NTNU
Norwegian University of
Science and Technology

Using optimization and simulation to evaluate differentiated stroke treatment

Andreas Bergstrøm Aarnseth
Erlend Moen Hov

Industrial Economics and Technology Management
Submission date: June 2018
Supervisor: Henrik Andersson, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Preface

This thesis concludes our Master of Science in Industrial Economics and Technology Management at the Norwegian University of Science and Technology. The objective of this thesis is to implement a model that enables differentiated treatment paths for stroke patients based on their diagnosis. At the same time, the general demand for emergency resources must be covered. By supplementing existing models and methods, optimization will be used to determine optimal allocation of the available emergency units. A comprehensive simulation program is also developed for this thesis. This is used to evaluate the performance of the different optimization solutions. The work on this Master's Thesis in Managerial Economics and Operations Research (TIØ4905) was carried out in the Spring 2018.

We would like to thank the personnel at the Ambulance Service at St. Olav Hospital for providing us with necessary data and valuable inputs.

We also would like to thank our supervisor, Professor Henrik Andersson, at the Department of Industrial Economics and Technology Management, for constructive discussions and his invaluable guidance.

Trondheim, 2018-06-10

Andreas Bergstrøm Aarnseth   Erlend Moen Hov

Sammendrag

En av samfunnets viktigste oppgaver er å tilby et velfungerende helsevesen til sine innbyggere. En stor del av helsevesenet omhandler akuttmedisinske tjenester, et samlebegrep som innbefatter blant annet AMK-sentraler, ambulanser og luftambulanser. Operasjonsanalyse har vært flittig i bruk siden 1960-tallet for å løse planleggingsutfordringer innen helse. Eksempler på bidrag spenner seg fra optimering av sykehuslokasjon til planlegging av timeplanen for operasjonsstuer. Denne masteroppgaven fokuserer på å løse et problem av strategisk og taktisk karakter. Problemet baserer seg på etablerte modeller for ambulanseplanlegging.

Optimeringsmodellen som presenteres er basert på en overlevelsesfunksjon. Målet er å finne optimal utplassering av medisinske ressurser med hensyn på slagbehandling. Samtidig skal modellen også ta hensyn til andre typer av medisinske nødsituasjoner som oppstår. Modellen har flere typer av mobile ressurser tilgjengelig, både ambulanser og helikoptre. I tillegg introduseres en ny type ambulanse - en "slagambulanse". Denne muliggjør diagnose av slagpasienter utenfor sykehuset. Ved å diagnostisere pasienter utenfor sykehuset kan en del av slagpasientene bli fraktet direkte til slagsenter som kan utføre trombektomi, en ny type slagbehandling, uten å måtte gå via lokalsykehuset først. Dette muliggjør differensierte ruter ut i fra diagnosen til pasienten. Ved å inkludere helikoptre i modellen vil resultatene bli mer realistisk og anvendelig, da flere områder i Norge i dag er helt avhengig av luftambulanse. Et helikopter kan både bli brukt til å rykke ut til pasienten, i tillegg til å møte en ambulanse for å frakte pasienten hurtigere til sykehus.

I masteroppgaven er det også utviklet et simuleringsprogram. Simuleringen er designet og endret fortløpende i samsvar med tilbakemeldinger fra Ambulansehelsetjenesten på St. Olavs. De har også vært behjelpelige med å tilby historiske data som har blitt brukt for å kalibrere simuleringsmodellen. Simuleringen bidrar med stokastisitet, og de fleste tider og etterspørsler generert i simuleringen baserer seg på sannsynlighetsfunskjoner. Simuleringen brukes for å validere løsningene fra optimeringen.

Simulerte løsninger viser at slagambulanser har effekt hos berørte pasienter. Siden hjerneslag utgjør en liten del av den totale mengden av nødsituasjoner, vil slagambulansen bli stående en del ledig å vente på oppdrag. Dette kan gi begrenset kost-nytte. Derimot, skulle teknologien utvikle seg slik at avansert bildediagnostikk blir tilgjengelig på alle ambulanser, vil slagpasienter kunne forvente en markant forbedring i tid til behandling.

Summary

One of the most important functions of society is to provide a well-functioning health care system. This includes delivering high quality emergency medical services (EMS) for those in need of acute medical care outside hospitals. Operations research has been widely used since the 1960s to contribute with valuable inputs to EMS planning. These contributions include models for a wide range of problems spanning from solving strategic problems such as hospital localization problems, to tactical and operational problems such as daily surgery planning. This thesis focuses on a problem of strategic and tactical character, building on established models for allocating EMS resources.

The optimization model presented is a survival function based approach that aims to allocate EMS units based on a stroke specific survival function, while at the same time considering the overall demand for EMS. The optimization model also enables use of multiple classes of EMS units, including ordinary ambulances, specially built stroke ambulances and helicopters. Stroke ambulances allow for sooner diagnosis, and hence increased survival for stroke patients. By getting early diagnosis of stroke patients, a subset of these patients can be transferred directly to specialized hospitals capable of providing a new form of treatment called thrombectomy. This constitutes the opportunity for differentiated treatment paths based on the patient's condition. The usage of helicopters in the model enables more realistic EMS planning for rural areas that rely heavily on air transport. In the model, helicopters can be used as first responding units, but are also used to meet ambulances to achieve faster treatment for the patient.

In this thesis a simulation program has also been developed. The simulation program is designed and altered based on feedback from personnel at the Ambulance Services at St. Olavs, giving it face validity. They have also provided us with valuable real data, which has been used as input for the simulation software. The simulation aims to complement the optimization model by adding randomness, such as variable demand both in respect of time and location. Additionally, the simulation is used to measure the performance of the solutions provided by the optimization model. This validation strengthens the credibility of the optimization model.

The combined result from optimization and simulation shows that stroke ambulances yield large savings in time to treatment for patients covered by such a unit. Since stroke emergencies constitute only a small portion of total demand for EMS, stroke ambulances should expect to have a substantial amount of standby time. This can result in a limited cost-benefit ratio. However, if stroke diagnosis equipment can become as practical and easy-to-operate as EKG is for heart diagnosis, the effects on time to treatment for stroke patients are significant.

List of Figures

2.1 Hospital locations in Norway . . . 8
2.2 Distribution of assignment priorities . . . 10
2.3 Important prehospital times . . . 10
2.4 Timeline acute stroke. . . . 11

3.1 Transportation time framework . . . 23
3.2 Predicted probability of mortality . . . 25
3.3 Odds ratio for favourable outcome of thrombolysis . . . 25
3.4 Next-event versus fixed-increment time advance methods . . . 29
3.5 Inverse transform method . . . 33

5.1 Figurative representation of the problem - phase 1 . . . 39
5.2 Figurative representation of the problem - phase 2: Ordinary ambulance responds 39
5.3 Figurative representation of the problem - phase 2: CT ambulance responds . . . 40
5.4 Visualization of the cost function. . . . 53
5.5 Visualization of an convex cost function. . . . 54
5.6 Visualization of an S-shaped cost function. . . . 55

6.1 Emergency generator step 1 . . . 62
6.2 Emergency generator step 2 . . . 63
6.3 Operator step 1 flowchart . . . 64
6.4 Operator step 2 flowchart . . . 65
6.5 Divert functionality flowchart . . . 67
6.6 Operator - Red emergency flowchart . . . 68
6.7 Operator - Stroke flowchart . . . 69
6.8 Simulation software - Map mode . . . 71
6.9 Simulation software - Units view . . . 72
6.10 Simulation software - Emergency view . . . 73
6.11 Simulation software - Summary view . . . 73

7.1 Center point locations . . . 79
7.2 Comparing simulation emergencies generated with real data . . . 85

8.1 Improvement in time to thrombolysis with three CT ambulances . . . 94
8.2 Effects of allowing two new station . . . 97
8.3 Comparison between different cases with CT ambulances and new stations . . . 98
8.4 Parallel solving flowchart . . . 101

List of Tables

2.1 Stroke centers in Norway. 9

3.1 Established EMS planning models and their properties 21
3.2 Conditional Prababilities Model 23
3.3 Modified Rankin Scale 24
3.4 Emergency priority probabilities. 31
3.5 Lookup table for emergency priorities. 31

7.1 Current allocation of ambulances and helicopters 83
7.2 Simulation results versus data from Ambulance Service at St. Olavs Hospital . . . 84

8.1 Comparing simulation results for current allocation with and without schedule . 87
8.2 Effects of changing stroke weight 88
8.3 Comparing optimization and simulation results 89
8.4 Correlation MIP solution and simulation results 90
8.5 Helicopter improvements on response time and time to hospital 91
8.6 Optimization effects of different objective functions 91
8.7 Results of CT ambulances. 92
8.8 Results when allowing new stations 96
8.9 Effects of new imaging technology on all ambulances 99

B.1 Effects of 3 CT ambulances versus regular base case with no CT ambulances. . . . 110

Contents

Preface i
Summary ii
List of figures v
List of tables vi

1 Introduction **1**

2 Background **4**
2.1 Stroke 4
2.1.1 The treatment chain: “Time is brain” 5
2.2 Health care in Norway 8
2.2.1 Hospitals 9
2.2.2 Emergency medical services units: ambulances and helicopters 9

3 Literature Review **13**
3.1 Emergency medical services planning models 13
3.1.1 Covering models 14
3.1.2 Survival function models 18
3.2 A treatment chain perspective 21
3.2.1 Stroke systems: a recommended field of research 21
3.2.2 The "Mothership" and "Drip and ship" models 22
3.2.3 Survival function and effects of stroke treatment 24
3.2.4 Summary 25
3.3 Simulation 26
3.3.1 Advantages of simulation 26
3.3.2 Simulation over time: continuous, discrete and mixed approaches 27
3.3.3 Discrete Event Simulation (DES) 28
3.3.4 Generating stochastic events 30
3.3.5 Use of simulation with optimization in EMS planning 33

4 Problem Description **35**

5 Mathematical model **38**
5.1 Figurative representation of the stroke treatment chain 38
5.2 Assumptions 40
5.3 Model formulation 42
5.3.1 Notation 42
5.3.2 Objective function 45
5.3.3 Restrictions 45
5.4 Model variations 50
5.4.1 Linearization of the objective function with a triangle method 50
5.4.2 Shapes of the cost function 53
5.4.3 Additional stations 56
5.4.4 Ensuring that no stations has more ambulances at night than day 57

6 Simulation **59**
6.1 Initialization Subprogram - object oriented approach 60
6.1.1 Zones 60
6.1.2 Hospitals, ambulance stations and helicopter bases 60
6.1.3 Emergency units 60
6.1.4 Operator 61
6.1.5 Simulation clock 61
6.2 Simulation clock and Timing subprogram 61
6.3 Library subprogram - Emergency generator 62
6.4 Event Subprogram - Operator 63
6.4.1 Handling an emergency - step 1 63
6.4.2 Handling an emergency - step 2 64
6.4.3 Emergency backlog 66
6.4.4 Diverting and aborting 67
6.4.5 Handling red emergencies - step 2A/B 67
6.5 Statistical Counter or Accumulator (debug mode and object attributes) 70
6.5.1 Map mode 70
6.5.2 Dashboard mode 71
6.6 Report Generator - aggregating results 73
6.7 Main program 74
6.7.1 Assumptions made in the simulation 75

7 Input data and simulation validation **77**
7.1 Describing the input data 77
7.1.1 Common input data 78

7.1.2 Optimization model only . . . 80
7.1.3 Simulation input data . . . 82
7.2 Current case . . . 83
7.3 Validating the simulation program . . . 84

8 Computational study **86**
8.1 Evaluating weighting between stroke and general cost function in the objective function . . . 87
8.2 Using simulation to evaluate the optimization model . . . 89
8.3 Different shapes of the cost function . . . 91
8.4 Effects of CT Ambulances . . . 92
8.4.1 Using current allocations to study effects of a CT ambulance . . . 92
8.4.2 Effects of more CT ambulances . . . 93
8.4.3 Boundary effects . . . 95
8.5 Establishment of new stations . . . 95
8.6 Imaging equipment on all ambulances: standard in the future? . . . 99
8.6.1 Solving several model instances in parallel . . . 100

9 Concluding remarks **103**
9.1 Recommendations for future research . . . 104

A Appendix - Input data **105**
A.1 Medical parameters . . . 105
A.2 Area of study . . . 106
A.2.1 Trøndelag . . . 106
A.2.2 Zones . . . 108
A.3 Simulation probability data . . . 108

B Appendix - Computational study **110**
B.1 Hardware and software configuration . . . 110
B.2 Comparing stroke treatment times: 3 CT ambulances and base case . . . 110

Bibliography **113**

Chapter 1

Introduction

This thesis is a continuation of the work done in the course TIØ4500: Managerial Economics and Operations Research, Specialization Project in the fall 2017.

Stroke is one of the leading causes of mortality and long term disability worldwide. In Norway alone, over 8600 people suffered from stroke in 2016, and the number is expected to rise. Many stroke survivors live with long term disabilities, and are in daily need of medical care and assistance. The costs of stroke are not only incurred in quality of life for those inflicted, but also results in a huge expenses for society as well. In recent years, endovascular thrombectomy, a new type of stroke treatment, has yielded some promising results for a particular group of patients suffering from one of the two main types of stroke. Results from clinical trials show that this treatment can help in saving functionality for those eligible. Unfortunately, the effects of both the primary treatment and the endovascular treatment is significantly reduced as time progresses from stroke onset. Moreover, only five hospitals in Norway can perform endovascular treatment. This means that early diagnosis and a highly efficient emergency medical services (EMS) are imperative. In 2016, only 2.2 % of the stroke patients in Norway received this treatment, even though some researchers claim that up to 15 % are eligible.

Since late 2014, The Norwegian Air Ambulance Foundation has conducted a pilot project with a "stroke ambulance". Until recently, the diagnosis technique could only be performed at hospitals with advanced medical imaging equipment, CT scanners. The foundation have now built a special ambulance with this capacity in order to provide early diagnosis for stroke patients. The project has showed promising results, and as technology continuously improves, this can represent the beginning of a paradigm shift in stroke treatment.

This thesis addresses the challenge of allocating resources along the entire patient chain, both geographically and capacity-wise. By 1) further developing existing models for EMS planning, 2) considering special capacities for a limited number of EMS units, and 3) considering the expected cost associated with the different outcomes, this thesis aims at giving correct treatment within the shortest possible time. By evaluating the patient chain and enabling differentiated treatment paths, the model presented in this thesis seek to minimize the expected costs related to medical care and stroke particularly.

More specifically, we present a survival function-based optimization model that implements a way of giving stroke patients different treatment paths based on their diagnosis and the available resources. We can do this, since some patients can now be prehospitally diagnosed if their responding unit has imaging equipment. At the same time, we will ensure that the general demand for EMS is satisfied. Some stroke patients will be taken to the closest hospital, while others will be taken directly to a stroke center for the aforementioned thrombectomy treatment. These two different paths constitute the "Drip and ship" and "Mothership" models, which are essential elements in our model. The model has also been developed to take ambulance helicopters into consideration. This allows faster response times for those living far away from ambulances, and enables transfer of patients between ambulances and helicopters. This extra element will make the model more realistic and consequentially yield more sensible results.

Additionally, to evaluate the model results, a simulation program has been developed. This is an essential part of this thesis, as the simulation adds stochasticity, availability concerns, and dynamics, making the results even more realistic. The simulation logic has been validated and found reasonable by the practitioners. The input data used is based on real world data from Trøndelag, which is the area used as the case application.

Outline of the thesis

This is by no means a medical article, but having an understanding of stroke, type of treatments, and the medical care system, will be useful when reading this thesis. For that reason, in Chapter 2 we present some background information on these topics. After each section, a brief summary with the key points are presented for the (non-medical) readers' convenience.

In Chapter 3 we provide an explanation of common models for EMS planning. Next, we look at some optimization contributions on stroke related problems. Then the "Drip and ship" and "Mothership" models for stroke treatment are presented and explained. Since simulation is an essential part of this thesis, a short introduction to the simulation theory used to build the software is presented. This provides the terminology, knowledge of building blocks and concepts needed to understand the simulation software.

With the foundation provided, we continue by presenting our problem description in Chapter 4. This is the basis for the model formulation presented in Chapter 5. After the base model is presented, some variations of the problem are described and discussed.

The simulation program is explained in Chapter 6. Each section of this chapter explains the different components of the simulation, based on the structure suggested in the Literature chapter. In the following chapter, Chapter 7, the input data used in both the optimization and simulation is presented. Gathering some parts of the data has been a time-consuming effort and in itself a demanding task. The process of data collection is therefore explained in more detail. We conclude this chapter by using the input data to validate and verify our simulation program by comparing our results to real world numbers.

The computational study is presented in Chapter 8. The results of the different implementations are presented and discussed. The purpose of this chapter is to highlight technical aspects of the model, as well as discussing the practical effects of the results. The thesis is concluded with some concluding remarks and suggestions for further research in Chapter 9.

Chapter 2

Background - Stroke and medical services

In this background, we provide general knowledge regarding stroke and describe current treatment guidelines. We also briefly describe the medical care services in Norway.

2.1 Stroke

Stroke is one of the leading causes of mortality and long term disability worldwide. According to the Norwegian Stroke Register 8650 people were in the year of 2016 admitted to hospitals in Norway as a result of an acute stroke. The average age the female stroke patients were 77 years and accounted for 46 % of the total stroke patients. 54 % of the patients were men, averaging 72 years old. (Norsk Hjerneslagregister, 2017) There are no completely accurate numbers regarding the prevalence, but it is assumed that there are over 55000 people alive in Norway today after surviving a stroke. Many of these patients have long term disabilities and need medical care and assistance in daily life activities. With Norway's aging population the number of post-stroke patients is expected to increase by 50 % by 2030 (Waaler, 1999). Undoubtedly this will lead to increased expenses on society.

There are two main types of stroke, haemorrhagic and ischaemic. Haemorrhagic stroke occurs when a weakened vessel rupture and bleeds into the brain. Ischaemic stroke occurs when a cerebral artery[1] is obstructed by a foreign mass. An obstruction leads to an abrupt loss of blood flow to the area of the brain which the artery supplies. This will cause damage to the brain tissue in that area, and even lead to permanent loss of brain cells. Ischaemic stroke can occur in

[1] Blood vessel within the brain.

several different ways. It can happen due to a blood clot forming in other parts of the circulation system and travel with the blood flow into the brain. This often happens due to atherosclerotic[2] plaques in the vessels, containing deposits of fat and calcification. When these plaques rupture they can cause a blood clot. An ischaemic stroke can also be caused by an obstruction formed locally in the brain, often due to plaques in cerebral arteries. Of the total number of strokes, 87 % were ischaemic, whilst 13 % were haemorrhagic. (Steiger and Cifu, 2016)

When blood flow in a cerebral artery is blocked, it causes acute loss of function of that particular part of the brain. The patient will experience an onset of different symptoms varying with location of the blockage and which parts of the brain the artery supplies. Typical symptoms can be acute onset of face dropping, weakness or paralysis in extremities, and difficulties speaking. Other symptoms include sudden loss of vision, confusion and loss of fine motor function. The brain cells, *neurons*, are highly sensitive to blood loss, more than other types of cells in the body. Some reports claim that when blood flow is obstructed, a typical patient will lose about 1.9 million brain cells every minute (Saver et al., 2016). The rapid loss of neurons can cause severe damage to the brain, leading to permanent disabilities or death. Rapid treatment which resumes blood flow can save daily life functions of the patient.

2.1.1 The treatment chain: “Time is brain”

Treatment of haemorraghic stroke is not discussed, as this is not the main focus of this thesis. Nevertheless, the prehospital treatment chain is the same, but the actual treatment differs. Treatment of an acute ischaemic stroke can be divided in two different axis. The first axis includes the specific treatment and intervention which aim to remove the occlusion of the cerebral artery to protect and save the affected cerebral cells. Treatments can be thrombolysis and endovascular thrombectomy. The second axis consists of more general treatment. This includes blood pressure control, cardiovascular control, ventilation and treatment of metabolic disturbances like hyperglycemia (high blood sugar levels) and hyperthermia (fever).

Time until treatment is initialized is the number one critical factor. An effective treatment chain is imperative. This includes both prehospital and inhospital services. When a patient with suspected stroke arrives at the hospital, the patient is taken straight to a CT-scanning[3] (or in some cases MRI[4]). This is done to distinguish an ischaemic stroke from a haemorrhagic stroke. If the scan identifies an ischaemic stroke, thrombolysis can be administered within 4.5 hours after

[2]Norwegian: *åreforkalkning*

[3]Magnetic resonance imaging: advanced X-ray machine that takes images from different angles to produce cross-sectional images.

[4]Another advanced medical imaging technique.

symptom onset.

Thrombolysis is a drug administered intravenously to the patient according to a specific protocol. It acts by inducing resolvent of the blood clot. (Lyden et al., 2015) Thrombolysis is proven to be most effective when administered within 3 hours after symptom start and are most efficient when given to patients with massive stroke symptoms. The effects of thrombolysis given to patients with mild symptoms are less and not well documented (Strbian et al., 2012). Important contraindications for thrombolysis include unknown symptom onset, previous haemorrhagic stroke or brain surgery, comatose patient or existing treatment with anticoagulants[5]. Thrombolysis can be administered to patients in all age groups.

In recent years, endovascular thrombectomy as treatment of ischaemic strokes has proven efficient for some stroke patients. This particular treatment can be administered for a large vessel ischaemic stroke. Recent randomised trials has shown substantially reduction in disability. The positive results has already led to changes in guidelines worldwide. Thrombectomy is proven efficient where the clot is situated in large vessels in the brain. Thrombectomy is an endovascular treatment where the circulation system is accessed by puncturing an artery and installing a catheter. With the help of intravenous contrast and x-ray, the catheter can be guided to the blood clot and the clot can be removed. This treatment demands highly competent systems which includes qualified personnel, equipment and labs. Today, it can be performed within six hours of stroke onset. However, recent studies show that thrombectomy yields promising results up to 24 hours. This might lead to changes in procedures. (Nilsen, 2018) In 2016 approximately 160 stroke patients in Norway were treated with thrombectomy, totally 2.2 % of all stroke patients. However, many more patients are probably eligible. Estimates by Norwegian Institute of Public Health state that 10 % of stroke patients can be eligible for this treatment (Stoinska-Schneider et al., 2016), while research by Tawil et al. (2016) suggest that up to 15 % can be eligible. When a patient with a large vessel ischaemic stroke arrives at the local hospital, thrombolysis is usually initiated at the local hospital before transporting the patient to the nearest hospital which offers thrombectomy.

