Logistical pathways in health care

J. Schakel BSc

Master's thesis

Applied Mathematics

October 2007

Supervisors:
dr. N. Litvak (University of Twente)
dr. M.J.D. Westerdijk (Capgemini Nederland B.V.)
Prof. dr. R.J. Boucherie (University of Twente)
dr. E.A. van Doorn (University of Twente)

Advanced Analysis
Financial Services, B57
Capgemini Nederland B.V.
Papendorpseweg 100
P.O. Box 2575
3500 GN Utrecht

Department of Applied Mathematics
Chair Stochastic Operations Research
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente, Enschede

University of Twente
Enschede - The Netherlands

Abstract

Mental health care is a relatively new research area. Over the years research was done in hospitals and other health care institutions. In The Netherlands a new reimbursement system was developed for health care institutions. And with this system a new registration system was born. Diagnosis and treatments were put together in a Diagnosis Treatment Combination (DTC), i.e. a DTC represents the whole care episode belonging to a specific demand of care of a patient. With this new reimbursement system DTCs can be used for other purposes than payments. Logistical pathways and other information like time spent between activities can be calculated.

In this research a methodology is developed to predict logistical pathways. We use a smoothing technique from language modelling to make a linear combination of relative frequencies with a variable number of history steps. The result is a transition matrix that can be used to predict the progress of a patient in the health care system. Statistically, there is a differenct between short, moderate, long and chronical DTCs. Hence, we make a refinement to the model and make four smoothed matrices instead of one. The validation is done by estimating the average length of a DTC and we compare this with the test results. We can conclude that the methodology predicts logistical pathways really well for non-chronical DTCs.

This research is the basis of the solution to logistical planning and capacity problems, not only in Mental Health Care institutions, but also in the Somatic Health Care.

Acknowledgement

During the six months writing on my Master's thesis at Capgemini Nederland B.V., I greatly enjoyed working and solving problems on the theoretical research about logistical pathways in health care. Although I was not complete comfortable with my studies Applied Mathematics, in the course of this period all the pieces came together. The research topic was about health care, humanly and of social interest, and really 'applied'. To throw myself into a new theory and investigate every step I took, truly give me satisfaction.

I would like to thank the people who contributed to the result of this thesis. First of all, my supervisors Nelly Litvak and Machiel Westerdijk, for their justified criticism and valuable advices. I also thank my collegues for their help when I got stuck in computer software, for given suggestions and the pleasant working atmosphere.

Of course I want to thank my family and friends for the moments of laughter and help. You really cheered me up when I needed it. Special thanks to my parents, who supported me through all these years!

Contents

1 Introduction **1**
- 1.1 Rationale . . . 1
- 1.2 Literature . . . 2
- 1.3 Problem formulation . . . 3
  - 1.3.1 Starting-point of the research . . . 3
  - 1.3.2 Objectives . . . 3
  - 1.3.3 Research questions . . . 4
  - 1.3.4 Research demarcation . . . 5
- 1.4 Research approach . . . 5
- 1.5 Overview . . . 5

2 Data analysis **7**
- 2.1 Length of DTC . . . 7
- 2.2 Frequencies of the activities . . . 8
- 2.3 Repeating activities . . . 9
- 2.4 Time between activities . . . 10
- 2.5 Duration of the activities . . . 11
- 2.6 Summary . . . 12

3 Trigram model for DTCs **13**
- 3.1 Model description . . . 13
- 3.2 Parameters evaluation . . . 14
- 3.3 Result of evaluated parameters . . . 16
- 3.4 Summary . . . 17

4 Model validation & refinement **19**
- 4.1 Model validation . . . 19
  - 4.1.1 Estimating the average length of a DTC . . . 19
  - 4.1.2 Results of the model validation . . . 21
- 4.2 Refinement of the model . . . 22
- 4.3 Average time in system . . . 25
- 4.4 Summary . . . 29

5 Results **31**
- 5.1 Examples of logistical pathways . . . 31
- 5.2 Zipf's law & most frequent DTCs . . . 31
- 5.3 Repeating activities . . . 33

6 Conclusions & recommendations **35**
6.1 Conclusions 35
6.2 Recommendations 35

A Abbreviations, definitions & symbols **37**
A.1 Abbreviations 37
A.2 Definitions 37
A.3 Symbols 38

B List of activities **39**

C Transition matrix **41**

D Proofs from the model **43**
D.1 Singularity points 43
D.2 Maximum 44

Chapter 1

Introduction

1.1 Rationale

Roughly four years ago the decision was made to fundamentally change the reimbursement system for mental health care (MHC). In the old situation, the MHC institutions were paid per individual activity. In the new situation, they will be paid for each complete treatment. Each MCH employee treating a patient registers every activity, date, medical attendant and admission. At the end of the treatment all the activities are combined and the insurer will pay for this complete treatment. Besides the change in payment of a treatment, the new system will also show explicitly the relationship between the demand of care and the payments. This relationship gives answers to questions like, which activity is brought in for which demand of care and against which costs? This product reimbursement increases transparency. And thus, makes the control of care and a free-market possible in MHC.

The system of Diagnosis Treatment Combination's (DTCs) is a form of product reimbursement. The DTC is a description of a 'product', which is the complete care episode belonging to a specific demand of care: the diagnosis. In most cases a DTC already starts in an outpatient situation. That means that already when a patient calls the doctor, that activity will be registered. And a DTC generally ends when the patient is discharged. The costs are allocated to the intermediate products, i.e. the medical activities, the sum gives the total cost of a DTC. The basic goal of the DTC case mix system is to establish a transparent reimbursement system for health care in the Netherlands [1].

Since the implementation of the DTC system in MHC institutions, it is possible to use the DTCs for other purposes than payments. The received transparency gives the institutions more options to control the care both on operational and strategical level. From the DTCs, logistical pathways of patients can be obtained. A logistical pathway of a patient is the sequence of all performed activities. For example: a patient is diagnosed first. Then he gets psychotherapy alone with his therapist and after that he gets psychotherapy with his family. He ends with psychotherapy again. A logistical pathway is then: diagnostic, psychotherapy, remaining psychotherapy, psychotherapy. Not only the activities are known. Also the time of the activities and the time between activities can be included in a logistical pathway.

A number of examples why logistical pathways can be of use for health care:

- Development of treatment protocols: treatment protocols can be developed by best-practices and scientific insight. By reflecting protocols inside the diagnosis groups both protocols and practice can be improved. Health processes can be optimized and different pathways will come forward, efficient pathways and less efficient pathways. A pathway is called efficient if the costs or the time is minimized. The less efficient pathways can be investigated closer to improve them.
- The logistics of MHC institutions and hospitals is complicated. Doctors need operating or appointment schedules. A ward where acute patients are hospitalized, needs a minimum number of beds and nurses to take care of the patients. Hence, logistical pathways is the starting-point of logistical planning: in order to develop a good schedule and have enough beds and nurses it is important to have a thorough insight in the requested capacity.
- Development of care programs: in many institutions there are initiatives to organize the control of care within so called care programs. Realistic insight in the current treatments can help both in defining care programs and to control the care within the care programs.

1.2 Literature

A lot has been written about logistical pathways in health care and include a lot of different theories and approaches to look at optimization problems. In this section we will give some of these theories and applications in practice.

The definition of a logistical pathway, also called (clinical) care pathway, is an outline of what is likely to happen on the patient's journey. It is as a time line, on which every event relating to treatment can be entered. Events such as consultations, diagnosis, treatment and medication can all be mapped on this time line.

In the UK and the USA, the pathways in hospitals are all constructed with the help of a multidisciplinary work-team [2]. They can help to find the bottlenecks in treatments protocols, reduce the length of stay and minimize costs without compromising patient care [3], [4]. However, in MHC it is more difficult to construct such pathways, since there is professional antagonism and a dearth of evidence-based practice [5]. With the introduction of the DTC system and received transparancy, it will be easier to obtain these pathways and use them to optimize for example, protocols and schedules [6]. Besides, the quality of care and the efficiency in care programs will be higher [7].

To reduce the length of stay of a patient in a hospital or MHC institution and obtain an optimal capacity planning [8] and schedules [9] (operating, appointment and doctors and nurses), queueing theory can help us [10]. Moreover, the data on lengths of stay can be analysed using phase-type distributions [11]. Simulation models are used a lot to simulate one or more wards or even a complete hospital [12].

With the increasing availability of health care data, all kind of new methods for data analysis arise. Bayesian networks, machine-learning, neural networks and regression trees are heavily discussed, since we want to know the best way to analyse all this data [13].

1.3 Problem formulation

In this section we will analyse the problems at the start of this research. After that we will formulate the objectives with the corresponding research questions and will give our approach to solve the problems.

1.3.1 Starting-point of the research

The model that has been developed by Capgemini Nederland B.V. is as in Figure 1.1, where each node represents an activity. The percentages at the branches tell us how many percent of the patients coming from the last node go to the next activity. For example: we start at the node 'Start DTC'. In total there are 18,216 patients. Of all the patients, 3,578 patients go to pre-intake, which equals 22.6%. From pre-intake 86.9% of the patients go to diagnostic etc. There are also loops in the current situation. That means the patient can return to a previous activity with a certain probability. For example: 306 patients are in diagnostic. About 50% of the patients go to crisis care and 77 patients (26%) returns to diagnostic.

This tree decomposition is not useful to correctly predict logistical pathways. One of the problems is that the pathways do not become clear, since there are only percentages at the branches and the number of branches will explode if we want to show all the possibilities. After a few activities, the percentages are also not reliable anymore, as the dataset has become much smaller. Moreover, the history of a patient is not taken into account. Only the last activity determines the next activity. The history of a patient is important, since for example, the probability of going to diagnostic is different if the patient was in psychotherapy or in physiotherapy before. Another problem is that nothing has been done about the time element. In this model it is not clear how long a patient stays in the medical system or between activities. The loops in Figure 1.1 only represents one repeating activity. However, if the activity is repeated more times, the percentages are different, which can not be shown in this figure.

1.3.2 Objectives

To continue from the present model, a couple of things have to be examined and developed:

- In DTCs there are activities which occur more than one time. For patients with a long pathway, this repeating activities show some kind of a pattern. One objective is to find this pattern and reflect this in a mathematical model.
- The present model is a tree decomposition, which gives the probabilities of going from one activity to another. However, after a couple of nodes

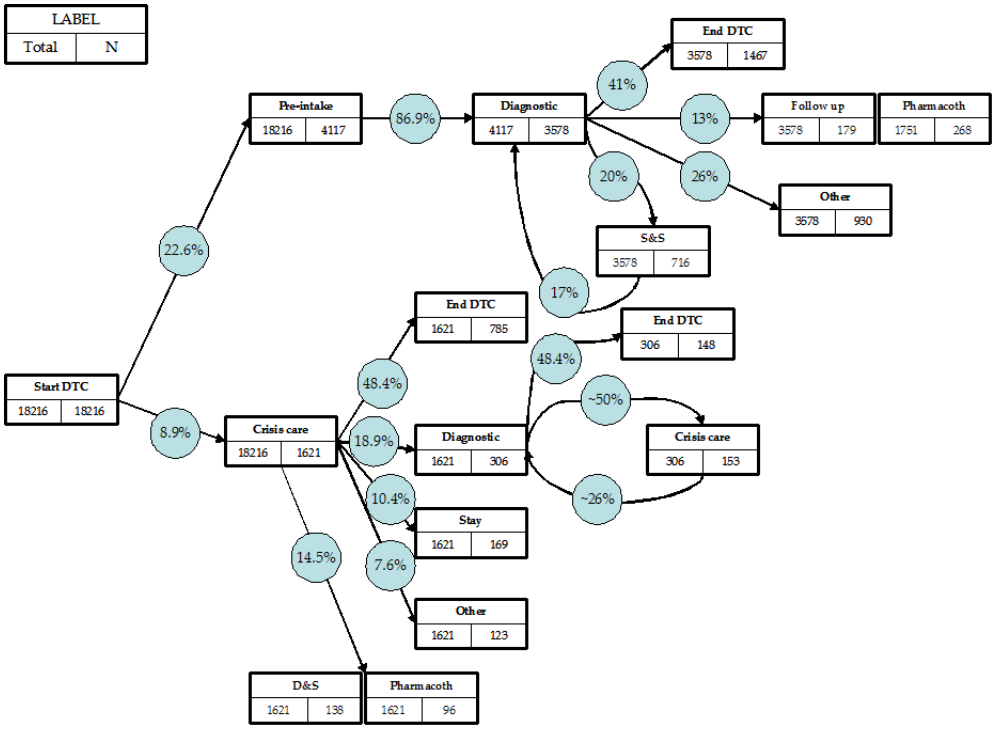

Figure 1.1: Tree decomposition at the starting-point of this research.

the number of branches will be very large and for the following nodes a different kind of model has to be found or a whole new model has to be developed.

