Scheduling the production of perishable patient-specific medication

Research of Mathematics

Author:
Jarco Slager (s1850881)

UT supervisor:
Dr. Ir. A. Braaksma

ZGT supervisor:
S. Selles, PharmD

May 8, 2022

UNIVERSITY OF TWENTE.

Voorwoord

Met deze scriptie sluit ik mijn Onderzoek van Wiskunde af; het laatste onderdeel van mijn master Educatie en Communicatie in de Bètawetenschappen aan de Universiteit Twente. Dit avontuur heeft langer geduurd dan ik had voorzien, mede door mijn keuze om vast aan de slag te gaan als wiskundedocent maar ook door factoren waar ik geen invloed op had. Het is mooi om mijn carrière als (wiskunde)student aan de UT af te sluiten met een afstudeeronderzoek bij de apotheek van de Ziekenhuisgroep Twente (ZGT). De toepassing van wiskunde binnen de Operations Research was voor mij op de middelbare school de reden om Applied Mathematics te gaan studeren. Ik ben dan ook blij en trots dat ik deze studie en dit onderzoek heb mogen doen.

Aleida, hartelijk dank voor al je hulp en steun. Je was een geweldige begeleider en ik heb veel aan al je feedback en meetings gehad. Ondanks dat alles wat langer duurde dan gepland stond jij altijd klaar om te helpen en bleef je enthousiast. Naast dat je me inhoudelijk veel hebt geholpen, waren onze meetings ook gewoon gezellig. Nogmaals bedankt en om mijn bericht voor jou trouw af te sluiten: groetnis!

Ook vanuit ZGT heb ik hulp gehad waar ik niet zonder kon. Suzanne, bedankt voor je meetings die je met me hebt gepland tussen je drukke baan door. Ik weet dat het voor jou niet altijd goed uitwkam, maar zonder jouw kennis en ideeën had ik het niet gered! Ook Jeroen, Anouk en Anita wil ik hartelijk bedanken voor hun hulp bij het verkrijgen van de data die ik nodig had om dit onderzoek te voltooien.

Jelle, bedankt voor je goede zorg! Vele avonden heb ik aan dit onderzoek gewerkt de afgelopen maanden, maar gelukkig was dat niet elke dag zo. In plaats daarvan gezellig 's avonds spelletjes doen was af en toe hard nodig! Deze onstpanning had ik wel nodig naast een full-time baan en dit onderzoek. Ik kijk uit naar de komende tijd om deze gezellige avonden voort te zetten!

Ook wil ik mijn familie bedanken. Hait, mim en Jeroen: na bijna zes jaar dan toch klaar met de studie! Het was in het begin niet duidelijk wat ik ging studeren, taalwetenschap of wiskunde. Jullie stonden altijd achter mijn keuze, welke het ook zou worden. Dank!

Annemarie, Femke, Jente, Leander, Lotte, Lucas, Nienke, Sven, Tessa en Wisse. Zonder deze tien mensen was mijn studententijd niet hetzelfde geweest. Het samen studeren, pauze houden, afspreken: het heeft er allemaal voor gezorgd dat ik een heel leuke tijd heb gehad. Ik kijk terug op heel leuke jaren en daar wil ik jullie dan ook heel graag voor bedanken.

Jarco

Hengelo, maart 2022

Abstract

In this research thesis, we investigate how to improve the production planning of perishable patient-specific medication at the pharmacy department of Ziekenhuisgroep Twente (ZGT). Each medication type uses a different type of raw material, which is stored in vials. Any remaining raw material after opening a vial is considered as spillage. Preparing medication of the same type at the same time (clustering) may reduce this spillage. In addition, a medical appointment may be cancelled, resulting in disposal of produced medication. Currently, the pharmacy department prepares medication one day before the patient's appointment and uses clustering. The pharmacy department wants to know at the start of the day which patients' medication should be prepared such that the spillage and disposal costs are minimal. We develop a Markov Decision Process to model this scheduling problem, and we use Approximate Dynamic Programming with the post-decision state variable and basis functions to find an approximate solution. The computational results show that a myopic policy, which we also developed, performs best in our research. This myopic policy postpones the production until the patient actually requires the medication now. If production of a certain medication type must take place, the policy says to produce medication of the same type for patients that do not require the medication now. This myopic policy provides a practical decision rule for the pharmacy department of ZGT.

Contents

1 Introduction

This chapter gives an introduction to this research. We explain the context of this research, give the main challenges and formulate the objective. We conclude this chapter by giving an outline of this thesis.

The ZGT pharmacy produces patient-specific medication. The production of this medication is done in a highly-controlled environment to eliminate contamination. The medication is administered to the patient at the hospital by medical personnel. However, an appointment can be cancelled (due to alarming blood levels for example). If the medication has been prepared, it cannot be used for different patients. The medication is then disposed.

Furthermore, raw material is used to produce the different types of medication. Raw material comes in vials, sometimes in different sizes. An opened vial cannot be used for later production. Any leftover raw material is considered as spillage. As several patients might require the same raw material, their production can be 'clustered'. This means that the medication is produced at the same time, to decrease the spillage of raw material. Some medication cannot be produced in advance due to a short shelf-life.

As we want to minimise the spillage and disposal costs, we should find an optimal planning that gives us the best tactics for the production of medication. Therefore, the goal of this research is to answer the following question:

"When should which medication be produced such that spillage and disposal costs are minimised?"

Stochastic processes play a prominent role in finding this optimal choice. These processes include the cancellation of appointments (as mentioned earlier) and the arrival of new patients with need for specific medication types.

The remainder of this report is organised as follows. In Chapter 2 we describe the research context in more detail. In Chapter 3 we review relevant literature. In Chapter 4 we introduce our mathematical model using a Markov Decision Process (MDP), which we (approximately) solve using the framework of Approximate Dynamic Programming (ADP), given in Chapter 5. In Chapter 6 we provide the results for our numerical experiments. Finally, in Chapter 7 we give the discussion and conclusions of our research.

2 Research context

In this chapter we introduce the context in which our research is conducted. First, we provide general information on ZGT. Afterwards, we focus on the aspects of the pharmacy department that are relevant to our research, including an overview of the current production procedure and the problems the pharmacy department experiences.

2.1 Ziekenhuisgroep Twente

ZGT is a hospital group situated in Twente, a region in the east of the Netherlands. They provide healthcare for approximately 390,000 inhabitants of the region (Ziekenhuisgroep Twente, 2018). ZGT has locations in Almelo and in Hengelo. The location in Almelo specialises in acute healthcare, whereas the location in Hengelo specialises in long-term outpatient healthcare (such as daily treatments). The pharmacy department in Hengelo is the focus of this research. Below, we elaborate on the production process at the pharmacy department.

2.2 Pharmacy department

At the pharmacy department of ZGT, medication is produced for administration within a hospital. In general the administration takes place in Hengelo itself, but it can also be transported to Almelo. The medication is not mass-produced, but is produced according to VTGM. VTGM is an abbreviation of Voor Toediening Gereedmaken (English: Prepare For Administration). The medication is parenteral (administration via injection) and is patient-specific. Examples of the produced medications are antibiotics, medication for arthritis and cytostatics.

To ensure high-quality medication without contamination, the medication is produced in a highly-controlled environment (clean room). The raw materials that are used for production are stored in sealed vials. If a vial is opened, any leftover material is spilled and cannot be used for later production unless a phaseal is used. A phaseal is a special cap that can be placed on an opened vial, such that the raw material can be used later on. The extension depends highly on the type of active substance and is not suitable for all raw materials.

In the current production procedure, the medication is prepared one day before the patient's appointment (if the medication's shelf-life allows) and the production follows the clustering protocol to reduce spillage. If the shelf-life is too short, the medication is prepared on the same day as the patient's appointment. Medication is also produced on the same day if the raw materials are considered expensive. This way, the pharmacy department tries to reduce disposal costs as the probability of same-day appointments being cancelled is low. If the patient's appointment is early in the morning, the medication is produced at the end of the day preceding the appointment.

The pharmacy department still experiences spillage even with the current clustering protocol. In addition, the procedure for expensive medication can result in inefficient production due to a lack of clustering. Therefore, the pharmacy wants to investigate the possibility of an improved production planning to reduce disposal and spillage costs.

3 Literature review

In this chapter we consider literature related to our research. We will look at scheduling problems in a medical context which include stochastic elements. Afterwards, we discuss literature related to scheduling using MDPs and we review literature on ADP, which provides a framework to find an approximate solution to the MDP model.

Given a set of patients, we want to find a subset of patients whose medication should be prepared now such that the costs are minimal. Finding this subset of patients is done in an on-line scheduling fashion: patients arrive over time. This is closely related to the process of on-line batch scheduling. Similar aspects of batch scheduling and our research are that we want to find a subset (batch) of our set of patients, there are different patient types, the scheduling is done on-line and patient characteristics are unknown until their arrival (Li et al., 2014; Nong et al., 2008; Poon and Yu, 2005a,b; Tian et al., 2012). Additionally, there can be a maximum on the size of the batches. Yuan et al. (2009) consider an on-line batch scheduling problem in which delivery times (i.e. deadlines) play a role. Their goal is to minimise the time until all jobs have been delivered, in which uncertain job arrivals are taken into account. Based on the processing time, the delivery time and the type of each job, they present an algorithm that minimises the time until all jobs have been delivered. They prove that their algorithm has the best possible competitive ratio, which is a measure for the quality of an on-line algorithm.

In batch scheduling, the objective is to minimise the makespan (i.e. the total time it takes to produce the items/products). Important aspects are release times and processing times in batch scheduling, whereas our problem does not focus on production times. In our problem, we aim to minimise disposal and spillage costs, for which we require more information than the information in batch scheduling as individual medication requests of the same type are not necessarily identical.

Some studies consider the theory of Markov Decision Processes (MDPs) to tackle on-line scheduling problems in healthcare (Hulshof et al., 2016; Patrick, 2012; Saure et al., 2012). Hulshof et al. (2016) consider a tactical planning in hospitals for patient admission planning and the allocation of hospital resource capacities. Their planning aims to achieve equitable access and treatment duration for patients. The MDP model considers different resource types, different patient queues and stochastic arrivals. They solve the problem by developing an ADP algorithm, which produces accurate approximations and has reasonable run times. Even though Hulshof et al. (2016) do not schedule in batches, their research shows that MDPs can be used to take different types of information and stochastic arrivals into account.

Patrick (2012) develops a policy using an MDP model for an outpatient clinic with no-show rates (i.e. cancellation). The goal of the policy is to find the optimal rewards/costs (consisting of revenue of servicing patients, idle time, patient access time and servicing through overtime) and provide a more consistent throughput. After the policy is found, he compares it to the current method called Open/Advanced Access (doing today's work today). He uses simulation to demonstrate that the MDP policy performs better than Open/Advanced Access.

Saure et al. (2012) produce a schedule for a cancer agency in which patients can have multiple appointment requests. They use a discounted infinite-horizon MDP as a model, with the goal to minimise the costs (based on waiting time, overtime and postponing of appointments). They take an uncertain number of new requests into account and indicate that cancellations and no-shows can be incorporated as well. To find the optimal schedule, they propose a linear program. Due to the size of the problem, they use approximate linear programming, in which they consider an affine function approximation for the costs. Their schedule works better than a myopic policy (which is in this case an approximation

of the current practice in the cancer agency).