The treatment of ischaemic stroke must be as efficient as possible because "time", literally "means brain". This suggests that delay in treatment leads to greater loss of brain tissue, increased disability for the patient, or in the worst case, death. Treatment can be described as a continuous chain from the onset of stroke symptoms and first medical contact, to prehospital services and inhospital services. The treatment chain must be efficient and well coordinated in order to decrease the negative outcome for the patient. At the same time, maintaining an efficient treatment chain with high-technological prehospital and in-hospital services, generates enormous

[5]Blood thinners

expenses.

In literature there are many terms used for these treatment types. The term "time to needle" (TTN) is often used as a metaphor of the time from onset of stroke symptoms until start of thrombolysis. We will call this time-to-thrombolysis. The "time to groin puncture" (TTGP) is time from onset of stroke symptoms until thrombectomy is initiated. We will refer to this as time-to-endovascular treatment.

Summary

Stroke can be divided into two main categories: those caused by an obstruction (ischaemic) and those caused by a bleeding (haemorraghic). Regardless of the type, time to diagnosis must be minimized in order to initiate the correct treatment - "time is brain". We have elaborated on the standard thrombolysis treatment for ischaemic stroke patients, and that the more advanced thrombectomy treatment can increase expected life quality for a subgroup of these patients. To minimize time to diagnosis and treatment, prehospital services and capacities are crucial. In the next section we will present an overview of the health care system in Norway, both inhospital and prehospital services.

2.2 Health care in Norway

Health care planning in Norway is a demanding task, due to the country's widespread geography and heterogeneously dispersed population. As of late 2017, there are 49 emergency hospitals in Norway. Reported figures indicate that there was 517 ambulances across the country, a decline from about 600 in the early 2000s. (SSB, 2016a)

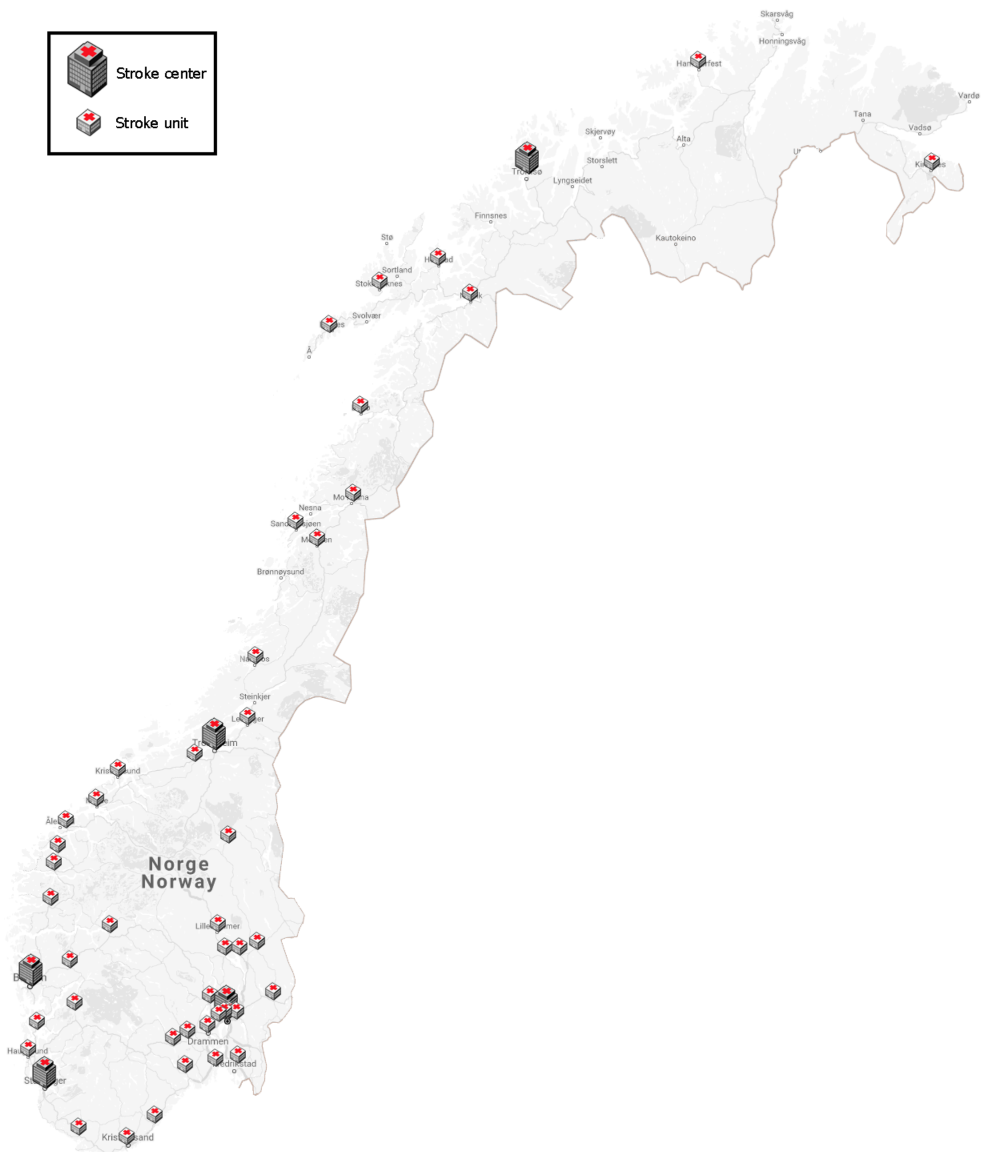


Figure 2.1: Hospital locations in Norway. Map data ©2018 Google.

2.2.1 Hospitals

All five university hospitals in Norway are stroke centers[6], see Table 2.1. All emergency hospitals have CT scanners enabling diagnosis of stroke and hence thrombolysis treatment. In addition to hospitals, there are also many emergency clinics[7]. In this thesis these clinics are not included as treatment or diagnosis facilities. These clinics does normally not have CT capacity, but because of the demographics in Norway, deployment of CT scanners at some of these clinics could potentially benefit patients that are far away from hospitals, or when helicopters are unavailable.

Table 2.1: Stroke centers in Norway.

Stroke centers
Universitetssykehuset Nord-Norge (Tromsø)
St. Olav Hospital (Trondheim)
Haukeland Universitetssykehus (Bergen)
Stavanger Universitetssykehus
Oslo Universitetssykehus

2.2.2 Emergency medical services units: ambulances and helicopters

In 2016, ambulances in Norway conducted almost 700 000 assignments. All assignments are given a priority based on its emergency: green, orange and red, and the distribution for each of them are approximately even, as shown in Figure 2.2.

[6]Hospitals specialized in stroke treatment with a specified set of services, including more advanced diagnosis, thrombectomy and other special interventions.

[7]Norwegian: *legevakt*

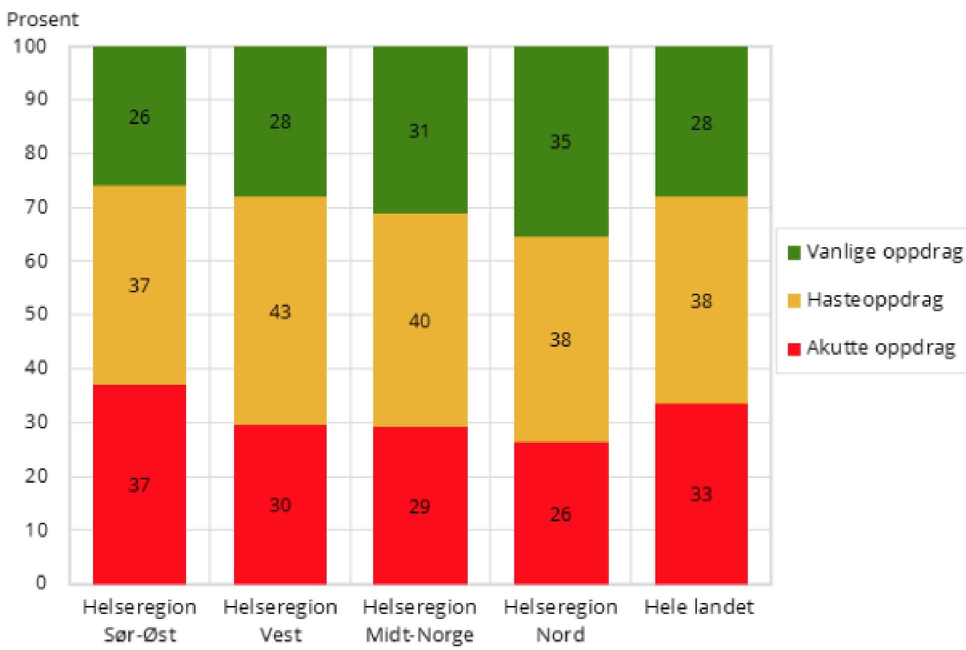


Figure 2.2: Distribution of assignment priorities for ambulances in 2016.(SSB, 2017a)

Stroke, or suspicion of stroke, is always treated as a red assignment, which implies transport to the closest hospital as quickly as possible.[8] The Norwegian Directorate of Health uses a quality indicator to measure prehospital response time (see Figure 2.3) for red emergencies. For the most emergent events, they measure how often an ambulance reaches the patient within the indicator limits from the emergency call has been made. This quality indicator is set to 12 minutes for urban areas and 25 minutes for rural areas. Both of these should be met in 90 percent of the occasions, according to the directions provided by the Norwegian Parliament. The quality indicator and these directions are however not required by law.

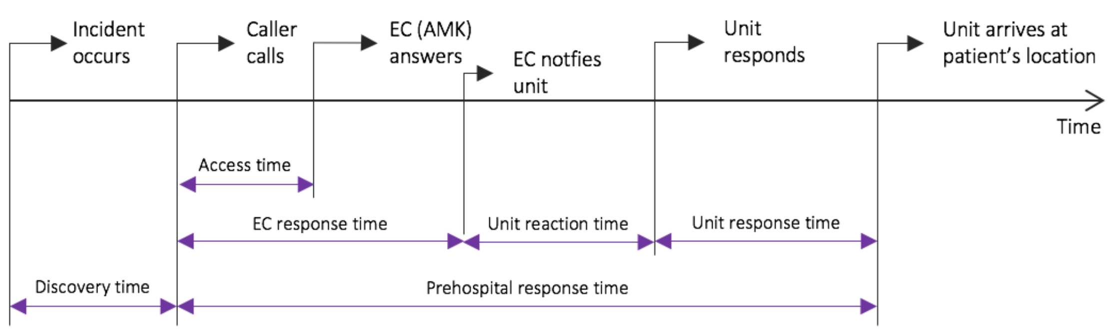
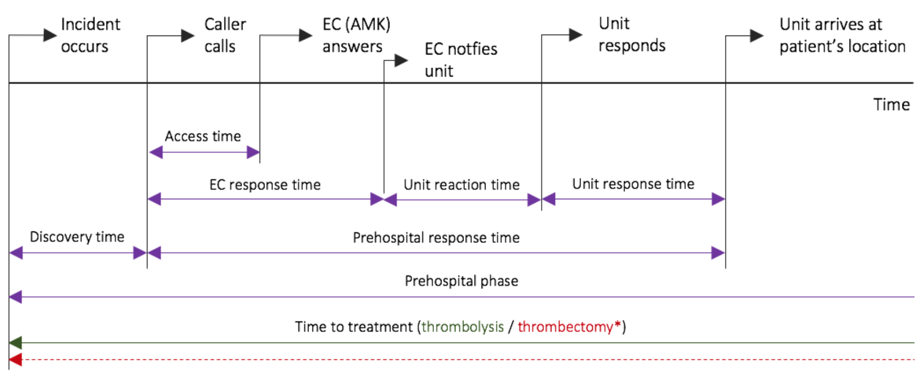


Figure 2.3: The most important prehospital times. AMK is the Norwegian medical Emergency Central. Figure based on Folkestad et al. (2004)

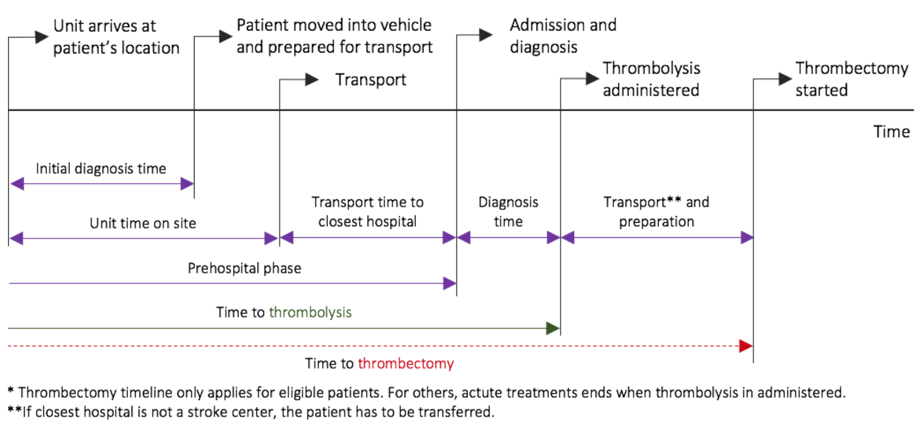
[8]When there is a minor time difference between time to closest hospital and a stroke center, the emergency medical services (EMS) can deviate from this rule. There are locally set guidelines for this.

In addition to ambulances, there are also other kinds of emergency vehicles. Along the coastline, there are several ambulance boats operating. There are also fixed wing air ambulances, that transport patients from remote areas or aids in transferring patients from one hospital to another. The second most used capacity (after ambulances) are helicopters. There are in total 13 helicopters in Norway, spread out on 12 bases. In addition, there are seven rescue helicopter bases. These can be requested when ambulance helicopters are unavailable, but other options should be evaluated first. Moreover, a rescue helicopter assigned to an ambulance assignment can be redirected, and hence cancelled, if it is needed for a search and rescue mission, which is it's main objective. (Uleberg et al., 2013) In 2016, helicopter ambulances in Norway conducted a total of 8249 assigments, of which 3144 were primary assignments (SSB, 2017b). These are assignments where patients need medical attention outside of a hospital. Helicopters can also transport patients between hospitals.

Based on Figure 2.3, we develop a timeline for a typical stroke case. Figure 2.4 is a simplified and general description of how stroke incidents are handled today.



(a) A rewritten timeline for stroke patients. First phase.



(b) Second phase. Transport, diagnosis and treatment.

Figure 2.4: Timeline for acute phase of ischaemic stroke

Note that the figure only illustrates the timeline for a patient with an ischaemic stroke. Haemorrhagic strokes are treated exactly the same until diagnosis is determined, but instead of thrombolysis and thrombectomy, other treatment is given.

"Stroke ambulance" - Ambulance with CT capacity

Since late 2014, The Norwegian Air Ambulance Foundation has conducted a pilot programme evaluating a specially built ambulance for prehospital stroke diagnosis. By having a CT machine in the ambulance, special trained crew can conduct a scan quickly after arriving to the patient. The scan can be interpreted by a doctor in the crew, or sent to a hospital for diagnosis. Meanwhile, the patient is transported to the hospital. When the diagnosis is determined, correct treatment is initiated in the ambulance if the patient has not already arrived at the hospital. The main function is to separate patients suffering a haemorraghic stroke from those having an ischaemic stroke, as discussed in Section 2.1. (SNL, 2017) Although the stroke ambulance is more than an ambulance with a CT, we refer to it as a *CT ambulance* in this thesis.

At the same time, another technology has been developed in order to help with the same diagnosis. The *Strokefinder* is a much simpler and smaller piece of equipment, similar to a helmet. (Bordvik, 2014) If proven successful, this can become standard equipment on any ambulance. This "helmet" uses microwaves to scan and automatically detect whether a stroke is haemorraghic or ishcaemic. The equipment is currently being tested in Norway, at Stavanger University Hospital. (Gjesdal, 2017)

Summary

If stroke is suspected, the emergency is given top priority by EMS. Today, as a rule of thumb, stroke patients are transported to the closest hospital for diagnosis. If the patient is diagnosed at the scene by a stroke ambulance, thrombolysis treatment can start once diagnosis is determined. This also enables the patient to be transported directly to a stroke center for thrombectomy treatment if he or she is an eligible candidate.

Although ambulances conduct the majority of assignments in Norway, proper EMS planning should include other vehicle types as well, especially helicopters since they are frequently used in this widespread country. Helicopters cover a much larger area, and extended use of helicopters can severely aid and relieve ambulances. However, helicopters are less reliable and more costly, and is considered only in the most emergent cases.

Chapter 3

Literature Review

First in this chapter, established literature on emergency medical services (EMS) planning is presented. Further, we review stroke related literature in Section 3.2. Some research is done on this matter from an operational research perspective, but optimization researchers encourage more study on the subject. In order to lay the groundwork for our model's cost assumptions, we will briefly present research on the relationship between the outcome for stroke patients and time to treatment. Lastly, in Section 3.3 some theory on simulation is reviewed.

3.1 Emergency medical services planning models

The problem of assigning locations to EMS vehicles has been popular among operations researchers (Erkut et al., 2008). In the literature, most of the existing models apply coverage or average response time as performance metrics for health outcomes. One of the first models by Toregas et al. (1971) on EMS planning focused on satisfying coverage. The goal was to minimize the number of facilities needed in order to satisfy the demand in given zones. The demand in a zone was considered satisfied when the distance or travel time from the medical facility to the zone fell below a certain threshold. Other coverage models attempts to maximize demand covered given a certain number of facilities. Erkut et al. (2008) lists several reasons why researchers have used coverage models so frequently: deterministic coverage models are fairly easy to solve using optimization software; the concept of a covering model can be easily communicated to decision makers/politicians; finally, many EMS systems already use percentage of calls covered as a performance metric. However, Erkut et al. (2008) argue that coverage models should not be used for EMS vehicle location problems, mainly because of its limited ability to take different response times into consideration. Therefore, after reviewing the covering models, an alternative approach, survival function models are reviewed.

3.1.1 Covering models

MCLP

Erkut et al. (2008) formulates the maximal covering location problem (MCLP) as a basic model one could use for locate EMS vehicles:

$$\max z = \sum_{i=1}^{m} d_i y_i \tag{3.1}$$

$$\text{s.t} \sum_{j=1}^{n} a_{ij} x_j \geq y_i \qquad i = 1,..,m \tag{3.2}$$

$$\sum_{j=1}^{n} x_j \leq q \tag{3.3}$$

$$x_j \in \{0,1\} \qquad j = 1,...,n \tag{3.4}$$

$$y_i \in \{0,1\} \qquad i = 1,...,m \tag{3.5}$$

where

d_i	: Demand in zone i.
y_i	: 1 if demand zone i is covered, 0 otherwise.
a_{ij}	: 1 if demand zone i is coverable by a station in zone j, 0 otherwise.
x_j	: 1 if a station is located in zone j, 0 otherwise.
q	: Number of stations to locate.

The objective (3.1) of the MCLP is to maximize the total demand covered. Restriction (3.2) ensure that the demand in zone i, d_i, is considered covered if a possible station in zone j covers it. Constraint (3.3) limits the number of stations to q. This model is defined so that only one ambulance is located at each station.

MCLP+PR

The Maximal Covering Location Problem with Probabilistic Response Time (MCLP+PR) formulated by Daskin and Haghani (1987), maximizes total demand covered, but incorporates the probability that an ambulance at station j can reach zone i within the limit.

$$\max z = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} P_{ij} y_{ij} \tag{3.6}$$

$$\text{s.t} \sum_{i=1}^{m} y_{ij} \leq m x_j \qquad j = 1, .., n \tag{3.7}$$

$$\sum_{j=1}^{n} y_{ij} = 1 \qquad i = 1, ..., m \tag{3.8}$$

$$\sum_{j=1}^{n} x_j \leq q \tag{3.9}$$

$$x_j \in \{0, 1\} \qquad j = 1, ..., n \tag{3.10}$$

$$y_{ij} \in \{0, 1\} \qquad i = 1, ..., m, j = 1, ..., n \tag{3.11}$$

where

d_i	: Demand in zone i
P_{ij}	: The probability that a vehicle at station j can reach zone i within the limit.
y_{ij}	: 1 if the demand zone i is closest to candidate zone j, 0 otherwise.
x_j	: 1 if a station is located in node j, 0 otherwise.
m	: Number of demand zones.
q	: Number of stations to locate.

The objective (3.6) of the MCLP+PR is to maximize the total demand covered while considering the probability of coverage. Restriction (3.7) state that a station must be located in zone j to be usable for covering demand in zone i. Restriction (3.8) ensures that zone i is served by a station in zone j. Finally, constraint (3.9) limits the number of stations.

MEXCLP

Erkut et al. (2008) mentions two formulations of the Maximal Expected Covering Location (MEXCLP) Problem, where both models extend the standard covering problem by taking the probability of an EMS vehicle being busy into consideration. It is also possible to locate multiple ambulances in one station. The following model is formulated by Daskin (1983):

$$\max z = \sum_{i=1}^{m} d_i \sum_{k=1}^{r} (1-p)p^{k-1}\hat{y}_{ik} \tag{3.12}$$

$$\text{s.t} \sum_{k=1}^{r} \hat{y}_{ik} \leq \sum_{j=1}^{n} a_{ij} z_j \qquad i = 1,..,m \tag{3.13}$$

$$\sum_{j=1}^{n} z_j \leq r \tag{3.14}$$

$$\hat{y}_{ik} \in \{0,1\} \qquad i = 1,...,m, k = 1,...,r \tag{3.15}$$

$$z_j \in \{0,1,...,c_j\} \qquad j = 1,...,n \tag{3.16}$$

where

d_i	: Demand in zone i
p	: Average fraction of time that the EMS vehicle is busy.
$\hat{y}_{ik}$	: 1 if zone i is covered by at least k EMS vehicles, 0 otherwise.
a_{ij}	: 1 if demand zone i is coverable by a station in zone j, 0 otherwise.
z_j	: Number of vehicles placed in the station in zone j
r	: Maximum number of vehicles to be allocated
c_j	: Maximum number of vehicles that can be stationed in candidate zone j.