- In the current model there is no time element. The institutions and the patients do not know how long it takes before a patient goes to the next activity. For the time in between activities depending on the activity where the patient is coming from, a model has to be developed.

1.3.3 Research questions

Given the scope of the problems described in Subsection 1.3.1 and 1.3.2 the following questions are important for this research:

- We want to choose a proper model, which can help to predict logistical pathways. To choose a proper model it has to be clear which elements are important. Elements like history of a patient and repeating activities have to fit in the chosen model.
- The chosen model will have parameters, which have to be evaluated. We have to decide which method we want to use to evaluate them.
- After we estimated the parameters and made the model we want to predict the pathways. Thus, we derive a transition matrix with the evaluated

parameters, since the probabilities of a transition matrix can help to predict logistical pathways and find a way to give insight in the pathways.

- Finally the time element has to be introduced. We have to decide how important the duration of the activities and the time between the activities are. We have to be able to implement these time elements to calculate for example the time a patient spends in a DTC.

1.3.4 Research demarcation

The research must be based on adequate data. The data used in this work contains about 45,000 DTCs. These DTCs are complete care episodes belonging to a specific demand of care. Together with information about the patient, like age and gender, a DTC is a complete data source. The information we use in this research is the sequences of activities. The treatments and the general time a doctor registers are ordered in time. Thus, the first treatment the patient gets is the first activity in the sequence. In this report, this sequence is called 'DTC'.

1.4 Research approach

Our approach is inspired by the techniques from language modelling. This choice is based on the following intuitive reasoning. It is not possible to model the activities as a Markov process. For a Markov process the Markov property has to hold. That means the conditional probability distribution of future activities of the process, given the present activity and all past activities, depends only upon the present activity and not on any past activity. We can not use a Markov process, since we want to take more history into account. Conversely, in language modelling the history is important, as the construction of sentences is determined by the grammar. A model used in language modelling is the trigram model. This model estimates the probability of a word, given the two previous words that is necessary to maintain the grammar. To get this probability we construct a linear combination of matrices with a different number of history steps. Eventually, since we have a linear combination, we get a smoothed transition matrix. If necessary, we will make a refinement to the model, since it will better predicts logistical pathways.

For the implementation of the time element we examine the continuous phase-type distribution. A continuous phase-type distribution can be defined as the time until absorption. We will investigate if we can use this distribution. The duration of the activities is negligible compared to the time between activities, since the first is in minutes or hours and the last is in days. With a refinement to the model, that means we get more than one smoothed transition matrix, a continuous phase-type is difficult to use. Therefore, we derive a different model with our smoothed matrices. Together with the mean time between activities we determine the average time in the system by a patient.

1.5 Overview

In Chapter 2, we will take a look at the data. The distribution of the length of a DTC, the frequencies of the activities, the duration of activities and the

time between activities will be discussed. The model is described in Chapter 3, together with the evaluation of the parameters. We will give a methodology to make a smoothed transition matrix with the trigram language model. The smoothed matrix can be used to predict logistical pathways and is the starting point of the solution to planning and capacity problems. In Chapter 4, the model will be validated by estimating the average length of a DTC and refinements will be made. In Chapter 5 the results will be given and the conclusions are drawn in the last Chapter 6 of this report.

Chapter 2

Data analysis

Before constructing a mathematical model, we have to get insight in the data. The data describes the sequences of activities. In this chapter we will take a look at the distribution of lengths of DTCs, the frequencies of the activities and the time element in activities. The information about this data will tell us for example, how long most of the DTCs are and how important history is.

2.1 Length of DTC

First of all, the length of a DTC is investigated. By summing the number of DTCs which have length n or smaller and dividing this by the total number of DTCs, a cumulative probability is calculated.

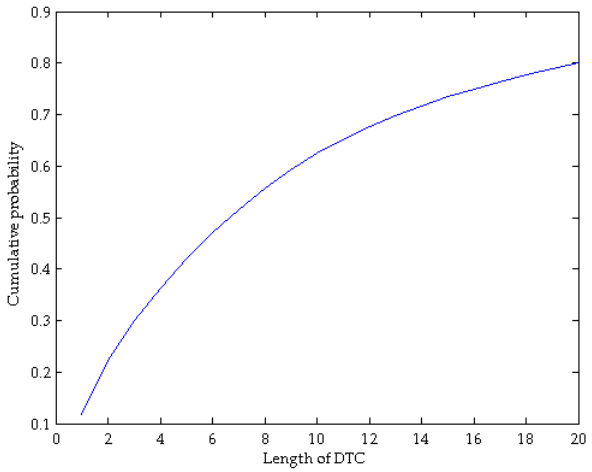

Figure 2.1: The distribution of the length of a DTC is given by the probability that a DTC has length n or smaller, that equals the cumulative probability.

The plot shows that 30% of the DTCs have length 1, 2 or 3. The DTCs with

length 9 or less count for 60% and 80% of the DTCs have length 20 or less. The other 20% varies from length 21 to a DTC with length 2,028.

2.2 Frequencies of the activities

Activities in MHC can be divided into levels. The highest level represents the main activities, compiled by managers of health care institutions and managers of Capgemini Nederland B.V.. The lowest level represents all the activities used for a patient. In this research we take the highest level, because this level is the most important for the managers in health care institutions. For example: diagnostic exists of all kinds of examinations, like physical and intelligence examination. If a patient gets a psychiatric examination, it will be registered at the lowest level. However, in this research, psychiatric examinition will be part of diagnostic at the highest level. Hence, the activity diagnostic will be registered. The list of activities of the highest level is given in Table 2.1. In total there are 19 states at this highest level. A state represents a kind of activity a patient gets for his or her diagnosis, except for state 0. That means the DTC ends.

Activity number	Activity
0	Finished
1	Pre-intake
2	Diagnostic
3	Psychodiagnostic examination
4	Follow up treatment contact
5	Supporting and structuring treatment contact
6	Psychotherapy
7	Other form of communicative treatment
8	Pharmacotherapy
9	Physical therapy
10	Course therapy
11	Physiotherapy
12	Occupational therapy
13	Accompanying
14	Nursing and caring
15	Crisis care
16	General indirect time
17	Stay (per day of stay)
18	Daily spending (per hour)

Table 2.1: List of activities at the highest level.

In Figure 2.2 the distribution of the activities is shown. The most frequent activity is general indirect time (activity 16), which means administration about patients and the like. The least frequent activities are physical therapy (activity 9), physiotherapy (activity 11) and occupational therapy (activity 12). In the model we will choose, we have to take into account that there are activities that are rare. As we build a model, we divide the data in a *training set* and a *test set*.

With the training set we build our model and with the test set we validate our model. In case where activities are rare, it is possible that activities or combination of activities appear in the test set. However, they will not appear in the training set. Especially, when the data set is small. Hence, we have to find a model that takes rare activities or combinations of activities into account.

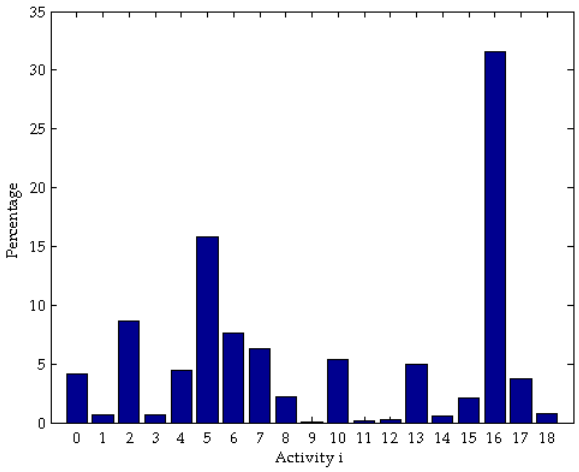

Figure 2.2: The distribution of activities. For example, diagnostic (activity 2) equals 8% of all the activities.

2.3 Repeating activities

In Appendix C the simplest transition matrix is given in Table C.1 to get an impression of the probabilities going from one activity to another.

After building this simplest transition matrix we take a look at a transition matrix where we take into account how many times a patient has been in that activity before. It is interesting to see whether the probability of going from activity i to the same activity i given that a patient has been there only once is different from the probability of going from activity i to the same activity i given that the patient has been there for three times already. We want to know if it is possible to see a connection among the number of times a patient has been in an activity before and the probability of returning to that activity again.

Every line in Figure 2.3 represents the probability of one activity given the number of times the patient has been in that activity before. That shows that most of the activities have the same distribution. The probability of activity i given that the patient has been a number of times in activity i before, gets larger as the number of activities i grows. We can also see that some probabilities fluctuate or drop to zero. There are two reasons for this. The first reason is that some activities appear not very often in the data. That is why repetitions do not appear very often. The second and most important reason, is that it is a feature of certain activities to not appear more times in succession. This kind

of activities is different from the other activities. Thus, the probability of an activity is dependent of the number of the same activities in the past.

The frequencies of the activities after 5 activities are different than after 45 activities. Therefore, the transition probabilities of going from activity i to activity j will change. Hence, we conclude that the history of a patient is important, if we want to model the repetitions of activities.

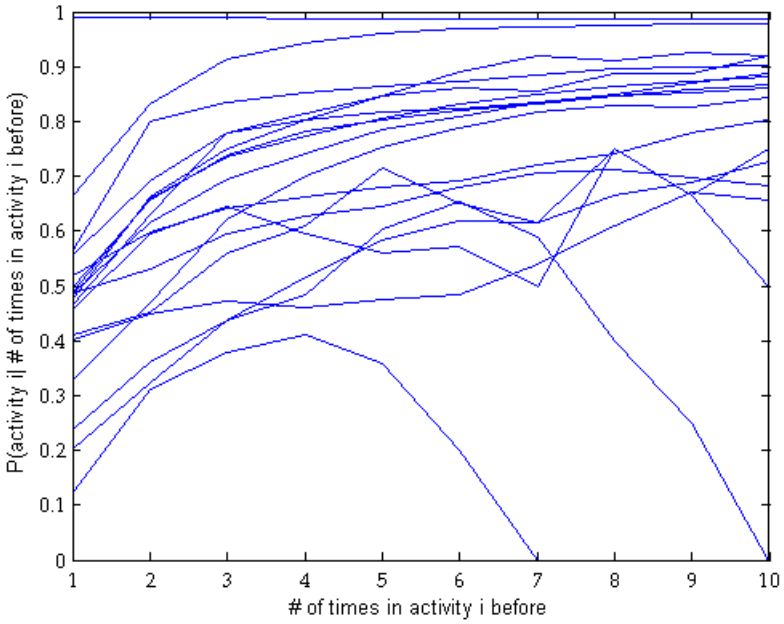

Figure 2.3: The probability of activity i given the number of times the patient has been there before.

