, with $Y_l \sim Y$, $\forall l$. Hence, V_m can be seen as the convolution of m *iid* variables, whose distribution can be obtained by the iterative procedure below, considering one convolution at a time, as follows:

$$P(V_m = i) = \sum_{j=(m-1)n_1}^{i} P(V_{m-1} = j)\, P(Y = i-j), \quad m{\cdot}n_1 \leq i \leq m{\cdot}n_2, \ \ \forall m \geq 2, \tag{3.3}$$

with

$$P(V_1 = i) = P(Y = i), \forall n_1 \leq i \leq n_2.$$

3.2 Considering Multiple Cancer Specialties

Multiple cancer specialties pose no significant additional difficulties for the calculation of the distribution of the total demand for tests. In that case, one just needs to repeat the procedure detailed in Section 3.1 for each cancer specialty. Like in the previous section, let us assume that a total of $n \geq 1$ cancer specialties make use of the considered test, and recall that T_k, $k = 1, \ldots, n$ is the random variable representing the total demand for tests from all incoming referrals for cancer specialty $1 \leq k \leq n$. In that case, the total demand for tests is

$$T = \sum_{k=1}^{n} T_k. \tag{3.4}$$

The distribution of T is now the convolution of n distinct and independent probability distributions, and the iterative procedure to find the distribution of T is rather similar to that presented in the last section. Let $W_m = \sum_{k=1}^{m} T_k$, $1 \leq m \leq n$, be a random variable representing the number of tests required by the first m cancer specialties. If follows that:

$$P(W_m = i) = \sum_{j=0}^{i} P(W_{m-1} = j)\, p_{T_m}(i-j), \ i \geq 0, \ \ \forall m \geq 2, \tag{3.5}$$

with

$$P(W_1 = i) = p_{T_1}(i), \forall i \in \Omega_{T_1},$$

and $T = W_n$.

4 Capacity Planning

This section discusses the elaboration of a capacity plan that makes use of the aggregate demand for tests obtained by the procedures in Sections 3.1 and 3.2. Let $E(T)$ be the average daily demand for a given test. A simple course of action for the decision maker is to offer a fixed number $C \in \mathbb{N}$ of daily slots, where $\mathbb{N}$ is the set of natural numbers. Considering that a slot that is not occupied on a given day cannot be kept in inventory for the next day, we can model the resulting system as a G/D/C queue. Furthermore, to ensure finite waiting times and long-term stability, we must have $C > E(T)$ (Shortle et al., 2018). Such a system can be solved analytically, by means of the $Z-$transform procedure detailed in Section 4.1. Alternatively, the solution can also be obtained via an embedded Markov chain (e.g., Shortle et al., 2018), as described in Section 4.2. The choice of C depends on the trade-off between unused capacity and quality of service to the end users, which can be modelled in a number of ways; for example as a function of the stationary distribution of the waiting times. Regardless of the long-term goal, the decision maker has to be able to evaluate the steady state distribution of the resulting queue, as described in the following subsections.

Another possible approach, which will be explored in detail in the Case Study - Section 5 - is to have some temporary extra capacity that is deployed with a given probability. In that case, an analytic solution by means of the $Z-$transform is no longer applicable. However, embedded Markov chains can still be employed, as detailed in Section 4.2.

4.1 Steady State Distribution for the Overall Demand using the $\mathcal{Z}$-Transform

The problem of finding the steady state distribution of the resulting queuing system for a given fixed capacity $C > E(T)$ can be solved by means of the $Z-$transform, making use of an equivalent signal processing formulation. Let $p_T(m) = P(T = m) = t_m$, denote the probability that exactly m tests are requested on a given day. The probability generating function (PGF) can be defined as

$$T(z) = \sum_{m=0}^{N} t_m z^m. \tag{4.1}$$

The PGF $T(z)$ is a polynomial of degree N, where N is the maximum possible number of test requests on a given day. The probability mass function

(PMF) can be expressed as a discrete sequence

$$t(m) = \sum_{r=0}^{N} t_r \delta[m - r] \tag{4.2}$$

that can be used to find the PGF

$$T(z) = \mathcal{Z}\{t(-m)\}, \tag{4.3}$$

where $\mathcal{Z}\{\cdot\}$ denotes the $Z-$transform operator (e.g., Oppenheim and Schafer, 2009). Let X_k, $k \geq 0$ be the stochastic process that describes the number of test requests in the system, i.e. the tests that are either being serviced or are waiting to be processed at any given period $k \geq 0$. Then, the steady state probability of process X_k, $k \geq 0$, denoted by vector π, such that

$$\pi_m = \pi(m) = \lim_{k \to \infty} P(X_k = m),\ m \geq 0,$$

has the PGF given by

$$\Pi(z) = \sum_{m=0}^{\infty} \pi_m z^m = \mathcal{Z}\{\pi_{-m}\}. \tag{4.4}$$

Bruneel and Wuyts (1994) and Chaudhry and Kim (2003) propose an explicit expression for the generating function, given by