$(1-p)p^{k-1}$ is thus the expected increase in coverage by increasing the number of ambulances that covers a demand zone by one. The objective in (3.12) is therefore to maximize expected demand coverage. Constraint (3.13) ensures that the number of vehicles covering zone i cannot exceed the number of vehicles in range(sum of a_{ij}) to zone i. Restriction (3.14) limits the total number of vehicles to be located.

MEXCLP+PR

The Maximal Expected Covering Location Problem with Probabilistic Response Time (MEXCLP + PR), formulated by Budge et al. (2008) incorporates uncertainty in both vehicle availability and vehicle response time. Erkut et al. (2008) summarizes the problem:

$$\max z = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} P_{i,k(i,j)} \prod_{u=1}^{j-1} \hat{p}_{k(i,u)}^{z_{k(i,u)}} (1 - \hat{p}_{k(i,j)}^{z_{k(i,j)}}) \tag{3.17}$$

$$\sum_{j=1}^{n} z_j \leq r \tag{3.18}$$

$$z_j \in \{0, 1, ..., c_j\} \qquad j = 1, ..., n \tag{3.19}$$

In this model, $k(i, j)$ is the j-th preferred EMS station to demand zone i. The objective function (3.17) maximizes total demand covered by weighting the combination of coverage for the demand zones. The probability of covering zone i from a station in zone j is calculated as the product of the probabilities that there is at least one EMS vehicle available at the j-th preferred station, and that EMS vehicles at the $j-1$ stations are busy. That product is multiplied by the probability that the response time from the j-th ($k(i, j)$) preferred station to demand zone i falls within limits. The result is multiplied with the demand in order to calculate the demand covered. If there is dependence between the stations regarding vehicle availability, this should be compensated for. However, in this model, no dependence is assumed. Restriction (3.18) ensures that the total number of vehicles located in the station zones, z_j, does not exceed the maximum number of EMS vehicles, r. c_j is the maximum number of vehicles that can be located in zone j.

P-median model

The p-median problem consists of minimizing the total weighted average distance between the demand nodes and the stations. The p-median is one of the classical location problems and was developed by Hakimi (1965). Daskin and Maass (2015) summarizes the p-median problem:

$$\min z = \sum_{i \in I} \sum_{j \in J} d_j c_{ij} x_{ij} \tag{3.20}$$

$$\tag{3.21}$$

$$\sum_{i \in I} x_{ij} = 1 \qquad j \in J \tag{3.22}$$

$$\sum_{i \in I} y_i = p \tag{3.23}$$

$$x_{ij} - y_i \leq 0 \qquad i \in I, j \in J \tag{3.24}$$

$$y_i \in \{0,1\} \qquad i \in I \tag{3.25}$$

$$x_{ij} \in \{0,1\} \qquad i \in I, j \in J \tag{3.26}$$

where

d_j	: Demand in node j
c_{ij}	: Unit cost of satisfying demand in node j from station i
x_{ij}	: The fraction of demand in node j that is covered by station i.
y_i	: 1 if a station is located in node i, 0 otherwise
p	: Number of EMS stations to allocate

The objective function (3.20) minimizes the total cost, weighted by the demand. Restriction (3.22) ensures that all demand in zone j must be met, while Restriction (3.23) limits the number of stations allocated. Lastly, Constraint (3.24) ensures that a station must be located in zone i before serving demand to zone j. In this model, demand in a zone can be covered by several stations/EMS vehicles.

3.1.2 Survival function models

Erkut et al. (2008) develops models where the concept of coverage is replaced with maximizing patient survival based on a survival function. The authors study several survival functions, focusing on out-of-hospital cardiac arrest. The models differ from each other in several ways. Factors that affect the survival function were among others:

- Was the collapse of the patient witnessed by others?
- Time duration between cardiac arrest and the call to EMS
- Emergency services response time

MSLP

Erkut et al. (2008) formulates the maximal survival location problem (MSLP). They assume that demand in a zone is served by an EMS vehicle at the closest station. In order to keep track of which EMS vehicle serves which demand zone, they define $y_{i,j}$, which equals 1 if the vehicle in zone j serves demand zone i. x_j equals to one if a station is located in node j.

The survival function s takes the sum of t_{ji} (travel time from station zone j to demand zone i) and t_d (pretravel delay) as arguments. d_i is the demand in node i. The probability p_{ij} that the patients in demand node i survives is given by the following survival function:

$$p_{ij} = \begin{cases} \mathrm{s}(\mathrm{t}_{ji} + t_d) & \text{if } \mathrm{y}_{ij} = 1 \\ 0 & \text{if } \mathrm{y}_{ij} = 0 \end{cases} \tag{3.27}$$

The MSLP is formulated:

$$\max \mathrm{z} = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} s(t_{ji} + t_d) y_{ij} \tag{3.28}$$

$$\text{s.t} \sum_{i=1}^{m} y_{ij} \leq m x_j \qquad j = 1, .., n \tag{3.29}$$

$$\sum_{j=1}^{n} y_{ij} = 1 \qquad i = 1, ..., m \tag{3.30}$$

$$\sum_{j=1}^{n} x_j \leq q \tag{3.31}$$

$$x_j \in \{0, 1\} \qquad j = 1, ..., n \tag{3.32}$$

$$y_{ij} \in \{0, 1\} \qquad i = 1, ..., m, j = 1, ..., n \tag{3.33}$$

where

d_i	: Demand in node i
y_{ij}	: 1 if demand in zone i is served by a station in zone j, 0 otherwise
m	: Number of demand zones to serve
x_j	: 1 if zone j has a station, 0 otherwise
q	: Maximum number of EMS stations to allocate

The objective function (3.28) maximizes the expected number of surviving patients. Restriction (3.29) ensures that in order for an EMS vehicle to respond from j to i, there must be a station in zone j. Constraint (3.30) enforces that every demand zone is served by exactly one station.

Restriction (3.31) limits the number of stations.

MSLP+PR

The Maximal Survival Location Problem with Probabilistic Response Time (MSLP+PR) is formulated as a MSLP, except a modification of the objective function:

$$\max z = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} E[s(R_{ij})] y_{ij} \tag{3.34}$$

where p_{ij} from Restriction (3.27) is expressed as $E[s(R_{ij})]$ and precomputed for every zone pair (i, j). R_{ij} is the response time between zone j and i.

MEXSLP

Erkut et al. (2008) formulate the Maximal Expected Survival Location Problem (MEXSLP) based on MEXCLP+PR's objective function (3.17):

$$\max z = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} p_{i,k(i,j)} = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} s(t_{ji} + t_d) \prod_{u=1}^{j-1} p^{z_{k(i,u)}} (1 - p^{z_{k(i,j)}}) \tag{3.35}$$

$$\text{s.t} \sum_{j=1}^{n} z_j \leq r \tag{3.36}$$

$$z_j \in \{0, 1, ..., j_j^c\} \qquad j = 1, ..., n \tag{3.37}$$

where $k(i, j)$ is the j-th preferred EMS station to demand zone i, and

p	: Average fraction of time that an EMS vehicle is busy.
z_j	: Number of EMS vehicles stationed in zone j
r	: Maximum number of allocated vehicles.

MEXSLP+PR

Modification of the objective function (3.17) in the MEXCLP+PR problem is sufficient to include uncertainty in response time and availability, and the survival function:

$$\max z = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} p_{i,k(i,j)} = \sum_{i=1}^{m} d_i \sum_{j=1}^{n} E[s(R_{i,k(i,j)})] \cdot \prod_{u=1}^{j-1} \hat{p}_{k(i,u)}^{z_{k(i,u)}} (1-\hat{p}_{k(i,j)}^{z_{k(i,j)}}) \tag{3.38}$$

Summary

We have so far discussed some of the existing models that has proven useful for EMS planning. Covering models and survival function models attempts to cover different important perspectives on the EMS planning. A summary of the key features of each model is listed in Table 3.1.

Table 3.1: Comparison of existing models with respect to properties.

Properties	MCLP	MCLP+PR	MEXCLP	MEXCLP+PR	MSLP	MSLP+PR	MEXSLP	MEXSLP +PR	P-median	Our model
Time discrimination	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Consider uncertain response time	No	Yes	No	Yes	No	Yes	No	Yes	No	No
Consider busy ambulances	No	No	Yes	Yes	No	No	Yes	Yes	No	No
Rerouting	No	No	No	No	No	No	No	No	No	Yes
Multiple vehicles cover zone	No	No	No	No	No	No	No	No	Yes	No

3.2 A treatment chain perspective

The first part of Section 3.1 concerns allocation of EMS vehicles based on coverage from a providers perspective. Survival function location problems aim to solve the location problem based on patient outcome. However, Aringhieri et al. (2017) claim that a shift from the provider perspective to the patient perspective is necessary. By utilization and further development of new technologies and existing models on different aspects of the emergency care pathway, a more efficient medical care can be provided.

3.2.1 Stroke systems: a recommended field of research

There are many fields within the stroke care system that can be studied from an OR perspective. Churilov and Donnan (2012) list some research conducted on the topic. Nonetheless, they claim that the OR contribution has been limited so far. They encourage more research on the matter and suggest an agenda for future work. In their work, several different problem areas within the stroke system are highlighted, and the authors suggest some OR interventions that can contribute within the different problem areas. These areas include amongst many others: prehospital stroke care, inhospital stroke care, and appropriate stroke care expertise allocation. The latter regards deciding location of stroke centers. Some of the suggested interventions related to these areas are stroke services planning and utilization models, stroke ambulance services models, and stroke units and thrombolysis facility location and layout models.

We are now going to study two common models used for stroke care. These models consider both prehospital and hospital capacities, and are therefore a way of studying a part of the stroke treatment chain.

3.2.2 The "Mothership" and "Drip and ship" models

As mentioned in the Background chapter, recent research favors endovascular treatment for large vessel ischaemic strokes in addition to thrombolysis treatment. However, first responders can not know that they are dealing with a large vessel stroke, or determine if it is stroke at all. For most countries this treatment can unfortunately not be provided at all hospitals due to costs, need for specialized personnel and insufficient patient volume. Schellinger et al. (2016) uses the European Society of Cardiology's guidelines that state that for PCI (specialized heart attack treatment), there is "a strong inverse volume-outcome relationship in emergency PCI", and claims that this also holds for endovascular treatment. Challenges with low patient volume applies especially for countries such as Norway, where large areas have low population density.

The first responders ability to make the correct diagnosis will, without specialized training and equipment, remain a challenge. Still, given a thrombectomy eligible patient, he or she may be closer to a stroke unit (SU) than a stroke center (SC). Then, there needs to be a decision on whether the patient should (a) be transported to the SU first for thrombolysis, and then relocated to a SC for further treatment, or (b) if the patient should be transported directly to a SC. These two different options are what is referred to as the "Drip and ship" and "Mothership" models respectively. The different paths a patient can take are depicted in Figure 3.1.

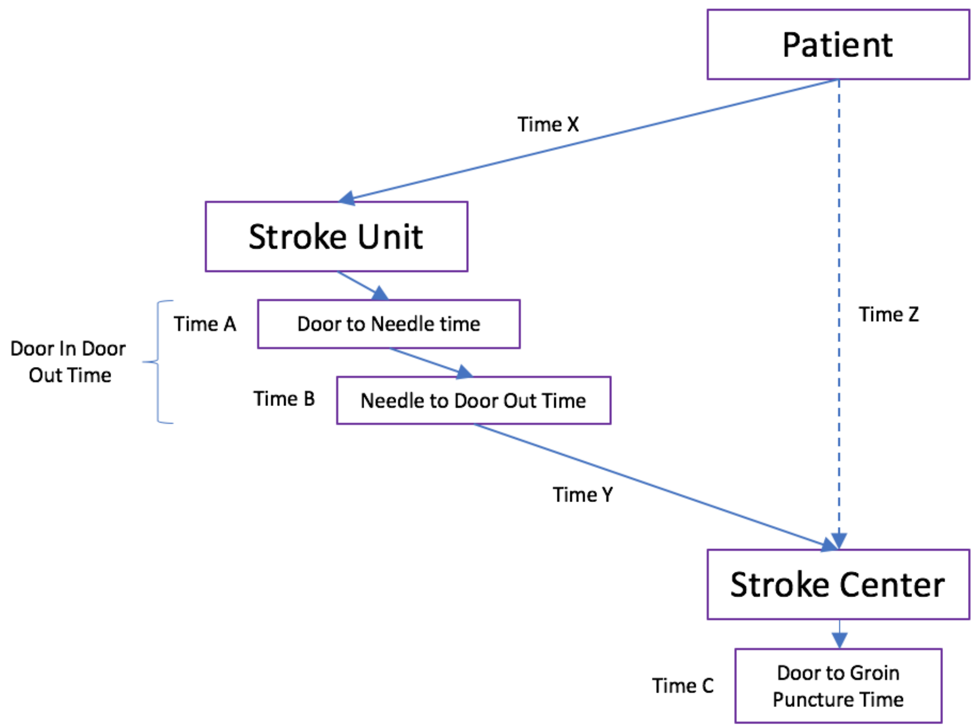


Figure 3.1: Transportation time framework. The dashed line represents the mothership model, the solid lines represent the drip and ship model. Time X is the transportation time from the patient to the SU. Time Y is the transportation time from the SU to the SC. Time Z is the transportation time from the patient to the SC. Time A is the time from the patient's arrival at the SU to the administration of alteplase (thrombolysis), and time B is the time from alteplase administration to leaving the SU. Time C is the time from the patient's arrival at the SC to the beginning of the endovascular procedure. (Holodinsky et al., 2017)

Holodinsky et al. (2017) have tried to model this problem assuming a large vessel ischemic stroke. The research aims to calculate the probability of a good outcome for model (a) and (b), or what is referred to as a "Drip and ship model" and a "Mothership model". The models are indicated in Table 3.2 below.

Table 3.2: Conditional Prababilities Model for endovascular treatment (EVT).

Model	Conditional Probabilities
Mothership	P(good outcome|mothership model) = P(reperfusion|EVT) · P(good outcome|reperfusion at ϕ mins) + P(no reperfusion|EVT) · P(good outcome|no reperfusion)
Drip and Ship	P(good outcome|drip and ship model) = P(early reperfusion|alteplase) · P(good outcome|reperfusion at ϕ mins) + P(no early reperfusion|alteplase) · [P(reperfusion|EVT) · P (good outcome|reperfusion at ϕ mins) + P(no reperfusion|EVT) · P(good outcome|no reperfusion)]

If one could determine the probabilities given in his model, this could be used to implement a stochastic model that enables rerouting for each individual patient.

3.2.3 Survival function and effects of stroke treatment

As shown in the previous section, to model the stroke treatment chain, the treatment and outcome relationship must be known. When analyzing stroke outcomes, medical researchers often use a grading system called the modified Rankin Scale (mRS). The grades and descriptions are given in Table 3.3.

Table 3.3: Grades and descriptions of the modified Rankin Scale (mRS). (Banks and Marotta, 2007)

Grade	Description
0	No symptoms at all
1	No significant disability: despite symptoms, able to carry out all usual duties and activities.
2	Slight disability: unable to perform all previous activities but able to look after own affairs without assistance.
3	Moderate disability: requiring some help but able to walk without assistance.
4	Moderate severe disability: unable to walk without assistance and unable to attend to own bodily needs without assistance.
5	Severe disability: bedridden, incontinent and requiring constant nursing care and attention.
6	Death.

When this scale is used in medical research, it is often based on measured functionality of patients 90 days after stroke onset. Mazighi et al. (2013) present a meta-analysis and attempt to describe the relationship between stroke onset-to-reperfusion time to mortality and favorable outcome (mRS 0-2) probabilities. These measures were taken for patients having received successful endovascular treatment. Moreover, Gumbinger et al. (2014) show that better outcome after thrombolysis treatment is time dependent, and non-linear. A graphical presentation of some of their findings are presented in the following figures.

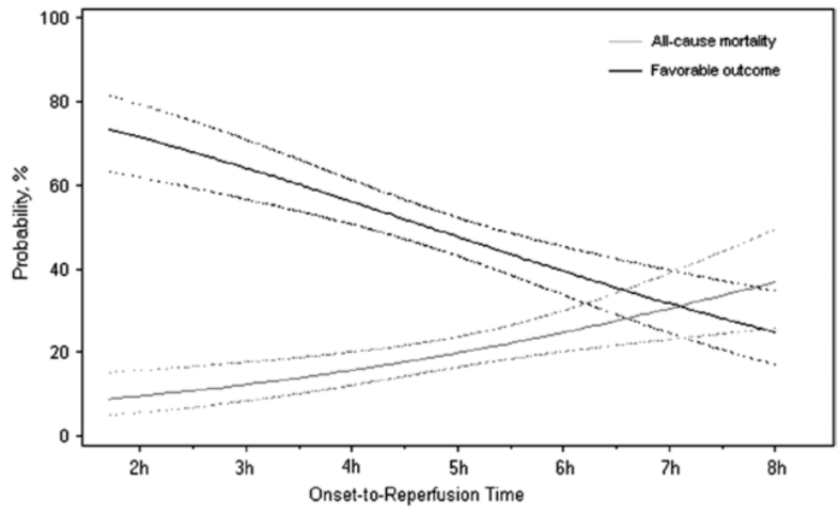


Figure 3.2: Unadjusted predicted probability of mortality and favorable outcome at 90 days by onset-to-reperfusion time. 95 % confidence intervals are showed in dashed lines. (Mazighi et al., 2013)

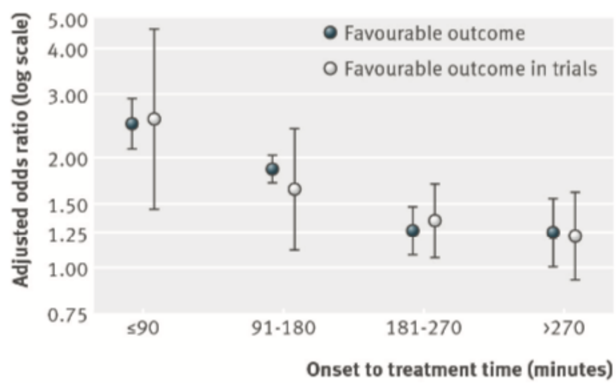


Figure 3.3: Odds ratio for favourable outcome (mRS 0-1) with time to treatment with recombinant tissue plasminogen activator (thrombolysis) after onset of stroke in binary logistic regression analysis and comparisons with pooled analysis of randomised clinical trials of alteplase for acute stroke. (Gumbinger et al., 2014)

3.2.4 Summary

While some location problems usually measure outcome in response time, we have seen that outcome in stroke related matters not necessarily depends solely on response time. Ambulances with equipment for diagnosis can enable use of a Mothership model, as described by Holodinsky et al. (2017). Planning with this option may be one step towards a shift to the patient perspective (instead of the provider perspective), as Aringhieri et al. (2017) promote. Likewise, by including more strategic decisions like facility/capacity location problems to the planning process, as Churilov and Donnan (2012) suggests, one might take a step even further. This is however outside the scope of this thesis.

We have also briefly presented some findings from medical articles on the relationship between time-to-treatment and outcome. The details are less important for this thesis. But they are included to support our assumption that this relationship is non-linear, and that a minute counts more within the first hour than after a day after stroke onset.

3.3 Simulation

Evaluating the effects of a solution produced by an optimization model can be challenging. A naive approach would be to simply implement the solution and hope for the best. However, in EMS planning unforeseen consequences may be fatal. Therefore, as an attempt to reveal the effects of a solution, one approach is to actually test the solution by simulating the behaviour and performance of the system, or by creating "an imitation of a system" (Robinson, 2004).

A simulation can be static, which means that it imitates a system at a specific point in time or that it is independent of the passage of time. The term simulation is, however, mostly used in the context of dynamic simulation. A dynamic simulation imitates the system as time progresses. (Law and Kelton, 2000). When the word simulation is used in this thesis, we are referring to a dynamic simulation performed on a computer.

Furthermore, simulations can be both deterministic and stochastic. While results from stochastic simulations will change from one simulation to the next, deterministic simulations will yield the same results given the same input.

3.3.1 Advantages of simulation

Robinson (2004) argues that simulation handle variability well, and divide variations into two categories: predictable and unpredictable. As time passes, the number of emergency vehicles in a system usually varies with a known time schedule. Such a variation is predictable. Often a system has unpredictable variations as well. These cannot be known in advance. In an EMS planning context, such variations include a fire starting or a person getting a heart attack. Both incidents occur randomly with respect to both location and time.

Furthermore, these incidents cannot always be treated independently of each other, as they tend to affect one another. Handling the interconnection between incidents is another strength of simulation, according to Robinson (2004). Again, in the case of EMS planning an example of interconnection could be two fires happening simultaneously, both requiring the same resources. Another example could be that an ambulance is already busy with a patient when another emergency occurs.

Complexity issues also make it very hard to predict system performance. Brooks and Tobias (1996) point out that "A complex model is probably more likely to contain errors as it is harder to verify that the model is working as intended". They further divide complexity into three elements: size, interconnection (as mentioned above) and calculation complexity. Robinson (2004) refers to the first and latter as combinatorial and dynamic complexity, respectively. The first relates to a combinatorial explosion, best demonstrated in the travelling salesman problem as the number of cities (size) increases. Dynamic complexity is not necessarily related to size. This type of complexity arises when components in a system interacts with each other, for instance when one ambulance transfers a patient to another ambulance or to a helicopter. This kind of action can have ripple effects that is hard to imagine and predict without simulation.

Furthermore, the alternative to simulation is in many occasions only to do experiments in the real world. In this regard, simulation has many practical advantages: it is usually less costly, quicker to implement and test, and it is easier to control the environments/experimental conditions. Lastly, in some cases, the system may not exist, so a real life experiment cannot be done at all without building the system to begin with.

However, simulation does have its downsides. While it can be less costly and quicker than implementing a real life experiment, building simulation software can be both time consuming and expensive. To build proper simulations, some models require large amount of data. In many cases, the data may not be available at all and must be collected. Even if it is available, much collation and analysis can be required to make it usable for the model. Another problem with building simulation models, Robinson (2004) claims, is that it requires a specific expertise that may not always be readily available. Such expertise scope computer programming (or using specific software), conceptual modelling, validation and statistics. Lastly, as with any modelling approach, there is a danger that the simulation can be interpreted as reality. Being aware of this overconfidence is imperative when interpreting results from simulation.