Additionally, some studies consider planning and scheduling in the context of perishable goods (Chen et al., 2009; Haijema, 2011; Lütke Entrup et al., 2005). Chen et al. (2009) distinguish two different types of perishable goods: goods that decay continuously and deteriorate over time (e.g. fruits) and goods that have a fixed shelf-life (e.g. blood). They consider stochastic demands, but they solely focus on perishable goods that decay continuously. Lütke Entrup et al. (2005) consider the planning of yoghurt production in a deterministic setting. They develop three mixed-integer linear programmes that provide near-optimal solutions for such a planning.

On the other hand, Haijema (2011) considers a stochastic setting for planning and scheduling related to perishable goods. In his research, he develops an MDP to provide a policy for a hospital to order blood platelet concentrates (BCPs) from blood banks. Those BCPs are used to help patients recover from various treatments. Furthermore, the transfusion does not depend on the patients' blood type (in contrast to the transfusion of red blood cells). The MDP uses vector-valued actions and takes the unknown demand for the next epoch into account. The state consists of the number of BCPs in stock per age type. He uses an approximate scheme based on value iteration to solve the optimality equations. The optimal policy may perform much better than the existing policies and Haijema (2011) indicates that similar results may be obtained for other products that have a short shelf-life. In case of a larger state space, he proposes aggregation techniques to reduce the number of states and thereby decrease the computation time.

In general, Markov Decision Processes are powerful tools for making decisions under uncertainty (Alagoz et al., 2010; Puterman, 2014). It has many applications, and as mentioned before, can also be used to tackle scheduling problems. It provides a policy that gives the best action given a certain state. Finding an optimal policy for small-scale MDPs can be done with methods such as value iteration and policy iteration (Puterman, 2014). As Hulshof et al. (2016) and Saure et al. (2012) show, the size of real-life problems can be large and thus the methods proposed by Puterman (2014) cannot be used anymore. These large-scale problems can become intractable due to the three curses of dimensionality: a large state space, a large action space and a large outcome space (Powell, 2007).

Hulshof et al. (2016) use techniques from the field of Approximate Dynamic Programming (ADP) to develop an algorithm to approximately find an optimal solution. They used techniques described by Powell (2007) to tackle the curses of dimensionality. Even for toy-sized instances, the ADP-algorithm developed by Hulshof et al. (2016) is significantly faster than the dynamic programming algorithm and produces accurate approximations.

A difference between our research and the literature above, both batch scheduling and scheduling with MDPs, is that in our case a patient is not completely defined by his/her (medication) type. Besides a medication type, the amount of medication that a patient requires is determined by their body surface area. Furthermore, the time of the appointment for administration and the disposal probability are not (only) determined by the patient's type. Hence the "uniqueness" of patients distinguishes our research from the literature, as in literature patients are determined by their type.

The research thesis by Medendorp (2021) also focused on improving the complete production processes that take place at the pharmacy department at ZGT. Medendorp (2021) also provides an improved production planning. However, this improved planning was designed for a period of time *in the past*, i.e. for a period of time in which all requests and cancellations are known. Where Medendorp (2021) provided a planning for such a deterministic setting, in this research we intend to take stochastic future arrivals and cancellations into account.

4 Mathematical model

In this chapter, we develop the mathematical model that we use to solve our optimisation problem. We choose the theory of MDPs to tackle this optimisation problem, as has been used in the context of perishable goods (see Chapter 3). Below, we will describe how we model the problem as an MDP.

For the MDP, we require decision epochs, a state and action space, transition probabilities and costs.

Decision epochs
This is an infinite horizon problem. At the start of the day, the decision is made which medications should be prepared. Medication can be prepared at most T days in advance, hence all patients that have an appointment in the next T days are take into account.

State space
We suppose that we have n medication requests in the system. The state of the system is given by $S = \{\vec{y}, \vec{m}, \vec{\tau}, \vec{\eta}, \psi\} \in \mathcal{S}$. It contains the following information:

- The type of the patients $\vec{y} = (y_1, y_2, \ldots, y_n)$. Patients can be of type $j \in \mathcal{J}$, where $\mathcal{J} = \{1, 2, \ldots, J\}$. The variable $y_i \in \mathcal{J}$ denotes the type of patient i. Only medication for patients of the same type can be clustered if the medication is produced on the same day. A medication request only requires one type of raw material.
- The amount of raw material needed for preparing medication $\vec{m} = (m_1, m_2, \ldots, m_n)$, with $m_i \in \mathbb{R}$. The value of m_i is the amount of raw material that is needed for the i-th medication request (of type y_i).
- The deadline of the preparation for the medication $\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_n)$, with $\tau_i \in \{0, 1, \ldots, T-1\}$. The value τ_i indicates the amount of days left before patient i's appointment of administration. When $\tau_i = 0$, the medication must be prepared immediately.
- The disposal probability $\vec{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)$, with $\eta_i \in [0, 1]$ the probability that the appointment of patient i is cancelled (i.e., the probability that the medication is disposed if it has been produced).
- The current day, $\psi \in \{0, 1, 2, \ldots, 6\}$. Here, $\psi = 0$ indicates it is Monday, $\psi = 1$ indicates it is Tuesday and so on. See the table below:

Day	**Values for ψ**
Monday	0
Tuesday	1
Wednesday	2
Thursday	3
Friday	4
Saturday	5
Sunday	6

In the remainder of this report, we use the words "medication request i" and "patient i" interchangeably. There is a difference between these two in practice, however, as a patient may have multiple medication requests. As these requests are processed individually and as in general patients only have one request, we use both terms to describe the same thing.

Action space
The action space $\mathcal{A}(S)$ is the set of possible actions that can be chosen in state S. An action is denoted by $\vec{a} = (a_1, a_2, \ldots, a_n) \in \mathcal{A}(S)$. The entries a_i are defined in the following way:

$$a_i = \begin{cases} 1, & \text{if the medication for patient } i \text{ is produced now,} \\ 0, & \text{otherwise.} \end{cases}$$

The set $\mathbb{A}$ denotes the set of patients for which $a_i = 1$. This definition will be useful for the explanation on how states change.

Transition probabilities
The deterministic part of the MDP consists of the removal of patient i which has $a_i = 1$. There are also stochastic processes that have to be taken into account for the change of the state. The first aspect that we take into account is that patients for which $a_i = 0$ may leave with a certain probability. The second aspect is that new patients may arrive to the system.

Cancellation
The appointment of patient i is cancelled with probability η_i. Let $\mathcal{L} \subseteq \vec{x}$ be the set of patients of which the appointment is cancelled. Then $\mathcal{L}^C \subseteq \vec{x}$ is the set of remaining patients. The probability that patients in $\mathcal{L}$ leave is given by

$$P(\mathcal{L}) = \prod_{i \in \mathcal{L}} \eta_i \cdot \prod_{i' \in \mathcal{L}^C} (1 - \eta_{i'}).$$

Arrival
We assume that the number of patients that arrive at the start of a decision epoch follows a Poisson distribution with rate λ. Let X be the random variable that records the number of arrivals. The probability that n patients arrive is then given by:

$$P(X = n) = \frac{e^{-\lambda} \cdot \lambda^n}{n!}. \tag{1}$$

We assume that the number of arriving patients is bounded to ensure that we have a finite state space and outcome space. Given that n patients arrive, we have to find the probability that n_1 patients of type 1 arrive, n_2 patients of type 2 and so on, denoted by $p(n_1, n_2, \ldots, n_J)$. Let p_j be the probability that an arriving patient is of type j. This situation can be modelled with the multinomial distribution, as we perform n experiments with j possible outcomes and we assume the outcome of any one of the n experiments does not affect the outcome of the other $n - 1$ experiments (Ross, 2007). Under the assumption that in total n patients arrive, the multinomial distribution for our problem looks like this:

$$p(n_1, n_2, \ldots, n_J) = n! \cdot \prod_{j=1}^{J} \frac{p_j^{n_j}}{n_j!},$$

with $n_1 + n_2 + \ldots + n_J = n$. The probability that n patients arrive and that n_j patients of type $j \in \mathcal{J}$ arrive is denoted by $\pi_n(n_1, n_2, \ldots, n_J)$ and can be computed by

$$\begin{aligned}\pi_n(n_1, n_2, \ldots, n_J) &= P(X = n) \cdot p(n_1, n_2, \ldots, n_J) \\ &= \frac{e^{-\lambda} \cdot \lambda^n}{n!} \cdot n! \cdot \prod_{j=1}^{J} \frac{p_j^{n_j}}{n_j!} \\ &= e^{-\lambda} \cdot \lambda^n \cdot \prod_{j=1}^{J} \frac{p_j^{n_j}}{n_j!}.\end{aligned}$$

We assume that arriving patients do not cancel at the same time as they arrive.

For a patient i that is still in the system, we have that τ_i decreases by one as the deadline comes one day closer. Furthermore, the value of ψ is increased by 1 as we move a day into the future. When $\psi = 7$, we set $\psi = 0$ (as explained in the description of the state space above).

Costs
The costs consist of two different parts: the spillage costs and the disposal costs. We will describe both of these separately below.

The spillage costs are denoted by $C_1(S, \vec{a})$. Let $Y_j(S, \vec{a})$ be the volume of spillage of raw material j when action $\vec{a}$ is chosen in state S, and let P_j be the costs of one unit of volume of raw material j. Then the costs of $C_1(S, \vec{a}) = \sum_{j=1}^{J} Y_j(S, \vec{a}) \cdot P_j$.

The expected disposal costs are denoted by $C_2(S, \vec{a})$. The disposal costs depend on the type of medication/patient i and are denoted by R_i. The probability that the prepared medicine of patient i is disposed is given by η_i. Then the expected disposal costs are given by $C_2(S, \vec{a}) = \sum_{i=1}^{n} a_i \cdot \eta_i \cdot R_i$, with $a_i = 1$ if and only if the medication for patient i is prepared now.

We let $C(S, \vec{a}) = C_1(S, \vec{a}) + C_2(S, \vec{a})$, the total spillage and (expected) disposal costs.

Bellman optimality equations
The optimality equations (Bellman's equations) are used to compute the value of being in a certain state (Puterman, 2014). As we are dealing with an infinite-horizon problem, we have to take the future costs into account. Therefore, we introduce a discount factor β. The optimality equations are given by

$$V(S) = \min_{\vec{a} \in \mathcal{A}(S)} \left(C(S, \vec{a}) + \beta \sum_{S' \in \mathcal{S}} p(S'|S, \vec{a}) V(S') \right).$$

In this equation, S' denotes a possible state that we transition to based on the current state S and the action $\vec{a}$. The quantity $V(S)$ for $S \in \mathcal{S}$ denotes the value of being in a certain state S. In Chapter 5 we explain how to determine the value of $V(S)$ for any state S. This concludes our MDP model.

5 ADP approach

As most real-life problems are too vast to solve with standard methods of MDP, an algorithm is developed within the framework of Approximate Dynamic Programming (ADP). This is necessary as the problem suffers from the three curses of dimensionality (Powell, 2007): a large state space, a large action space and a large outcome space. Below we briefly explain why our problems suffers from these three curses. The state of the system is vector-valued. Disregarding the values of the entries of the vectors themselves, this gives a lower bound on the number of states, which equals $n^4 \cdot 1$. For $n = 50$ and $J = 50$, which is a small number of patients and patient types, we would have that each of the 50 patients could have one of 50 patient types, leading to $\frac{50^{50}}{50!} = 2.9 \cdot 10^{20}$ combinations. As we have not taken into account the possibilities for $\vec{m}$, $\vec{\tau}$, $\vec{\eta}$ and ψ, the actual amount of states will be even larger.

For the action space, suppose that at most 10 patients can be scheduled in one day (in practice this amount is much larger). Then the number of actions (with $n = 100$) equals $\sum_{i=0}^{10} \binom{100}{i} \approx 1.9 \cdot 10^{13}$, providing a lower bound on the size of our action space.