$$\Pi(z) = T(z)Q(z), \tag{4.5}$$

where

$$Q(z) = \prod_{i=1}^{L} \frac{1 - \beta_i}{z - \beta_i}, \tag{4.6}$$

and β_i is a root of $z^C - T(z) = 0$. The polynomial $z^C - T(z)$ has $L = (N - C)$ roots outside of the unity circle $|z| \leq 1$ (Bruneel and Wuyts, 1994; Chaudhry and Kim, 2003). One can find $\Pi(z)$ using the inverse $Z-$transform, and modelling Eq. (4.5) as a digital causal stable filter whose output gives π_{-m}. We have

$$\Pi(z^{-1}) = T(z^{-1})H(z), \tag{4.7}$$

where $H(z)$ is the system function with L poles $p_i = \frac{1}{\beta_i}$, $i = 1, \ldots, L$, within the unit circle, and

$$H(z) = \prod_{i=1}^{L} \frac{1 - p_i}{1 - p_i z^{-1}} = \frac{\nu_0}{1 + \sum_{i=1}^{L} \eta_i z^{-i}}. \tag{4.8}$$

The steady state distribution for the requested tests in the system, π, is the inverse $Z-$transform of $\Pi(z^{-1})$. The digital filter output can be implemented using the following difference equation (Oppenheim and Schafer, 2009):

$$\pi_m = \nu_0 t_m - \sum_{i=1}^{L} \eta_i \pi_{m-i}, \tag{4.9}$$

where ν_0 is the numerator of Eq. (4.8), and η_i are the coefficients of the polynomial in the denominator of Eq. (4.8).

4.2 Steady State Distribution for the Overall Demand via Embedded Markov Chains

Recall from Section 4 that the system is modelled as a G/D/C queue, where C is the fixed number of available daily tests. This queue can be described by an embedded Markov chain (Shortle et al., 2018) with state space $S = \{0, 1, 2, \ldots\}$ that represents the evolution of process X_k, $k \geq 0$, defined in the previous section as the number of test requests in the system. Let $p_T(l) = P(T = l)$ be the probability that l test requests are received on a given day. The elements of the transition matrix P^C for a given capacity C are given by:

$$p_{ij}^C = \begin{cases} p_T(l), & \text{if } j = i - C + l \text{ and } i > C; \\ p_T(l), & \text{if } j = l \text{ and } i < C; \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

for all $i, j \in S$. The first line in Eq. (4.10) models the transition when there is already a queue of tests waiting to be processed. The second line considers the system with no waiting test requests; observe that all requests already in the system will be served in the current period, whilst the incoming requests will be processed on the following day. Classical Markov chain theory yields that the steady state distribution of X_k, $k \geq 0$ is the solution of the following system of equations (Brémaud, 1999):

$$\begin{aligned} \pi P^C &= \pi, \\ \sum_{i=0}^{\infty} \pi(i) &= 1. \end{aligned} \tag{4.11}$$

Now consider the case where the capacity is a random variable, i.e. C is a random variable taking values from the set of positive integers $\Omega_C = \{c_1, c_1 + 1, \ldots, c_2\}$, with $P(C = c) = P_C(c)$. In that case, the $Z-$transform method of the previous section is no longer applicable. However, the Markov chain approach is still valid, and the system is stable if $E(T) < E(C)$. In that case, we have

$$P^C = \sum_{l \in \Omega_C} p_C(l) P^l, \tag{4.12}$$

where P^l is evaluated by means of Eq. (4.10), with $C = l$. Once again, the steady state behaviour is obtained by solving (4.11).

The variable capacity setting is interesting to model the case when extra capacity is made available with a given probability, for example, by granting access to some shared resources. This possibility will be explored in the case study in Section 5.

5 Case Study

We illustrate our novel methdological approach using data on lung cancer referrals from one health board (region) of South Wales. We consider data on

lung cancer referrals from the Cwm Taf Morgannwg University Health Board in Wales, covering the 26-week period from 1st April 2016 to 30th September 2016. A total of 341 patients were referred to the lung cancer pathway in the studied period.