The corresponding probabilities are given in Table 2.2. We see the probabilities dropping to zero after 7 and 10 times are the activities pre-intake and psychodiagnostic examination respectively. For example, a patient is sent to a MHC institution and gets the activity pre-intake. However, the first time the institution does not know exactly what is wrong with the patient. Thus, the patient gets a pre-intake for a couple of times and if they find the problem he gets treatment. If not, the patient will be sent to another MHC institution or will be sent home and the DTC stops for this patient in this MHC institution.

2.4 Time between activities

To get an impression of the time element in this research, an investigation of time between activities and the duration of the activities is made. Table 2.3 shows the mean time between activities, all in days.

	Number of times the patient has been before in activity i									
Activity	1	2	3	4	5	6	7	8	9	10
1	0.1228	0.3114	0.3778	0.4118	0.3571	0.2	0	0	0	0
2	0.4019	0.4496	0.4736	0.4599	0.4758	0.4842	0.5399	0.6107	0.6703	0.6557
3	0.2364	0.3612	0.4384	0.4831	0.6047	0.6538	0.5882	0.4	0.25	0
4	0.4884	0.6607	0.7367	0.7727	0.8045	0.8317	0.8499	0.8632	0.8718	0.8817
5	0.487	0.6147	0.6931	0.7425	0.7856	0.8093	0.8348	0.8499	0.8693	0.8868
6	0.5562	0.6916	0.778	0.8018	0.818	0.8236	0.8356	**0.8457**	0.8537	0.8615
7	0.4981	0.6568	0.7378	0.7808	0.802	0.8197	0.832	0.8469	0.8595	0.8657
8	0.2023	0.3245	0.4382	0.5167	0.5833	0.619	0.6154	0.6667	0.6875	0.7273
9	0.4599	0.5963	0.6462	0.5952	0.56	0.5714	0.5	0.75	0.6667	0.5
10	0.3297	0.4687	0.621	0.7009	0.7528	0.7887	0.817	0.8279	0.8275	0.8436
11	0.4766	0.6626	0.75	0.8025	0.8462	0.8909	0.9184	0.9111	0.9268	0.9211
12	0.4155	0.4505	0.561	0.6087	0.7143	0.65	0.6154	0.75	0.6667	0.75
13	0.666	0.8325	0.9147	0.9421	0.9602	0.9683	0.9727	0.9753	0.9772	0.9779
14	0.468	0.6311	0.7786	0.8147	0.8471	0.8611	0.8548	0.8868	0.8865	0.92
15	0.4855	0.5314	0.596	0.6275	0.6436	0.6789	0.7048	0.7117	0.6962	0.6818
16	0.5196	0.5994	0.6416	0.6622	0.6785	0.6913	0.7205	0.7426	0.7778	0.8032
17	0.5654	0.799	0.8339	0.8529	0.8639	0.8716	0.8855	0.8959	0.8994	0.902
18	0.6203	0.7302	0.8056	0.8214	0.8498	0.8628	0.8624	0.8759	0.8988	0.9144

Table 2.2: The probability of going to activity i given the number of times the patient has been there before. For example: a patient has been in psychotherapy (activity 6) for 8 times. Then the probability that he gets psychotherapy again will be 0.8457.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	12.2	21.1	27.8	31.2	25.3	18.6	34.0	28.0	1.0	26.3	0	0	33.3	0	8.2	15.8	17.7	40.5
2	5.5	13.3	15.3	20.3	21.5	21.8	21.1	18.2	16.5	6.8	3.1	5.8	22.2	15.6	8.9	12.1	10.9	7.8
3	17.9	11.5	7.1	15.0	15.6	15.5	13.6	26.0	6.0	5.9	2.0	0	10.3	6.0	3.9	10.0	4.5	1.6
4	18.0	12.3	11.7	12.6	12.1	8.1	4.4	12.9	3.1	3.5	2.4	4.4	5.8	4.3	6.9	11.0	2.6	3.8
5	15.4	13.3	10.0	12.0	9.2	4.6	5.3	12.4	10.1	1.9	2.7	3.7	5.7	2.2	8.3	4.9	4.0	4.9
6	5.8	8.8	5.8	11.2	5.5	10.8	4.5	11.0	16.7	2.7	1.5	4.9	2.1	2.0	6.0	5.1	2.1	3.9
7	7.3	13.8	10.5	5.2	4.7	4.5	10.3	10.4	4.0	2.2	1.8	2.9	5.2	3.1	11.8	6.2	3.1	3.0
8	5.0	12.5	8.5	14.3	12.2	12.2	10.1	29.7	2.0	4.8	1.3	5.5	11.2	3.4	7.6	13.3	5.8	3.5
9	1.0	4.0	8.0	3.2	2.9	1.8	3.2	7.5	2.5	1.8	2.2	27.3	4.1	1.6	4.0	3.0	15.7	0
10	2.3	3.7	3.0	3.5	2.0	2.5	2.1	3.2	1.8	3.0	2.2	2.0	2.7	1.5	4.1	1.9	2.5	3.0
11	0	2.3	1.4	2.8	2.3	**7.0**	2.0	2.5	3.0	3.1	2.9	3.5	1.8	0	0	3.2	3.8	0
12	0	13.8	0	5.6	3.0	3.9	2.6	4.3	1.0	2.3	1.8	2.4	2.8	1.6	6.0	7.0	5.9	1.0
13	20.7	10.0	7.7	6.8	7.0	2.3	6.6	11.9	2.1	2.5	3.3	5.0	5.3	1.8	8.0	6.6	1.8	3.7
14	1.0	8.8	3.7	3.7	2.3	3.0	4.6	3.2	1.3	1.6	0	1.2	2.2	6.1	3.5	2.6	2.1	0
15	4.6	8.8	12.6	8.1	8.3	6.3	10.6	7.0	2.0	6.6	0	3.0	8.0	2.6	5.2	5.7	2.6	3.3
16	6.4	10.2	10.5	11.1	4.7	4.6	6.0	14.1	1.8	1.8	2.2	6.9	6.0	2.7	6.1	3.8	3.7	3.6
17	10.0	4.8	2.5	4.7	4.2	2.3	3.2	3.2	16	2.2	7.7	6.7	1.8	2.6	1.8	3.2	2.4	4.9
18	1.0	5.7	2.0	3.4	4.3	2.9	7.0	4.9	0	3.0	0	1.0	3.1	0	3.3	3.1	3.5	3.7

Table 2.3: Mean time between activities in days. For example: a patient gets physiotherapy (activity 11). Then it takes 7.0 days before he gets psychotherapy (activity 6).

2.5 Duration of the activities

In Table 2.4 the mean time of the activities is given. Activities consist of direct and indirect time. Direct time is the time that the doctor is in contact with the patient, i.e. the patient is treated for his diagnosis. Indirect time is the time the doctor needs to work on his documentation or the time to write a prescription and the like. Time like travel time of a doctor is not included. Hence, total time is an additive sum of the direct and indirect time. Most of the activities are in minutes, except for stay and daily spending. These are in days and hours respectively. They do not have direct or indirect time, since the patient is 'hos-

pitalized'.

	Mean time per activity		
Activity	Direct time	Indirect time	Total
1		24 minutes	24 minutes
2	47 minutes	16 minutes	63 minutes
3	80 minutes	42 minutes	121 minutes
4	31 minutes	10 minutes	41 minutes
5	23 minutes	5 minutes	28 minutes
6	32 minutes	8 minutes	**40 minutes**
7	24 minutes	4 minutes	29 minutes
8	23 minutes	6 minutes	29 minutes
9	26 minutes	6 minutes	32 minutes
10	17 minutes	5 minutes	22 minutes
11	16 minutes	3 minutes	19 minutes
12	25 minutes	13 minutes	38 minutes
13	39 minutes	2 minutes	41 minutes
14	31 minutes	3 minutes	33 minutes
15	87 minutes	7 minutes	94 minutes
16		16 minutes	16 minutes
17			3.6 days
18			4.9 hours

Table 2.4: Duration of the activity. For example: a patient gets treated by a psychotherapist (activity 6) for 32 minutes. After the treatment the psychotherapist has to work on the patient's administration for 8 minutes. The total time the psychotherapist worked on the patient equals 40 minutes.

2.6 Summary

We conclude that 80% of the DTCs has length 20 or less. Some activities, like diagnostic and general indirect time appear very often and some activities, like physical and physiotherapy are very rare. In the model that we are going to choose, we have to take this into account. Besides, we have to take the history of the patient into account, since the transition probabilities depend on previous activities.

Chapter 3

Trigram model for DTCs

In language modelling the grammar in a sentence is important. Therefore, the history of a word matters. If we include history of words, some combinations of words will become rare in the data. According to Jelinek [14] the trigram model takes two history steps into account. Besides, he reckons with the combination of words that are rare, by using linear combinations of probabilities with less history. Hence, the model is commonly used in language modelling and it can predict a word from previous words. The same model can be applied in health care, since the preceding activities of a patient are important for the upcoming activities. Moreover, some combinations will be rare. In this chapter we will describe the model and evaluate the parameters of the model.

3.1 Model description

In language modelling a common model used is a so called trigram model to estimate a transition matrix. The transitions represent the probability of word, given the two previous words. This trigram model is a linear smoothing model, since in text samples the number of words to predict from is not big enough. Hence, to estimate the probability $p(w_3|w_1, w_2)$, where (w_1, w_2, w_3) is a string of three words, a linear combination of the probabilities $f(w_3)$, $f(w_3|w_2)$ and $f(w_3|w_1, w_2)$ is made.

In our DTCs some activities are also rare and some appear very often. To get a good estimation of the transition probabilities, the smoothing model is applied. In stead of two previous activities we use three previous activities, since our data set is quite large. Hence, we want to estimate the probability $P(w_4|w_1, w_2, w_3)$, where (w_1, w_2, w_3, w_4) forms a string of four activities and the first three activities are the previous activities. We construct a smoothed model by making a linear combination of the probabilities $f(w_4)$, $f(w_4|w_3)$, $f(w_4|w_2, w_3)$ and $f(w_4|w_1, w_2, w_3)$. Every probability has a factor λ_i which has to be evaluated. The evaluation is done by using the maximum likelihood estimation.

3.2 Parameters evaluation

To predict a word from previous words in a sample of a text, Jelinek [14] uses n-gram models. Let w_i be a word from a fixed and known vocabulary $\mathcal{V}$. Then the number of words in a sample of text is huge. However, still a couple of words do not appear very often. With a training set twice as big as the test set, we can calculate the following probabilities:

$$f(w_i) = \frac{C(w_i)}{\sum_i C(w_i)}\,,$$

$$f(w_i|w_{i-1}) = \frac{C(w_{i-1}, w_i)}{C(w_{i-1})} \quad \text{and}$$

$$f(w_i|w_{i-2}, w_{i-1}) = \frac{C(w_{i-2}, w_{i-1}, w_i)}{C(w_{i-1}, w_i)}\,.$$

where

$$C(w_{i-1}, w_i) = \#\,\text{of transitions from } w_{i-1} \text{ to } w_i\,.$$

The first probability $f(w_i)$ is called unigram, where $C(w_i)$ is the number of words w_i and $\sum_i C(w_i)$ is the total number of words. The second probability $f(w_i|w_{i-1})$ bigram and the last transition $f(w_i|w_{i-2}, w_{i-1})$ is called trigram. Naturally, it is possible to take more history into account and construct a fourgram or even a fivegram.

In the test set some combination of words may never appear, since some phrases are rare. That is where the smoothing comes in. A linear combination of uni-, bi- and trigrams gives:

$$P(w_3|w_1, w_2) = \lambda_1 f(w_3) + \lambda_2 f(w_3|w_2) + \lambda_3 f(w_3|w_1, w_2)\,, \tag{3.1}$$

where,

$$\lambda_1 + \lambda_2 + \lambda_3 = 1\,.$$

Now, instead of words activities are used. The λ_i's are computed using the maximum likelihood method as described below. To determine the λ_i's, the same data on which the relative frequencies $f(\cdot|\cdot)$ are computed, can not be used, since in that case the estimates would be $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = 1$. Hence, the total training data must be divided into two parts. The first part, called *kept* data which is twice as big as the second part, is used to estimate the relative frequencies $f(\cdot|\cdot)$. The second part, called *held-out* data is used to estimate the weights λ_i. We can imagine that this frequency $f(w_3|w_1, w_2)$ approximates $P(w_3|w_1, w_2)$ better, if it is based on a larger count $C(w_1, w_2)$. Hence, the λ_i's should depend on the counts $C(w_1, w_2)$ and $C(w_2)$.