For the outcome space, we disregard the computations for cancellations as they will increase the number of outcomes. Suppose that between 0 and 10 patients could arrive. So, assuming that $J = 50$, we have $\sum_{i=0}^{10} \frac{50^i}{i!} \approx 3.3 \cdot 10^{10}$ possible outcomes.

As we have established that our problem is too large to solve with the standard methods of MDPs, we will explain how ADP is used. We aim to find a policy $A^\pi(S)$, which is a function that returns a decision $\vec{a} \in \mathcal{A}(S)$, under the policy $\pi \in \Pi$. The set Π refers to the set of potential policies. $\mathcal{A}(S)$ denotes the set of feasible decisions for a given state S. The notation for the policy π should not be confused with the notation for the arrival probability of n patients, π_n, as explained in Chapter 4.

The evolution of the system over time as a result of the decision $\vec{a}$ and the random information W is captured by the transition function S^M, which has the Markov property. It is defined by

$$S' = S^M(S, \vec{a}, W). \tag{2}$$

The Post-Decision State
To tackle the large outcome space, the post-decision state variable S^a is introduced. This post-decision state is reached directly after making a decision, but before the arrival of any new information W. Hence, it is only based on the current pre-decision state S and the action $\vec{a}$. This reduces the complexity of the future costs. Note that we omit the vector notation $\vec{a}$ in the superscript of the post-decision state variable, where we simply write a for readability purposes.

A new transition function is introduced, which captures the transition from the pre-decision state S to the post-decision state S^a. This function is given by

$$S^a = S^{M,a}(S, \vec{a}). \tag{3}$$

After choosing action $\vec{a}$, the post-decision state is $S^a = \{\vec{y}^a, \vec{m}^a, \vec{\tau}^a, \vec{\eta}^a, \psi^a\}$. The action $\vec{a}$ influences the state in the following ways:

- $\vec{y}^a$, $\vec{m}^a$ and $\vec{\eta}^a$ are obtained by removing the entries that correspond to the patients in $\mathbb{A}$ from the pre-decision state. Hence, for all $i \in \mathbb{A}$, we remove the i-th entries from all vectors.

- $\vec{\tau}^a$ is obtained by removing the times for the patients in τ that correspond to the patients in $\mathbb{A}$. The values for the patients that are not in $\mathbb{A}$ are reduced by 1.
- $\psi^a = \psi + 1 \mod 7$, as this will only be updated when moving to a different stage, independent of the action. As soon as the new week begins on Monday, we set $\psi = 0$

The post-decision state is used to provide a value estimate from a particular combination of a pre-decision state and a decision. So, it is used as an estimate of the future pre-decision state.

Now, the DP formulation can be written as

$$V(S) = \min_{\vec{a} \in \mathcal{A}(S)} (C(S, \vec{a}) + \beta V^a(S^a)).$$

The value function of the post-decision state is given by

$$V^a(S^a) = \mathbb{E}\{V(S)|S^a\}.$$

The value function of the post-decision state $V^a(S^a)$ will be approximated. This approximation is denoted by $\bar{V}^k(S^a)$, in which k denotes the iteration counter as the approximation will be computed iteratively.

Then, the optimal decision should be made by solving

$$\tilde{A}^k = \arg \min_{\vec{a} \in \mathcal{A}(S)} (C(S, \vec{a}) + \beta \bar{V}^{k-1}(S^a)). \tag{4}$$

This will give the decision that minimises the value $\hat{v}^k$ for state S in the k-th iteration. This function is given by

$$\hat{v}^k = \min_{\vec{a} \in \mathcal{A}(S)} (C(S, \vec{a}) + \beta \bar{V}^{k-1}(S^a)).$$

After making decision $\tilde{A}^k$, and finding an approximation for the value (which is $\hat{v}^k$), the value function approximation $\bar{V}^{k-1}(S^a)$ is updated.

To find an appropriate approximation for the value function, the goal is to minimise the difference between the future cost approximation (so $\bar{V}^k(S^a)$) and the approximation $\hat{v}^k$ with the updating function, as k increases.

5.1 Value function approximation

The state space is too large to compute the value function for each different state. In this section we explain how we approximate the value of being in a certain state, by using basis functions.

5.1.1 Basis functions

The large number of states prohibits us to estimate the value for each particular state. For the problem to become tractable, we introduce basis functions. Basis functions are used to estimate the value $V(S)$ of being in a certain state S (in the case of the post-decision state we estimate the value of being in state S^a). To do this, we come up with certain features that might contribute to the value of being in a state.

First of all, we expect that $V(S^a)$ depends on the amount of patients of each type. More patients of type j might give the pharmacy the opportunity to cluster the production and with that, minimise spillage.

Additionally, another feature we distinguish is the amount of medication that all patients need (again per patient type). For example, if a vial of type j contains 100 mL of raw material it is useful to know if the total amount of raw material requested is 40 mL or 90 mL as this affects spillage.

The next feature takes the deadlines into account. In case the patients have a deadline 'far' into the future, we have more freedom to choose when to schedule them and reduce spillage. Therefore, the next feature takes into account the number of patients that have a 'close' deadline. We define a close deadline to be less than or equal to ς days into the future.

Due to the limited shelf-life of medication, it is important to know if the medication is prepared on a Monday or on a Thursday for example. On Monday, it is possible to prepare medication with a shelf-life of two days and with that, cluster the production. On Thursday, however, we do not have the freedom to prepare medication for two days as the administration of the medication generally takes place on weekdays. Therefore, the day of the week is also one of the features.

As every set of basis functions requires a constant basis function as well, we add this as our last feature (Powell, 2007). The summary of our features and basis functions is given in Table 1.

Table 1: *The features and the corresponding basis functions.*

f	Feature	Basis function	# basis functions
f_1	Constant	$\phi_{f_1}(S^a) = 1$	1
f_2	Amount of patients	$\phi_{f_2}(S^a) = n_j, j \in \mathcal{J}$	J
f_3	Amount of medication	$\phi_{f_3}(S^a) = \sum_i m_j, j \in \mathcal{J}$	J
f_4	Patients with close deadline	$\phi_{f_4}(S^a) = n(\tau), \tau \leq \varsigma$	ς +1
f_5	Day	$\phi_{f_5}(S^a) = \psi + 1 \mod 7$	1

5.2 Finding the decision

To solve for which action we have minimal direct and future costs, we introduce a mixed-integer program (MIP) which is given below. We introduce the set $\mathcal{I}_j$, where $i \in \mathcal{I}_j$ if $y_i = j$. Hence, the set $\mathcal{I}_j$ contains all patients of type j.

$$\min_{\vec{a}\in\mathcal{A}(S)} \left(C(S,\vec{a}) + \mathcal{C}z + \beta \sum_{f\in\mathcal{F}} \theta_f \phi_f(S^a) \right) \tag{5}$$

$$\text{s.t.} \quad \sum_i a_i - u_\psi \leq z \tag{6}$$

$$b_j = -\sum_{i\in\mathcal{I}_j} a_i \cdot m_i \mod M_j \qquad \forall j \tag{7}$$

$$a_i = 0 \qquad \forall i \text{ with } \tau_i > l_i \tag{8}$$

$$a_i = 1 \qquad \forall i \text{ with } \tau_i = 0 \tag{9}$$

$$a_i \in \{0,1\} \qquad \forall i \tag{10}$$

$$z \geq 0 \tag{11}$$

$$\vec{y}^a = \vec{y} \setminus \mathbb{A} \tag{12}$$

$$\vec{m}^a = \vec{m} \setminus \mathbb{A} \tag{13}$$

$$\vec{\eta}^a = \vec{\eta} \setminus \mathbb{A} \tag{14}$$

$$\vec{\tau}^a = \vec{\tau} \setminus \mathbb{A} - \mathbb{1}^{n-\sum_i a_i} \tag{15}$$

$$\psi^a = \psi + 1 \mod 7 \tag{16}$$

This essentially finds the decision that gives the minimal costs and minimal value for the post-decision state. The objective function and the constraints of the MIP are explained below:

- The objective function (5): We minimise the expected costs with the immediate costs $C(S,\vec{a})$ and the future costs, which are taken into account with the post-decision state, consisting of the basis functions. The term $\mathcal{C}z$ indicates the penalty costs. These costs are used to take production constraints into account (see the point below).
- Constraint (6): No more than u_ψ medicines can be prepared in one day, which depends on ψ. When $\psi \in \{5,6\}$, then $u_\psi = 0$ as no production takes place in the weekend. Production that exceeds u_ψ receives a penalty based on the value of z.
- Constraint (7): The amount for spillage of raw material j is b_j. This variable is part of the costs $C(S,\vec{a})$ in the objective function, as the spillage costs equal $\sum_{j=1}^{J} b_j \cdot P_j$ (see Chapter 4). The parameter M_j denotes the amount of raw material in an unopened vial of type j.
- Constraint (8): The medication of patient i cannot be prepared in case shelf-life l_i is too short. The shelf-life for patient i is determined by its medication type and it is measured in the amount of days. If the prepared medication of patient i is not administered for l_i days, it expires and is disposed.
- Constraint (9): The medication must be prepared immediately in case the deadline has been reached.
- Constraint (10): The decisions are binary.
- Constraint (12): The post-decision variable for the patient type is computed.
- Constraint (13): The post-decision variable for the amount of medicine is computed.

- Constraint (14): The post-decision variable for the cancellation probability is computed.
- Constraint (15): The post-decision variable for the time until patients' appointments is computed.
- Constraint (16): The post-decision variable for the day is increased by 1. If the value of 7 is reached (a new Monday), it goes to day 0.

To find an (approximate) optimal policy for our ADP, we have to find optimal decisions for each state (which is done with the MIP) and we have to update our approximation of the value of being in a state. In Algorithm 1 we propose the steps to solve the problem.

Algorithm 1: The ADP algorithm to solve the problem.

1. **Initialisation:** Choose an initial approximation $\bar{V}^0$. Set $k = 1$ and set the maximum number of iterations N. Set the initial state to S_1.
2. **Update:** Solve the MIP (5) to obtain $\vec{a}^k$.
   Update the approximation $\bar{V}^k(S^{k,a}) = \sum_{f \in \mathcal{F}} \theta_f^k \phi_f(S^{k,a})$, where θ^k can be obtained by Algorithm 2.
   Find the post-decision state with $S^{k,a}$.
   Obtain a sample realisation W^{k+1} and compute the new pre-decision state.
3. **Increment/Stop:** Increment k by 1. If $k \leq N$, go to Step 2.
4. **Return:** Return $\bar{V}^N$.

The value function is updated in Step 2. A method to do this is "Recursive Least Squares" (Powell, 2007), where we use non-stationary data. This way, we are able to put more weight on recent observations instead of having equal weight on each observation.

Algorithm 2: The recursive least squares algorithm to update θ.

1. **Error:** Compute the error $\hat{\epsilon}^k$:

$$\hat{\epsilon}^k = \bar{V}^{k-1}(S^a) - \hat{v}^k.$$

2. **Update:** Update γ^k and B^k:

$$\gamma^k = \nu_k + (\phi(S^a))^T B^{k-1} \phi(S^a),$$
$$B^k = \frac{1}{\nu_k}(B^{k-1} - \frac{1}{\gamma^k}(B^{k-1}\phi(S^a)(\phi(S^a))^T B^{k-1}).$$

3. **Update:** Update H^k:

$$H^k = \frac{1}{\gamma^k} B^{k-1}.$$

4. **Update** Update θ^k:

$$\theta^k = \theta^{k-1} - H^k \phi^k(S^a)\hat{\epsilon}^k.$$

5. **Return:** Return $\bar{V}^k(S^a) = \sum_{f \in \mathcal{F}} \theta_f^k \phi_f(S^a)$.