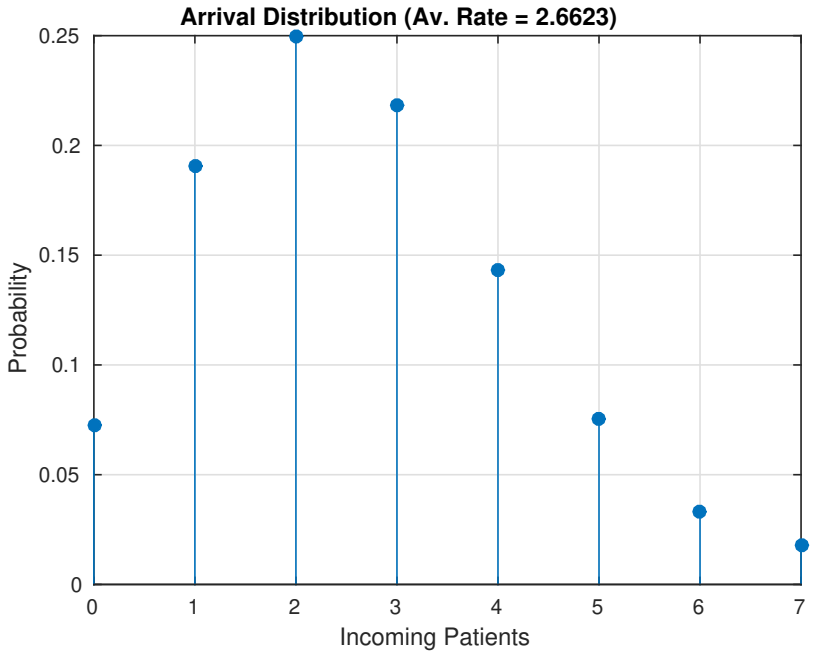

Figure 1: Incoming Distribution of Lung Cancer Suspicions

In this example, we will consider CT test requests. Considering five business days per week, we have a rate $\lambda = \frac{341}{26 \cdot 5} \approx 2.623$ incoming patients per day. We assume that the number of daily incoming referrals is described by a Poisson process with rate λ. Imposing a boundary of 0.99 in the cumulative distribution $F_A : A \to [0, 1]$, we limit the maximum number of arrivals to $n = 7$, and we make

$$P(A = 7) = \sum_{k=7}^{\infty} P(A = k),$$

to obtain the probability distribution depicted in Figure 1.

The total number of CTs performed over the period is 393. For the sake of modelling, and considering the *Optimal Lung Cancer Pathway* guidelines (Lung Clinical Expert Group, 2017), we assume that every patient undergoes at least one CT test. Since there were 341 lung cancer referrals in the database, we assume that 52 patients had a repeated CT scan. Hence, we have $P(Y = 1) = \frac{341}{393} \approx 0.868$, and $P(Y = 2) \approx 0.132$. Applying Eq. (3.2) to the available data, we obtain the distribution of the number of daily CT requests, depicted

in Figure 2.

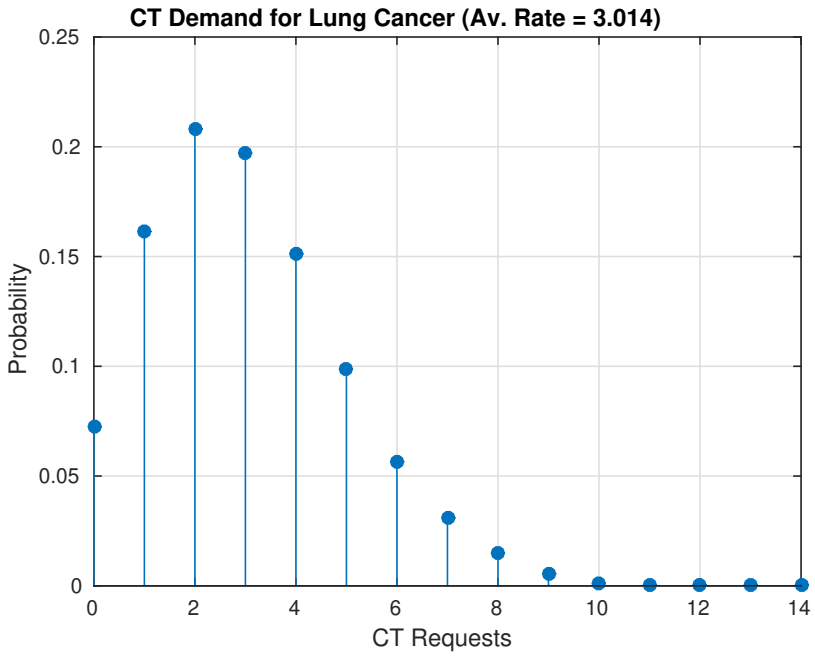

Figure 2: Distribution of CT Requests for Lung Cancer Referrals

The cumulative distribution of CT requests per business day is depicted in Figure 3 below. One can see that, for a service rate $\rho \geq 0.8$ a supply of five CT slots per business day would suffice, whereas at least four CT slots per business day are required for stability. Furthermore, depending on the required service rate, the daily available capacity should range between the average number of daily CT requests $\lambda_{CT} \approx 3.014$ and 14 CT slots per day. A fractional capacity can be attained by providing a fixed number of daily slots, for example 3 slots, and offering some extra slots on specific days. For example, by having three regular slots plus one extra slot every Monday, we would have an average of 3.2 slots per business day.

5.1 Optimal Capacity Planning

As mentioned in Section 4, if we keep a fixed capacity, that is, a fixed number of daily CT slots, our model becomes a G/D/C queuing system. In that case, we may define C as a function of a compromise between the perceived costs of delayed tests and unused capacity. However, more general inventory policies can be pursued, which would produce G/D/C queuing systems with removable servers. By convention, we assume that no incoming request can be processed at the time of arrival. All requests have to wait at least one period to be processed.