In Section 2.2 we concluded that the activities in the past are important for the probability of activity i. In this model we will use instead of two, three previous activities. Thus, $P(w_4|w_1, w_2, w_3)$ is smoothed in three steps. First we get:

$$P^*(w_4|w_3) = \gamma(C(w_3))f(w_4) + (1 - \gamma(C(w_3)))f(w_4|w_3), \tag{3.2}$$

where $f(w_4)$ and $f(w_4|w_3)$ are relative frequencies from the kept data. Then,

$$\begin{aligned} P^*(w_4|w_3,w_2) &= \theta(C(w_2,w_3))P^*(w_4|w_3) + \\ &\quad (1-\theta(C(w_2,w_3)))f(w_4|w_3,w_2) \end{aligned} \tag{3.3}$$

and the final step

$$\begin{aligned} P(w_4|w_1,w_2,w_3) &= \zeta(C(w_1,w_2,w_3))P^*(w_4|w_2,w_3) + \\ &\quad (1-\zeta(C(w_1,w_2,w_3)))f(w_4|w_1,w_2,w_3) \end{aligned} \tag{3.4}$$

The coefficient γ in Equation (3.2) can be estimated by using the maximum likelihood estimation for each different value of the count $C(w_3)$. The value γ should actually depend only on ranges into which $C(w_3)$ falls, since very few activities w_3 will have high values of $C(w_3)$. According to Jelinek [14], we have to split our activities. For instance, ranges can be used as follows: two ranges $\mathcal{R}$ can be derived by splitting the activities. The first range $\mathcal{R}_1$ contains the activities that appear more than average and the second range $\mathcal{R}_2$ contains activities appearing less than average. With the help of some numerical computations the estimation can be done with the following steps.

1. The total training data is divided into kept and held-out data sets. The kept data set is twice as big as the held-out data set.
2. The relative frequencies $f(w_4|w_1,w_2,w_3)$, $f(w_4|w_2,w_3)$, $f(w_4|w_3)$ and $f(w_4)$ are calculated from the kept data set.
3. $N(w_4,w_3,w_2,w_1)$, $N(w_4,w_3,w_2)$ and $N(w_4,w_3)$ are derived. $N(w_4,w_3)$ is the number of times the bigram (w_4,w_3) takes place in the held-out data set.
4. Finding γ maximizing the value

$$\sum_{v\in\mathcal{R}}\sum_{w_4} N(v,w_4)\log[\gamma f(w_4)+(1-\gamma)f(w_4|w_3)]\,. \tag{3.5}$$

As said, in step 4 above, γ is estimated by using the maximum likelihood method. To apply this method we first have to define the likelihood function:

$$L(\gamma;a_1,\ldots,a_n,b_1,\ldots,b_n)=\prod_{i=1}^{n}\prod_{j=1}^{n}P(w_4=a_i|w_3=b_j;\gamma)\,.$$

With 18 activities we get

$$\begin{aligned} L(\gamma;a_1,\ldots,a_{18},b_1,\ldots,b_{18}) &= \prod_{i=1}^{18}\prod_{j=1}^{18}P(w_4=a_i|w_3=b_j;\gamma) \\ &= \prod_{i=1}^{18}\prod_{j=1}^{18}\gamma(C(b_j))f(a_i)+(1-\gamma(C(b_j)))f(a_i|b_j)\,. \end{aligned}$$

We search for the maximum of $\log L$, since it is easier to work with sommations than products. Clearly, $\log L$ and L have the same extreme points, since

$L(\gamma) > 0$ for $0 \leq \gamma \leq 1$ and because the logarithm is a monotone increasing function. This is how we obtain Equation (3.5) in step 4:

$$\begin{aligned} L^*(\gamma; a_1, \ldots, a_{18}, b_1, \ldots, b_{18}) &= log\, L(\gamma; a_1, \ldots, a_{18}, b_1, \ldots, b_{18}) \\ &= log \prod_{i=1}^{18} \prod_{j=1}^{18} \gamma(C(b_j)) f(a_i) + (1 - \gamma(C(b_j))) f(a_i|b_j) \\ &= \sum_{i=1}^{18} \sum_{j=1}^{18} log \left[\gamma(C(b_j)) f(a_i) + (1 - \gamma(C(b_j))) f(a_i|b_j)\right] . \end{aligned}$$

To maximize Equation (3.5), we take the derivative with respect to γ and set the result to 0:

$$\sum_{v \in \mathcal{R}} \sum_{w_4} N(v, w_4) \left[\gamma + \frac{f(w_4|v)}{f(w_4) - f(w_4|v)}\right]^{-1} = 0 . \tag{3.6}$$

We have to find $\gamma \in [0, 1]$ satifying Equation (3.6). We know that although the expression on the left hand side of Equation (3.6) has singularity points, the singularity points are not in [0,1]. The proof is given in Appendix D.1. To solve for γ we substitute γ's $\in [0, 1]$ in Equation (3.6) and take the one which approaches the maximum the most. Once γ is found, the probability $P^*(w_4|w_3)$ in Equation (3.2) can be determined. With $P^*(w_4|w_3)$ known, finding the values θ for Equation (3.3) and ζ for Equation (3.4) is the same as finding γ's for Equation (3.2).

Eventually, we have all the γ's, θ's and ζ's. They can be used to determine the λ_i's. Substituting Equation (3.2) and (3.3) in Equation (3.4) gives:

$$\begin{aligned} P(w_4|w_1, w_2, w_3) &= \zeta\Big(\theta\big(\gamma f(w_4) + (1 - \gamma) f(w_4|w_3)\big) + \\ &\quad (1 - \theta) f(w_4|w_2, w_3)\Big) + (1 - \zeta) f(w_4|w_1, w_2, w_3) \end{aligned}$$

and the λ_i's become:

$$\begin{aligned} \lambda_1 &= \zeta \cdot \theta \cdot \gamma \\ \lambda_2 &= \zeta \cdot \theta \cdot (1 - \gamma) \\ \lambda_3 &= \zeta \cdot (1 - \theta) \\ \lambda_4 &= 1 - \zeta . \end{aligned}$$

Hence, the smoothed transition matrix becomes:

$$\begin{aligned} P(w_4|w_1, w_2, w_3) &= \lambda_1 f(w_4) + \lambda_2 f(w_4|w_3) + \lambda_3 f(w_4|w_2, w_3) + \\ &\quad \lambda_4 f(w_4|w_1, w_2, w_3) . \end{aligned}$$

3.3 Result of evaluated parameters

In the last subsection we showed how the parameters could be evaluated. For simplicity we only used one range. Now, we apply the maximum likelihood

Linear smoothing parameters λ_i
$\lambda_1 = 0.0008$
$\lambda_2 = 0.0467$
$\lambda_3 = 0.3836$
$\lambda_4 = 0.5691$

Table 3.1: The λ_i's for all the activities.

estimation to evaluate the parameters of the model. Hence, the λ_i's become as in Table 3.1.

With these λ_i's we are able to build our smoothed transition matrix. We are not able to print the matrix, since the dimensions of the matrix are 5832 by 19. However, with this matrix we can predict the pathways of patients.

3.4 Summary

In this chapter we presented a solid model for the prediction of activities and therefore, the prediction of logistical pathways. We evaluated the parameters by applying the maximum likelihood estimation. As a result, we got 4 smoothing parameters and our smoothed transition matrix.

Chapter 4

Model validation & refinement

An important methodological issue is to insure that a model is credible, i.e. to evaluate the model in terms of its performance. In Chapter 3 we used the *training set* to evaluate the parameters for our model. We need the *test set* to measure its performance. In this chapter we use the training set to validate our model by estimating the average length of a DTC and compare this with the average length of the test set.

4.1 Model validation

To see whether our model works correctly and predicts accurately, we have to validate the model. This is done by estimating the average length of a DTC using the smoothed transition model. We chose to estimate the average length as validation, since the average length is easy to calculate for both test set and training set.

To calculate the average length of the test set we just add all the lengths of the DTCs and divide this by the number of DTCs. The average length of the training data is calculated by summing the probability of the length of a DTC being larger than n, i.e. the probability that a DTC will not end. The estimated average length calculated with the training set, is compared with the average length calculated with the test set. Hence, if the probabilities of a DTC that ends after n activities is equal for both the training set as the test set, we may assume that our model is correct and can be used to predict logistical pathways.

4.1.1 Estimating the average length of a DTC

In Figure 2.1 a cumulative probability is given to show the distribution of the lengths of DTCs. In this section we use our model to determine the average

length of a DTC. The random variable X is the length of a DTC.

$$\begin{aligned}\mathbb{E}(\text{length of DTC}) &= \sum_{n=0}^{\infty} P(X > n) \\ &= \sum_{n=1}^{\infty} P(X \geq n) \\ &= \sum_{n=1}^{\infty} P(\text{no "end" till } n) \\ &= 1 + \sum_{n=2}^{\infty} P(\text{no "end" till } n)\,. \end{aligned} \tag{4.1}$$

The probability a DTC is longer than n activities equals: $P(\text{no "end" till } n)$. That means the DTC will not end at activity n. Thus, $P(\text{no "end" till } 1) = 1$, since there are no DTCs of length 0. Hence, they are always 1 or longer.

$$\begin{aligned}P(X \geq 2) &= \sum_{i,j} \pi_1(i) A_1(i,j) \\ &= \sum_{i,j} \pi_2(ij)\,, \end{aligned} \tag{4.2}$$

where,

$$\pi_1(i) = P(\text{DTC starts with } i)\,.$$

$A_1(i,j)$ is the same as the bigram transition matrix in section 3.2:

$$A_1(i,j) = P(\text{transition to } j \,|\, \text{started at } i)\,.$$

For $P(X \geq 3)$, $\pi_2(ij)$ and the trigram $A_2(ij, jk)$ are used. $\pi_2(ij)$ is the probability that a DTC starts with (ij). Using the same Equation as (4.2), $\pi_3(ijk)$ follows. The same can be done for $P(X \geq 4)$: $\pi_3(ijk)$ and the smoothed transition matrix $A_3(ijk, jkr)$ are used. In Equation (4.1) we sum over n from two to infinity. Thus, our next step would be determining a fivegram. However, instead of a fivegram, we use our smoothed transition matrix again. To reach infinity a recursive equation in vector-matrix notation emerges:

We obtain the following result, since we use $A_{n-1} = A_3 \quad \forall n \geq 4$.

Proposition 1

$$\begin{aligned} P(X \geq n) = \pi_n &= \pi_{n-1}\, A_{n-1} \\ &= \pi_{n-2}\, A_{n-2}\, A_{n-1} \\ &= \pi_{n-3}\, A_{n-3}\, A_{n-2}\, A_{n-1} \\ &\;\;\vdots \\ &= \pi_3\, A_3\, A_4 \,\ldots\, A_{n-1} \\ &= \pi_3\, A_3^{n-3} \qquad\qquad \forall n \geq 4\,. \end{aligned}$$

Using Equation (4.1) and Proposition 1, we can validate our model, by calculating the average length of a DTC.

4.1.2 Results of the model validation

The results of our approach validation are shown in this subsection. The validation is done by estimating the average length of a DTC.

We calculated the average length of a DTC using the test set. After that, we calculated the average length using Proposition 1. In Table 4.1 the average length of a DTC, calculated with the test set and calculated with our approach, is given.