3.3.2 Simulation over time: continuous, discrete and mixed approaches

According to Pidd (1994), there are basically three approaches to handle the progress of time when dealing with a (dynamic computer) simulation. In the first approach, continuous simulation, one can track system changes continuously through time. This approach is widely used for economists and engineers. Such systems are usually directly dependent on time, i.e. its objects' behaviour can be described by functions of time. For instance a system with fluids, where some of its attributes can be described by differential equations.

Computers can approximate continuous change by evaluating the system for every small time increment. With smaller increment sizes, the approximation becomes more precise (Robinson, 2004). However, with reduced increment size the simulation run time increases as more calculations must be done.

The second approach described is discrete event simulation (DES). In contrast to continuous simulation, DES only cares about those points in time where the system state changes. Since these events may (and probably will) occur at irregular time intervals, the simulation flow in DES is not as smooth as in the method mentioned above. DES is being used increasingly in health-care services (Fone et al., 2003).

Lastly, a hybrid of the two can be implemented. Some systems may require both discrete and continuous simulations to give the best results. This is also known as a mixed discrete/continuous simulation. "An example of this is a factory in which there is a cooking process controlled by known physics which is modelled by continuous equations. Also in the factory is a packing line from which discrete pallets of products emerge." (Pidd, 1994).

In this thesis, the continued focus will be on the DES approach.

3.3.3 Discrete Event Simulation (DES)

Time advance methods

In dynamic simulation, a simulation-clock keeps track of time. As discussed above, DES only cares about the points in time where an event occurs. In order to implement this, a time-advance mechanism is needed. For DES, there are two well-known approaches: fixed-increment advance and next-event time advance. In fixed-increment advance, simulation starts at time zero, and then advances with a given increment. This approach is also called "time-slice" or "periodic scan" (Garrido, 1998). For every increment, a check is ran to check whether an event has occurred or not. With this approach, an event may occur in the interval between two "beats", thus leading to a disjuncture between the clock beat and the actual moment of the event. Since the fixed-increment approach only deals with discrete points in time in accordance to the increments, the events are forced to happen at the nearest future beat. This might lead to unwanted consequences. It is possible to reduce this disjuncture by decreasing the fixed increment, but this leads to more increments where nothing happens, thus slowing down the simulation. Having many "empty" beats is one of the major drawbacks of this method. (Diaz and Behr, 2010).

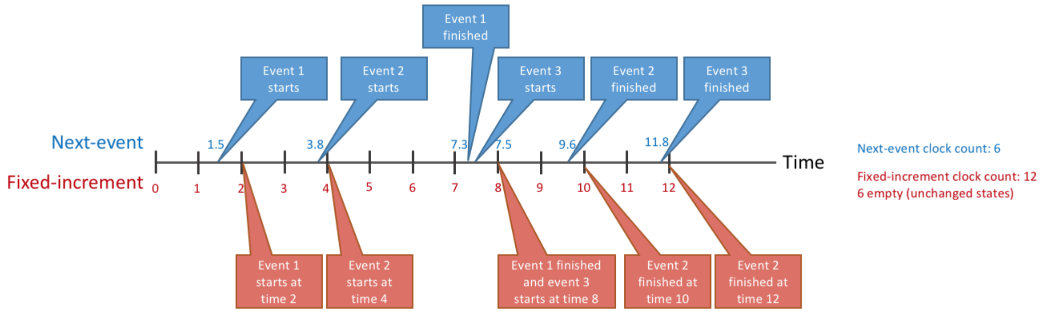


Figure 3.4: Next-event versus fixed-increment time advance methods. As we can see, next-event method generates no unnecessary beats, but has an irregular beat, in contrast to the even fixed-increment clock with half of the beats with no state changes. The figure also shows the small time discrepancy in fixed-increment method, as events are recorded at the next clock beat.

The issue with determining the increment size is resolved in the alternative approach, by using the next-event time advance technique. Here, for every beat on the simulation clock, new events are generated along with its time an added to an event list. Then, when one event is finished simulated, the simulation clock advances to the time of the next event in the event list that has not yet been simulated.

Main components

Diaz and Behr (2010) list and describe the following components associated with DES:

1. System State. A characterization of the state of the system at a particular moment; expressed as state variables.
2. Statistical Counter or Accumulator. A tool that both records and expresses a system's evolving performance; it accumulates information as the simulation unfolds.
3. Initialization Subprogram. A protocol utilized in the initialization of the simulation, usually setting the start time to zero.
4. Simulation Clock. A tool that provides the elapsed real-world time.
5. Timing Subprogram. A protocol that progresses the simulation clock to the moment when an event is to happen.
6. Event Subprogram. A protocol that launches a routine that updates the state of the system with the occurrence of each event.

7. Library Subprogram. A protocol used to produce random observations drawn generally from predetermined probability distributions. (More on this in Subsection 3.3.4)

8. Report Generator. A tool that calculates and reports statistics that describe the performance of the system.

9. Main Program. A routine that coordinates the concert of subordinate routines, executing these in the correct sequence. It initializes the timing subprogram that determines the subsequent event, passes control to the related event subprogram, and updates the system state. This routine verifies for termination and triggers the report generator once the simulation ends.

There are many ways of dividing and defining components of a simulation program. In practical implementations, some components can be combined into one, or split into several parts. For this thesis the terminology listed above is used when explaining the simulation program.

3.3.4 Generating stochastic events

When building a simulation that should handle unpredictable (stochastic) events, a method to generate these events is needed. To generate random events, we need to generate random numbers. With computers, generating (pseudo)random numbers are easy, and different programming languages have ways of doing it. We will not dive into the different computer algorithms for generating random numbers, but instead describe what random numbers are, and their properties. Then, we examine how these random numbers can be used to generate events from a probability distribution.

Random numbers

Random numbers can be either integer numbers or real numbers. The simplest way of generating a random integer on the scale [1,6], is by throwing a dice. Another physical example is the top hat method: To generate random numbers between 0-99, 100 pieces of paper are placed into a hat, each with a number from 0-99 written on it. By drawing one number, replacing it and repeating this, a random sequence of numbers in the scale [0, 99] is created. These random numbers have two properties: (Robinson, 2004)

- Uniform: it is the same probability of any number occurring at any point in the sequence;
- Independent: once a number has occurred this does not affect the probability of it occurring again or the probability of another number occurring.

In a digital world, creating random numbers is not as trivial as it may sound. In fact, "the very idea of a deterministic machine, like a computer, creating random numbers seems to be an oxymoron." (Nahin, 2002). Actually, computers can only (by itself) create what are called pseudo-random numbers; i.e. numbers that appear to be random. But in fact, these numbers are created with an algorithm, that sooner or later will repeat itself. In this thesis, the modules used to create random numbers, are in fact pseudo-random, but will be referred to as random numbers.

Drawing events from a probability distribution

Here, two methods for drawing events by using random numbers are described. The first method utilizes a lookup table containing a stochastic variable X and the cumulative probability of X. Assume that the probability of an emergency with a given priority is as follows:

Table 3.4: Emergency priority probabilities.

Priority (X)	Probability, $P(X = x)$
Green	0.3
Orange	0.4
Red	0.3

Then the lookup table would be as follows:

Table 3.5: Lookup table for emergency priorities.

Priority (X)	Cumulative Probability, $F_X(x) = P(X \leq x)$
Green	0.3
Orange	0.7
Red	1.0

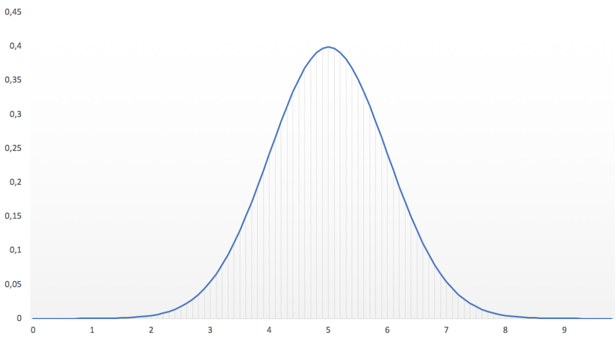
Now, in order to generate an emergency with a priority, a random real number $u \in [0, 1]$ is drawn. If $0 \leq u \leq 0.3$, then a green emergency is generated. Similarly, if $0.3 < u \leq 0.7$, the emergency is orange, and $0.7 < u \leq 1$ gives a red priority emergency.

This approach is easy to implement, and common when dealing with discrete events/inputs, like in the example above.

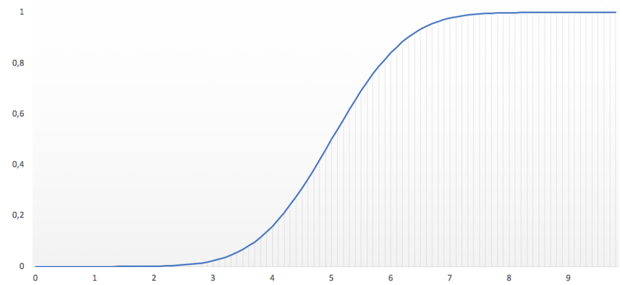
When generating events or inputs that can be continuous, another approach can be used. Inverse transform sampling, or the inverse transformation method, assumes that a random variable X has a known probability density function, and can be described with the cumulative distribution function F_X. Then, the method is as follows:

1. Generate a random real number $u \in [0,1]$ according to the random properties described above.
2. Compute the value x, such that $F_X(x) = u$, by using the inverse function: $F^{-1}(u) = x$
3. x is the randomly selected sample.

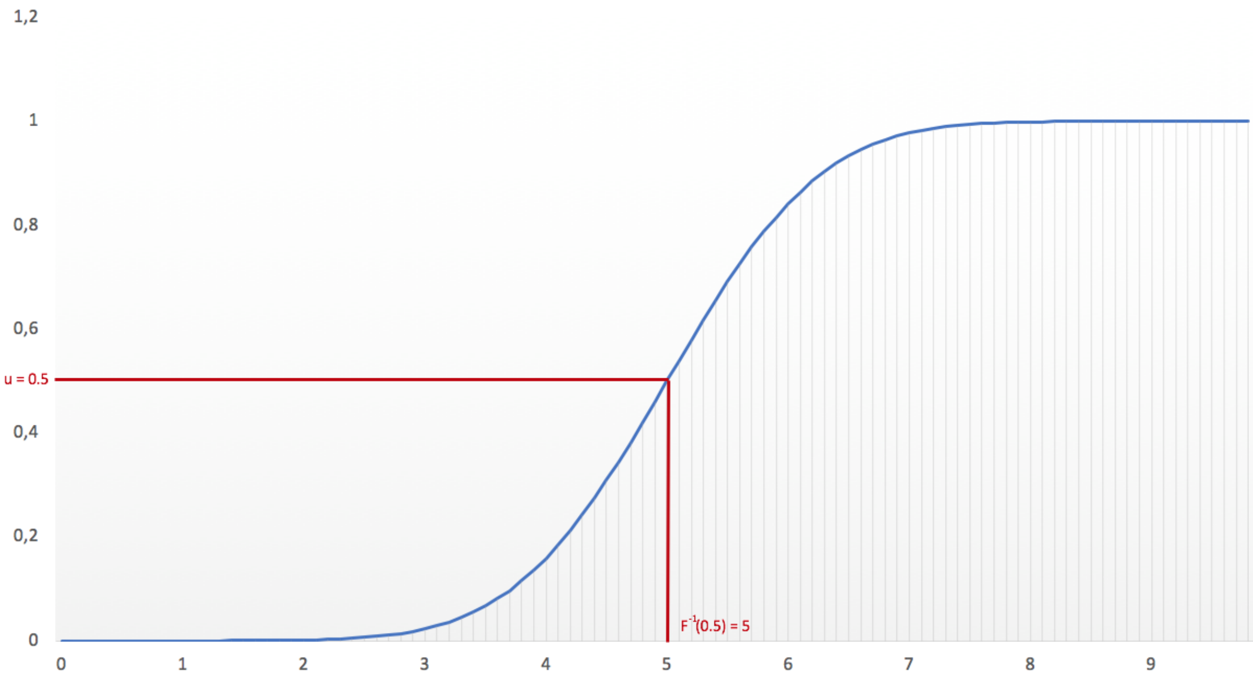
For example, assume $X \sim N(5,1)$. The probability density function is shown in Figure 3.5a. From this function, we can find the cumulative distribution function, shown in Figure 3.5b. Assume that our random number, $u = 0.5$, then we know that $F_X(x) = 0.5$, which means that $F^{-1}(0.5) = X$. Finding X is illustrated in Figure 3.5c. In reality the inverse function must be derived from the integral of the probability density function, if the cumulative distribution function is not known. This is not further explained in this thesis.



(a) Probability density function



(b) Cumulative distribution function



(c) Cumulative distribution function

Figure 3.5: The inverse transform method utilizes only the cumulative distribution function. This can be derived from the probability density function.

3.3.5 Use of simulation with optimization in EMS planning

Simulation for evaluating EMS planning decisions is considered a valuable tool. It can be used to evaluate specific models and OR problems, like Andersson et al. (2016) have done for their ambulance relocation problem. Another application is to use simulation to evaluate performance of entire systems. Mccormack and Coates (2015) evaluates a survival based OR ambulance allocation and station location model for multiple vehicle classes (ambulances and rapid response cars). In their work, simulation is used to evaluate performance for London using real data provided by the London EMS. This work portrays a more comprehensive use of simulation. Their arguments for using simulation over alternative modelling techniques is that it offers increasing realism and accuracy. This accuracy can be achieved by replaying real call data (trace-driven), giving a real demand. Hence, they avoid the trouble and simplifications needed to model demand. However, simply replaying real data raises the question whether anomalies, or registration errors, should be included in the evaluation.

Chapter 4

Problem description

In this chapter the Differentiated Stroke Treatment Problem is presented. The problem consists of three parts. Firstly, mobile resources should be allocated within a specific geographic area. Secondly, we want to choose the best suited unit to respond to incidents based on the patient's symptoms, differentiating between stroke and non-stroke patients. Thirdly, given a stroke, we want to transport the patient to the hospital which gives the best prospect of outcome.

A geographic area is defined and divided into zones. Between each zone, there is a travel time for all types of mobile resources. Each zone is categorized as either an urban or a rural zone. For each zone, there is information available on the demand for general medical attention, and for stroke related medical attention. These are called the general demand, and the stroke demand. For the stroke demand, a given proportion is eligible for thrombectomy treatment. In the area we consider, there is a given number of resources that can satisfy both of these demands. The resources are both stationary and mobile. The stationary resources consist of hospitals, ambulance stations and helicopter stations. Mobile resources consists of ambulances and helicopters, that can be allocated to the stations. An individual zone can hold several stationary and mobile resources.

A hospital is located in a given zone and cannot be moved. There are two types of hospitals: stroke units and stroke centers. Both types can satisfy general demand. Both can also diagnose stroke patients, as well as perform thrombolysis treatment. But only stroke centers can perform thrombectomy treatment. A zone is covered by the hospital which is closest. If the closest hospital is a stroke unit, the zone also has a secondary hospital, which is a stroke center. If the closest hospital is a stroke center, the zone does not need a secondary hospital, since that hospital can cover all types of patients in the zone. Therefore, for each zone the following distances are known: time to the closest hospital, time to the closest stroke center, and time to the closest stroke center via a stroke unit.

Ambulance stations and helicopter stations can be located in a subset of given zones within the geographic area. An ambulance station can hold several ambulances, and a helicopter station can hold several helicopters. They can also be empty. Each station must be assigned a set of zones which they are responsible for. If medical attention is needed in a zone, an ambulance/helicopter from the station responsible for that zone responds. Each zone must be covered by one station. The mobile unit transports the patient from the demand zone where the patient is, to the closest hospital. If an ambulance initially responds to the patient, a helicopter may also be dispatched to meet the ambulance, and pick up the patient. Meeting is only considered when it saves more time to hospital than a given threshold. Meetings can take place in any zone. After the patient has been transported to the hospital for diagnosis and received correct treatment, the unit returns to its station.

A subset of the ambulances can perform diagnosis prehospitally by taking a CT scan when they arrive at the patient. These CT ambulances can also administer thrombolysis when the patient has been diagnosed before arrival at the hospital. If the patient is eligible for thrombectomy treatment, the patient will be transported directly to the assigned stroke center, not via the closest hospital if that is a stroke unit.

All mobile resources can satisfy a given amount of stroke demand. The response time from the zone where the mobile unit is stationed, to the zone it serves affects the capacity. Longer travel times reduce the potential demand a unit can satisfy. A station has the accumulated capacity of all its units. This means that mobile units at the same station share the workload among each other.

The cost associated with general demand is influenced by the response time and time to closest hospital for patients in each zone. Urban and rural zones have different time to treatment thresholds for general emergencies. For stroke incidents in each zone, the time to treatment starts with the emergency call, but does not end until final treatment is conducted at the best choice of hospital for that zone. The choice of medical facility for a zone is determined by the type of vehicle that responds to the zone as well as the distances to the medical facilities. If the responding unit is a CT ambulance and the stroke patient is not eligible for thrombectomy, the appropriate treatment can be given at the patient's location.The cost of stroke treatment for each zone is calculated based on a survival function and the stroke demand in the zone. The survival function is affected by the time to treatment.

The total cost incurred is the sum of all cost functions, both stroke and general treatments, for all zones. The goal is to allocate the mobile resources in the geographic area such that the overall cost related to thrombolysis, thrombectomy and general emergencies are minimized. Implicitly, this will maximize the overall survival.

Chapter 5

Mathematical model

In this chapter we suggest an optimization model that aims to solve the problem described in Chapter 4. Our base model, which is presented in Section 5.3, serves as a general approach to minimize the time to treatment for patients based on a survival function. Many of the simplifications and assumptions needed to make the problem solvable are done on the input data and not in the mathematical model. Because of this, these assumptions are discussed in Section 5.2, before the model implementation.

After the model is presented, we discuss some variations of the problem. The variations are based on issues that we have discussed, and others are based on issues that became apparent when running the model. Some of these variations are implemented in Chapter 8. Since this problem has many parameters, some figures will be presented first to clarify all components.

5.1 Figurative representation of the stroke treatment chain

The stroke treatment chain is composed of many parts. Understanding the chain can be hard, especially when stroke ambulances are introduced. This alters the chain of events. Hopefully, these conceptual figures will help. In the figures, most of the time components (both parameters and variables) used in the optimization model (and simulation) are presented. Although these are not defined yet, the figures can be used for reference when we do define them and use them in the mathematical model. You will recognize some parts of these figures from the Background chapter, but mathematical notation is added.

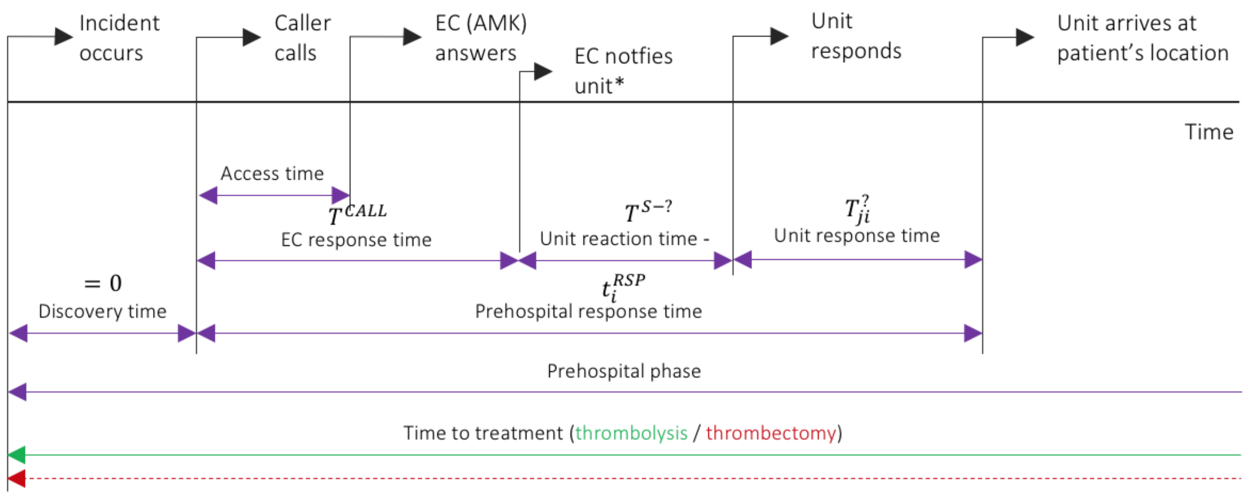


Figure 5.1: First phase of the acute stroke treatment chain. * If there are CT ambulances available, the emergency central (EC) may choose to dispatch one of these for certain zones. ? indicates that these parameters varies for ambulance and helicopter.

The first phase, which is depicted in Figure 5.1 describe the chain of events for responding to a stroke patient regardless of vehicle class. Figure 5.2 describes the continuation for the stroke treatment chain if the responding unit is an ordinary ambulance, or helicopter.

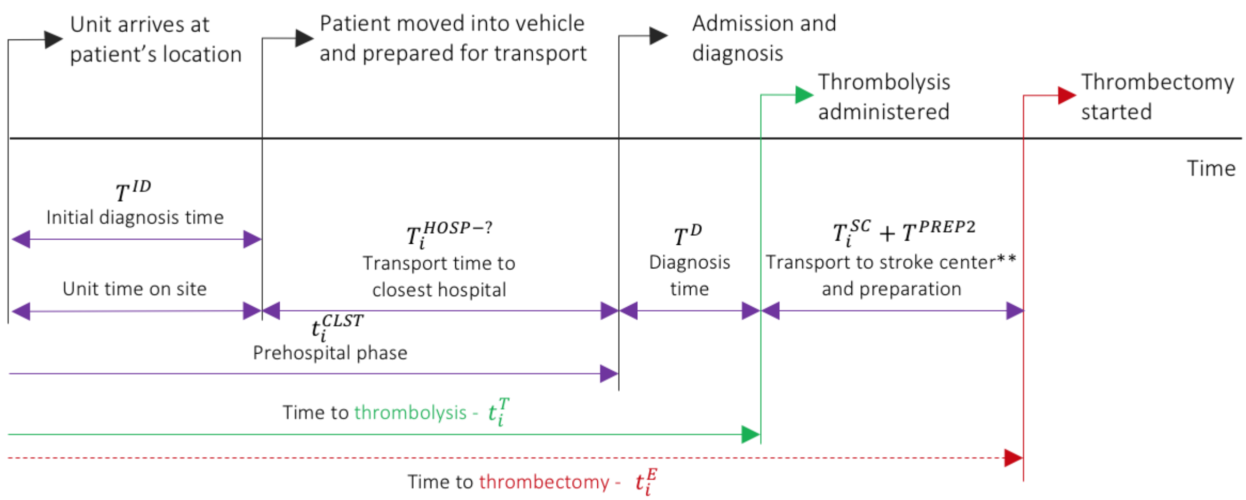


Figure 5.2: Chain of events if a regular medical emergency unit (no CT) arrives to a stroke patient. ** Only happens if patient is eligible for thrombectomy. This happens with a probability of P^E. All other stroke patients are treated and considered finished when thrombolysis is administered.