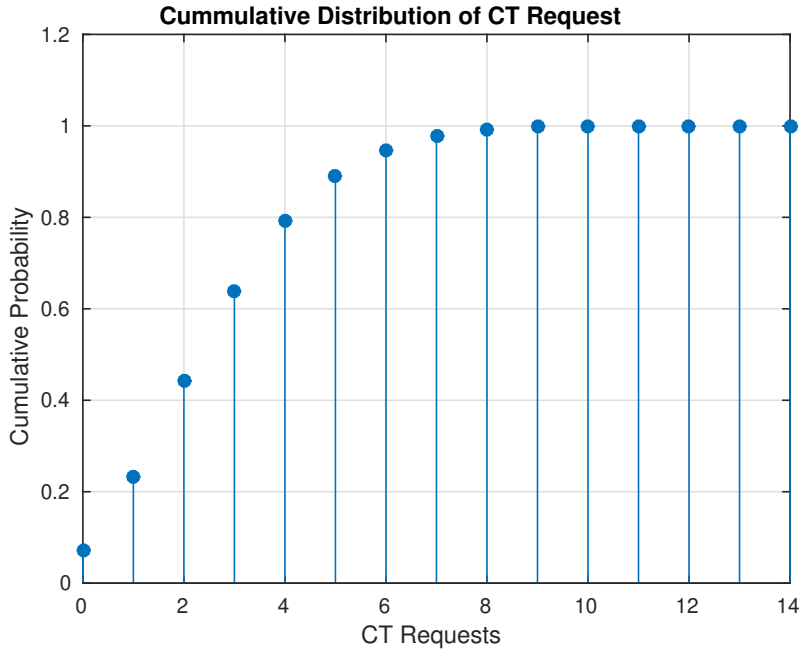

Figure 3: Cumulative Distribution of CT Requests for Lung Cancer Referrals

The motivation is that a cancer test typically requires an appointment and very seldom the appointment will be available for the same day.

Consider the example in the previous section, whose cumulative distribution of requests is depicted in Figure 3, and whose average daily request rate is $\lambda_{CT} \approx 3.014$. Let us assume we have a fixed capacity of $C = 3$ daily CT slots and an extra slot that can be used with a probability $\alpha \in [0, 1]$. For a given fixed capacity C, the system can be described by a Markov chain (e.g., Brémaud, 1999) with state space $S = \{0, 1, \dots\}$ and transition matrix P^C defined in Section 4.2.

In the example, we have:

$$P^C = \begin{bmatrix} p_T(0) & p_T(1) & \dots & p_T(14) & 0 & 0 & 0 & 0 & \dots \\ p_T(0) & p_T(1) & \dots & p_T(14) & 0 & 0 & 0 & 0 & \dots \\ p_T(0) & p_T(1) & \dots & p_T(14) & 0 & 0 & 0 & 0 & \dots \\ 0 & p_T(0) & \dots & p_T(13) & p_T(14) & 0 & 0 & 0 & \dots \\ 0 & 0 & \dots & p_T(12) & p_T(13) & p_T(14) & 0 & 0 & \dots \\ 0 & 0 & \dots & p_T(11) & p_T(12) & p_T(13) & p_T(14) & 0 & \dots \\ 0 & 0 & \dots & p_T(10) & p_T(11) & p_T(12) & p_T(13) & p_T(14) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

for $C = 3$. Observe in the transition matrix that, if there is no request which

cannot be immediately processed i.e. $X_k = i \leq C$, then the next state equals the demand in period $k \geq 0$, regardless of the current state $i \leq C$. This happens because the system will process all the i pending requests during period k and will be left with only the incoming requests in the current period. In contrast, when $i > C$, some of the requests will be left in the queue and the number of impending requests at the onset of the following day will be $X_{k+1} = X_k - C + l$, where l is a realization of the random variable T, which represents the number of incoming tests at any given time.