The matrix $B^0 = \epsilon I$, where ϵ is a small constant. This constant is chosen to be $\epsilon = 10^{-6}$, which is suitable for our problem.

The value for ν_k indicates the weight that is put on the prior observations. It ranges between 0 and 1, which decreases the weight on prior observations. It is defined in the following way:

$$\nu_k = \begin{cases} 1, & \text{stationary} \\ 1 - \frac{\delta}{k}, & \text{non-stationary} \end{cases},$$

where k is the iteration counter. The constant δ is determined in the experiments. The experiments showed that $\delta = 0.6$ is suitable for this problem.

6 Results

In this chapter, we describe the results for the numerical experiments.

We implemented Algorithms 1 and 2 in Python. The sample realization W^{n+1} is obtained as follows. For each patient i, we generate a random number between 0 and 1 from the uniform distribution. If the random number is larger than η_i, the patient stays in the system. Otherwise, patient i is removed and the appointment is cancelled.

For the arrival of new patients, we generate a random amount of arrivals according to a Poisson distribution with rate λ. In the weekends, we assume there are no new arrivals. Then, based on the number of arrivals, we use the multinomial distribution to determine which patient types arrive. For each patient i, we generate random integers (from the uniform distribution) that give m_i and τ_i.

Firstly, we investigate a toy-sized instance in Section 6.1. Then in Section 6.2 we provide the numerical results for the ZGT case. Using the toy-sized instance enables us to generate results faster than the ZGT case as the latter has large computational time. Furthermore, the toy-sized instance can ease our search for specific behaviour or patterns in the decision making process and with that, provide useful information for the ZGT case.

6.1 Experiments with toy-sized instance

For the toy-sized instance, the parameters that have been used are given and explained below. These parameters have been chosen such that the problem is not too large (less patient types and less arrivals) and that it does resemble the ZGT case (horizon length for example). We chose $\beta = 0.95$ as this is standard in Markov Decision Theory and the value for $N = 10,000,000$ was required to achieve convergence. The maximum production is chosen to be approximately 1.6 times larger than the arrival rate as has been used for the ZGT case as well (which is based on data and discussion with the ZGT pharmacy). The close deadlines and the penalty costs are the same values as for the ZGT case.

Parameter	**Value**
Horizon length in days (T)	4
Number of patient types (J)	5
Maximum production on a weekday ($u_\psi, \psi \leq 4$)	16
Close deadlines in days (ς)	3
Discount factor (β)	0.95
Number of iterations (N)	10,000,000
Arrival rate per day (λ)	10
Penalty costs for exceeding maximum production ($\mathcal{C}$)	1,000

As $J = 5$, we have selected 5 different medications to be part of the experiments for the toy-sized instance. These medications have the following characteristics:

Table 2: *Medication parameters for the toy-sized instance.*

Medication	Volume (vial)	Price (p/u)	Shelf-life	Min	Max	Arrival fraction	Disposal probability
1	100	3.644	0	139	291	0.487	0.015
2	500	2.58	3	698	1152	0.035	0.005
3	150	0.05	7	139	900	0.212	0.010
4	1000	0.00321	4	294	6303	0.133	0.014
5	30	77.33	14	210	840	0.133	0.008

We briefly explain the information given in each column. Each medication has a name, an amount that is in a vial, the price per unit of the medication (in euro), the amount of days the medication can be made in advance, the minimum and maximum amount of units of medication a patient could need, the fraction of arrivals that correspond to this medication and the probability that the request is cancelled.

We clarify that these characteristics in Table 2 are based on five real medication types. They have been selected from the complete list of medication of ZGT. The reason for their selection is such that there is some variability in all of the characteristics. Finally, we do note that the arrival fractions have been normalised in this toy-sized instance such that the total arrival fraction equals 1.

6.1.1 Convergence

To approximate the value of a state, we introduced basis functions in Section 5.1.1. Each basis function ϕ_f has a coefficient, θ_f. Using the recursive least squares algorithm (see Algorithm 2), we try to determine the value of these coefficients.

As the number of iterations is very large, we store the values of θ_f, $f \in \mathcal{F}$, every 100 iterations. In Figure 1, the development of each value θ_f is shown as the number of iterations increases. Each graph seems to converge to a steady value, even though some have not reached this steady state yet as a nonzero slope can be observed from this graph.

The graph that stands out is the one that has a value of approximately 8000. This graph corresponds to f_1, the constant feature. This implies that the minimal costs always amount to 8000, independent of the action that is chosen.

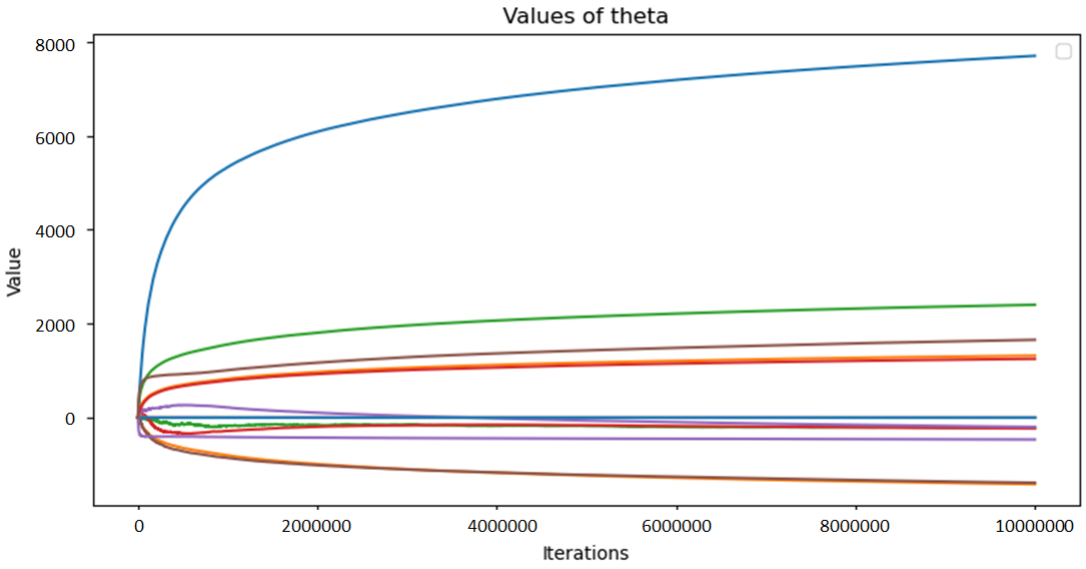

Figure 1: *Values of θ_f for each feature.*

To understand how much the coefficients are updated in each iteration, we can compute the value of $\sum_{f \in \mathcal{F}} |\theta_f^k - \theta_f^{k-1}|$, i.e. the total absolute difference of the value of θ_f in iteration k (θ_f^k) and in iteration $k-1$ (θ_f^{k-1}). The value of this total absolute difference can be seen in Figure 2 for each iteration.

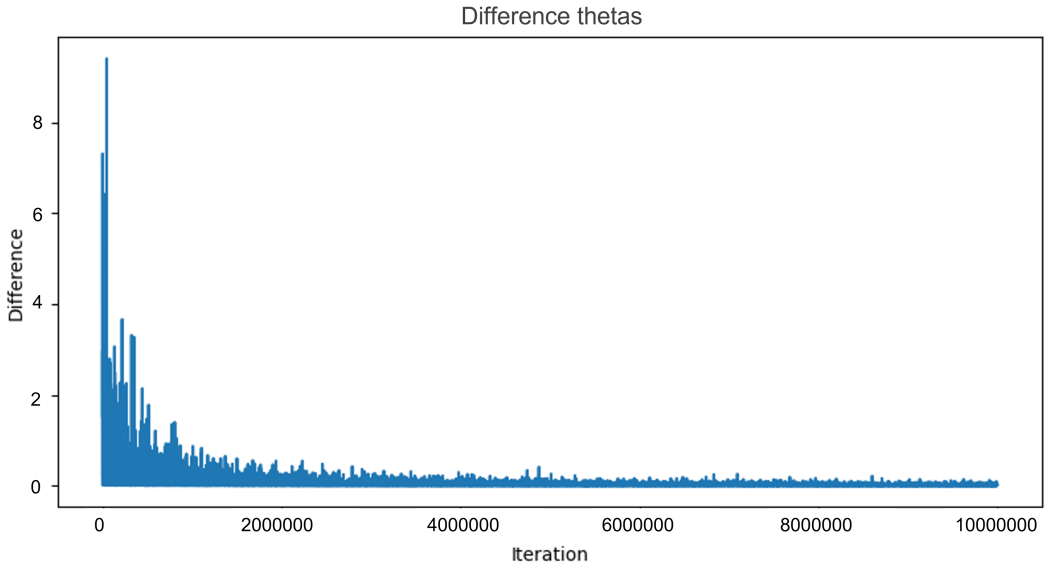

Figure 2: *The value of $\sum_{f \in \mathcal{F}} |\theta_f^k - \theta_f^{k-1}|$.*

As can be seen in Figure 2, the differences of subsequent values of θ_f^k are relatively large. As the iteration counter k increases, this sum of differences seems to decrease. For large values of k, we

want to know the value of this sum. In Figure 3, we plotted this sum for the last 250,000 iterations. Note, as this difference is only plotted every 100 iterations, we see 2,500 values. None of these values exceeds the value of 0.10, which implies that the values of theta change slightly between iterations. In combination with Figure 1, it seems the values of θ_f are converging.

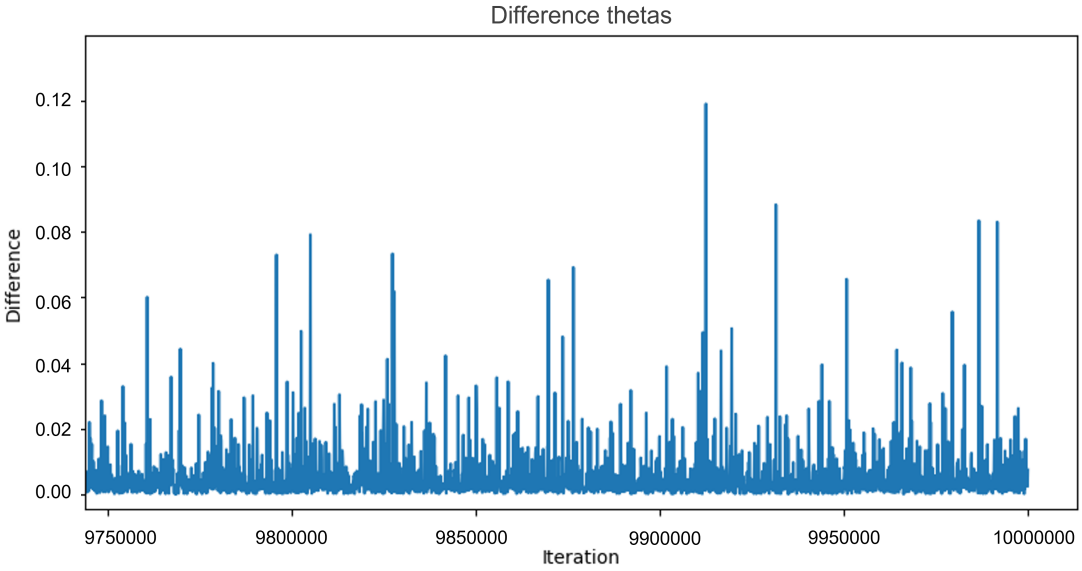

Figure 3: *The value of* $\sum_{f\in\mathcal{F}} |\theta_f^k - \theta_f^{k-1}|$ *for* $k = 9,750,000$ *to* $k = 10,000,000$.