If the responding unit is a CT ambulance, the chain of events changes as thrombolysis can be administered prehospitally. Then the patient can be taken directly to a stroke center if he or she is eligible for thrombectomy. This is shown in Figure 5.3.

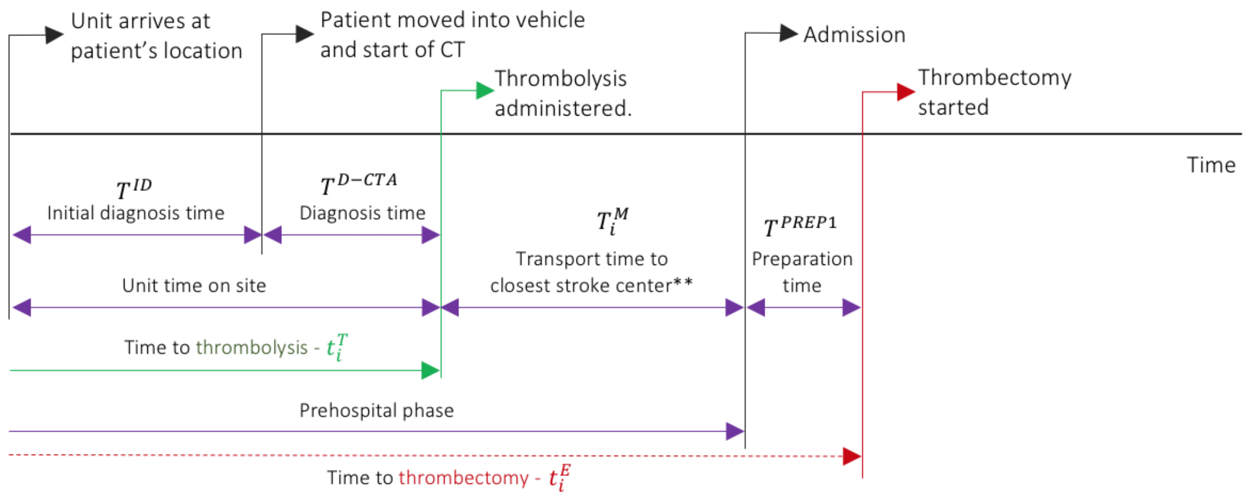


Figure 5.3: Chain of events if a medical emergency unit with CT arrives to a stroke patient.** Only happens if patient is eligible for thrombectomy. This happens with a probability of P^E. All other stroke patients are treated and considered finished when thrombolysis is administered.

For patients not eligible for thrombectomy, we consider the patient treated when thrombolysis is administered. Notice that use of CT ambulance lead to a longer unit time on site, hence prolonging the prehospital phase. However, the time to thrombolysis is reduced.

5.2 Assumptions

We have made the following modelling assumptions when formulating the problem:

Discovery time As depicted in Figure 5.1, all incidents start with a Discovery time. This time is sometimes also referred to as "bystander time". As this time is highly random and often unknown, we disregard it, and set to 0.

Availability In the model all medical response units are assumed to be available at their stations at all times. This assumption can be justified on the basis that red emergencies and stroke have the highest priority for emergency services. This means that low-priority missions will be canceled if an emergency unit is needed for a red/stroke patient.

Zone and its location For every zone, we consider its location to be a point location. This also means that its demand is located in this point.

Zones covered by only one unit and unit type In the model, it is assumed that only one station can cover a zone - and that this is satisfactory. If a station has several vehicle types, each zone can be covered by only one of these. In reality, some ambulances are busy, and other ambulances need to take over.

Change of transportation type along the chain The base model takes into consideration that an ambulance can respond to the patient, complete the initial diagnosis and transport the patient to a meeting zone where a helicopter relieves the ambulance. If a patient is diagnosed thrombectomy eligible at a stroke unit, helicopter will be used to transport the patient to the closest stroke center.

Predetermined choice of hospital It is assumed that any zone belongs to the hospital which gives the lowest transportation time. By doing this, we can calculate travel times in advance. For thrombectomy, we have calculated the travel time to the closest stroke center.

Demand and unit capacities Ideally, the demand in each zone should reflect the demographics of the zone. Specifically, a zone with an older population should have a high expected demand for medical services. As we showed in the background, stroke is highly correlated with age. Hence, zones with an higher average age should have a higher demand than an equally large zone with a younger population. We have however assumed that the population is homogeneously distributed, so that demand in each zone only depends on the population size. Our implementation of demand and capacities are further discussed in Chapter 7.

CT ambulance usage We assume that if a CT ambulance is stationed together with ordinary ambulances in a zone j, this CT ambulance responds mainly to strokes. This gives synergistic effects, that increases the CT capacity of the station. This assumption is justified since ordinary ambulances will handle a greater proportion of the non-stroke emergencies. If a CT ambulance is the only unit in a station, it covers all emergencies.

Hospital capacity We assume that there are no capacity restrictions at the hospitals. If a patient gets the fastest possible treatment by travelling to a hospital, that hospital will accept the patient. Additionally, a stroke center must perform the thrombectomy procedure a number of times during a year to maintain a certain skill level. This restriction is also disregarded in this report.

5.3 Model formulation

In this section we will start by presenting the notation. Then, we present the objective function and the restrictions. For the readers convenience, we have grouped similar restrictions and commented these as we move along. Again, the figures in Section 5.1 might be useful for this and the following sections.

5.3.1 Notation

Indices

i, j, k, h	: Zones

Sets

N	: All zones
N^A	: All zones where ambulances can be stationed
N^H	: All zones where helicopters can be based

Parameters

α	: Weight of stroke cost function in the objective function, $\alpha \in [0,1]$
$C^T(t_i^T)$	: Non-linear function returning a cost based on time to thrombolysis
$C^E(t_i^T, t_i^E)$	: Non-linear function returning a cost based on time to endovascular procedure
$C^G(s_i, t_i^{CLST})$	: Linear function returning a cost based on response time and time to closest hospital
D_{ij}^A	: Demand covered by an ambulance responding from zone j to i
D_{ijk}^A	: Demand covered by an ambulance responding from zone j to i, meeting helicopter in k
D_{ih}^H	: Demand covered by a helicopter responding from zone h to i
D_i^S	: Stroke demand: demand for EMS caused by stroke patients zone i
D_i^G	: General demand: demand for EMS caused by all other (non-stroke) patients in zone i
K^A	: Ambulance capacity
K^H	: Helicopter capacity
K^{CTA}	: CT Ambulance capacity

K^{SYN}	: Additional synergic capacity from having a CT ambulance in a station
L^{A}	: Maximum number of ambulances available
L^{H}	: Maximum number of helicopters available
L^{CTA}	: Maximum number of CT ambulances available
P^{T}	: Probability that a patient can only be treated with thrombolysis
P^{E}	: Probability that a patient is eligible for the endovascular procedure
T^{CALL}	: Emergency call time (EC response time)
T^{ID}	: Time needed for initial diagnosis when arriving at patient location
T^{D}	: Time needed for admission, CT and diagnosis for stroke patients at hospital
T^{D-CTA}	: Time needed for CT and diagnosis for stroke patients in CT ambulance
T^{A}_{ji}	: Transport time from zone j to zone i for ambulances
T^{H}_{ji}	: Transport time from zone j to zone i for helicopters
T^{HOSP-A}_{i}	: Transport time to closest hospital by ambulance from zone i
T^{HOSP-H}_{i}	: Transport time to closest hospital by helicopter from zone i
T^{PREP1}	: Endovascular treatment preparation time for a patient who has been diagnosed in CT ambulance
T^{PREP2}	: Endovascular treatment preparation time for a patient who has not been pre-hospitally diagnosed
T^{LIMIT}	: Minimum improved transport time needed to justify helicopter transport
T^{MAX}_{i}	: Recommended maximum response time for zone i
T^{S-A}	: Scramble time for ambulances (unit reaction time)
T^{S-H}	: Scramble time for helicopters (unit reaction time)
T^{S-CTA}	: Scramble time for CT ambulances (unit reaction time)
T^{M}_{i}	: Transport time directly to stroke center from zone i ("Mothership")
T^{SC}_{i}	: Transport time from zone i's closest hospital to stroke center
T^{W}_{ijkh}	: Waiting time needed in connection with changing from ambulance to helicopter, when ambulance responds from zone j to i, meets a helicopter in k that responded from h
T^{C}	: Time needed to transfer patient from ambulance to helicopter

Decision variables

General time variables:

t^{CLST}_{i}	: Total time to closest hospital for zone i
t^{RSP}_{i}	: Response time for zone i

Time-to-thrombolysis:

t^{T}_{i}	: Time-to-thrombolysis for zone i

t_i^{T-A} : Time-to-thrombolysis for zone i if covered by an ordinary ambulance
t_i^{T-CTA} : Time-to-thrombolysis for zone i if covered by a CT ambulance
t_i^{T-H} : Time-to-thrombolysis for zone i if covered by a helicopter
t_i^{T-AH} : Time-to-thrombolysis for zone i if covered by an ordinary ambulance that meets a helicopter
t_i^{T-CTAH} : Time-to-thrombolysis for zone i if covered by a CT ambulance that meets a helicopter

Time-to-endovascular treatment:
t_i^{E} : Time-to-endovascular treatment for zone i
t_i^{E-A} : Time-to-endovascular treatment for zone i if covered by an ordinary ambulance
t_i^{E-CTA} : Time-to-endovascular treatment for zone i if covered by a CT ambulance
t_i^{E-H} : Time-to-endovascular treatment for zone i if covered by a helicopter
t_i^{E-AH} : Time-to-endovascular treatment for zone i if covered by an ordinary ambulance that meets a helicopter
t_i^{E-CTAH} : Time-to-endovascular treatment for zone i if covered by a CT ambulance that meets a helicopter

Other decision variables:
s_i : Slack variable, response time in excess of T^{MAX}
x_j^{CTA} : 1 if a CT ambulance is stationed in zone j, 0 otherwise
y_j^{A} : Number of ordinary ambulances stationed in zone j
y_j^{H} : Number of helicopters based in zone j
w_{ij}^{A} : 1 if an ordinary ambulance covers zone i from zone j, 0 otherwise
w_{ij}^{H} : 1 if a helicopter covers zone i from zone j, 0 otherwise
w_{ij}^{CTA} : 1 if a CT ambulance covers zone i from zone j, 0 otherwise
w_{ijkh}^{AH} : 1 if an ordinary ambulance covers zone i from zone j, meets helicopter from h in k, 0 otherwise
w_{ijkh}^{CTAH} : 1 if a CT ambulance covers zone i from zone j, meets helicopter from h in k, 0 otherwise

5.3.2 Objective function

$$\min \quad \alpha \cdot \left(P^T \cdot \sum_{i \in N} D_i^S \cdot C^T(t_i^T) + P^E \cdot \sum_{i \in N} D_i^S \cdot C^E(t_i^T, t_i^E) \right) + (1-\alpha) \cdot \left(\sum_{i \in N} D_i^G \cdot C^G(s_i, t_i^{CLST}) \right) \quad (5.1)$$

The objective function (5.1) is a weighted sum of the total cost of stroke treatment and the total cost of general emergencies. The cost of stroke treatment is divided into cost of thrombolysis and cost of endovascular treatment.

5.3.3 Restrictions

$$t_i^T = t_i^{T-A} + t_i^{T-CTA} + t_i^{T-H} + t_i^{T-AH} + t_i^{T-CTAH} \qquad i \in N \quad (5.2)$$

Restriction (5.2) assigns a value to the time-to-thrombolysis variable based on the type of coverage for zone i.

$$t_i^E = t_i^{E-A} + t_i^{E-CTA} + t_i^{E-H} + t_i^{E-AH} + t_i^{E-CTAH} \qquad i \in N \quad (5.3)$$

Restriction (5.3) assigns a value to the time-to-endovascular procedure variable based on the type of coverage for zone i.

$$t_i^{T-A} = \sum_{j \in N^A} (T^{CALL} + T^{S-A} + T_{ji}^A + T^{ID} + T_i^{HOST-A} + T^D) \cdot w_{ij}^A \qquad i \in N \quad (5.4)$$

$$t_i^{T-CTA} = \sum_{j \in N^A} (T^{CALL} + T^{S-CTA} + T_{ji}^A + T^{ID} + T^{D-CTA}) \cdot w_{ij}^{CTA} \qquad i \in N \quad (5.5)$$

$$t_i^{T-H} = \sum_{h \in N^H} (T^{CALL} + T^{S-H} + T_{hi}^H + T^{ID} + T_i^{HOSP-H} + T^D) \cdot w_{ih}^H \qquad i \in N \quad (5.6)$$

$$t_i^{T-AH} = \sum_{j\in N^A}\sum_{k\in N}\sum_{h\in N^H}(T^{CALL} + T^{S-A} + T_{ji}^{A} + T^{ID} + T_{ik}^{A} + T_{ijkh}^{W} + T^{C} + T_{k}^{HOSP-H} + T^{D}) \cdot w_{ijkh}^{AH} \qquad i \in N \tag{5.7}$$

$$t_i^{T-CTAH} = \sum_{j\in N^A}\sum_{k\in N}\sum_{h\in N^H}(T^{CALL} + T^{S-CTA} + T_{ji}^{A} + T^{ID} + T^{D-CTA}) \cdot w_{ijkh}^{CTAH} \qquad i \in N \tag{5.8}$$

Restrictions (5.4) to (5.8) assign values to the time-to-thrombolysis variable for each type of coverage for zone i.

$$t_i^{E-A} = t_i^{T-A} \quad + \sum_{j\in N^A}(T_i^{SC} + T^{PREP2}) \cdot w_{ij}^{A} \qquad i \in N \tag{5.9}$$

$$t_i^{E-CTA} = t_i^{T-CTA} \quad + \sum_{j\in N^A}(T_i^{M} + T^{PREP1}) \cdot w_{ij}^{CTA} \qquad i \in N \tag{5.10}$$

$$t_i^{E-H} = t_i^{T-H} \quad + \sum_{h\in N^H}(T_i^{SC} + T^{PREP2}) \cdot w_{ih}^{H} \qquad i \in N \tag{5.11}$$

$$t_i^{E-AH} = t_i^{T-AH} \quad + \sum_{j\in N^A}\sum_{k\in N}\sum_{h\in N^H}(T_k^{SC} + T^{PREP2}) \cdot w_{ijkh}^{AH} \qquad i \in N \tag{5.12}$$

$$t_i^{E-CTAH} = t_i^{T-CTAH} \quad + \sum_{j\in N^A}\sum_{k\in N}\sum_{h\in N^H}(T_{ik}^{A} + T_{ijkh}^{W} + T^{C} + T_k^{M} + T^{PREP1}) \cdot w_{ijkh}^{CTAH} \qquad i \in N \tag{5.13}$$

Restrictions (5.9) to (5.13) assign values to the time-to-endovascular procedure variable for each type of coverage for zone i.

$$\begin{aligned}
t_i^{RSP} = T^{CALL} &+ \sum_{j\in N^A} \left(w_{ij}^{A} \cdot (T^{S-A} + T_{ji}^{A}) + w_{ij}^{CTA} \cdot (T^{S-CTA} + T_{ji}^{A}) \right) \\
&+ \sum_{h\in N^H} w_{ih}^{H} \cdot (T^{S-H} + T_{hi}^{H}) \\
&+ \sum_{j\in N^A} \sum_{k\in N} \sum_{h\in N^H} \left(w_{ijkh}^{AH} \cdot (T^{S-A} + T_{ji}^{A}) + w_{ijkh}^{CTAH} \cdot (T^{S-CTA} + T_{ji}^{A}) \right) \qquad i \in N
\end{aligned} \tag{5.14}$$

Restriction (5.14) assigns values to the response time based on coverage type for zone i.

$$\begin{aligned}
t_i^{CLST} &= t_i^{RSP} + T^{ID} \\
&+ \sum_{j\in N^A} T_i^{HOSP-A} \cdot (w_{ij}^{A} + w_{ij}^{CTA}) \\
&+ \sum_{h\in N^H} T_i^{HOSP-H} \cdot w_{ih}^{H} \\
&+ \sum_{j\in N^A} \sum_{k\in N} \sum_{h\in N^H} (T_{ik}^{A} + T_{ijkh}^{W} + T^{C} + T_k^{HOSP-H}) \cdot (w_{ijkh}^{AH} + w_{ijkh}^{CTAH}) \qquad i \in N
\end{aligned} \tag{5.15}$$

Restriction (5.15) assigns values to the time to closest hospital variable based on coverage type for zone i.

$$\begin{aligned}
&\sum_{i\in N} D_{ij}^{A} \cdot (w_{ij}^{A} + w_{ij}^{CTA}) + \\
&\sum_{i\in N} \sum_{k\in N} \sum_{h\in N^H} D_{ijk}^{A} \cdot (w_{ijkh}^{AH} + w_{ijkh}^{CTAH}) \le K^{A} \cdot (y_j^{A} + x_j^{CTA}) \qquad j \in N^A
\end{aligned} \tag{5.16}$$

$$\begin{aligned}
&\sum_{i\in N} D_{ij}^{A} \cdot w_{ij}^{CTA} + \\
&\sum_{i\in N} \sum_{k\in N} \sum_{h\in N^H} D_{ijk}^{A} \cdot (w_{ijkh}^{CTAH}) \le K^{CTA} \cdot x_j^{CTA} + K^{SYN} \cdot y_j^{A} \qquad j \in N^A
\end{aligned} \tag{5.17}$$

Restrictions (5.16) and (5.17) ensure that a ground station cannot cover more than its total capacity. The synergistic effects implemented in Constraint (5.17) was discussed in Section 5.2.

$$\sum_{i \in N} D_{ih}^{H} \cdot w_{ih}^{H} + \sum_{i \in N} \sum_{j \in N^A} \sum_{k \in N} D_{hk}^{H} \cdot (w_{ijkh}^{CTAH} + w_{ijkh}^{AH}) \leq K^{H} \cdot y_{h}^{H} \qquad h \in N^{H} \tag{5.18}$$

Restriction (5.18) ensures that a helicopter base cannot cover more than its total capacity.

$$t_i^{RSP} - s_i \leq T_i^{MAX} \qquad i \in N \tag{5.19}$$

Restriction (5.19) connects the response time and slack variable for zone i. This ensures that the response time above a certain threshold is penalized in the objective function.

$$\sum_{j \in N^A} y_j^A \leq L^A \tag{5.20}$$

$$\sum_{j \in N^H} y_j^H \leq L^H \tag{5.21}$$

$$\sum_{j \in N^A} x_j^{CTA} \leq L^{CTA} \tag{5.22}$$

Restrictions (5.20) to (5.22) state that the number of EMS vehicles in use cannot be greater than the number of available vehicles.

$$\sum_{j \in N^A} (w_{ij}^{A} + w_{ij}^{CTA}) + \sum_{j \in N^A} \sum_{k \in N} \sum_{h \in N^H} (w_{ijkh}^{AH} + w_{ijkh}^{CTAH}) + \sum_{h \in N^H} w_{ih}^{H} = 1 \qquad i \in N \tag{5.23}$$

Restriction (5.23) enforce that a zone i is assigned one EMS vehicle to handle the stroke incidents.

$$w_{ij}^{CTA} + w_{ijkh}^{CTAH} \leq x_j^{CTA} \qquad i \in N, j \in N^A, k \in N, h \in N^H \tag{5.24}$$

$$w_{ih}^{H} + w_{ijkh}^{AH} + w_{ijkh}^{CTAH} \leq y_h^H \qquad i \in N, j \in N^A, k \in N, h \in N^H \tag{5.25}$$

$$w_{ij}^{A} + w_{ijkh}^{AH} \leq y_j^A + x_j^{CTA} \qquad i \in N, j \in N^A, k \in N, h \in N^H \tag{5.26}$$

Restrictions (5.24) to (5.26) connect w, x and y-variables.

$$t_i^T, t_i^E, t_i^{RSP}, t_i^{CLST}, s_i \geq 0 \qquad i \in N \tag{5.27}$$

$$t_i^{T-A}, t_i^{T-CTA}, t_i^{T-H}, t_i^{T-AH}, t_i^{T-CTAH} \geq 0 \qquad i \in N \tag{5.28}$$

$$t_i^{E-A}, t_i^{E-CTA}, t_i^{E-H}, t_i^{E-AH}, t_i^{E-CTAH} \geq 0 \qquad i \in N \tag{5.29}$$

$$y_j^A \geq 0 \text{ and integer} \qquad j \in N^A \tag{5.30}$$

$$y_j^H \geq 0 \text{ and integer} \qquad j \in N^H \tag{5.31}$$

$$x_j^{CTA} \in \{0,1\} \qquad j \in N^A \tag{5.32}$$

$$w_{ij}^A, w_{ij}^{CTA} \in \{0,1\} \qquad i \in N, j \in N^A \tag{5.33}$$

$$w_{ij}^{AH} \in \{0,1\} \qquad i \in N, j \in N^A, k \in N, h \in N^H \tag{5.34}$$

w_{ij}^{AH} is created when $T_k^{HOSP-A} - T_{ijkh}^W + T^C + T_k^{HOSP-H} \geq T^{LIMIT}$. This ensures that a helicopter can only meet an ambulance if the alternative way of travelling with ambulance to the appropriate hospital takes more than T^{LIMIT} minutes longer.

$$w_{ij}^{CTAH} \in \{0,1\} \qquad i \in N, j \in N^A, k \in N, h \in N^H \tag{5.35}$$

w_{ij}^{CTAH} is created when $T_k^M - T_{ijkh}^W + T^C + T_k^{M-H} \geq T^{LIMIT}$. This ensures that a helicopter can only meet an ambulance if the alternative way of travelling with ambulance to the appropriate hospital takes more than T^{LIMIT} minutes longer. It is assumed that a helicopter only considers meeting a CT ambulance in thrombectomy cases, since the patient is already receiving thrombolysis treatment in the CT ambulance.

$$w_{ij}^H \in \{0,1\} \qquad i \in N, j \in N^H \tag{5.36}$$

w_{ij}^H is created when $T_i^{HOSP-A} - T_i^{HOSP-H} \geq T^{LIMIT}$. This ensures that a helicopter can only be used if the alternative way of travelling with ambulance to the closest hospital takes more than T^{LIMIT} minutes longer.