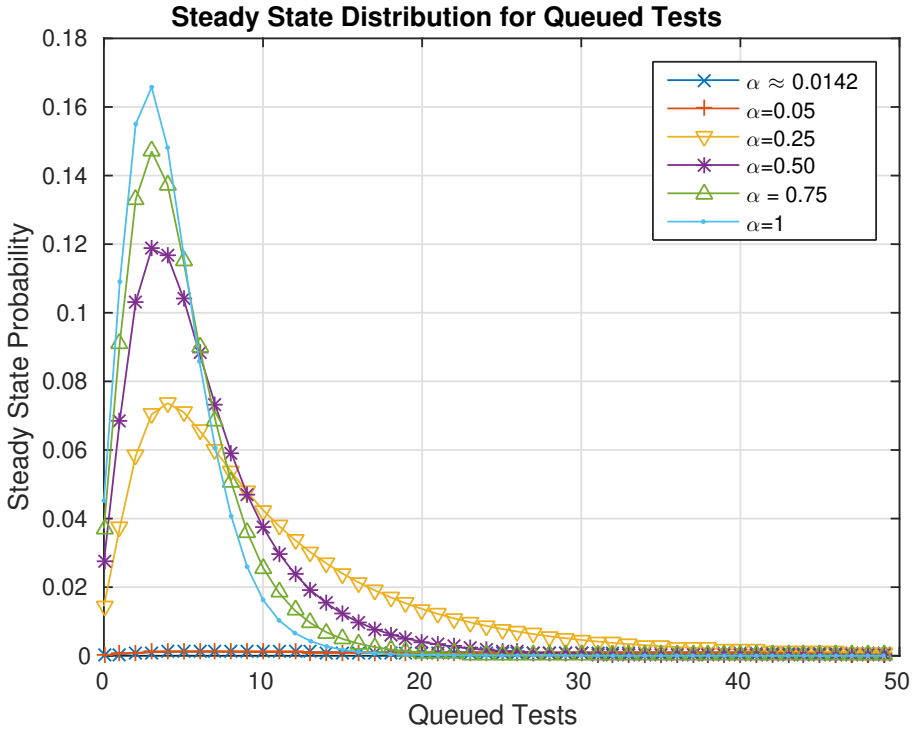

Figure 4: Steady state probabilities for the number of pending tests

Now, let us assume in the example that we allocate an extra capacity with a given probability α. Hence, we have $P(C = 3) = 1 - \alpha$ and $P(C = 4) = \alpha$. In that case, the transition matrix for process X_k, $k \geq 0$, becomes:

$$P^C = \alpha P^4 + (1 - \alpha)P^3.$$

To evaluate the steady state behaviour of the system under any value of α, it suffices to find the limiting distribution of X_k, $k \geq 0$, by solving the system in (4.11). Figure 4 depicts the steady state distribution for distinct values of α. One can notice in Figure 4 that small values of α tend to keep a large queue of

impending tests. As we increase the value of α we observe a steady decrease in the system occupation, which is steeper for larger values of α.

Since the service is deterministic, we can easily obtain the expected waiting time of an incoming patient. Suppose, for example, that $X_k = n$ just after the arrival of the patient's request. Then, the patient will have to wait $\lceil \frac{n}{C} \rceil$ days for his/her request to be processed, where

$$\left\lceil \frac{n}{C} \right\rceil = \min \left\{ u \in \mathbb{Z} : u \geq \frac{n}{C} \right\},$$

is the ceiling of $\frac{n}{C}$, i.e. the closest value in the set of integers $\mathbb{Z}$ that exceeds $\frac{n}{C}$. Hence, we can easily calculate the steady state distribution of the waiting time from the distribution of pending requests.

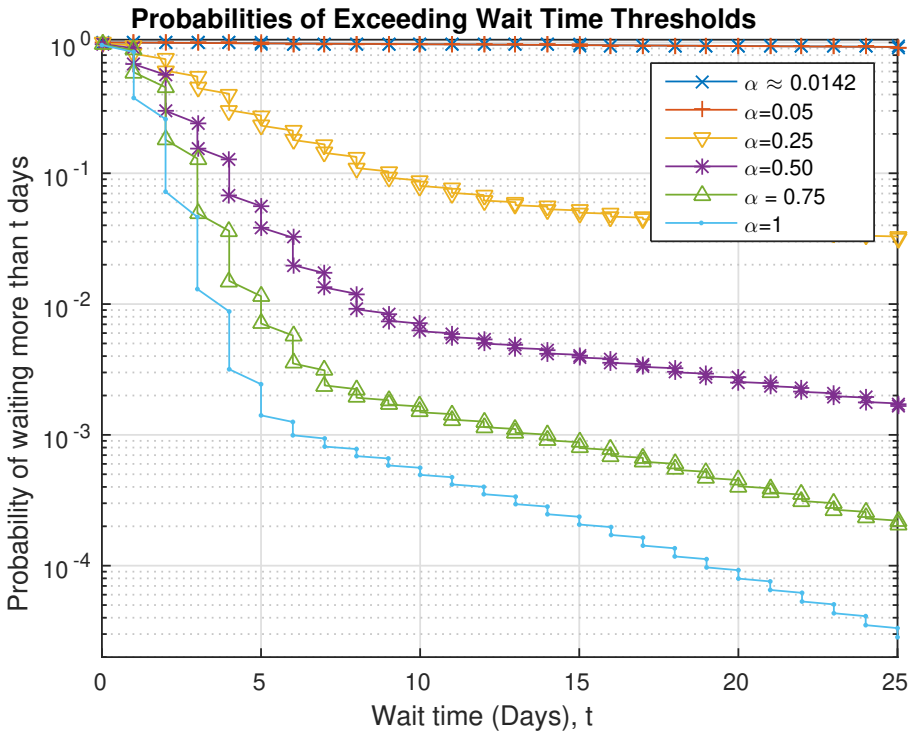

Figure 5: Waiting time thresholds and probabilities

Figure 5 depicts the probability that an incoming request will have to wait more than t days to be processed for distinct integer values $t \geq 0$. With the results in Figure 5 the decision maker can, for example, establish a target time t and a target probability $\bar{p}$, and determine a suitable value of α such that $P(\omega > t) \leq \bar{p}$. For example, if we accept that at most 10% of patients are allowed to wait more than 10 days for a test result, then $\alpha = 0.25$ suffices. On

the other hand, $\alpha = 1$ ensures that no more than one in a thousand requests will wait in excess of six days. In order to facilitate the reproducibility of the results, the code in R (R Core Team, 2018) for the example is provided as supplementary material.