Further in the report we will give the conditional expectation of the length of a DTC and conclude that with a refinement to the model we can predict logistical pathways very good for DTCs which are not chronical. That means, DTCs shorter than 75 activities.

Dataset	$\mathbb{E}[X]$
Test set	24.63
Smoothed model	16.38

Table 4.1: Average length of DTC, calculated with the test set and our smoothed model.

In Figure 4.1 we see the model validation. The lines describe the probability of the length of a DTC being smaller or equal than n, i.e. $P(X \leq n)$ with variable $X =$ length of DTC. We can see that the distance between the smoothed data and the test data is large, and not acceptable. Whereas, if the smoothed data is that much underneath the test set, the model has to be adjusted in order to predict a logistical pathway.

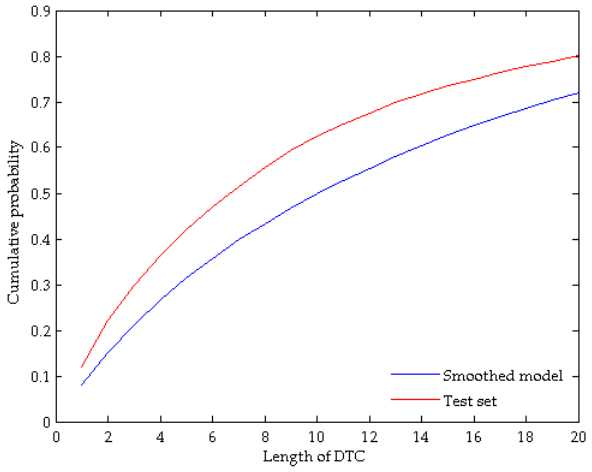

Figure 4.1: The distribution of the length of DTCs. We compare our smoothed model with the test set.

4.2 Refinement of the model

From the last section we saw that there is a difference between the average length derived with our model and with the test set. This makes us wonder if we can make a refinement to the model and if there are more indications that point out a refinement is necessary. In this section we make a refinement and estimate the average length of a DTC much better.

In Figure 2.1 we see that 80% of the DTCs have length 20 or less. The question arises if the transitions of the first 20 activities of a DTC are different than the transitions after 20 activities. In Figure 4.2 we can see that if we divide a DTC into two intervals the distribution of the percentages of the activities is different. The distribution in the figure on top is of the first 20 activities and the figure below is of activity 20 until the end of the DTCs. We can conclude that we have to use more than one smoothed matrix, since the distribution of the percentages between the first 20 activities and the remaining activities and also the number of times an activity happens is different in the stages of the patient's progress in MHC.

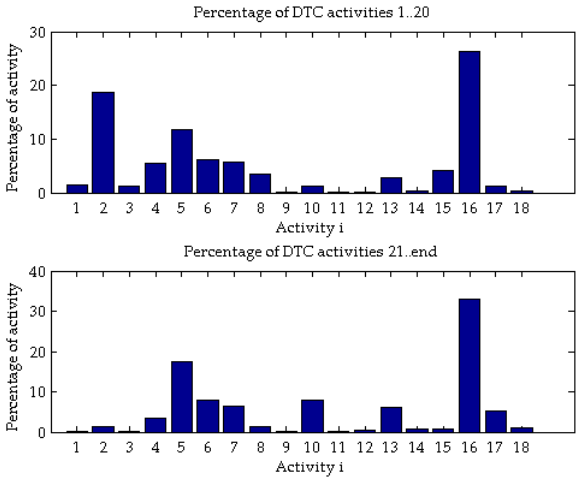

Figure 4.2: Percentage of DTC, divided into two intervals. One interval contains the first 20 activities and the other interval contains activity 21 until the end of the DTCs.

Let us try to split the activities after 5 activities and after 10 activities. The result in Figure 4.3 is not as good as in the right picture in Figure 4.4, where we make four intervals. By trial and error we come to the conclusion that the cumulative probabilities with the refinement of making four intervals at activity 20, 45 and 75 follow the cumulative probabilities of the test set much better.

In the beginning of the DTC the cumulative probability of our refinement follows the cumulative probability of the test set perfectly. Which means that with four intervals the prediction of the pathways will be much better than with one interval. Hence, we make 4 smoothed matrices. The first matrix is A_3

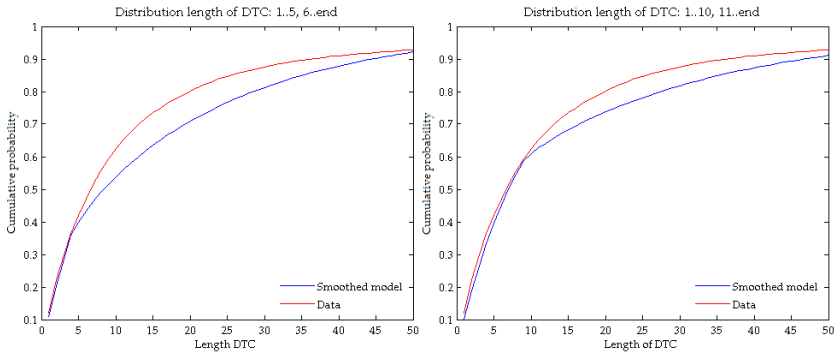

Figure 4.3: Percentage of DTC, divided into two intervals at activity 5 and 10.

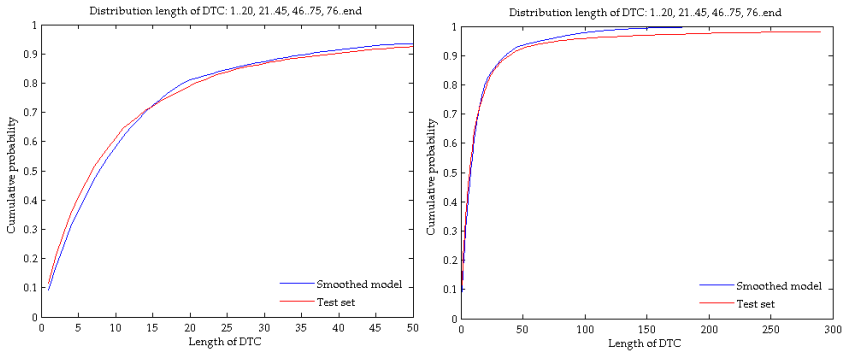

Figure 4.4: Percentage of DTC, divided into two intervals, after 20 activities.

with short DTCs, computed with the data of the first 20 activities. The second matrix is A_4 with moderate DTCs, with the data of activity 21 until activity 45. The third matrix A_5, long DTCs, with the data of activity 45 until activity 75. And the last matrix A_6 of activity 76 until the end with DTCs of chronical patients. All intervals need their own transition matrix with their own corresponding λ_i's.

The result is shown in Table 4.2. The parameter λ_i is the factor that determines how much weight the relative frequencies get. For example: the larger λ_4, the more the three previous activities are important to predict the next activity. The smaller λ_1, the less important the relative frequency $f(w_4)$ is to predict $P(w_4|w_1, w_2, w_3)$.

The differences between the λ_i's are pretty large. Together with the relative frequencies $f(w_i)$, $f(w_i|w_{i-1})$, $f(w_i|w_{i-1}, w_{i-2})$ and $f(w_i|w_{i-1}, w_{i-2}, w_{i-3})$ and the number of times a bigram $N(w_4, w_3)$, trigram $N(w_4, w_3, w_2)$ and fourgram $N(w_4, w_3, w_2, w_1)$ takes place, four smoothed matrices are derived. We observed that the four matrices are considerably different. For example, there are more zeros where a DTC ends, in the probability matrix for short DTCs. Therefore, the smoothed matrix for short DTCs will be different.

Short DTCs	Moderate DTCs	Long DTCs	Chronical DTCs
$\lambda_1 = 0.0010$	$\lambda_1 = 0.0003$	$\lambda_1 = 0.0006$	$\lambda_1 = 0.0028$
$\lambda_2 = 0.0791$	$\lambda_2 = 0.1529$	$\lambda_2 = 0.2132$	$\lambda_2 = 0.0913$
$\lambda_3 = 0.4469$	$\lambda_3 = 0.4878$	$\lambda_3 = 0.4782$	$\lambda_3 = 0.4348$
$\lambda_4 = 0.4730$	$\lambda_4 = 0.3590$	$\lambda_4 = 0.3080$	$\lambda_4 = 0.4711$

Table 4.2: The λ_i's for the four intervals.

The smoothed transition matrices are too large to represent in this report. Because of the total number of activities being equal to 18 and the three previous activities, we get $18 \cdot 18 \cdot 18 = 5832$ combinations of activities, which are states in the smoothed transition matrices. These states can make a transition to 19 states, since we get one extra state, called 'finished'. These matrices are completely filled with elements between 0 and 1, since the elements are all probabilities.

In Table 4.3 we see the difference between the model and the refinement to the model. We take the smoothed transition matrices and add all the probabilities for each activity a patient can undergo. We see that the smoothed matrix of short DTCs is different from the other matrices. The refinement to the model is recommended and applied in this research.

	Probability mass in percentage %				
Activity	Complete DTCs	Short DTCs	Moderate DTCs	Long DTCs	Chronical DTCs
0	34.47	39.57	34.14	33.38	34.10
1	0.59	0.70	0.42	0.02	0.31
2	4.45	5.23	3.10	1.93	2.79
3	1.52	1.38	1.30	1.05	1.22
4	4.89	4.93	4.81	4.87	4.71
5	6.66	6.06	7.08	7.37	6.96
6	4.47	4.16	4.51	4.89	4.47
7	4.68	4.00	5.25	5.05	5.00
8	2.25	2.52	2.06	1.56	1.77
9	1.54	1.23	1.14	1.60	1.54
10	3.75	3.10	4.55	4.23	4.36
11	1.73	1.05	1.72	1.86	1.88
12	2.15	1.62	1.58	1.95	2.19
13	4.30	3.76	4.33	5.55	4.76
14	1.96	1.95	1.94	1.91	1.88
15	2.05	2.15	2.06	1.62	1.78
16	12.83	11.64	13.93	14.49	14.12
17	3.02	2.29	3.24	3.91	3.49
18	2.69	2.67	2.83	2.74	2.69

Table 4.3: Probability mass per activity. We used the smoothed matrices and for each matrix we added all the probabilities going to activity i.

Proposition 1 transforms into the expression given in the next proposition.

Proposition 2

$$P(X \geq n) = \begin{cases} \pi_3 \, A_3^{n-3} & \text{for } 4 \leq n \leq 20 \\ \pi_3 \, A_3^{n-3} \, A_4^{n-20} & \text{for } 21 \leq n \leq 45 \\ \pi_3 \, A_3^{n-3} \, A_4^{n-20} \, A_5^{n-45} & \text{for } 46 \leq n \leq 75 \\ \pi_3 \, A_3^{n-3} \, A_4^{n-20} \, A_5^{n-45} \, A_6^{n-75} & \text{for } 76 \leq n < \infty \end{cases} .$$

In the left picture in Figure 4.4 we can see that if the DTC becomes longer than 75 activities, the probability the length of a DTC is shorter than n, calculated with our smoothed matrices, goes to one much faster than the test set shows. If a patient gets more than 75 treatments, the test set implies that the patient will stay much longer than 75 activities. A patient will not leave the system as fast as our model predicts. This explains why the average length of our model is smaller than calculated with the test set. In Table 4.4 we can see that the average length of our refined model is smaller than the average length of the test set. We conclude that chronical DTCs should be taken into account separately.

Dataset	$\mathbb{E}[X]$
Test set	24.63
Smoothed model	16.38
Refined smoothed model	16.40

Table 4.4: Average length of DTC, calculated for the test set, the smoothed model and our refined smoothed model, where the data was split in four intervals.

Because of the difference in lengths between the test set and the refined model, we calculate the average length of a DTC given that it is smaller than y. If the difference is small, we can say our refined model will predict logistical pathways very well for DTCs smaller than y activities.