Restrictions (5.27) to (5.36) are non-negativity constraints, and define variable types.

5.4 Model variations

In this section different variants of the mathematical model are presented. Subsection 5.4.1 describes a linearization of the in principle non-linear cost function, followed by a subsection containing several alternative cost functions. Subsection 5.4.3 introduces an extension that allow the model to open new ambulance stations. This section ends with a description of how to optimize for different periods of day.

5.4.1 Linearization of the objective function with a triangle method

The cost incurred when having a stroke that is eligible for endovascular treatment is influenced by two factors: The time-to-thrombolysis, t_i^T, and the time-to-endovascular treatment, t_i^E. The cost function for each zone, i, can be seen as a function C^E with the arguments t_i^T and t_i^E: $C^E(t_i^T, t_i^E)$. The idea is that the time-to-thrombolysis will affect the shape of the cost function of time-to-endovascular treatment. We have chosen to linearize the function as follows:

The cost function is evaluated for certain values of time to thrombolysis and time to endovascular treatment. Examples of cost functions are presented in section 5.4.2. Our model must choose a combination of t_i^T and t_i^E for each zone, and this choice is done by choosing a linear combination of three points. In order to keep the resulting linear combination on the surface, we must ensure that these three points are neighbours. The surface is therefore divided into triangles. Our model must choose one triangle, and can then triangulate a point using the three corner points of the triangle.

Additional notation

Indices

r	: Point of time for t_i^T
s	: Point of time for t_i^E

Sets

R^T	: Set of time points for t_i^T
R^E	: Set of time points for t_i^E
V	: Set of all triangles
V_{rs}	: Set of triangles with time points (r, s) as one of their corner points

Parameters

C_r^T	: Cost of time r for t_i^T
C_{rs}^E	: Cost of time r and s for t_i^E
R_r^T	: The associated time value for time point r for t_i^T
R_s^E	: The associated time value for time point s for t_i^E

Decision variables

δ_{irs}	: Weight on point r,s for zone i
$\mu_{i\tau}$	: 1 if triangle τ is activated for zone i, 0 otherwise

Objective function

$$\begin{aligned} \min \alpha \cdot \left(P^T \cdot \sum_{i\in N} \sum_{r\in R^T} \sum_{s\in R^E} C_r^T \cdot D_i^S \cdot \delta_{irs} + P^E \cdot \sum_{i\in N} \sum_{r\in R^T} \sum_{s\in R^E} C_{rs}^E \cdot D_i^S \cdot \delta_{irs} \right) \\ + (1-\alpha) \cdot \left(\sum_{i\in N} D_i^G \cdot (s_i + t_i^{CLST}) \right) \end{aligned} \tag{5.37}$$

The cost functions $C^T(t_i^T)$ and $C^E(t_i^T, t_i^E)$ have been sampled at times (r,s). E.g. C_{rs}^E is the sampled cost for endovascular treatment given that thrombolysis was started at time r, and endovascular treatment was initiated at time s. δ_{irs} represents the weights of the triangle corners (r,s) in the linear combination. We use the same δ_{irs} variables to calculate the cost of time to thrombolysis. However, we allow the model to sample from a seperate cost function, C_r^T, related to non-thrombectomy eligible strokes. The cost of general response $C^G(s_i, t_i^{CLST})$ is assumed to have a linear relationship with the demand for general treatment in zone i, the excess response time s_i for zone i, in addition to the time to closest hospital for zone i.

Restrictions

$$\sum_{r\in R^T} \sum_{s\in R^E} R_r^T \cdot \delta_{irs} = t_i^T \qquad i \in N \tag{5.38}$$

$$\sum_{r\in R^T} \sum_{s\in R^E} R_s^E \cdot \delta_{irs} = t_i^E \qquad i \in N \tag{5.39}$$

$$\sum_{r\in R^T} \sum_{s\in R^E} \delta_{irs} = 1 \qquad i \in N \tag{5.40}$$

$$\sum_{\tau\in V} \mu_{i\tau} = 1 \qquad i \in N \tag{5.41}$$

Restrictions (5.38) and (5.39) connects the weights of the points with the corresponding time values to convert the linear combinations back to a treatment time. Constraint (5.40) enforce the linear combination of the points sums to one. Restriction (5.41) makes sure that only one triangle is active, so only that triangle's corner points can be used for triangulation of a position on the cost function.

$$\delta_{irs} \leq \sum_{\tau \in V^{rs}} \mu_{i\tau} \qquad i \in N, r \in R^T, s \in R^E \text{ and } s \geq r \tag{5.42}$$

Constraint (5.42) ensures that a point can be used if and only if one of the triangles attached to that point is active. This will make sure that only neighboring points can be active at the same time.

$$\delta_{irs} \in [0,1] \qquad i \in N, r \in R^T, s \in R^E \tag{5.43}$$

$$\mu_{i\tau} \in \{0,1\} \qquad i \in N, \tau \in V \tag{5.44}$$

Restrictions (5.43) and (5.44) are non-negativity and binary constraints.

5.4.2 Shapes of the cost function

Concave cost function

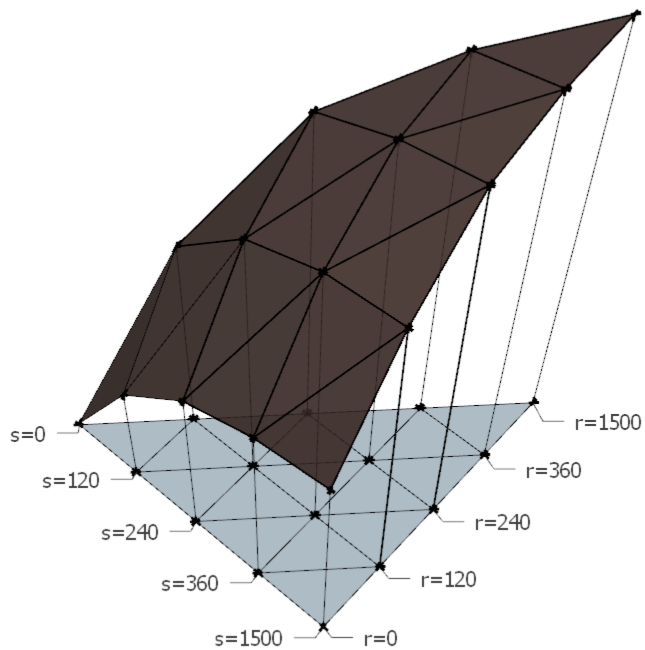


Figure 5.4: Visualization of the cost function, evaluated for time points r and s. r represent time points for time-to-thrombolysis, t_i^T, while s represent for time-to-endovascular t_i^E. Since $t_i^E \geq t_i^T$, the cost function is not evaluated for the points where $s < r$. That is, thrombectomy cannot be performed before thrombolysis is adminstered.

The time-to-thrombolysis and time-to-endovascular treatment are assumed to have a declining marginal cost as the time to treatment increases. That way, reducing time to treatment within the first minutes after the stroke happens is much more beneficial for the objective value than saving time several hours after the stroke happens. This shape tries to resemble the relationship between the survival function and the effects of stroke treatment as described in the literature (Subsection 3.2.3). For this reason the concave cost function is chosen as the default function in this thesis.

Convex cost function

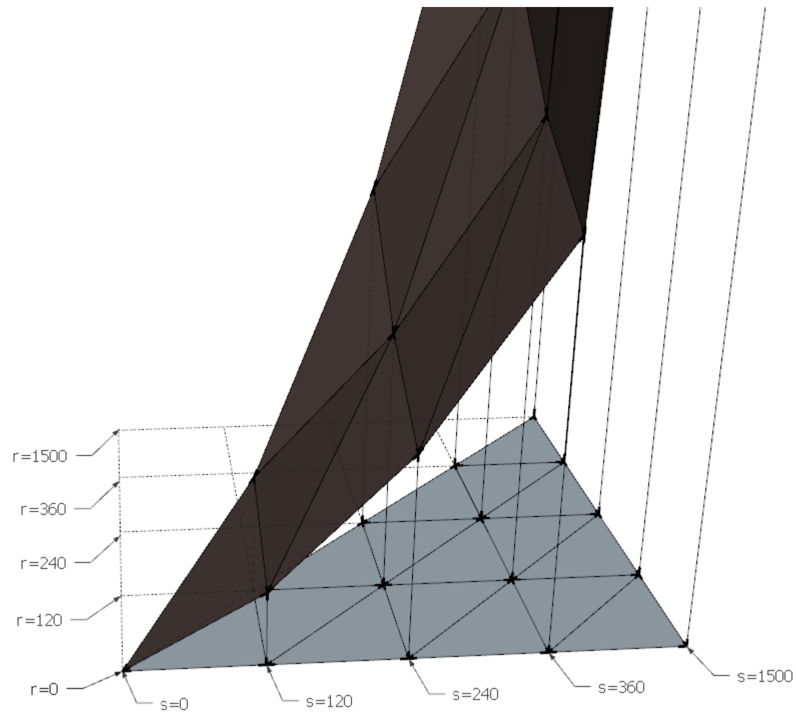


Figure 5.5: Visualization of a convex cost function, evaluated for time points r and s. r represent time points for time-to-thrombolysis, t_i^T, while s represent for time-to-endovascular t_i^E.

Here, an alternative objective function is proposed. This assume that the cost associated with the time to treatment is following a convex function. The cost function is evaluated for certain values of time-to-thrombolysis and time-to-endovascular treatment. The function has a surface as seen in Figure 5.5. Our model must choose a combination of t_i^T and t_i^E in the same manner as described in Subsection 5.4.1. We have only changed the cost associated with each time point, and therefore no additional changes in notation are required. We have used the concave cost function as a basis when crafting this convex function. The two cost functions intersect at $t_i^T = 0$, $t_i^E = 360$ and $t_i^T = 360$, $t_i^E = 360$

S-shaped objective function

The time-to-thrombolysis and time-to-endovascular treatment were in the objective functions above assumed to induce a cost starting from stroke onset. We have created an alternative objective function, which assumes that treatment within 30 minutes after the stroke happens does not induce any cost. If time to treatment exceeds 30 minutes, the time-to-thrombolysis follows a non-linear cost function, whereas time-to-endovascular treatment follows a linear cost function dependent on the time-to-thrombolysis. The time-to-endovascular cost function is linear, as opposed to the previous objective functions presented, where both time-to-thrombolysis and time-to-endovascular cost functions were non-linear.

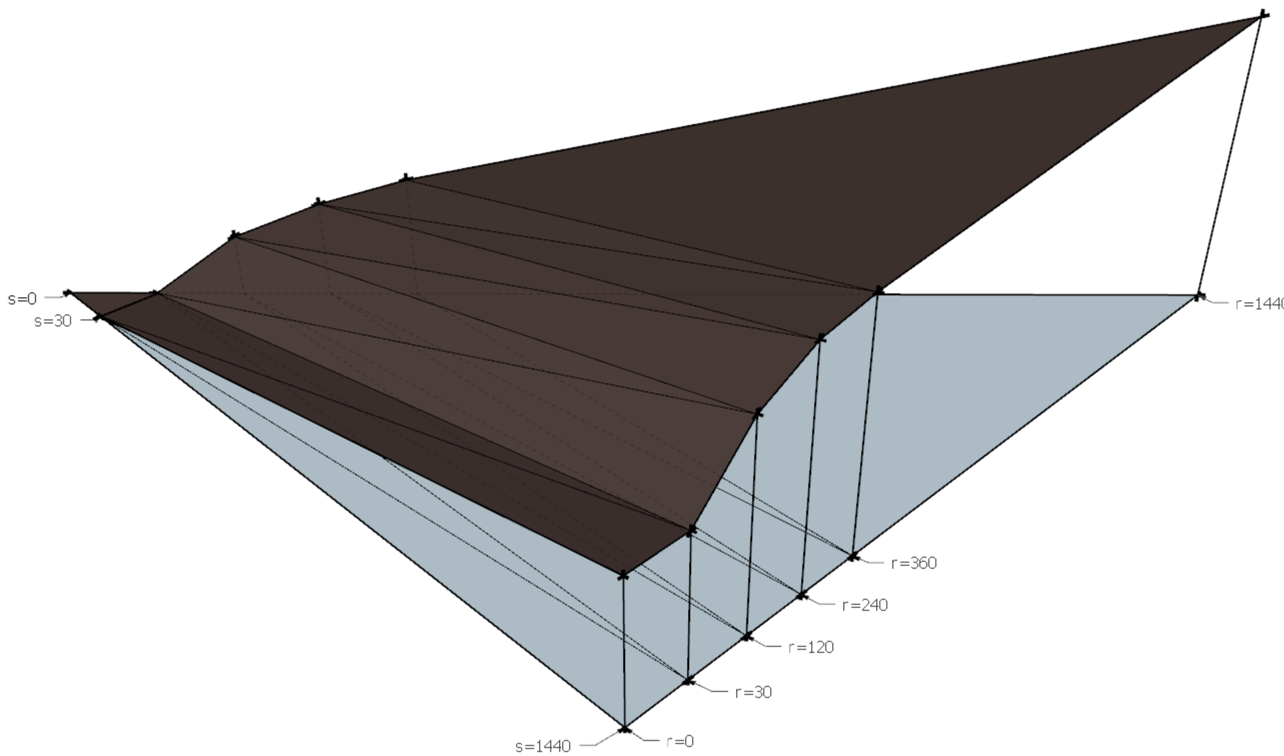


Figure 5.6: Visualization of the S-shaped cost function, evaluated for time points r and s. r represent time points for time-to-thrombolysis, t_i^T, while s represent for time-to-endovascular t_i^E.

The cost function is evaluated for certain values of time-to-thrombolysis and time-to-endovascular treatment. The function's surface is shown in Figure 5.6. Our model must again choose a combination of t_i^T and t_i^E in the same manner as described earlier. Only the number of points and the cost associated with each time point are changed, and therefore no additional changes in notation are required.

One-dimensional cost function

The cost of endovascular treatment is affected by the time-to-thrombolysis and the time-to-endovascular procedure. In order to capture this correlation, the model incorporates a two-dimensional cost function $C_{r,s}^E$. The linearization of this function adds several new variables and complexity to the model. A simplification would be to use a one-dimensional cost function C_s^E, only affected by the time-to-endovascular treatment. In order to generate a one-dimensional cost function, we sample the two-dimensional cost function where $t_i^T = t_i^E$. The objective function is changed accordingly:

$$\min \quad \alpha \cdot \left(P^T \cdot \sum_{i \in N} D_i^S \cdot C^T(t_i^T) + P^E \cdot \sum_{i \in N} D_i^S \cdot C^E(t_i^E) \right) + (1-\alpha) \cdot \left(\sum_{i \in N} D_i^G \cdot C^G(s_i, t_i^{CLST}) \right) \tag{5.45}$$

where $C^T(t_i^T)$ and $C^E(t_i^E)$ are linearized by implementing special ordered sets of type 2 (SOS2).

This makes the model disregard the correlation between the cost of time-to-endovascular treatment and the time to thrombolysis. We replace the three-dimensional variable δ_{irs} with two two-dimensional variables, which could improve the solution time. Since the original formulation consideres endovascular treatment to be eligable in only 10 % of the cases, the reduction of complexity by disregarding the two-dimensional cost function may not affect the result much.

5.4.3 Additional stations

The base model formulated in Section 5.3 is able to locate resources among the zones that already have a station. The model can be extended to allow for additional zones (candidate zones) to have a station. With this implementation new stations can be established. Existing stations are not required to be used. This means that one or several new stations can lead to existing station(s) being empty (closed down).

For this variant, the objective function is left unchanged. Let the set of existing station zones, $N^A \subset N$, where N is the set of all zones. Then, let N^C be the set of all other nodes, or candidate zones. That is, $N^C = N \setminus N^A$. The limit on how many new stations that are allowed is defined as L^C. Finally, a new variables is introduced:

$$o_j = \begin{cases} 1 & \text{if an ambulance station is placed in zone } j \\ 0 & \text{otherwise} \end{cases} \tag{5.46}$$

Ambulances cannot be placed in a zone j unless $o_j = 1$. This is formulated:

$$M \cdot o_j \geq x_j^{CTA} + y_j^A \qquad j \in N^A \tag{5.47}$$

where a "Big M"-formulation is used. A suitable $M = L^{CTA} + L^A$. The final restriction needed ensures that no more new stations than the set limit is possible:

$$\sum_{j \in N^C} o_j \leq L^C \tag{5.48}$$

Note that this variant easily could be modified to include helicopter bases as well.

"Wildcard heuristic" for CT vehicles

Opening for new station(s) present a considerable relaxation of the model. It does not get any easier by allowing several vehicle types (ordinary ambulances and stroke ambulances). Solving this to optimality can be a time consuming effort. A simple 2-step heuristic that deals with the complications caused by multiple vehicle types is therefore suggested.

Say we want to solve the model for L^A ordinary ambulances, L^{CTA} stroke ambulances, and L^C possible new stations:

1a: Solve model with $(L^A + L^{CTA})$ ordinary ambulances, and L^C free stations.

1b: Add new station zones to the set of station zones in the original model (with no free stations).

2: Solve original model with the new set of possible station zones, with L^A ambulances and L^{CTA} stroke ambulances.

Division of solution space

As discussed, complexity of the problem increases significantly as different types of vehicles and new stations can be utilized. An alternative approach to handle the complexity is to divide the problem into several smaller subproblems, such that the set consisting of all subproblems spans the entire solution space of the original problem. Each subproblem handles a small part of the total solution space, and is therefore able to solve more quickly. The best solution of all subproblems is the optimal solution for the total problem.

5.4.4 Ensuring that no stations has more ambulances at night than day

Assume that the ambulances work in two shifts, daytime and nighttime. Since daytime has the highest demand, the total number of ambulances on duty at daytime are always greater than at night. Therefore, it must be ensured that no station has more ambulances during night time, since this would lead to ambulances being moved before night, which is impractical.

To ensure this, let y_j^A*, $x_j^{CTA}*$ and y_h^H* be the optimal solution with the maximum number of ambulances and helicopters available. This resembles the daytime capacity. Then, when solving for night time solutions, y_j^A, x_j^{CTA} and y_h^H , add restrictions:

$$x_j^{CTA} \leq x_j^{CTA}* \qquad j \in N^A \tag{5.49}$$

$$y_j^{A} \leq y_j^{A}* \qquad j \in N^A \tag{5.50}$$

$$y_h^{H} = y_h^{H}* \qquad h \in N^H \tag{5.51}$$

Chapter 6

Simulation

In this chapter, the simulation program created for this thesis is described. The main components (as listed in Subsection 3.3.3) of the software will be described in detail. In Section 6.1, the initialization subprogram, which creates all necessary objects with their associated attributes is described. Then, the reasoning behind the setup of the timing subprogram, or the "driver" of the simulation, will be explained. Thirdly, the library subprogram, which generates emergencies and their attributes according to a probability distribution will be described. The bulk of the logic and functionality in the simulation is within the event subprogram. This part emulates the emergency operator in a real life situation. It receives emergencies from the library subprogram and based on the system state decides how the emergency should be handled in the best possible manner.

In developing this simulation software, a graphical software has been built in order to easily follow one emergency, or one emergency vehicle as the time passes. This has made debugging and verification simpler, hence making the software more reliable. Both a map and a "dashboard" approach will shortly be described and showed. Next, a simplified report generator is described. The final component of the simulation is the main program. In this section, it is showed how all components are put together to form the simulation program. Lastly, we will briefly highlight some of the most important assumptions that are made in the simulation software. The specific input data used to simulate our chosen case is described in the next chapter.

6.1 Initialization Subprogram - object oriented approach

Before the simulation can start, the environment in which it should run must be configured. For the particular problem in this thesis, this includes creating: zones, ambulance stations, helicopter bases, hospitals, ambulances (and CT ambulances), and helicopters. Additionally, all of these objects have attributes which are essential for the simulation. We will now describe the different objects and their attributes.

6.1.1 Zones

Each zone is created as a separate object. Each zone has a name and an unique ID associated to it. It also has a location given in latitude and longitude. Each zone has a population, which is later used in the emergency generator algorithm. Furthermore, for every zone there is information about its closest hospital, its closest stroke center (can be the same as closest hospital). This information contains the zone in which these hospitals are, and the travel distance both by ground and by air. Lastly, each zone has an array with distances (distance matrix) that contains distances (also both by ground and air) to all its nearest zones. Zones and distances are equal to the input to the optimization model.

6.1.2 Hospitals, ambulance stations and helicopter bases

Next, ambulance stations and helicopter bases are created. These objects are placed in a zone and enables ambulances and helicopters to be stationed in these zones. Both types contains information regarding the number of ambulances/helicopters belonging to that station/base.

In the same manner, hospitals are created, and placed in a zone. Hospitals can be created as stroke centers or regular hospitals. This enables patients to be transported to these zones and get treatment.

6.1.3 Emergency units

All emergency units, ambulances, helicopters and CT ambulances have some common features. As mentioned, they have a "home" station/base. Each unit has a name, a current location (latitude/longitude) and an origin and destination if they are dispatched. Additionally, they have a state (on/off), a status (at station, scrambling, enroute to patient, arrived at patient, etc.), duty hours, and the ability to be marked as responsible for an emergency.

Additionally, all units can move around and change their state and status depending on what they are doing. All units can be assigned an emergency, move to its location, pick up the patient, perform diagnosis, and transport the patient to a hospital. All units have the ability to be redirected.

6.1.4 Operator

The last crucial object needed to be initialized is the operator. While the emergency units can move around by themselves, they need someone to assign them missions. This is done by the operator. This object is explained in more details in Section 6.4.

6.1.5 Simulation clock

Lastly, the initialization subprogram sets up some more technical configuration. This includes creating and setting the simulation clock to 0. Next, the simulation clock and the timing subprogram will be described.