As $\sum_{f\in\mathcal{F}} |\theta_f^k - \theta_f^{k-1}|$ is smaller than 0.12 for the last 250,000 iterations, we believe that we can use θ_f^N to investigate the performance of the ADP approach versus a myopic policy in a toy-sized instance. For real-life problems, we might set a stricter stopping criterion. As we cannot guarantee that $\sum_{f\in\mathcal{F}} |\theta_f^k - \theta_f^{k-1}|$ will reach a stricter stopping criterion and as the current value of N already results in a computation time of over 10 hours, we believe that the current estimate is adequate (taking into account that some values of θ_f are of the order of 1,000). Below, we give the values of θ that are used in the investigation of the performance of the ADP approach. For those features that have multiple "categories", we introduce new notation. For example, our second feature f_2 is the amount of patients. As this is divided into the amount of patients of each type we write $f_{2,j}$, indicating feature 2 and category (in this case patient type) j.

Table 3: *Values for θ_f for all $f \in \mathcal{F}$.*

Feature f	**Value of** θ_f
Constant f_1	7,698.05260
Amount of patients of type 1 $f_{2,1}$	-1,416.86861
Amount of patients of type 2 $f_{2,2}$	-219.653036
Amount of patients of type 3 $f_{2,3}$	-229.614452
Amount of patients of type 4 $f_{2,4}$	-198.830478
Amount of patients of type 5 $f_{2,5}$	-1,385.78986
Amount of medication of type 1 $f_{3,1}$	2.94319537
Amount of medication of type 2 $f_{3,2}$	0.582988875
Amount of medication of type 3 $f_{3,3}$	0.113707810
Amount of medication of type 4 $f_{3,4}$	0.000461934444
Amount of medication of type 5 $f_{3,5}$	0.247931238
Amount of patients with deadline today $f_{4,0}$	1,319.45230
Amount of patients with deadline tomorrow$f_{4,1}$	2,401.00030
Amount of patients with deadline the day after tomorrow $f_{4,2}$	1,251.45593
Amount of patients with deadline in three days $f_{4,3}$	-465.648721
Day f_5	1,655.80571

6.1.2 Performance compared to myopic policy

The values of θ_f we computed in Section 6.1.1 are used in our MIP to take future states into account. We hope to select the optimal action that minimises the direct costs and the costs that future states might bring. We compare the ADP-approach (called look-ahead) to a myopic policy, which only takes into account the direct costs and thus ignores the contribution of any future states an action might bring us.

We start with an initial state S_0, which is generated randomly according to the data that is also used to find the values of θ. Then we solve the MIP from Section 5.2 in two ways: one with the values of θ_f we computed and one with $\beta = 0$. The latter is the myopic policy as this ensures the future costs of states are not taken into account.

Both approaches find their respective optimal decision. Based on this decision, the states are updated. Using statistically equivalent external arrival information, the state is updated again and the MIP computes a new optimal decision. We simulate this process for one year and see how the two approaches perform. Note that only the initial state of the two MIPs is the same, as their different decisions will result in different post-decision states. In Figure 4 the costs both policies give in one year are given.

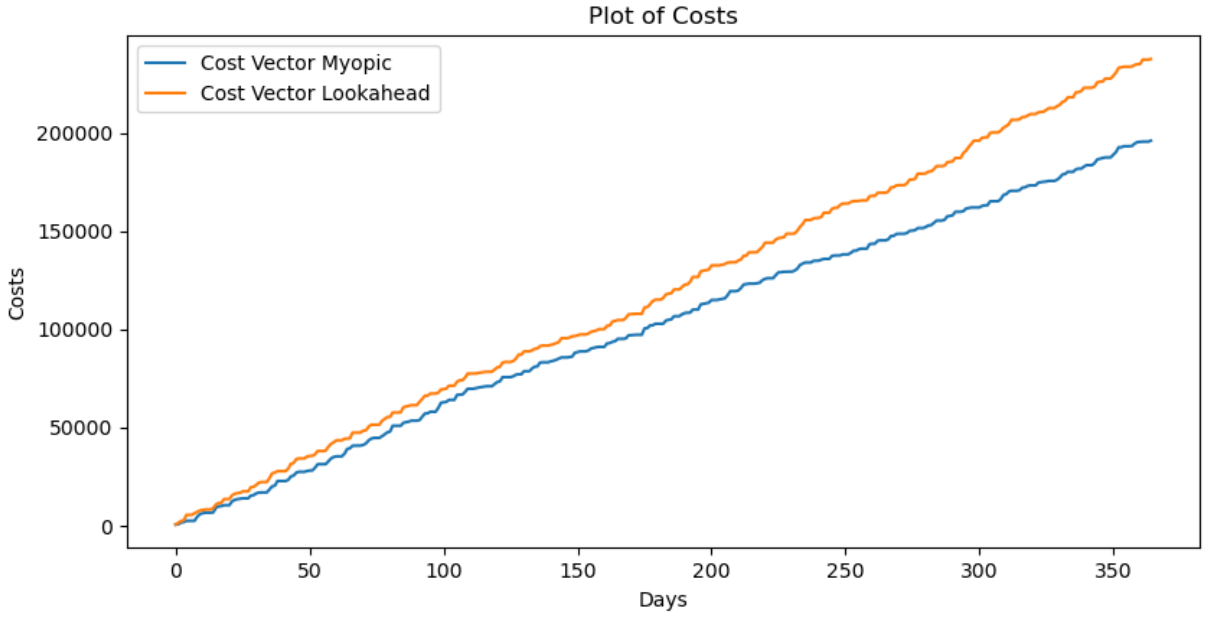

Figure 4: *The costs of one year of the myopic policy and the look-ahead policy with* $\beta = 0.95$.

As can be seen from the graph, the myopic policy actually performs better than the ADP approach (i.e. look-ahead policy). One of the possible factors is that the look-ahead policy values the future costs excessively. To decrease this dependency on the future costs, we used different values of β to analyse its effect. In Figure 5 we plot the costs of the myopic and look-ahead policy using $\beta = 0.05$.

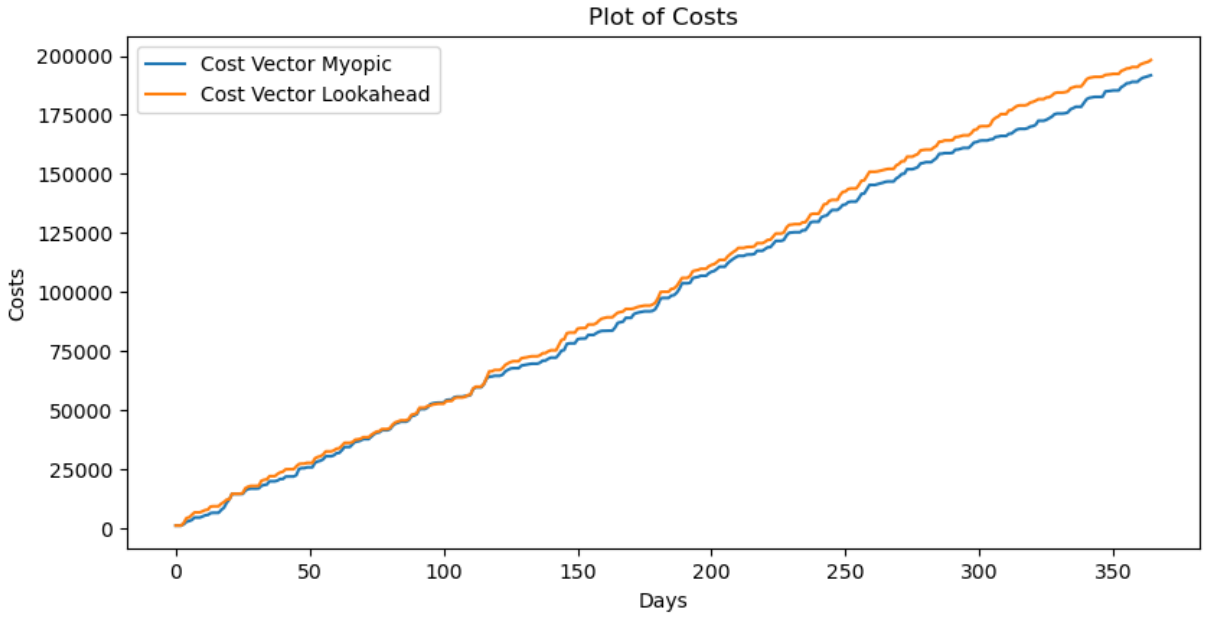

Figure 5: *The costs of one year of the myopic policy and the look-ahead policy with* $\beta = 0.05$.

Decreasing the value of β shows that the look-ahead policy gets closer to the myopic policy (as the myopic policy uses $\beta = 0$). In this particular graph, the myopic policy outperforms the look-ahead policy. However, for the same value of β, in other attempts the look-ahead policy outperformed the myopic policy. We can see that decreasing the value of β improves the performance of the look-ahead policy, but it does not result in a policy that performs better than a myopic policy. Hence, using a myopic policy would be preferable as there is no need to go through the process of ADP.

6.1.3 Performance myopic policy compared to ZGT pharmacy policy

Below, we briefly describe the policy that the pharmacists currently use to prepare the medication. Afterwards, we will compare this to the myopic policy.

In general, medication is prepared one day before the patient's appointment, if the shelf-life of the medication is large enough. Otherwise, the medication is prepared on the same day as the patient's appointment. The production is always done according to the clustering protocol. There is an exception to this rule: if the medication is too expensive (of which the pharmacy has a list), the medication is also prepared on the same day as the patient's appointment, regardless of the shelf-life. In rare cases, if the patient's appointment is early in the morning, the pharmacy does prepare the medication one day in advance (after 14:00). Due to the model taking into account days instead of day-segments, these rare cases are not considered in this research. In the toy-instance there are two medication types that are considered expensive: the first and the last one (see Table 2).

We do mention that the actual procedure at ZGT might perform slightly better than the policy we described in this research. In some cases, if during the day a medication request arrives for the next day, the pharmacist assistants ask the pharmacist to add the request to their software so it can be added for clustering and reduce spillage costs that way.

The performance of the current ZGT policy and the myopic policy is compared in the same way that the look-ahead policy and myopic policy were compared. The initial state is generated randomly and both policies make a decision on how to prepare the medication. Then, statistically equivalent external arrival information is used to update the state. The process continues for one year. The costs can be seen in Figure 6.