6 Concluding Remarks

Motivated by the need to improve the delivery of diagnostic services in cancer pathways in Wales, this study introduced a novel recursive procedure to obtain the probability distribution of the overall demand for a given diagnostic service in a disaggregated manner. The approach is general in that it does not impose constraints on the distributions of both incoming referrals and the number of repeated diagnostic tests for each incoming patient. Furthermore, the proposed model is easy to implement and simple enough to allow the design of an optimal policy with respect to a given performance criterion. But it is also general enough to warrant application in analogous problems in healthcare and beyond. This illustrates the power and impact of analytical approaches in healthcare systems.

By disaggregating the problem for each diagnostic test, we are able to model the resulting system as a perishable inventory problem that can be solved by means of a G/D/C queuing model for a given capacity C. In that case, the problem can be solved by means of analytic signal processing techniques or Markov chain techniques, and the decision maker has to select a capacity C that yields a good compromise between service quality and unused capacity. However, when solved by means of Markov models, the approach is more general and enables the decision maker to define random capacities by deploying shared resources with a prescribed probability. The case study illustrates the flexibility of the approach and demonstrates how the decision maker can use the results to enforce bounds on the service time with prescribed probabilities.

Finally, while the model illustrates the aggregation of the demand for tests coming from different cancer specialties, it is worth mentioning that the resources needed are generally shared with other medical specialties and pathways, and are often scarce. Fortunately, the approach introduced in this paper can be seamlessly applied to aggregate the demand for any number of medical specialties and pathways, which makes it an ideal tool for general capacity planning studies that involve all of the potential users of any given healthcare resource. Such studies may ensure fairness as opposed to prioritisation rules which improve the access of certain segments of patients at the expense of a deteriorated service for the segments left out.

Future research directions include introducing more flexibility to the decision maker by modelling the system as a Markov decision process and thereby defining the service level as a function of the current state of the system. While flexible, such an approach requires the decision maker to prescribe a cost function, which may be very difficult in the setting, considering that the trade-off between waiting times and the cost of extra capacity is not easily quantifiable.

Moreover, since the actual system capacity is generally fixed, a flexible use of such capacity by cancer pathways needs to be offset by an effective resource sharing protocol that enables the extra capacity to be deployed in other pathways when it is not needed for cancer care. Hence, a number of research problems can be defined to tackle each of these issues in future studies.

Acknowledgements

We would like to take this opportunity to thank all those staff in Cwm Taf Morgannwg University Health Board, Aneurin Bevan University Health Board, and Velindre Cancer Centre, who have supported this work through their clinical advice and supply of data.

The authors would like to thank the referees and editors for their many invaluable and constructive comments and suggestions, which led to a substantially improved manuscript.

Funding

This work has resulted from research funded by a Cancer Research UK grant 'Analysis and Modelling of a Single Cancer Pathway Diagnostics' (Early Diagnosis Project Award A27882) and from a KESS2 grant under the project title 'Smart Simulation and Modelling of Complex Cancer Systems'. Knowledge Economy Skills Scholarships (KESS) is a pan-Wales higher level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West Wales and the Valleys. The authors thank the Brazilian Research Council - CNPq for grant #311075/2018-5.

References

E. Ahmadi, D. T. Masel, A. Y. Metcalf, and K. Schuller. Inventory management of surgical supplies and sterile instruments in hospitals: a literature review. *Health Systems*, 8(2):134–151, 2019. doi: 10.1080/20476965.2018.1496875.

O. Alagoz, T. Ayer, and F. S. Erenay. *Operations Research Models for Cancer Screening*, pages 1–14. American Cancer Society, 2011. ISBN 9780470400531. doi: 10.1002/9780470400531.eorms0597.

E. F. Arruda, B. B. Pereira, C. A. Thiers, and B. R. Tura. Optimal testing policies for diagnosing patients with intermediary probability of disease. *Artificial Intelligence in Medicine*, 97:89 – 97, 2019. ISSN 0933-3657. doi: 10.1016/j.artmed.2018.11.005.

E. Aspland, D. Gartner, and P. Harper. Clinical pathway modelling: a literature review. *Health Systems*, 0(0):1–23, 2019. doi: 10.1080/20476965.2019.1652547.

H. M. Bidhandi, J. Patrick, P. Noghani, and P. Varshoei. Capacity planning for a network of community health services. *European Journal of Operational Research*, 275(1):266 – 279, 2019. ISSN 0377-2217. doi: 10.1016/j.ejor.2018.11.008.