6.2 Simulation clock and Timing subprogram

As reviewed in Section 3.3.3, timing advance in DES can be handled in two ways. For this simulation program, a fixed-increment approach is chosen. The reasons for this choice is the following:

- "Always" shifting states. As mentioned, one of the drawbacks of this approach is the simulation of "empty" beats where no states are changed, thus leading to a slower simulation. However, when a large system with many objects is simulated, as it can be in this program, there is very little time in which nothing happens. For this simulation, there are so many changes in states, especially in high demand hours, that the number of updates on clock probably would be very similar with next-event based approach. If one are to conduct a qualitative evaluation on unit movement (that is to track on or more ambulances as they "do their job"), which is possible in the software, a time-slice approach like we have implemented makes it much easier to follow.

- Insignificant disjuncture. Another drawback is that the approach may lead to a disjuncture between event time and registered time in simulation. This issue is countered with reducing the fixed-increment size so that the disjuncture is insignificant. For this case, the increment size is set to equal the resolution of the input data (one minute). This means that the maximum disjuncture between an actual event an its simulated time is < 1 minute.

- Subjective increment size. Lastly, the approach is criticized because determining the increment size is a subjective decision, as changes in increment size might lead to different results. This is again resolved here by having a very low increment size, so that the consequences become much lower.

6.3 Library subprogram - Emergency generator

For every period (according to the clock increment size), all objects are updated according to their states. Also, for every period a check is conducted to see if an emergency has occurred. If so, we need to establish what priority the emergency has, and if it is a stroke. This is done by drawing two (or three) random numbers and check according to the probability distributions for these events. For these particular checks, a look-up table approach (described in the literature) is used. The algorithm is described in Figure 6.1.

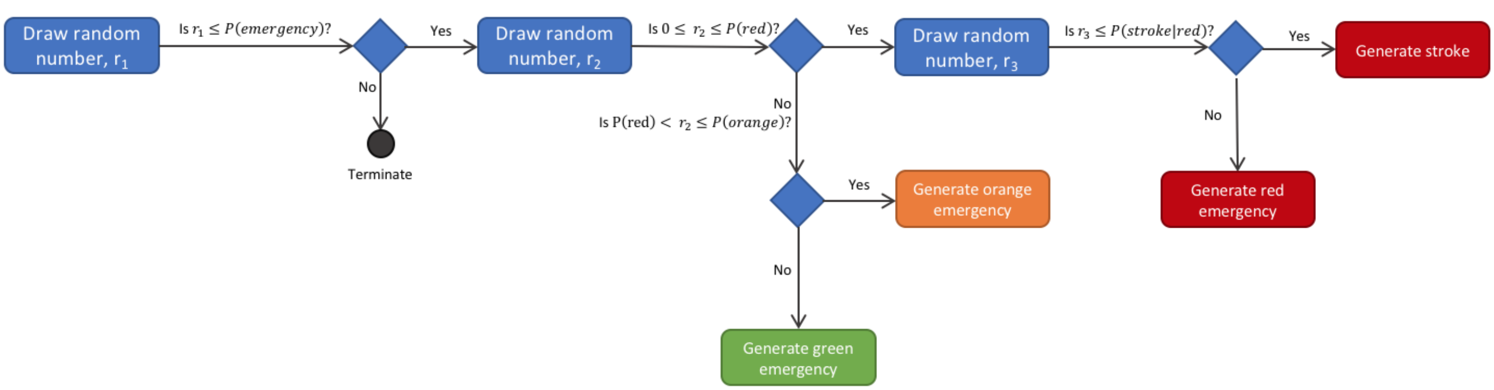


Figure 6.1: Emergency generation, step 1. For every period on the simulation clock, this emergency generation algorithm is run.

If an emergency has occurred, its attributes needs to be determined as well. Firstly, its location needs to be assigned. This is also done by drawing a random number, and using a look-up table. Secondly, different emergencies requires different amounts of time for the emergency personnel to conduct an initial diagnosis of the patient at the patients location. Similarly, different emergencies requires different amounts of time to clean, restock and prepare the ambulance for duty after the patient has been delivered to the hospital. Both of these times are drawn from a probability distribution function using the inverse transform method.

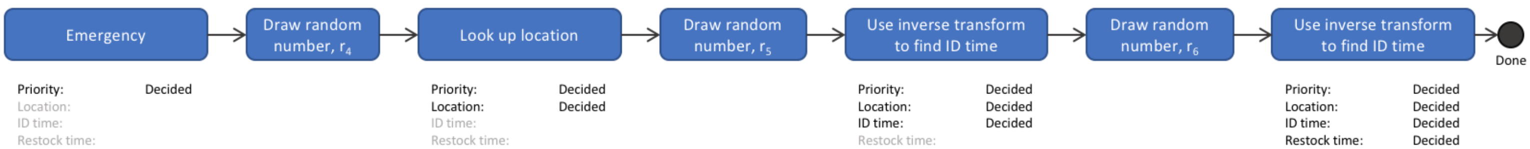


Figure 6.2: Emergency generation, step 2. Sequentially setting emergency attributes.

When the emergency is properly generated, it needs to be handled. This is done by the event subprogram, or what we call "the operator", which we now will describe.

6.4 Event Subprogram - Operator

As mentioned, emergency units are able to update their own status when an emergency is assigned to them. But in order to be assigned an emergency, someone needs to asses the different possibilities. In real life, the emergency central (AMK) does this. In the simulation program, an operator does this. When an emergency come about, the operator receives it.

6.4.1 Handling an emergency - step 1

The operator can access information regarding all emergency units, their current status, state and location. Different priorities are handled differently, and only the most acute emergencies ("red" priority) can receive help from the most scarce and valuable resources, like helicopters. CT ambulances can only be dispatched to strokes. The first steps when the operator receives an emergency is depicted in Figure 6.3.

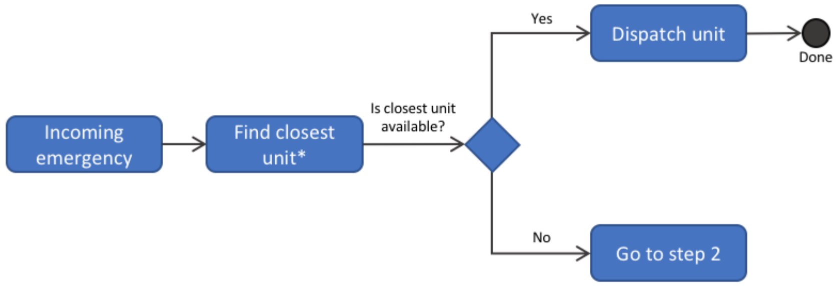


Figure 6.3: Operator step 1. When the operator receives an emergency, a basic check is run to see if the closest unit is available. If it is, this unit is dispatched. * Actually, if several units can respond in the same lowest time, all of these units are evaluated, and if only one is available, that one is dispatched.

6.4.2 Handling an emergency - step 2

If the incoming emergency's closest unit is available, this unit is given the responsibility for this emergency and dispatched to the patient. Note that this is not necessarily the case for red emergencies, as these can use more resources, like helicopter and CT ambulances. These resources do however require that the benefits/gain is good enough. More on this later in this section. If there are several units that have the same lowest response time, the algorithm automatically finds an available one, if any. If the closest unit(s) is not available, the operator needs to evaluate different options depending on the emergency's priority. This second step is depicted in Figure 6.4.

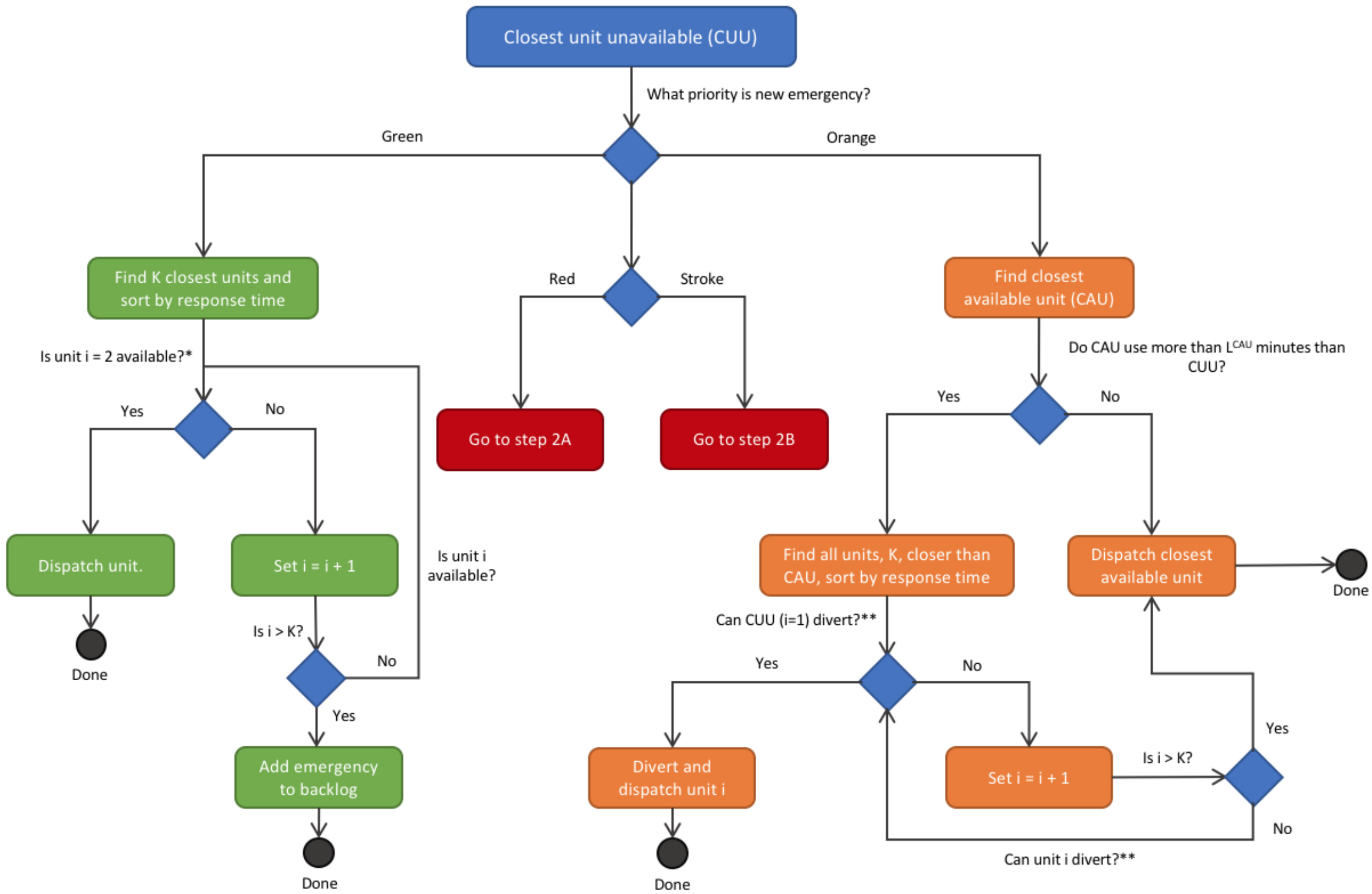


Figure 6.4: Operator step 2. If the closest unit is not available, the operator acts differently based on the emergency's priority. If regular emer * Unit i = 1 is already evaluated (and not available). ** Divert assessment is explained later in seperate figure.

If the emergency in question has green priority, this will never lead to any ongoing responses being cancelled (since they can't be lower prioritized). Instead, we evaluate the K closest units to the emergency location, and see if any of them are available. We start with the closest ones, and move further and further away. Without this limit on K units, available units from very far away may be dispatched. This will lead to several unwanted consequences: First and foremost, the dispatched ambulance may leave another area uncovered, leading to lower response time in its area. This may eventually lead to a "butterfly effect" where no ambulances are where they are supposed to be, leading to long response times. Secondly, another ambulance may finish its current mission and respond to the emergency before the available ambulance arrives, leading to a longer response time for that emergency. Lastly, ambulances will probably also use a long time to return to their station, possibly leading to issues concerning overtime and other employment regulatory aspects. Another variation of this would be to only consider units within a given range of the emergency location. As far as we understand[1], there are no absolute rules for these kind of situations, but the decision is based on a total assessment done by the operators at AMK. If none of the K closest ambulances are available, the emergency is sent to the "Emergency backlog".

For orange emergencies, we may want to consider aborting another ambulance (if it is on a green emergency). However, before we do that, we check if the closest available unit (CAU) is close. If the CAU is not more than L^{CAU} minutes away, this will be dispatched directly. By defining this constant, there is no point in considering aborting and diverting ambulances closer if the benefit is less than L^{CAU}. Moreover, we only consider diverting and aborting those ambulances closer than the CAU, as those farther away will always be worse than the CAU.

For red emergencies, more alternatives must be taken into consideration since a greater set of resources may be used. But before we go into this, we will explain the emergency backlog, and the divert check.

6.4.3 Emergency backlog

All the time, but especially in high demand hours, the operator needs to prioritize between different emergencies. This might lead to putting some emergencies on hold. When emergencies cannot be handled at once, they are put into a backlog. This is a list containing all emergencies that has not been delegated to an emergency unit. In some cases, if all units (according to the operator logic) are busy, the emergency is put on this list at once it is reported in. In other cases, an emergency with a lower priority is cancelled and the assigned unit is diverted and given a more important emergency. In these cases, the cancelled emergency is put into the backlog. The backlog is a sorted list based on priority and time occurred. This means that high priority emergencies that happened early makes the top of the list.

In most cases, the list will only contain green emergencies. The reason for this is that orange and red emergencies occurring will usually lead to a green emergency's unit being cancelled. Moreover, if the K closest units from a green emergency are busy, the emergency will be put on to the backlog, as described on Figure 6.4.

[1] After speaking with the personnel at Ambulansetjenesten St. Olavs

6.4.4 Diverting and aborting

As mentioned, units with an assigned emergency can be diverted. This means that their current emergency response is cancelled and they are sent to a more important assignment instead. Some criteria must be met in order for this to happen. These criteria, once again, depend on the priorities. The process is described in Figure 6.5.

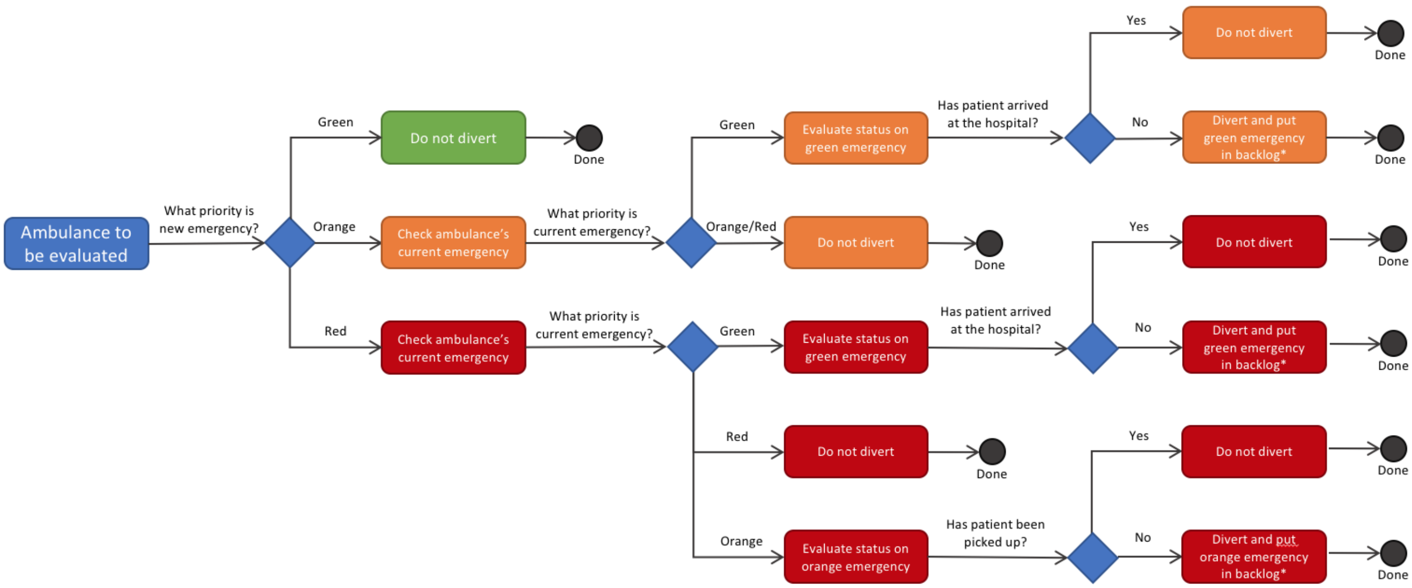


Figure 6.5: If a red or orange emergency occurs, both available and unavailable units will be considered to respond to the emergency, while green emergencies will only consider available units.

6.4.5 Handling red emergencies - step 2A/B

As mentioned, delegating red emergencies is a more complex matter. Since these are the most important emergencies to respond quickly to, there must be more options of cancelling occupied units, as described above. Additionally, these acute emergencies can be taken care of by helicopters (and in our thesis, CT ambulances). However, there must be some way of limiting the usage of helicopters, since this is a scarce resource. Also, a CT ambulance must only be dispatched for strokes where the reduced time to diagnosis is great enough.

For all red emergencies, the response time is of utmost importance, as minutes can be the difference between life and death. Therefore, the unit with the lowest response time (we have called this unit for the "preliminary best first responding unit", PBFRU), will always be dispatched as the first responder. The helicopter can only be "preliminary" since another constraint must be satisfied: the total time saved until the patient arrives to its closest hospital must be L^H or more minutes. Moreover, if an ambulance is considered the PBFRU, it must be assessed whether or not to dispatch a helicopter to meet up with the ambulance and transport the patient to the closest hospital. Again, the helicopter usage must be limited, and the benefit of this meet-approach must be greater than L^{A+H}. The decision tree is visualized in Figure 6.6.

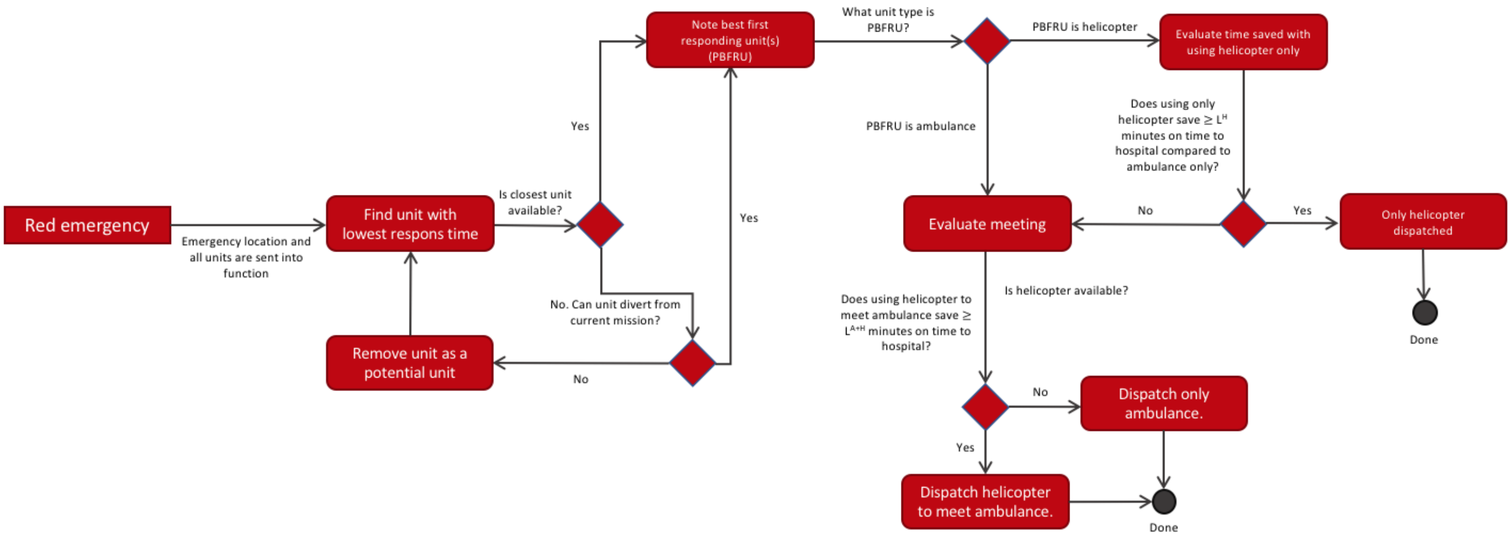


Figure 6.6: Flowchart of decisions made by the operator for a red emergency response.

For suspected strokes, a CT ambulance can also be utilized, either as a first responding unit or to meet up with an ordinary ambulance. This makes the decision of what unit(s) to respond intricate. The easiest path is when the CT ambulance is PBFRU, i.e. closest to the patient. If this is the case, the CT ambulance will respond, perform CT and start treatment for ischaemic stroke at site, and transport haemorraghic stroke patients to the closest hospital. If the patient has an ishcaemic stroke and is thrombectomy eligible, the patient will be transported directly to a stroke center. As with other red emergencies, there must be a benefit of using helicopter to meet the CT ambulance to transport the patient to a stroke center. Here, helicopter usage must lead to time savings of L^{CTA+H} minutes or more.

If the PBFRU is a helicopter, the decision is equal to those in other red emergencies: in order to use helicopter, one needs to save L^H minutes or more.

In most cases, PBFRU will be ordinary ambulances. This gives three alternative paths (or actually four): ordinary ambulance all the way, ambulance being met by a CT ambulance, and ambulance being met by a helicopter. In order for one of the two latter to happen, the best meeting alternative is evaulated first. This is the alternative that provides the shortest time to treatment. For the helicopter this means to transport the patient to a hospital and conduct CT scan there, and of the CT ambulance to meet up with the ambulance and start a prehospital CT scan. Once the best alternative of the two is decided, the benefit of using extra resources must be evaluated. The criteria for using helicopter is the same as before, L^H, and the criteria for using CT Ambulance to meet is that the time saved is L^{CTA} minutes or more. The scheme of events are depicted in Figure 6.7.