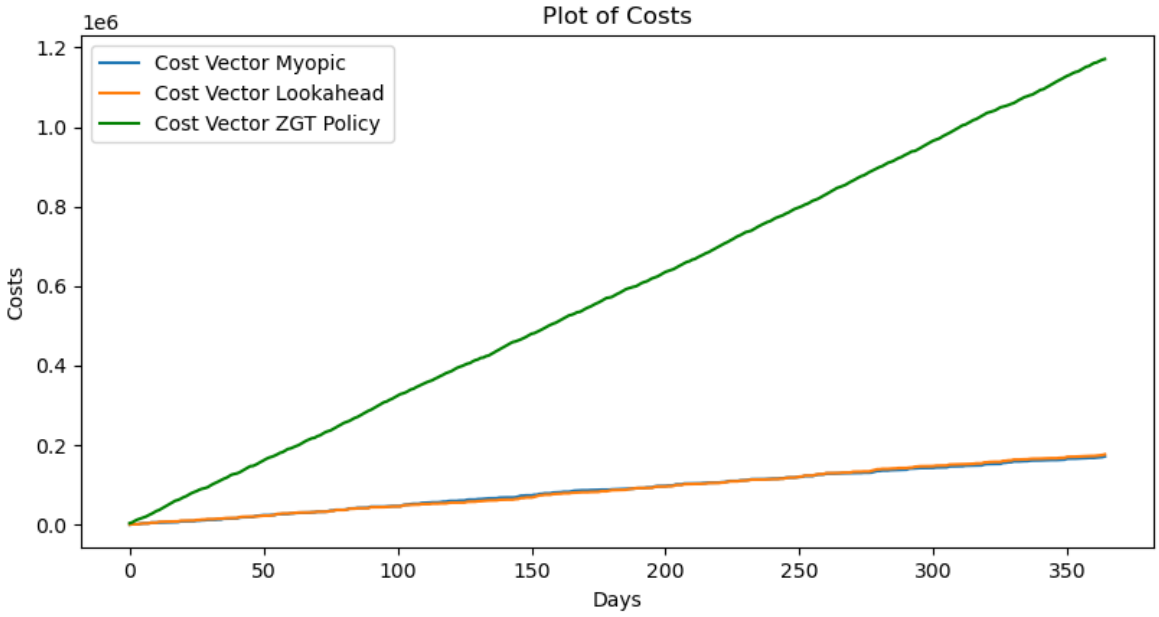

Figure 6: *The costs of one year of the ZGT policy, the look-ahead policy ($\beta = 0.05$) and the myopic policy.*

This figure shows that the myopic policy as well as the ADP policy much perform better than the current ZGT policy. A possible explanation is that our toy-instance contains two medication types that are considered expensive. As the myopic policy focuses on decreasing spillage, this gives a major reduction in the total costs. Furthermore, the arrival of these "expensive patients" is quite frequent compared to the real-life situation and can therefore give a major reduction in the total costs.

To show that both the myopic policy and the look-ahead policy reduce spillage compared to the ZGT policy, we computed the daily spillage over a year. This data can be seen in the boxplot in Figure 7.

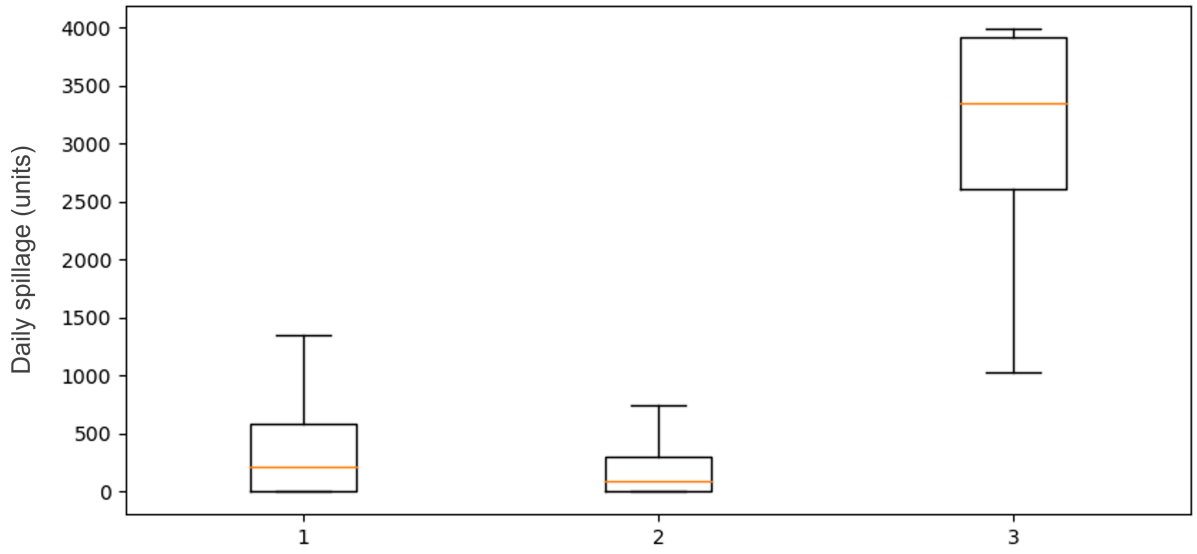

Figure 7: *The total daily spillage of the look-ahead (1), myopic (2) and ZGT (3) policies.*

As Figure 7 shows, the overall spillage of the myopic policy is the lowest out of these three policies. To understand this, we take a look at the decisions of the myopic policy.

It seems that the myopic policy focuses on reducing the spillage costs, as the disposal costs are low due to the small disposal probabilities. In general, the myopic policy chooses to prepare medication for patient i if $\tau_i = 0$, i.e., when the patient needs the medication today. However, we also observe that in some cases, a patient i for which $\tau_i \neq 0$ is scheduled.

By further examination of these cases, we see that that happens when medication of the same type is produced for a patient who has an appointment for today. By preparing this medication in advance, the pharmacy can use clustering to reduce spillage. This is only done when the shelf-life of the medication allows the preparation.

Choosing which medication to prepare besides the ones that must be prepared today is an optimisation problem that is solved by the myopic policy. However, we can investigate whether a "simple" policy based on our observations can reduce the costs.

Hence, we extend our current ZGT policy by adding a new rule. This rule is as follows: if the spillage for medication type j does not equal zero, we check if there is another patient i with $y_i = j$ and $\tau_i \geq 1$. If m_j is such that we do not have to open a new vial of raw material j, we add this to the current production moment. Note that we only do this if the shelf-life of this medication type allows this. The costs for this new policy ("ZGT New Rule") can be seen in Figure 8. Also, the costs for the myopic, look-ahead and the current ZGT policy are plotted.

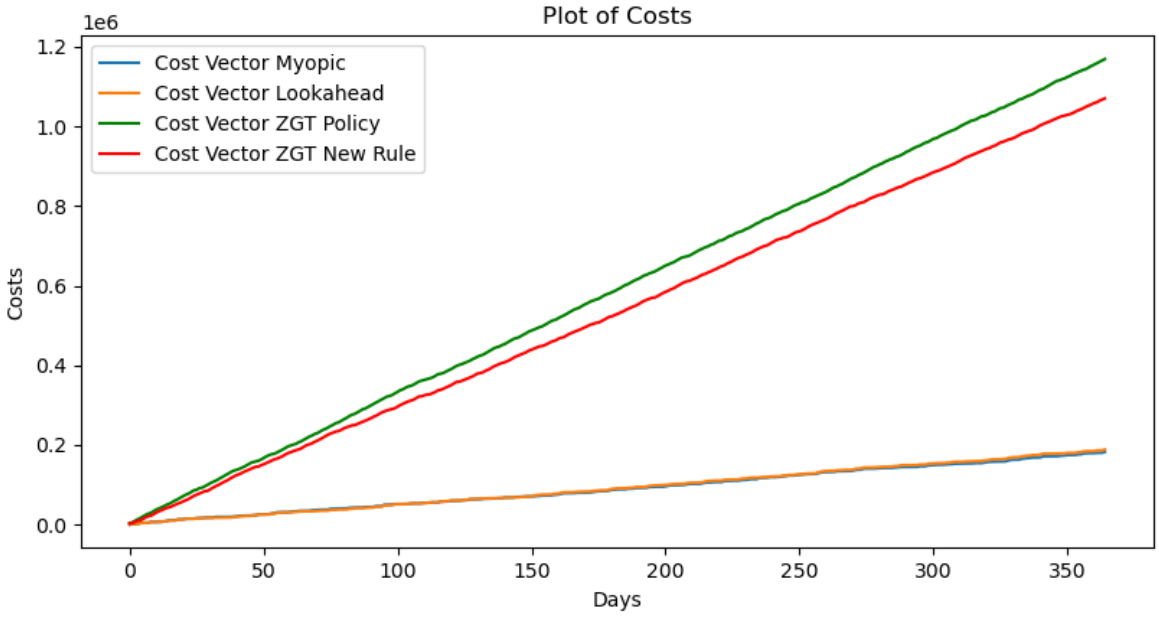

Figure 8: *The costs of one year of all four policies.*

As can be seen in this figure, the new decision rule does lower the costs. The decrease is not as major as the costs of the myopic and the look-ahead policies, as in those policies an optimal decision is made to reduce the (spillage) costs. It shows that adding a "simple" rule to the current ZGT policy has a positive effect on the reduction of spillage.

6.2 Experiments with ZGT case

From the results in Section 6.1, we see that the look-ahead policy does not perform better than the myopic policy. As computing the coefficients θ_f in the look-ahead policy is time consuming, we decide to disregard the look-ahead policy in the experiments for the ZGT case. Similar to the toy-sized instance, we simulate the situation of ZGT for a period of one year. The general parameters that were used are given in the table below:

Parameter	**Value**
Horizon length in days (T)	4
Number of patient types (J)	68
Maximum production on a weekday ($u_\psi, \psi \leq 4$)	130
Close deadlines in days (ς)	3
Arrival rate per day (λ)	83
Penalty costs for exceeding maximum production ($\mathcal{C}$)	1,000

The horizon length in days has been chosen as in general medication cannot be prepared more than four days in advance. The penalty costs have been chosen to allow the pharmacists to work overtime, but only in cases when there is no other way.

The characteristics of the medication types and the arrival rate per day has been based on the data provided by the ZGT pharmacy. This data contained all medication requests to the ZGT pharmacy in

the period of 14-12-2020 until 14-12-2021. This data is given in Table 4 in Appendix A. We note that the disposals are only considered if the medication had already been prepared, as ZGT removes the requests from their database in case it has not yet been prepared. Based on the research of Medendorp (2021) and in consultation with a ZGT pharmacist, we decided to set the disposal probability of a patient depending on the day they arrive. Let $\hat{\eta}_j$ be the disposal probability of a patient of type j in Table 4. Then, for a patient i of type j:

- Patient i arrives with $\tau_i = 0$: disposal probability is $\hat{\eta}_i$.
- Patient i arrives with $\tau_i = 1$: disposal probability is $1.25 \cdot \hat{\eta}_i$.
- Patient i arrives with $\tau_i > 1$: disposal probability is $1.5 \cdot \hat{\eta}_i$.

Each day, we make a new decision on which medication should be prepared. The costs can be seen in Figure 9.