If we take a look at the average length given that a DTC is smaller than y activities, we get the following formula

$$\begin{aligned}\mathbb{E}[X|X \leq y] &= \sum_{n=1}^{\infty} n\mathbb{P}(X = n|X \leq y) \\ &= \sum_{n=1}^{\infty} \frac{n\mathbb{P}(X = n, X \leq y)}{\mathbb{P}(X \leq y)} \\ &= \frac{1}{\mathbb{P}(X \leq y)} \sum_{n=1}^{y} n\mathbb{P}(X = n)\,.\end{aligned}$$

In Table 4.5 the results of $\mathbb{E}[X|X \leq y]$ with different y's are given. The average length $\mathbb{E}[X|X \leq 100] = 16.58$ for the testset and calculated with the refined model $\mathbb{E}[X|X \leq 100]$ equals 15.66. The average length of our refined model approaches that of the test set with a difference of 5.5%. In future work another model for DTCs of chronical patients has to be developed.

We conclude that the methodology predicts the pathways really well for short, moderate and long DTCs.

4.3 Average time in system

In the current situation a time element is not implemented. It is not known how long a patient stays in a DTC. In Section 2.4 we saw the mean time a patient spends between two activities. In Section 2.5 we presented the mean duration of the activity. However, we want to know how long a patient stays in

y	$\mathbb{E}[X \mid X \leq y]$ of test set	$\mathbb{E}[X \mid X \leq y]$ of refined model	Difference in %
80	15.77	15.31	2.9
90	16.19	15.51	4.2
100	16.58	15.66	5.5
110	16.93	15.78	6.8

Table 4.5: Average length of DTC given the maximum number of activities.

a DTC. Therefore, we have to investigate how to implement the time elements. First, we consider a continuous phase-type distribution. After that, we derive a new equation to calculate the total time in a DTC, since we have our refined smoothed model.

With the time a patient spends in the system, we have two time elements which we have to consider. The first element is the time the patient spends in the activities, also called: the duration of the activities. The second element is the time the patient spends between the activities. To derive the expected time of the patient in the system we neglect the duration of an activity, since this time is registered in minutes and hours and the time in between activities in days.

A continuous phase-type distribution can be defined as the time until absorption in a continuous time Markov chain with one absorbing state and all other states are transient [15]. The transition rate matrix Q of that process can be written as:

$$Q = \begin{bmatrix} t & T \\ 0 & 0 \end{bmatrix},$$

where the first entry in the state space represents the absorbing state. The absorbing state in our case is the end state, when the DTC has finished. The vector t is given by the following equation, since the sum of the elements on each row must equal zero:

$$t = -T \cdot 1,$$

where 1 is a $m \times 1$ column vector of ones and T a $m \times m$ matrix:

$$T = \begin{bmatrix} -\nu_1 & \lambda_{1,2} & \lambda_{1,3} & \cdots & \lambda_{1,m} \\ \lambda_{2,1} & -\nu_2 & \lambda_{2,3} & \cdots & \lambda_{2,m} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda_{m,1} & \cdots & \cdots & \cdots & -\nu_m \end{bmatrix}.$$

In this research we focus on the mean time between activity i and all the other activities. The time in each activity can be investigated in future research.

In our case $\nu_i = p_{i,i}\lambda_i$, since the patient stays between activity i and the other activities for a mean time and can return to this activity with probability

$p_{i,i}$. λ_i equals $\frac{1}{\mu_i}$, that is the arrival rate of patients at activity i. Where, μ_i is the mean time a patient stays between activity i and the other activities. The $\lambda_{i,j}$ equals the rate going from i to j, which can be calculated through $\lambda_{i,j} = \frac{p_{i,i}\lambda_i p_{i,j}}{1-p_{i,i}}$. With $\frac{p_{i,j}}{1-p_{i,i}}$ the probability of going to activity j given that the patient does not go to activity i again. Hence,

$$T = \begin{bmatrix} -p_{1,1}\lambda_1 & \frac{p_{1,1}\lambda_1 p_{1,2}}{1-p_{1,1}} & \cdots & \cdots & \frac{p_{1,1}\lambda_1 p_{1,m}}{1-p_{1,1}} \\ \frac{p_{2,2}\lambda_2 p_{2,1}}{1-p_{2,2}} & -p_{2,2}\lambda_2 & \frac{p_{2,2}\lambda_2 p_{2,3}}{1-p_{2,2}} & \cdots & \frac{p_{2,2}\lambda_2 p_{2,m}}{1-p_{2,2}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \frac{p_{m-1,m-1}\lambda_{m-1} p_{m-1,m}}{1-p_{m-1,m-1}} \\ \frac{p_{m,m}\lambda_m p_{m,1}}{1-p_{m,m}} & \cdots & \cdots & \frac{p_{m,m}\lambda_m p_{m,m-1}}{1-p_{m,m}} & -p_{m,m}\lambda_m \end{bmatrix}.$$

One state of matrix T represents the transition rate of activity i to activity j. Therefore,

$$t = -\begin{bmatrix} (-p_{1,1}\lambda_1 + \sum_{i\setminus\{1\}} \frac{p_{1,1}\lambda_1 p_{1,i}}{1-p_{1,1}}) \\ (-p_{2,2}\lambda_2 + \sum_{i\setminus\{2\}} \frac{p_{2,2}\lambda_2 p_{2,i}}{1-p_{2,2}}) \\ \vdots \\ (-p_{m,m}\lambda_m + \sum_{i\setminus\{m\}} \frac{p_{m,m}\lambda_m p_{m,i}}{1-p_{m,m}}) \end{bmatrix} = \begin{bmatrix} -\frac{p_{1,1}\lambda_1 p_{1,0}}{1-p_{1,1}} \\ -\frac{p_{2,2}\lambda_2 p_{2,0}}{1-p_{2,2}} \\ \vdots \\ -\frac{p_{m,m}\lambda_m p_{m,0}}{1-p_{m,m}} \end{bmatrix}.$$

The probability $p_{i,0}$ equals the probability of going from activity i to 0, that is where the DTC ends.

The process has an initial probability of starting in any of the $m+1$ phases given by the probability vector

$$\begin{bmatrix} \alpha_0 & \alpha \end{bmatrix},$$

where α is the probability of starting the process in one of the non-absorbing activities and $\alpha_0 = 0$, since we always begin the DTC in a non-absorbing activity.

The distribution of a continuous phase-type variable Y is then completely determined by the parameters (α, T) and its cumulative distribution function becomes:

$$F(y) = 1 - \alpha e^{Ty}1 \qquad y \geq 0\,.$$

It follows that the probability density function can be computed as

$$f(y) = \alpha e^{Ty}a \qquad y > 0\,.$$

The derivation of the cumulative distribution function and the probability density function can be found in Neuts [16]. Using the Laplace-Stieltje transform of $F(\cdot)$, the non-centered moments become:

$$\mathbb{E}[Y^k] = (-1)^k k!\alpha T^{-k}1 \qquad \text{for k} \geq 1\,.$$

Hence, the first moment, which equals the total time spent in the system by a patient until his DTC has finished, becomes

$$\mathbb{E}[Y] = -\alpha T^{-1}1\,. \tag{4.3}$$

For the refined model, we have to modify Equation (4.3) in the following way:

Let $\overline{\mu}$ be the mean time spent between one activities and the others, as shown in Table 4.6. The dimension of $\overline{\mu}$ equals 18, since there are 18 activities.

Activity	Average in between time in days μ_i
1	18.8
2	13.2
3	9.8
4	10.2
5	6.7
6	7.4
7	7.5
8	14.5
9	3.1
10	2.4
11	2.9
12	3.3
13	5.1
14	3.6
15	4.9
16	4.5
17	2.7
18	3.6

Table 4.6: Mean time between activity i and the other activities.

With probability π_1 the patient starts in an activity and spends some time $\overline{\mu}$ between that activity and the next activity. Hence, the time spent after one activity is:

$$\mathbb{E}[Y_1] = \pi_1 \overline{\mu}\,.$$

After the first activity the patient goes to second activity with probability $\pi_1 A_1$ and we also have to multiply this probability with $\overline{\mu}$ to get the time he stays between the second and the third activity

$$\mathbb{E}[Y_2] = \pi_1 A_1 \overline{\mu}\,.$$

To get the expected value of our total time spent in the system we have to add all the probabilities of going to the next activity multiplied by the probability that we spend some time in between the activities. We use the other smoothed matrices A_4, A_5 and A_6, since we made a refinement where we devided the DTCs in intervals, short, moderate, long and chronical DTCs, respectively. The expected time spent in the system by a patient is given as follows.

Proposition 3

$$\begin{aligned}
\mathbb{E}[Y_{total}] &= \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \cdots + \mathbb{E}[Y_\infty] \\
&= \pi_1\overline{\mu} + \pi_1 A_1\overline{\mu} + \pi_1 A_1 A_2\overline{\mu} + \pi_1 A_1 A_2 A_3\overline{\mu} + \\
&\quad \pi_1 A_1 A_2 A_3^2\overline{\mu} + \cdots + \pi_1 A_1 A_2 A_3^{17}\overline{\mu} + \\
&\quad \pi_1 A_1 A_2 A_3^{17} A_4\overline{\mu} + \cdots + \pi_1 A_1 A_2 A_3^{17} A_4^{25}\overline{\mu} + \\
&\quad \pi_1 A_1 A_2 A_3^{17} A_4^{25} A_5\overline{\mu} + \cdots + \pi_1 A_1 A_2 A_3^{17} A_4^{25} A_5^{30}\overline{\mu} + \\
&\quad \pi_1 A_1 A_2 A_3^{17} A_4^{25} A_5^{30} A_6\overline{\mu} + \cdots + \pi_1 A_1 A_2 A_3^{17} A_4^{25} A_5^{30} A_6^{\infty}\overline{\mu}.
\end{aligned}$$

Once we know the vector π_1 and the matrices A_1, A_2, A_3, A_4, A_5 and A_6 in Proposition 3 we are able to calculate the total expected time spent in the system by a patient. The result is that a patient stays 114.7 days in the system against the 153.5 days the patient stays in the system calculated with the test set.

4.4 Summary

As we validated our model estimating the average length of a DTC, we concluded that we had to refine it. We made four smoothed transition matrices instead of one and we could see that they predicted the average length perfectly if the DTCs do not have chronical patients.

Chapter 5

Results

After we validated and refined our model we want to show how it can be applied. In this chapter we show a couple of trees with the most important logistical pathways of patients. Also the repeating activities will be discussed.

5.1 Examples of logistical pathways

In Figure 5.1 we give an example of how a logistical pathway can be made. The beginning of a DTC starts with the first node 'Start'. In the first level we show only activities that appear at the first step in at least 5% of the DTCs. Thus, in the first level more than 5% of the patients goes to pre-intake (activity 1), diagnostic (activity 2), supporting and structuring treatment contact (activity 5), crisis care (activity 15) and general indirect (activity 16). In the second level we use 2.5% threshold. That means, that more than 2.5% of the patients go to the next activities. For example, the patient in diagnostic (activity 2) gets diagnosed again, he goes to supporting and structuring treatment contact (activity 5) or the doctor writes general indirect time (activity 16) to prescribe some medication, etc.

Another example: if the patient already had 3 activities, than there are 18^3 possible DTCs he could have had. The 10 most occuring DTCs, are the starting points of this tree in Figure 5.2. With a probability of 0.15 the patient goes to the next node. The probability for each node is above the figure.