I. A. Bikker, N. Kortbeek, R. M. van Os, and R. J. Boucherie. Reducing access times for radiation treatment by aligning the doctor's schemes. *Operations Research for Health Care*, 7:111 – 121, 2015. ISSN 2211-6923. doi: 10.1016/j.orhc.2015.06.005. ORAHS 2014 - The 40th international conference of the EURO working group on Operational Research Applied to Health Services.

J. Bowers, B. Lyons, G. Mould, and T. Symonds. Modelling outpatient capacity for a diagnosis and treatment centre. *Health Care Management Science*, 8(3):205–211, 2005. ISSN 1572-9389. doi: 10.1007/s10729-005-2011-0.

P. Brémaud. *Gibbs fields, monte carlo simulation, and queues*. Springer-Verlag, New York, 1999.

Herwig Bruneel and Ilse Wuyts. Analysis of discrete-time multiserver queueing models with constant service times. *Operations Research Letters*, 15(5):231 – 236, 1994. doi: 10.1016/0167-6377(94)90082-5.

M. Capan, A. Khojandi, B. T. Denton, K. D. Williams, T. Ayer, J. Chhatwal, M. Kurt, J. M. Lobo, M. S. Roberts, G. Zaric, S. Zhang, and J. S. Schwartz. From data to improved decisions: Operations research in healthcare delivery. *Medical Decision Making*, 37(8):849–859, 2017. doi: 10.1177/0272989X17705636. PMID: 28423982.

E. Castro and S. Petrovic. Combined mathematical programming and heuristics for a radiotherapy pre-treatment scheduling problem. *Journal of Scheduling*, 15(3): 333–346, 2012. ISSN 1099-1425. doi: 10.1007/s10951-011-0239-8.

M. L. Chaudhry and N. K. Kim. A complete and simple solution for a discrete-time multi-server queue with bulk arrivals and deterministic service times. *Operations Research Letters*, 31(2):101 – 107, 2003. ISSN 0167-6377. doi: 10.1016/S0167-6377(02)00214-6.

G. Culpan, A.-M. Culpan, P. Docherty, and E. Denton. Radiographer reporting: A literature review to support cancer workforce planning in england. *Radiography*, 25 (2):155 – 163, 2019. ISSN 1078-8174. doi: 10.1016/j.radi.2019.02.010.

D. B. Fuller, E. F. Arruda, and V. J. M. Ferreira Filho. Learning-agent-based simulation for queue network systems. *Journal of the Operational Research Society*, pages 1–17, 2019. doi: 10.1080/01605682.2019.1633232.

L. Garg, S. McClean, B. Meenan, and P. Millard. Non-homogeneous Markov models for sequential pattern mining of healthcare data. *IMA Journal of Management Mathematics*, 20(4):327–344, 2008. ISSN 1471-678X. doi: 10.1093/imaman/dpn030.

D. Gartner, M. Frey, and R. Kolisch. Hospital-wide therapist scheduling and routing: Exact and heuristic methods. *IISE Transactions on Healthcare Systems Engineering*, 8(4):268–279, 2018.

R. Guidotti, G. Rossetti, L. Pappalardo, F. Giannotti, and D. Pedreschi. Personalized market basket prediction with temporal annotated recurring sequences. *IEEE Transactions on Knowledge & Data Engineering*, 31(11):2151–2163, 2019. ISSN 1558-2191. doi: 10.1109/TKDE.2018.2872587.

P. J. E. Hulshof, R. J. Boucherie, E. W. Hans, and J. L. Hurink. Tactical resource allocation and elective patient admission planning in care processes. *Health Care Management Science*, 16(2):152–166, 2013. ISSN 1572-9389. doi: 10.1007/s10729-012-9219-6.

Peter JH Hulshof, Nikky Kortbeek, Richard J Boucherie, Erwin W Hans, and Piet JM Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175, 2012.

Lung Clinical Expert Group. National optimal lung cancer pathway, 2017. URL `https://www.cancerresearchuk.org/sites/default/files/national_optimal_lung_pathway_aug_2017.pdf`.

M. Martins, C. Araujo, L. Marques, and C. Araújo. The healthcare supply network: current state of the literature and research opportunities. *Production planning & control*, pages 1–20, 2019. ISSN 0953-7287. doi: 10.1080/09537287.2019.1663451.

J. Marynissen and E. Demeulemeester. Literature review on multi-appointment scheduling problems in hospitals. *European Journal of Operational Research*, 272 (2):407 – 419, 2019. ISSN 0377-2217. doi: 10.1016/j.ejor.2018.03.001.

R. Meskarian, T. Monks, R. Chappell, and C. Kipps. Regional capacity planning for dementia clinic diagnosis. *Alzheimer's & Dementia*, 13(7, Supplement):P1159, 2017. ISSN 1552-5260. doi: 10.1016/j.jalz.2017.06.1695. 2017 Abstract Supplement.