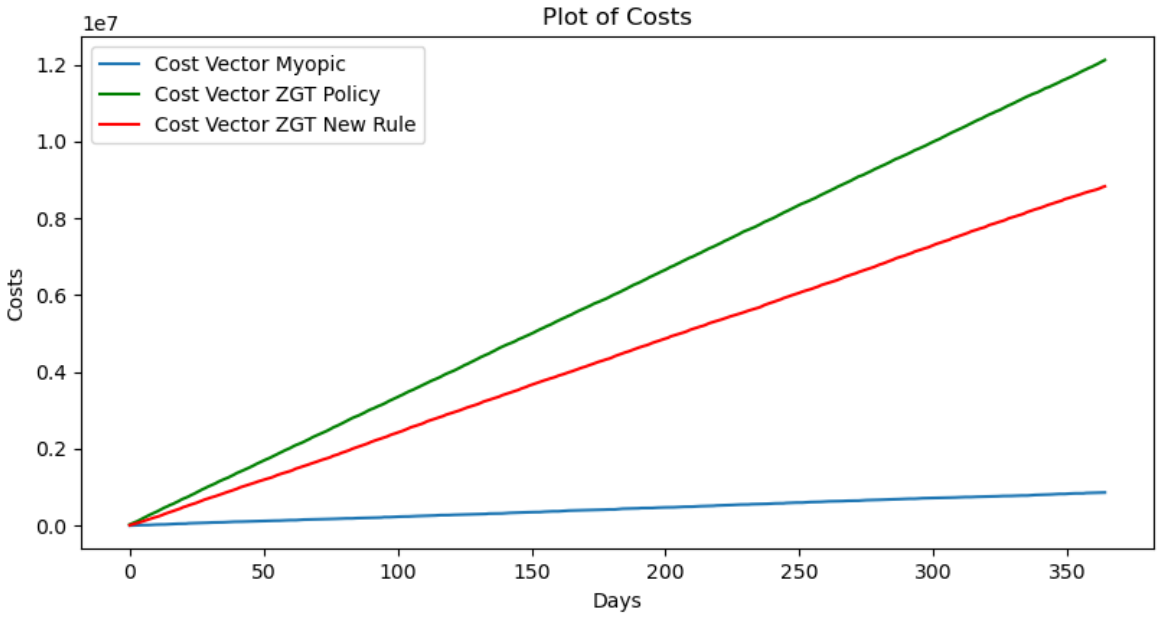

Figure 9: *The costs of one year for the myopic, ZGT and the ZGT New Rule policies.*

Regarding costs, we see a similar situation as for the toy-sized instance. Again, the myopic policy outperforms the current ZGT policy and the current ZGT Policy with our added rule. We also observe that again, the ZGT New Rule policy outperforms the current ZGT policy. This is due to the reduction of spillage, of which the results are given in Figure 10.

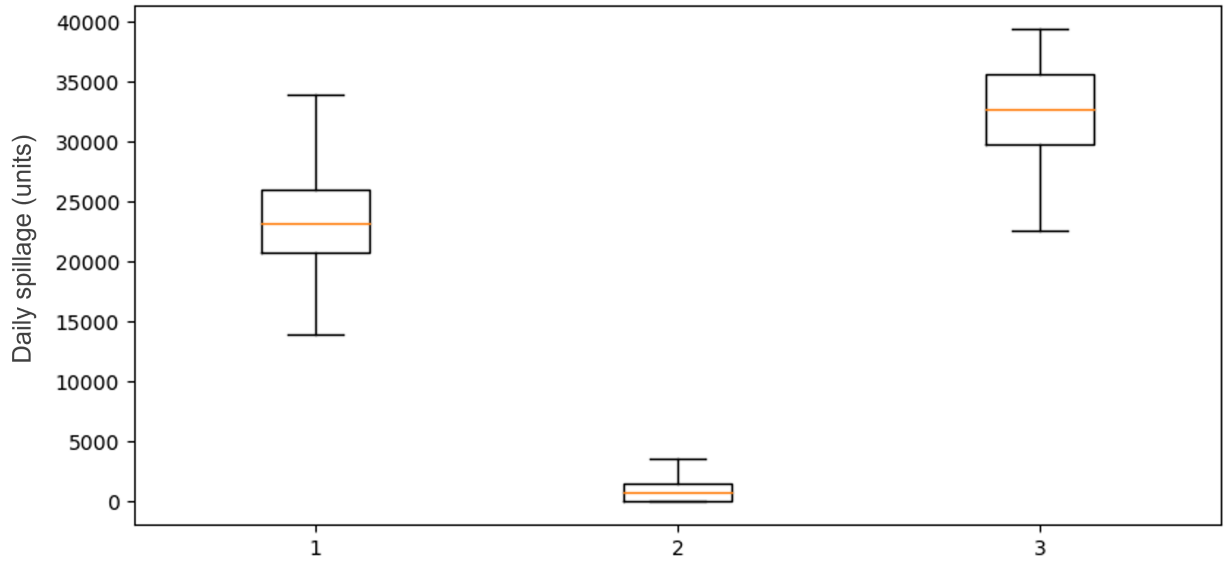

Figure 10: *The total daily spillage of the ZGT New Rule (1), myopic (2) and ZGT (3) policies.*

Again, it must be said that the ZGT policy performs better in practice than it does in our simulations as we cannot model the exact approach that the pharmacy uses to reduce spillage cost as described in the previous section.

To sum up the results for the ZGT case, we see that the myopic policy outperforms the current ZGT policy. Also, the additional rule to the current ZGT policy that was introduced in the former section shows a reduction in spillage and total costs for the ZGT pharmacy.

7 Discussion and conclusions

In this chapter, we provide the discussion on the results and the limitations of this research (Section 7.1). Afterwards, in Section 7.2, we give conclusions and further implications of this research.

7.1 Discussion

In this section we discuss the results of our research and what they imply. Furthermore, we look into the limitations of our research.

Regarding the ADP-approach, we observed that the coefficients θ_f converged, implying that we used enough iterations to get an idea of the future costs based on our actions and current state. We note that the coefficients for the amount of medication patients need (f_3) are much smaller than the other coefficients. There could be two reasons for this. Firstly, it could be that the amount of medication that patients need is not important for estimating the value of being in a certain state. However, it is also possible that the amount of medication does matter. We note that the values of ϕ_{f_3} range between 139 and 6303, whereas the values of ϕ_f are usually below 10 for each $f \neq f_3$. This way, the importance of the amount of medication can be justified and thus the feature f_3 should not be discarded.

Looking at the performance of the look-ahead (ADP) policy compared to the myopic policy, we see that the myopic policy outperforms the look-ahead policy. One of the reasons might be that the ADP approach provides an inaccurate value function approximation. As described in Chapter 5, the MDP has an enormous state space. Therefore, not all states can be visited during the training (i.e., the process of determining the values of θ_f) and we cannot guarantee that we visit the best possible states.

Additionally, our problem contains a high amount of uncertainty. The outcome space of our MDP is, like the state space, incredibly large. Making decisions based on the current state and the immediate costs only (i.e., making decisions myopically) seems to be favourable, as there is very little chance that the "desired" states are actually reached. This leads to a high number of iterations necessary to reach converging coefficients. Furthermore, the low quality of the data might prohibit us from obtaining a solid prediction of the future. First of all, the ZGT pharmacy does not record every time a request is cancelled. This only happens when the medication has already been prepared. Furthermore, with more accurate data we might be able to estimate disposal probabilities better using patient characteristics. For example, initially, we had the additional research objective to determine patient- and time-specific disposal probabilities. We did this using multiple linear regression, see Appendix B. In the hospital, recording data is done in various departments. Therefore, recording data is not "streamlined" and impacts the quality of the data, resulting in incorrect predictions (see Section B.3). This led to us making the assumption to base the disposal probabilities only on the medication type instead of other factors.

Additionally, the data we used for the ZGT case (see Appendix A) was also incomplete. As medication types change frequently, information is outdated quickly (e.g., removing or adding medication types). We could therefore only use the medication types of which we had all the information. The pharmacy works with approximately 90 medication types, where we were only able to use 68 due to outdated information. Finally, we note that the data is mostly from the year 2021 and we have to take into account that the Covid-19 pandemic had an impact on the healthcare system, including the pharmacy department as they were also involved in the vaccination process in the Netherlands. In general, the long-term healthcare was put on hold, resulting in less requests for medication. However, in case data

is available that resembles the normal situation more closely, our method might provide better results for the ZGT pharmacy.

Compared to the ZGT policy, we see that the myopic and the look-ahead policies both outperform the ZGT policy. That is due to two things: the myopic and look-ahead policies generally prepare the medication on the day that the patient needs it, whereas the ZGT policy prepares it one day in advance. This way, the disposal costs are virtually eliminated. In addition, the myopic and look-ahead policies also focus on minimising the spillage costs by preparing medication that is not needed on the same day, but that can be stored for a patient in the days to come.

However, in practice the myopic and look-ahead policies might not be completely suitable. In theory, we can prepare medication on the same day as the appointment, but if the appointment is early in the morning, the pharmacy has to produce that much earlier. In terms of logistics, this will impact the current procedure drastically. We did model it this way, as ZGT requested decision moments at the start of the day. Dividing the day into different blocks, we could make a decision at the beginning of each block to ensure we can prepare the medication for the patients on that day.

As we showed that the look-ahead policy does not perform better than the myopic policy, we decided to not take the look-ahead policy into account for the ZGT case. Obtaining the look-ahead policy for the toy-sized instance took 11 hours and therefore, it was not worth using it further in this research. In the remainder of this discussion, we will only compare the myopic policy with the ZGT policy and the ZGT policy with the new rule. Note that the myopic policy does, to certain extent, take the "future" into account as the scheduled appointments for the coming days are known.

One of the characteristics of the myopic policy is the preparation of medication for patients a few days in advance, if that decreases the spillage costs on the current day. Due to time constraints, we chose to implement a (sub-optimal) rule to the current ZGT policy. It is not optimal as we always choose to prepare the medication of the first patient we encounter that reduces today's spillage costs. We see that this simple addition to the current ZGT policy can already improve the performance. Note that the addition of this rule would not impact the current procedure much. To improve this rule, it is possible to create a linear program to find the set of medication requests with the objective to minimise today's spillage.

For the sub-optimal rule, the software at the ZGT pharmacy should be adjusted slightly such that it either outputs the medication requests that should be prepared, or that it shows a list of possible medication requests in such a way that the pharmacists can make the selection of which requests should be prepared. Note that, in case a linear program is developed to find the optimal set of medication requests, the software should output the list of medication requests.

Furthermore, due to the size of the MDP, we were not able to find an optimal policy using Markov Decision Theory. We were only able to use the heuristic method of ADP and therefore, we cannot compare the performance of the ADP policy to an optimal policy computed by our MDP model.

Finally, due to time constraints we had to exclude some parts of the optimisation problem. The first problem that we had to exclude was the use of phaseals, which can extend the shelf-life of raw material. Furthermore, we used only one size of vials, whereas the pharmacy has different vial sizes for some medication types (also to reduce spillage). Choosing which vials require phaseals and choosing which vials should be opened drastically increases the size of our optimisation problem.

7.2 Conclusions

The goal of this research was to answer the following question: "When should which medication be produced such that the spillage and disposal costs are minimised?". We developed a Markov Decision Process to find the optimal policy to prepare medication such that the costs are minimal. Due to the size of the problem, we turned to the theory of Approximate Dynamic Programming.

The ADP algorithm provided a policy to minimise the costs, which we compared to a myopic policy (not taking the future uncertainties into account) and the current ZGT policy. We conclude from the results that the myopic policy performed better than the policy provided by ZGT and that of the ADP algorithm. Based on the decision rules from the myopic policy, we developed a simple decision rule to add to the ZGT policy which can already reduce the total costs.

In practice, for the ZGT pharmacy this means they can reduce spillage by adjusting their working procedure slightly. In case a vial contains some leftover raw material, the pharmacists should check if this leftover material can be used for a medication request for two (or more days) from now. Of course, if it is impossible to prepare this medication due to the limited shelf-life of the medication (or if there are no medication requests), the leftover material in the vial is considered spillage.

In future research, it is possible to extend the ZGT New Rule policy by developing a linear program that selects the optimal medication requests (instead of the first one that is encountered). This way, there is no drastic change in the current production procedure at ZGT. We expect this linear program to be similar to mixed-integer program in Section 5.2. In addition, phaseals and different vial volumes can be taken into account. Finally, for ZGT it might be useful to carefully keep track of data to obtain a better view of the stochastic processes and with that, provide a better view of the uncertainty.

References

Alagoz, O., Hsu, H., Schaefer, A. J., and Roberts, M. S. (2010). Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*, 30(4):474–483.

Chen, H.-K., Hsueh, C.-F., and Chang, M.-S. (2009). Production scheduling and vehicle routing with time windows for perishable food products. *Computers & operations research*, 36(7):2311–2319.