5.2 Zipf's law & most frequent DTCs

We can sort the pathways after for instance, three activities. The most frequent pathways are on top of the ranking. We make a cumulative probability with the most frequent pathways. In Figure 5.3 we plotted the cumulative probabilities against the number of most frequent pathways. We can see that with three activities 66% of the pathways are covered by the 100 most frequent pathways. With three history steps, there are 5832 activities. Therefore, we can say there

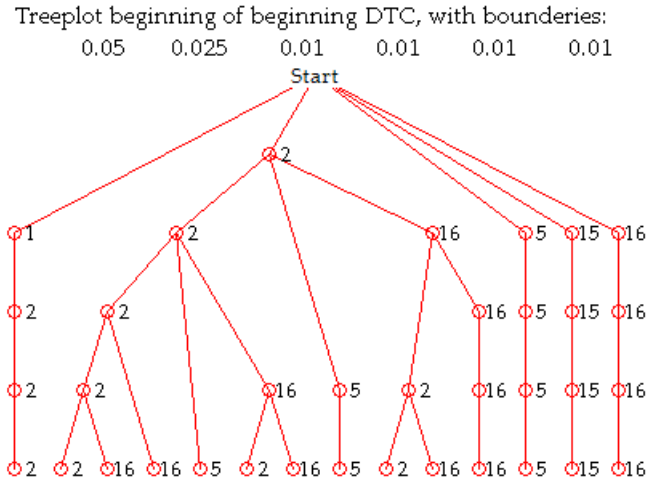

Figure 5.1: The beginning treeplot of a DTC. For example: on top the DTC starts. 5% of the patients go to pre-intake and after that 2.5% of all the patients go to diagnostic. 1% of the patients stays in diagnostic.

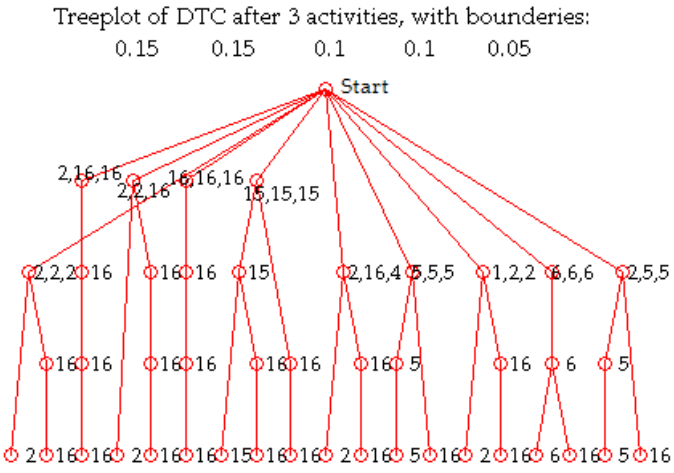

Figure 5.2: The treeplot of a DTC after three activities. After 3 activities we take the 10 pathways with the highest probability. After that, we set bounderies for going to the next activity. So, 15% of the patients go to the next node.

are definitely vast patterns in the patient's sequencing activities in MHC. We do the same for the most frequent pathways after 6, 9, 12, 15, 18 and 20 activities.

In most of the natural systems, as in language, website popularity, cities, a well-known Zipf's law is observed [17]. Zipf's law originally states that, in a corpus of natural expressions, the frequency of any word is roughly inversely proportional to its rank in the frequency table. In our case we want to investi-

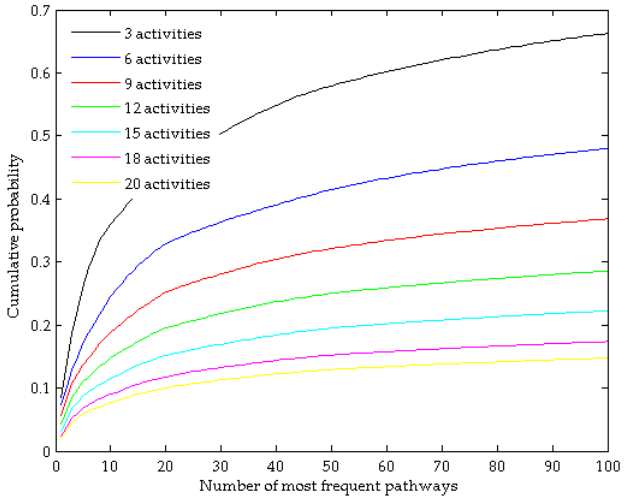

Figure 5.3: The cumulative probability of the 100 most frequent pathways after 3, 6, 9, 12, 15, 18 and 20 activities.

gate if this law holds for pathways. Thus, the most frequent pathway will occur approximately twice as often as the second most frequent pathway, which occurs twice as often as the fourth most frequent pathway, etc.

In Figure 5.4 we depict on the horizontal axis the logarithm of the rank of the most frequent pathway. On the vertical axis we see the logarithm of the probability of the most frequent pathways. Instead of frequencies, we used the probabilities. This does not change the picture, since the difference between frequencies and probabilities is only in a multiplicative constant. We see that Zipf's law holds for the ten most frequent pathways. After that, the function fluctuates. However, still has a trent downwards, more or less following Zipf's law.

5.3 Repeating activities

In Section 2.2 we saw the probability of going to activity i given the number of times the patient has been there before. Using the smoothed transition matrix, we see in Figure 5.5 that the probabilities are increasing.

The probabilities are the same after the patient has been there four times before, since we work with the smoothed transition matrix. This does not predict the probabilities dropping to zero, since the activities will not happen more times after each other and the probabilities of the other activities will not increase any further.

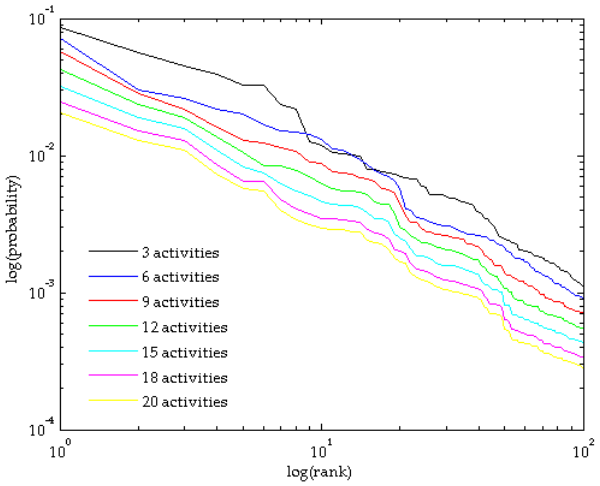

Figure 5.4: Zipf's law tested on our smoothed model.

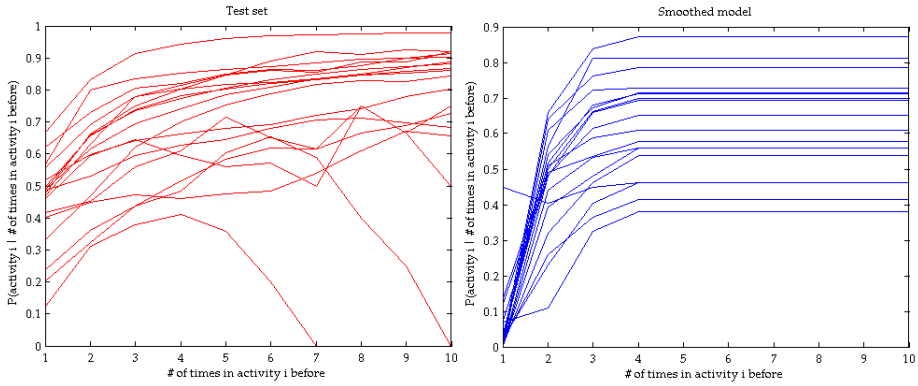

Figure 5.5: Repeating activities: the probability of going to activity i given the number of times the patient has been there before. In the left picture we see the repeating activities build with the test set. On the right we see the prediction with our smoothed model.

Chapter 6

Conclusions & recommendations

6.1 Conclusions

In order to be able to predict a logistical pathway of a patient, we need a transition matrix to know the probabilities of going from one activity to another. The data used in this research is very large. However, if we take only the data of one MHC institution our dataset is not that large. Especially if we want to take into account were the patient is coming from and which activities he already had in the past. A language smoothing model is the solution to this problem.

The refinement of the model, dividing the DTCs in intervals gives us a better prediction of the DTCs. Despite the corresponding average length of a DTC being shorter than the average length calculated with the test set, the division into four intervals gives a more realistic picture of the pathways of patients. This is especially true for the non-chronical patients.

The methodology used in this research gives us insight in DTCs. The average length of a DTC and average time spent in a DTC can be calculated. Moreover, we can predict logistical pathways and see if these pathways

6.2 Recommendations

In this research we made some choices to make the methodology work. Of course other choices are possible. In this section we make some recommendations for further research and give some suggestions were that can take us.

The idea of ranges is that activities appearing very often have other λ_i's than activities appearing less often. As we saw in our data we have activities that appear very often and other activities that appear less often. We recommend to apply ranges to get a more sophisticated model.

We decided to make an adaptation to the model we developed and we

chose to split the activities in intervals; short, moderate, long and chronical DTCs. With these intervals we made four smoothed matrices. We saw that the logistical pathways could be perfectly predicted for DTCs smaller than 75 activities. However, the difference between the average length of the test set and training set was large. A recommendation will be to further investigate chronical DTCs.

The data used is very general, since the data is of more than one institution. This methodology takes 3 history steps into account and therefore the methodology described in this research can also be applied to the SHC. If we want more specific information we can use clustering techniques to split patients in groups that are diagnosed to the same disorder. By deriving pathways of patients we are able to predict where the bottlenecks in capacity will arise. Schedules for surgeries, doctors and nurses, bed capacity and length of stay can be derived for MHC institutions as well as for SHC institutions.

Appendix A

Abbreviations, definitions & symbols

In this chapter some tables with abbreviations, definitions and symbols.

A.1 Abbreviations

DTC	Diagnosis treatment combination
MHC	Mental health care
SHC	Somatic health care

A.2 Definitions

Activity	Treatment a patient gets or action a doctor does.
Bigram	Transition matrix with probabilities of going to activity j, given the previous activity i.
Direct time	Time a doctor is in contact with the patient.
Indirect time	Time a doctor is not direcly in contact with the patient. However, he has to work on the patient's documentation.
Logistical pathway	Time line, on which every event relating to treatment can be entered.
Trigram	Transition matrix with probabilities of going to activity k, given activity (i, j).
Unigram	Transition matrix with probabilities of going to activity r, given activitiy (i, j, k).

A.3 Symbols

$\lightning$	Contradiction!
$\square$	Quod erat demonstrandum, that which had to be proven.
$\mathbb{E}[Y]$	Expected time the patient stays in the system calculated with continuous phase type distribution.
$\mathbb{E}[Y_1]$	Expected time the patient is between activity 1 and 2.
$\mathbb{E}[Y_2]$	Expected time the patient is between activity 2 and 3.
$\mathbb{E}[Y_{total}]$	Total expected time the patient stays in the system.
α	Probability of starting the process in one of the non-absorbing states.
α_0	Probability of starting the process in the absorbing state.
γ	Coefficient for smoothing parameter.
λ_i	Linear smoothing factor.
$\lambda_{i,j}$	Transition rate from activity i to activity j.
μ_i	Mean time between activity i and the other activities.
ν_i	
$\pi_1(i)$	Probability the DTC starts with i.
$\pi_2(ij)$	Probability the DTC starts with (ij).
$\pi_3(ijk)$	Probability the DTC starts with (ijk).
θ	Coefficient for smoothing parameter.
ζ	Coefficient for smoothing parameter.
$A_1(i,j)$	Probability of the transition from i to j.
$A_2(ij,jk)$	Probability of the transition from (ij) to (jk).
$A_3(ijk,jkr)$	Probability of the transition from (ijk) to (jkr).
$C(w_2,w_3)$	Number of counts of words/activities (w_2,w_3).
$f(w_4\|w_3)$	Relative frequency of word/activity w_4, given the previous word/activity w_3.
$f(w_4\|w_2,w_3)$	Relative frequency of word/activity w_4, given the two previous words/activities (w_2,w_3).
$f(w_4\|w_1,w_2,w_3)$	Relative frequency of word/activity w_4, given the three previous words/activities (w_1,w_2,w_3).
$L(\gamma;a_1,\ldots,a_n,b_1,\ldots,b_n)$	Likelihood function.
n	Length of the DTC.
$N(w_1,w_2,w_3,w_4)$	Number of counts the fourgram (w_1,w_2,w_3,w_4) appears in the held out data.
$p_{i,j}$	Probability of going from activity i to activity j.
$P(w_4\|w_1,w_2,w_3)$	Probability of word/activity w_3, given the three words/activities (w_1,w_2,w_3).
Q	Transition rate matrix.
$\mathcal{R}$	Range.
t	Transition rate vector of absorbing state 0.
T	Transition rate matrix of all activities, except absorbing state 0.
w_i	Word or activity at time i.
$\mathcal{V}$	Vocabulary.
y	Length of the DTC.