T. B. T. Nguyen, A. I. Sivakumar, and S. C. Graves. A network flow approach for tactical resource planning in outpatient clinics. *Health Care Management Science*, 18(2):124–136, 2015. ISSN 1572-9389. doi: 10.1007/s10729-014-9284-0.

T. B. T. Nguyen, A. I. Sivakumar, and S. C. Graves. Capacity planning with demand uncertainty for outpatient clinics. *European Journal of Operational Research*, 267 (1):338 – 348, 2018. ISSN 0377-2217. doi: 10.1016/j.ejor.2017.11.038.

Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. Pearson, New Jersey, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

N. Richers, M. Amorim, E. F. Arruda, and H. Rey. Optimal inventory policies for surgical supplies through a decision making model using markov decision process for a minimum service level. In *51st Brazilian Symposium on Operational Research*, Limeira SP, 2019.

H. L. Romero, N. P. Dellaert, S. van der Geer, M. Frunt, M. H. Jansen-Vullers, and G. A. M. Krekels. Admission and capacity planning for the implementation of one-stop-shop in skin cancer treatment using simulation-based optimization. *Health Care Management Science*, 16(1):75–86, 2013. ISSN 1572-9389. doi: 10.1007/s10729-012-9213-z.

C. R. Rosales, M. Magazine, and U. Rao. The 2bin system for controlling medical supplies at point-of-use. *European Journal of Operational Research*, 243(1):271 – 280, 2015. ISSN 0377-2217. doi: 10.1016/j.ejor.2014.10.041.

A. Sauré, J. Patrick, S. Tyldesley, and M. L. Puterman. Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research*, 223(2):573 – 584, 2012. ISSN 0377-2217. doi: 10.1016/j.ejor.2012.06.046.

C. E. Saville, H. K. Smith, and K. Bijak. Operational research techniques applied throughout cancer care services: a review. *Health Systems*, 8(1):52–73, 2019. doi: 10.1080/20476965.2017.1414741.

Scottish Government. Demand optimisation in diagnostics: Best test, best care, 2017. URL `https://www.gov.scot/binaries/content/documents/govscot/publications/advice-and-guidance/2017/02/demand-optimisation-diagnostics/documents/00514237-pdf/00514237-pdf/govscot%3Adocument/00514237.pdf`.

Jinghua Shi, Oguzhan Alagoz, Fatih Safa Erenay, and Qiang Su. A survey of optimization models on cancer chemotherapy treatment planning. *Annals of Operations Research*, 221(1):331–356, 2014.

J. Shortle, J. Thompson, D. Gross, and C. Harris. *Fundamentals of Queueing Theory*. Wiley Series in Probability and Statistics. Wiley, New York, 5 edition, 2018. ISBN 9781118943526. doi: 10.1002/9781119453765.

G. A. Silvestri, A. V. Gonzalez, M. A. Jantz, M. L. Margolis, M. K. Gould, L. T. Tanoue, L. J. Harris, and F. C. Detterbeck. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest*, 143(5, Supplement): e211S – e250S, 2013. ISSN 0012-3692. doi: https://doi.org/10.1378/chest.12-2355.

South Yorkshire, Bassetlaw & North Derbyshire Cancer Alliance. CT & MRI diagnostic demand & capacity final report, 2019. URL `https://www.healthandcaretogethersyb.co.uk/application/files/5615/5255/9149/Diagnostic_demand_and_capacity_report_140319.pdf`.

R. Suárez-Vega, J. L.Gutiérrez-Acuña, and M. Rodríguez-Díaz. Locating a shopping centre by considering demand disaggregated by categories. *IMA Journal of Management Mathematics*, 29(4):435–456, 2017. ISSN 1471-678X. doi: 10.1093/imaman/dpx006.

Welsh Government. NHS Wales planning framework 2019/2022, 2019. URL `http://www.wales.nhs.uk/sitesplus/documents/862/Item%205.1.1%20NHS%20Wales%20Planning%20Framework%202019-22.pdf`.

N. Woznitza, R. Steele, K. Piper, S. Burke, S. Rowe, A. Bhowmik, S. Maughn, and K. Springett. Increasing radiology capacity within the lung cancer pathway: centralised work-based support for trainee chest X-ray reporting radiographers. *Journal of Medical Radiation Sciences*, 65(3):200–208, 2018. doi: 10.1002/jmrs.285.

X. Wu, J. Li, and C. Chu. Modeling multi-stage healthcare systems with service interactions under blocking for bed allocation. *European Journal of Operational Research*, 278(3):927 – 941, 2019. ISSN 0377-2217. doi: 10.1016/j.ejor.2019.05.004.

L. Xiao, M. Dridi, A. Hajjam El Hassani, W. Lin, and H. Fei. A solution method for treatment scheduling in rehabilitation hospitals with real-life requirements. *IMA Journal of Management Mathematics*, 30(4):367–386, 2018. ISSN 1471-678X. doi: 10.1093/imaman/dpy009.