Davies, M. L., Goffman, R. M., May, J. H., Monte, R. J., Rodriguez, K. L., Tjader, Y. C., and Vargas, D. L. (2016). Large-scale no-show patterns and distributions for clinic operational research. In *Healthcare*, volume 4, page 15. Multidisciplinary Digital Publishing Institute.

Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. springer open.

Glover IV, M., Daye, D., Khalilzadeh, O., Pianykh, O., Rosenthal, D. I., Brink, J. A., and Flores, E. J. (2017). Socioeconomic and demographic predictors of missed opportunities to provide advanced imaging services. *Journal of the American College of Radiology*, 14(11):1403–1411.

Haijema, R. (2011). Optimal issuing of perishables with a short fixed shelf life. In *International Conference on Computational Logistics*, pages 160–169. Springer.

Hulshof, P. J., Mes, M. R., Boucherie, R. J., and Hans, E. W. (2016). Patient admission planning using approximate dynamic programming. *Flexible services and manufacturing journal*, 28(1-2):30–61.

Kaplan-Lewis, E. and Percac-Lima, S. (2013). No-show to primary care appointments: why patients do not come. *Journal of primary care & community health*, 4(4):251–255.

Li, W., Yuan, J., and Yang, S. (2014). Online scheduling of incompatible unit-length job families with lookahead. *Theoretical Computer Science*, 543:120–125.

Lütke Entrup, M., Günther, H.-O., Van Beek, P., Grunow, M., and Seiler, T. (2005). Mixed-integer linear programming approaches to shelf-life-integrated planning and scheduling in yoghurt production. *International journal of production research*, 43(23):5071–5100.

Medendorp, L. (2021). Optimisation of the production process of patient-specific medication, MSc Thesis, University of Twente.

Mohammadi, I., Wu, H., Turkcan, A., Toscos, T., and Doebbeling, B. N. (2018). Data analytics and modeling for appointment no-show in community health centers. *Journal of primary care & community health*, 9:2150132718811692.

Nong, Q., Yuan, J., Fu, R., Lin, L., and Tian, J. (2008). The single-machine parallel-batching on-line scheduling problem with family jobs to minimize makespan. *International Journal of Production Economics*, 111(2):435–440.

Patrick, J. (2012). A markov decision model for determining optimal outpatient scheduling. *Health care management science*, 15(2):91–102.

Perron, N. J., Dao, M. D., Kossovsky, M. P., Miserez, V., Chuard, C., Calmy, A., and Gaspoz, J.-M. (2010). Reduction of missed appointments at an urban primary care clinic: a randomised controlled study. *BMC family practice*, 11(1):1–8.

Poon, C. K. and Yu, W. (2005a). A flexible on-line scheduling algorithm for batch machine with infinite capacity. *Annals of Operations Research*, 133(1-4):175–181.

Poon, C. K. and Yu, W. (2005b). On-line scheduling algorithms for a batch machine with finite capacity. *Journal of Combinatorial optimization*, 9(2):167–186.

Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Ross, S. (2007). *Introduction to probability models*. Academic Press, Amsterdam Boston.

Saure, A., Patrick, J., Tyldesley, S., and Puterman, M. L. (2012). Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research*, 223(2):573–584.

Shaparin, N., White, R., Andreae, M., Hall, C., and Kaufman, A. (2014). A longitudinal linear model of patient characteristics to predict failure to attend an inner-city chronic pain clinic. *The Journal of Pain*, 15(7):704–711.

Tian, J., Cheng, T., Ng, C. T., and Yuan, J. (2012). An improved on-line algorithm for single parallel-batch machine scheduling with delivery times. *Discrete Applied Mathematics*, 160(7-8):1191–1210.

Yuan, J., Li, S., Tian, J., and Fu, R. (2009). A best on-line algorithm for the single machine parallel-batch scheduling with restricted delivery times. *Journal of Combinatorial Optimization*, 17(2):206–213.

Ziekenhuisgroep Twente (2018). ZGT richting 2022: Onze keuze voor de toekomst. https://www.zgt.nl/media/19728/brochure-onze-keus-voor-de-toekomst-zonder-vw-voor-website-gecomprimeerd.pdf, last accessed on 12-07-2021.

A Parameters ZGT case

In this appendix, we provide the data of the medication we used in the ZGT case. In Table 4 we give the detailed information for each medication type, including the volume in a vial, the price, the shelf-life, the minimum and maximum amount of raw material necessary, the arrival fraction and the cancellation probability.

Table 4: *Data for the ZGT case.*

Medication	Volume (vial)	Price (p/u)	Shelf-life	Min	Max	Arrival fraction	Disposal Probability
1	100	3.644	0	139	291	0.1086402	0.015287099
2	50	59.9254	1	100	100	0.0002836	0.142857143
3	500	2.58544	3	698	1152	0.0081014	0.005
4	1200	2.999166667	2	840	1200	0.0012152	0
5	100	0.8974	2	120	156	0.0095597	0.038135593
6	25	0.448	7	120	188	0.0053064	0.015267176
7	10	0.825	4	6	18	0.0139345	0.014534884
8	15	1.688666667	4	15	21	0.0010532	0.038461538
9	3	199.9	1	1	3	0.0316766	0.003836317
10	100	1.4	7	5	500	0.0908575	0.010699955
11	100	0.0365	5	312	312	0.0019849	0
12	20	0.03	1	80	80	0.0002836	0
13	20	15	4	62	95	0.0008101	0.05
14	150	0.05	7	139	900	0.0484871	0.010025063
15	1000	0.000625	7	3000	6000	0.0147851	0
16	2000	0.001025	8	6000	8000	0.0044153	0
17	2000	0.00048	7	2000	2000	0.0010937	0
18	1500	0.000486667	0	4500	4500	0.0006481	0
19	10	51.226	4	225	225	0.0005671	0.071428571
20	10	0.217	4	32	178	0.0099648	0.016260163
21	10	37.5	1	376	701	0.005671	0.028571429
22	500	0.1045	1	490	801	0.0013367	0
23	20	2.9285	4	77	229	0.0213068	0.009505703
24	2	0.995	1	33	150	0.0267347	0.009090909
25	5	0.97	1	10	110	0.0066027	0.006134969
26	300	7.03	4	300	300	0.0123142	0.003289474
27	10	0.488	4	50	203	0.0023899	0
28	100	0.0772	4	134	351	0.0295702	0.015068493
29	200	9.49	4	6000	12000	0.0106534	0
30	1000	0.00321	4	294	6303	0.0323652	0.013767209
31	1000	0.00225	7	386	934	0.0045368	0.017857143
32	2400	0.034620833	14	5400	5400	0.0015798	0
33	1000	2.865	4	1000	1000	0.0048204	0
34	5	74.9	4	259	259	0.0023899	0.016949153
35	500	0.054	3	350	450	0.0138534	0
36	1000	3.2	4	100	1000	0.0023899	0.016949153
37	40	0.10875	4	40	2522	0.0204156	0.007936508
38	1	384.8068182	1	1	2	0.0018633	0.043478261
39	50	10.9	3	550	1500	0.0040102	0.01010101
40	100	0.95	4	50	1200	0.0799206	0.007095793

Medication	Volume (vial)	Price (p/u)	Shelf-life	Min	Max	Arrival fraction	Disposal Probability
41	40	0.09125	1	121	748	0.0110585	0.014652015
42	40	15	1	40	60	0.0026735	0.015151515
43	100	16.49	4	204	453	0.0055495	0.01459854
44	25	103.92	14	200	400	0.0202941	0.013972056
45	5	0.08166	7	500	500	0.0006076	0
46	50	0.018	4	500	500	0.0000405	0
47	20	2	4	20	46	0.0085065	0
48	100	16.99	14	4	1000	0.0094787	0.025641026
49	100	0.685	4	273	1125	0.0273018	0.022255193
50	10	0.978	1	26	56	0.002714	0.044776119
51	10	40.503	1	240	480	0.005671	0.021428571
52	250	1.3332	4	750	1000	0.003038	0
53	50	0.085	7	82	358	0.0330538	0.023284314
54	50	114.64	4	5	10	0.0371451	0
55	30	0.185	2	87	384	0.1043059	0.013592233
56	300	0.537333333	1	300	300	0.0000405	0
57	500	0.0093	1	2000	2000	0.000162	0
58	30	77.33333333	14	210	840	0.0309475	0.007853403
59	80	1.56975	1	240	800	0.0013367	0
60	50	0.037	4	150	150	0.000162	0
61	10	3.2	0	500	1000	0.0305424	0.0066313
62	500	1.95	0	1000	2000	0.0002836	0
63	5	0.05	1	260	660	0.0005266	0.076923077
64	15	8.248	4	15	15	0.0000405	0
65	500	0.0032	1	500	2500	0.0107749	0.033834586
66	100	3.51	4	320	610	0.0019443	0.020833333
67	10	3.535	4	5	13	0.0022279	0.036363636
68	2	10.57	4	1	2	0.01118	0

B Disposal probabilities with Multiple Linear Regression

In this appendix, we explain how we investigated the possibilities of multiple linear regression to find patient-specific disposal probabilities.

B.1 Literature review for disposals

Disposal of medication happens due to three reasons. Firstly, it is possible that a patient does not show up. Secondly, the physician or the patient may cancel the appointment (e.g. due to deviating blood levels or due to illness). Lastly, the dose of medication may have changed.

One of the major challenges in this research is to determine the cancellation/disposal probability of an appointment. Reasons for patients to cancel the appointment are feeling better or worse, transport problems and misunderstanding about the time of consultation (Kaplan-Lewis and Percac-Lima, 2013; Perron et al., 2010). Feeling better or worse can be influenced by the medication that the patient is already taking.

Shaparin et al. (2014) have researched failure to attend medical appointments in Newark, New Jersey. They took several characteristics into account, such as spoken language, age, gender and referring physician. They found that there are several reasons why for instance speaking Spanish as a primary language can result in the cancellation of an appointment. Age and gender are common predictors for no-show rates (Davies et al., 2016; Glover IV et al., 2017; Mohammadi et al., 2018). Other predictors are for example visit type, insurance, tobacco use and ethnicity.

B.2 Mathematical framework

Suppose we would like to estimate a variable y based on n data points and k explanatory variables. The explanatory variables can also be qualitative with the use of dummy variables (Friedman, 2017). We regard y as the outcome of a random variable Y and we write the following model equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i. \tag{17}$$

In Equation (17) the β_i's are unknown parameters and the disturbances ϵ_i follow a normal distribution with mean 0 and variance σ^2.

For convenience, we define the vectors $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$, $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$ and the matrix $X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}$.

Now, the estimator $\hat{\beta}$ can be found by the method of least squares (Friedman, 2017) and is given by

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

In case Y is a binary variable, the result will be a linear probability model for estimating $Pr(Y = 1|X)$.

B.3 Experimental results

Based on the literature review, the advice from ZGT and the available data, our explanatory variables are taken as the patients' age, gender, specialism, referring physician and medication type.

Using data of the whole of ZGT for 5 years (2015 - 2019), we used our model to predict the probability that the medication will be administrated. The value of $R^2 = 0.009$, which implies that the linear regression model does not fit the data well.

Simple calculations showed that in these 5 years, the probability that an appointment was cancelled was approximately 0.33%. After consulting ZGT, it appeared that the data contained flaws. The cancellation of appointments was not always recorded in the data provided by ZGT and therefore, we decided to exclude this data in our research and use a different data set to determine the disposal probability. Due to the restrictions of that data set, we were not able to use multiple linear regression.