Appendix B

List of activities

Table B.1 gives the list of activities, which belong to the level that experts find most important for this research.

Activity number	Activity
0	Finished
1	Pre-intake
2	Diagnostic
3	Psychodiagnostic examination
4	Follow up treatment contact
5	Supporting and structuring treatment contact
6	Psychotherapy
7	Other form of communicative treatment
8	Pharmacotherapy
9	Physical therapy
10	Course therapy
11	Physiotherapy
12	Occupational therapy
13	Accompanying
14	Nursing and caring
15	Crisis care
16	General indirect time
17	Stay (per day of stay)
18	Daily spending (per hour)

Table B.1: List of activities

Appendix C

Transition matrix

In Table C.1 we see the transitions from activity i to activity j. One element of the matrix represents the probability of going from activity i to activity j, calculated by counting the number of pairs (ij) divided by the total number of pairs (ik), where k can be every activity.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.0414	0.1092	0.5599	0.0199	0.0093	0.0341	0.0271	0.0033	0.0106	0.0003	0.0033	0	0	0.005	0	0.0093	0.1651	0.002	0.0003
2	0.0747	0.019	0.4046	0.0196	0.0335	0.0616	0.0294	0.0345	0.024	0.0002	0.0063	0.0004	0.0002	0.0087	0.0005	0.0081	0.2688	0.0052	0.0008
3	0.0494	0.0101	0.1994	0.2597	0.0381	0.0535	0.0314	0.0252	0.011	0.0006	0.0082	0.0013	0	0.0123	0.0009	0.0057	0.2774	0.0151	0.0009
4	0.0807	0.0006	0.0362	0.0048	0.5176	0.0531	0.0216	0.0282	0.0342	0.0007	0.0124	0.0021	0.0034	0.0123	0.0036	0.0052	0.1632	0.019	0.0013
5	0.0662	0.0012	0.0297	0.003	0.0236	0.4819	0.0291	0.0286	0.0383	0.0005	0.0415	0.0004	0.0013	0.0144	0.0013	0.007	0.2191	0.0064	0.0064
6	0.052	0.0016	0.0203	0.0027	0.0177	0.0543	0.5406	0.0299	0.0208	0.0001	0.0283	0	0.0004	0.0261	0.0008	0.0012	0.1914	0.0111	0.0007
7	0.0695	0.0008	0.0321	0.0034	0.0255	0.0626	0.0351	0.5024	0.0186	0.0004	0.0329	0.0023	0.0023	0.0155	0.0027	0.0048	0.1684	0.0177	0.0031
8	0.0864	0.0018	0.0553	0.003	0.0684	0.1746	0.0552	0.0413	0.2299	0.0001	0.0115	0.0001	0.0005	0.0458	0.0015	0.017	0.1976	0.0056	0.0043
9	0.0309	0	0.0241	0.0034	0.0412	0.0687	0.0069	0.0378	0.0103	0.6082	0.0069	0	0.0034	0.0069	0.0034	0	0.1375	0.0103	0
10	0.0162	0.0003	0.0186	0.0023	0.0212	0.1608	0.0602	0.0581	0.0098	0.0002	0.3192	0.0033	0.0036	0.0168	0.0021	0.0007	0.2789	0.0179	0.0096
11	0.0116	0	0.0256	0.0093	0.1023	0.0465	0.0023	0.0907	0.0047	0	0.0884	0.393	0.0209	0.007	0	0	0.1884	0.0093	0
12	0.0189	0	0.0208	0	0.0871	0.0966	0.0208	0.0871	0.0114	0	0.0947	0.0057	0.4413	0.0246	0.0038	0.0019	0.0398	0.0398	0.0057
13	0.0279	0.0006	0.0152	0.0023	0.0176	0.0399	0.0386	0.0208	0.0276	0.0003	0.0134	0.0003	0.0005	0.6583	0.0139	0.0051	0.0837	0.022	0.012
14	0.025	0	0.0106	0.0045	0.0477	0.0416	0.0151	0.0394	0.0106	0.0008	0.0174	0	0.0015	0.1544	0.4799	0.003	0.1287	0.0197	0
15	0.178	0.0019	0.0449	0.0024	0.0116	0.0374	0.003	0.0131	0.0166	0	0.0013	0	0.0001	0.0081	0.0003	0.4916	0.1645	0.0241	0.0011
16	0.0668	0.0033	0.091	0.0076	0.0318	0.1056	0.0494	0.0393	0.0211	0.0004	0.0348	0.0009	0.0002	0.0137	0.0018	0.0136	0.5089	0.0074	0.0023
17	0.0331	0.0004	0.0253	0.0057	0.0456	0.0431	0.0367	0.0457	0.0084	0.0003	0.0267	0.0018	0.0026	0.0397	0.0035	0.0206	0.0901	0.5641	0.0065
18	0.0159	0	0.008	0.001	0.0104	0.0911	0.0048	0.0211	0.0114	0	0.0436	0	0.0007	0.0554	0	0.0035	0.0776	0.0145	0.6409

Table C.1: Transition matrix of going from activity i to activity j

Appendix D

Proofs from the model

D.1 Singularity points

Proof that singularity points $\notin [0,1]$:

$$f(w_t) \in [0,1] \quad \text{and} \quad f(w_t|v) \in [0,1]$$

Singularity points from (3.6):

$$\frac{1}{\gamma+\alpha} = 0 \quad \text{gives singularity points} \Rightarrow \gamma + \alpha = 0 \Rightarrow \gamma = -\alpha$$

Where

$$\alpha = \frac{f(w_t|v)}{f(w_t) - f(w_t|v)}\,.$$

If

$$\gamma \notin [0,1], \quad \text{then} - \alpha \notin [0,1] \Rightarrow \alpha \notin [-1,0] \Rightarrow \frac{f(w_t|v)}{f(w_t) - f(w_t|v)} \notin [-1,0]\,.$$

Assume:

$$-1 < \frac{f(w_t|v)}{f(w_t) - f(w_t|v)} < 0\,.$$

First we take a look at:

$$\frac{f(w_t|v)}{f(w_t) - f(w_t|v)} < 0\,.$$

Then

$$f(w_t) - f(w_t|v) > 0 \quad \text{and} \quad f(w_t|v) < 0$$

↯

or

$$f(w_t|v) > 0 \quad \text{and} \quad f(w_t) - f(w_t|v) < 0 \Rightarrow f(w_t) < f(w_t|v)\,.$$

If

$$f(w_t) < f(w_t|v) \quad \text{and both} \in [0,1]$$
$$\Rightarrow \frac{f(w_t|v)}{f(w_t) - f(w_t|v)} < -1\,.$$

↯

Secondly:

$$\frac{f(w_t|v)}{f(w_t) - f(w_t|v)} > -1\,.$$

Then

$$f(w_t|v) > -f(w_t) + f(w_t|v) \Rightarrow f(w_t) > 0$$

If

$$f(w_t) > f(w_t|v) \quad \text{and both} \in [0,1]$$
$$\Rightarrow \frac{f(w_t|v)}{f(w_t) - f(w_t|v)} < -1\,.$$

↯

If

$$f(w_t|v) > f(w_t) \quad \text{and both} \in [0,1]$$
$$\Rightarrow \frac{f(w_t|v)}{f(w_t) - f(w_t|v)} > 0\,.$$

↯

Thus:

$$\frac{f(w_t|v)}{f(w_t) - f(w_t|v)} \notin [0,1]\,.$$

Hence:

$$\gamma \notin [0,1]\,.$$

□

D.2 Maximum

We have to proof that the values found by applying the maximum likelihood estimation is a maximum and not a minimum.

Equation (3.5) is defined for $\gamma \in [0,1]$. We will repeat this equation and call it $g(\gamma)$:

$$g(\gamma) = \sum_{w_3 \in \mathcal{R}} \sum_{w_4} N(w_3, w_4) \log[\gamma f(w_4) + (1-\gamma) f(w_4|w_3)]\,. \tag{D.1}$$

The first derivative of $g(\gamma)$:

$$g^{(1)}(\gamma) = \sum_{v \in \mathcal{R}} \sum_{w_4} N(v, w_4) \left[\gamma + \frac{f(w_4|v)}{f(w_4) - f(w_4|v)}\right]^{-1} \tag{D.2}$$

and the second derivative will be:

$$g^{(2)}(\gamma) = -\sum_{v \in \mathcal{R}} \sum_{w_4} N(v, w_4) \left[\gamma + \frac{f(w_4|v)}{f(w_4) - f(w_4|v)}\right]^{-2}\,. \tag{D.3}$$

The second derivative is always smaller or equal than zero. That means the function $g(\gamma) \in [0,1]$ is a concave function. Therefore, the extremum on $[0,1]$ will always be a maximum and not a minimum.

□

Bibliography

[1] DTC MHC, http://www.dbcggz.nl, retrieved April 2007

[2] Panella M., Marchisio S. and Di Stanislao F., 2003, "Reducing clinical variations with clinical pathways: do pathways work?", *International Journal for Quality in Health Care*, 15(6):509-521

[3] Warner B.W. et al., 1998, "An evidenced-based clinical pathway for acute appendicitis decreases hospital duration and cost", *Journal of Pediatric Surgery*, 33:1371-1375

[4] Takegami K. et al., 2003, "Impact of a clinical pathway and standardization of treatment for acute appendicitis", *Surgery Today*, 33:336-341

[5] Jones A., 1999, "A modernized mental health service: the role of care pathways", *Journal of Nursing Management*, 7:331-338

[6] Bakker H.E. and Hofdijk J., 2005, "DBC's kunnen meer dan je denkt", *Zorgadministratie en Informatie*, 122 32:9-13

[7] Lenssen P., 2005, "Naar verbetering van het totale behandelproces", *Prismant magazine*, 20:15-16

[8] Nguyen J.M. et al, 2003, "A universal method for determining intensice care unit bed requirements", *Intensive Care Medicine*, 29(5):849-854

[9] Denton B., Viapiano J. and Vogl A., 2007, "Optimization of surgery sequencing and scheduling decisions under uncertainty", *Health Care Management Science*, 10(1):13-24

[10] Preater J., 2002, "A bibliography of queues in health and medicine", *Health Care Management Science*, 5(4):283

[11] Faddy M.J. and McClean S.I., 1999, "Analysing data on lengths of stay of hospital patients using phase-type distributions, *Applied stochastic models in business and industry*, 15:311-317

[12] Cochran J.K. and Bharti A., 2006, "A multi-stage stochastic methodology for whole hospital bed planning under peak loading", *International Journal of Industrial and Systems Engineering*, 1(2):8-36

[13] Lucas P., 2004, "Bayesian analysis, patterns analysis and data mining in health care", *Current Opinion in Critical Care* 10:399-403

[14] Jelinek F., 1998, ”Statistical Methods for Speech Recognition”

[15] Asmussen S., 1987, ”Applied Probability and Queues”

[16] Neuts M.F., 1994, ”Matrix-geomatric solutions in stochastic models, an algorithmic approach”

[17] Zipf G.K., 1935, ”Psycho-biology of languages”