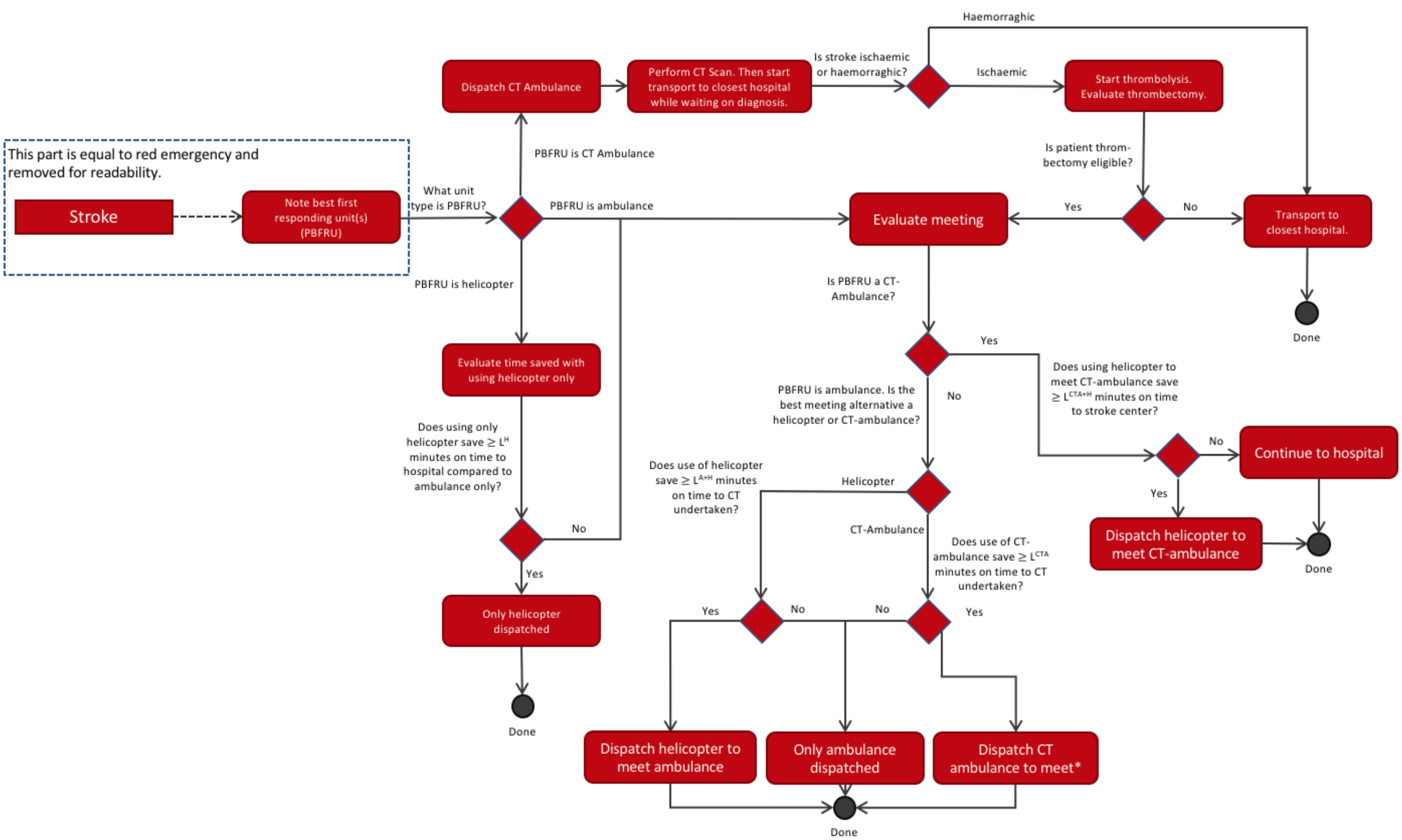


Figure 6.7: The use of CT ambulance makes the handling of strokes (or suspected strokes in real life) a more intricate process than other red emergencies. Without CT ambulances, strokes will be treated more or less similar to other red emergencies. * This path can be further split into two if the patient is thrombectomy eligible. If so, helicopter will once again be evaluated as a meeting unit to the CT ambulance to transport the patient directly to a stroke center.

The figure has some simplifications in order to keep the chart readable. Actually, once a CT ambulance meets up with an ordinary ambulance, this enables prehospital diagnosis. And if the patients turns out to be thrombectomy eligible, helicopter is again considered to meet up with the CT ambulance in order to transport the patient directly to a stroke center.

Moreover, the figure has another simplification. For those patients eligible for thrombectomy that are not diagnosed prehospitally and are taken to a *stroke unit*, helicopter will be used to transfer these patients to the stroke center.

6.5 Statistical Counter or Accumulator (debug mode and object attributes)

For the simulation to run effectively, the evolving performance of the system is not expressed as this will increase runtime significantly. However, the simulation can give feedback on performance continuously. This will enable and ease debugging, validation and verification of the system. For this simulation, there are two main ways this can be done. We call these for map mode, and dashboard mode.

6.5.1 Map mode

The map mode displays emergencies on a map as they occur. All units, both ambulances and helicopter are also plotted at their current locations. With this mode, one can follow a patient all the way from the time that the emergency occurs until the patient arrives at the hospital. A screenshot from the map mode is presented in Figure 6.8.

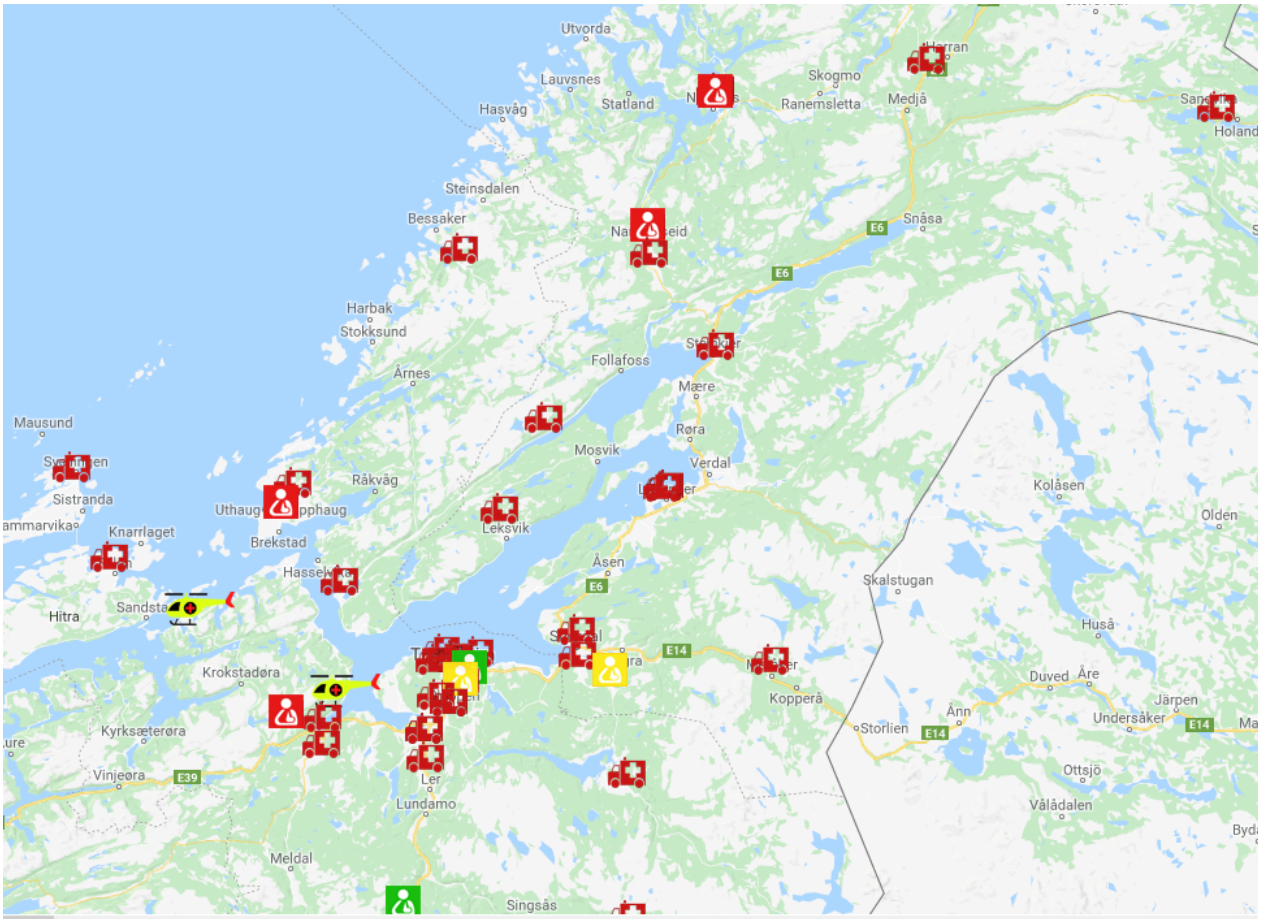


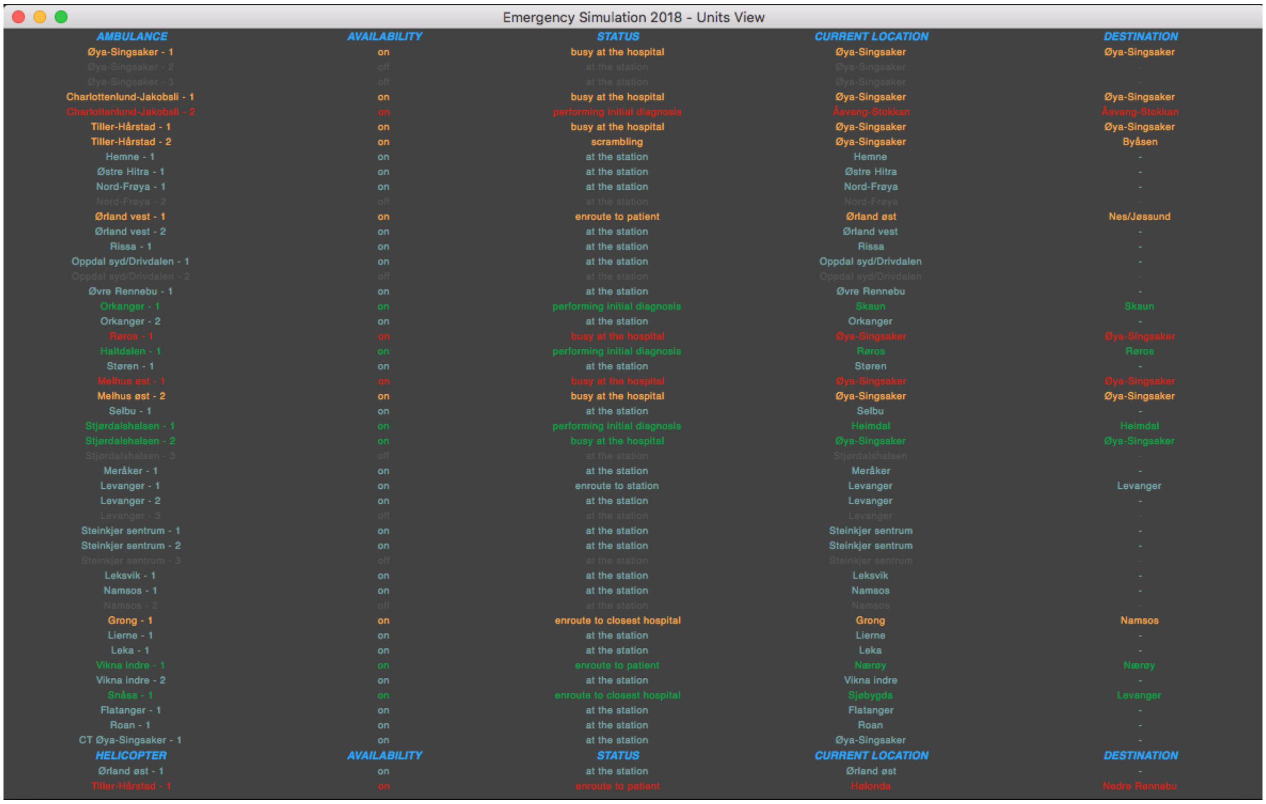
Figure 6.8: The Map mode: Ambulances and helicopters are responding to emergencies happening in Trøndelag. Map data ©2018 Google.

6.5.2 Dashboard mode

The dashboard mode has three different views. One shows all units and their current state (availability), status, position and destination. The second shows all emergencies as they happen, and gives information on its destination, responsible unit(s), priority and the most important times. The final view displays the simulation clock, the real world date/time and a summary of all emergencies.

Units view

In this window, all units are listed with their names in the first column. In the second column, the units availability is showed. If the unit is off duty, this will say "off". Next, the status of the unit is shown. The color of the line also indicates whether or not a unit is busy or available. Units standing by at the station or units returning to station are available, while units with green, orange or red colors are handling an emergency with the respective priorities. Finally, current location and their destination is listed.



Emergency Simulation 2018 - Units View

AMBULANCE	AVAILABILITY	STATUS	CURRENT LOCATION	DESTINATION
Øya-Singsaker - 1	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Øya-Singsaker - 2	off	at the station	Øya-Singsaker	-
Øya-Singsaker - 3	off	at the station	Øya-Singsaker	-
Charlottenlund-Jakobsli - 1	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Charlottenlund-Jakobsli - 2	on	performing initial diagnosis	Åsvang-Stokkan	Åsvang-Stokkan
Tiller-Hårstad - 1	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Tiller-Hårstad - 2	on	scrambling	Øya-Singsaker	Byåsen
Hemne - 1	on	at the station	Hemne	-
Østre Hitra - 1	on	at the station	Østre Hitra	-
Nord-Frøya - 1	on	at the station	Nord-Frøya	-
Nord-Frøya - 2	off	at the station	Nord-Frøya	-
Ørland vest - 1	on	enroute to patient	Ørland øst	Nes/Jøssund
Ørland vest - 2	on	at the station	Ørland vest	-
Rissa - 1	on	at the station	Rissa	-
Oppdal syd/Drivdalen - 1	on	at the station	Oppdal syd/Drivdalen	-
Oppdal syd/Drivdalen - 2	off	at the station	Oppdal syd/Drivdalen	-
Øvre Rennebu - 1	on	at the station	Øvre Rennebu	-
Orkanger - 1	on	performing initial diagnosis	Skaun	Skaun
Orkanger - 2	on	at the station	Orkanger	-
Røros - 1	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Haltdalen - 1	on	performing initial diagnosis	Røros	Røros
Støren - 1	on	at the station	Støren	-
Melhus øst - 1	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Melhus øst - 2	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Selbu - 1	on	at the station	Selbu	-
Stjørdalshalsen - 1	on	performing initial diagnosis	Heimdal	Heimdal
Stjørdalshalsen - 2	on	busy at the hospital	Øya-Singsaker	Øya-Singsaker
Stjørdalshalsen - 3	off	at the station	Stjørdalshalsen	-
Meråker - 1	on	at the station	Meråker	-
Levanger - 1	on	enroute to station	Levanger	Levanger
Levanger - 2	on	at the station	Levanger	-
Levanger - 3	off	at the station	Levanger	-
Steinkjer sentrum - 1	on	at the station	Steinkjer sentrum	-
Steinkjer sentrum - 2	on	at the station	Steinkjer sentrum	-
Steinkjer sentrum - 3	off	at the station	Steinkjer sentrum	-
Leksvik - 1	on	at the station	Leksvik	-
Namsos - 1	on	at the station	Namsos	-
Namsos - 2	off	at the station	Namsos	-
Grong - 1	on	enroute to closest hospital	Grong	Namsos
Lierne - 1	on	at the station	Lierne	-
Leka - 1	on	at the station	Leka	-
Vikna indre - 1	on	enroute to patient	Nærøy	Nærøy
Vikna indre - 2	on	at the station	Vikna indre	-
Snåsa - 1	on	enroute to closest hospital	Sjøbygda	Levanger
Flatanger - 1	on	at the station	Flatanger	-
Roan - 1	on	at the station	Roan	-
CT Øya-Singsaker - 1	on	at the station	Øya-Singsaker	-
HELICOPTER	AVAILABILITY	STATUS	CURRENT LOCATION	DESTINATION
Ørland øst - 1	on	at the station	Ørland øst	-
Tiller-Hårstad - 1	on	enroute to patient	Hølonda	Nedre Rennebu

Figure 6.9: The units view. Ambulances are listed on top, and helicopters in the bottom. CT ambulances will be indicated with "CT" in their names.

Emergencies view

While the units window shows the current situation from the units' perspective, the emergency window offers a different view. Here, each emergency is listed with all details necessary to verify that the emergency is handled correctly. For one, this shows the time the emergency was reported to the operator, where it happened, whether it is a stroke and the responsible units. If several units are listed on the same emergency, this indicates that a meet has found place between those units according to the rules described earlier. Lastly, some key metrics are listed: Time to closest hospital, the number of times a response to the emergency has been cancelled (if any), the first responding unit's scramble time, time spent at site for initial diagnosis, and lastly, the time needed for restocking/cleaning at the hospital for the responsible unit.

Emergency Simulation 2018 - Emergency View

Date/Time	Emergency Location	Stroke	Responsible Ambulance	Responsible Helicopter	Responsible CTAmbulance	TT CLST	Cancel count	Scramble time	IDTime	Time@Hosp
2018-01-01 08:51:00	Rosenborg-Møllenberg		Øya-Singsaker - 2				0	7	17	15
2018-01-01 08:47:00	Hemne	YES	Hemne - 1				0	0	18	4
2018-01-01 08:38:00	Buvik		Melhus øst - 1				0	0	31	125
2018-01-01 08:36:00	Namsos		Namsos - 2				0	8	26	51
2018-01-01 08:34:00	Stjørdal		Stjørdalshalsen - 1				0	5	33	47
2018-01-01 08:29:00	Ranheim		Charlottenlund-Jakobsli - 1				0	0	25	25
2018-01-01 08:26:00	Steinkjer sentrum		Steinkjer sentrum - 1				0	5	81	106
2018-01-01 08:19:00	Hundhamaren-Saksvik bygdelag		Charlottenlund-Jakobsli - 2				0	3	85	62
2018-01-01 08:18:00	Hundhamaren-Saksvik bygdelag		Tiller-Hårstad - 2				1	6	32	51
2018-01-01 08:14:00	Geitastrand og Orkdal vestre		Orkanger - 1				0	8	51	32
2018-01-01 08:11:00	Buvik		Melhus øst - 2				1	11	25	68
2018-01-01 07:59:00	Soknedal		Støren - 1				0	8	41	79
2018-01-01 07:55:00	Klæbu		Tiller-Hårstad - 1			49	0	0	13	33
2018-01-01 07:34:00	Namsos		Namsos - 1			42	0	5	34	48
2018-01-01 07:30:00	Berg-Tyholt		Øya-Singsaker - 1				0	26	52	33
2018-01-01 06:07:00	Hemne			Ørland øst - 1		69	0	5	35	23
2018-01-01 06:01:00	Oppdal syd/Drivdalen		Oppdal syd/Drivdalen - 1	Tiller-Hårstad - 1		81	0	5	9	59
2018-01-01 05:46:00	Steinkjer sentrum		Levanger - 1			142	0	4	73	33
2018-01-01 05:23:00	Charlottenlund-Jakobsli		Charlottenlund-Jakobsli - 1			44	0	3	29	43
2018-01-01 05:14:00	Strindheim		Tiller-Hårstad - 1			87	1	29	30	24
2018-01-01 03:56:00	Sparbu/Henning		Steinkjer sentrum - 1			108	0	5	68	16
2018-01-01 03:10:00	Byåsen		Øya-Singsaker - 1			34	0	4	13	54
2018-01-01 02:12:00	Skatval		Stjørdalshalsen - 1			55	0	3	17	45
2018-01-01 01:35:00	Namsos		Namsos - 1			21	0	17	1	62
2018-01-01 01:32:00	Flatåsen-Saupstad		Tiller-Hårstad - 1			68	0	38	15	92
2018-01-01 01:19:00	Ila-Trolla		Øya-Singsaker - 1			30	0	2	19	13
2018-01-01 01:17:00	Skogn		Levanger - 1			54	0	2	21	4

Figure 6.10: The emergency view lists all emergencies as they occur and continuously updates the information to reflect its current state as the simulation run.

Summary view

The final view in the dashboard, is the summary view. Here, the key figures are presented. The simulation clock count, the real world date/time and current emergency count. Lastly, the current number of emergencies in the backlog is showed.

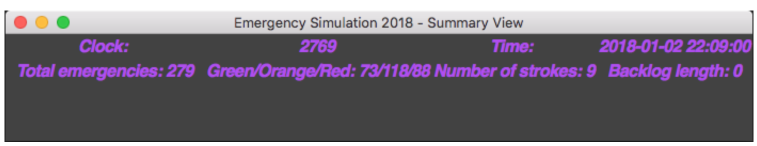


Figure 6.11: The summary view shows the most aggregated numbers and the simulation clock.

6.6 Report Generator - aggregating results

Finally, when the simulation is completed, the results need to be gathered and stored. With the object oriented approach, this is simply a function that go through every object and lists the metrics of study. For this simulation, iterating through all emergencies and get all details gives the necessary information. These details include amongst others, time emergency was reported, priority of emergency, unit(s) assigned to handled emergency, number of times emergency was cancelled, scramble time for units, response time, time to hospital, time to thrombolysis/thrombectomy (if stroke) and more.

Instead of aggregating the data in this step, all this "raw data" is stored, so that it can be analyzed at a later stage.

6.7 Main program

The main program runs and coordinates everything that happens in the simulation. When executed, the main program first calls the initialization subprogram. The initialization subprogram also sets the simulation clock to 0. This part was described in more detail in Section 6.1. When all objects are created, the simulation starts and goes on as long as the clock is within the configured time horizon.

For every "tick" on the simulation clock, the first part that executes is the library subprogram (generating emergencies). Next, all units have their states' updated, this means that every unit continue their current task, or finishes and starts the next. When all units are updated, the backlog is checked. If this contains any emergencies, these will be attempted to be assigned to available units. When this is done, the emergency created is handled by the operator according to the logic described earlier. If the emergency generator do not return any emergency, this part is skipped. Lastly, the main program updates the simulation clock, and starts over again until the while condition (`clock < simulationHorizon` returns false.

Algorithm 1 main.py

```
function MAIN
    initializationSubprogram()
    while clock < simulationHorizon do
        emergency = emergencyGenerator()
        for unit in allUnits do
            unit.updateState()
        operator.checkBacklog()
        if emergency then
            operator.handleEmergency(emergency)
        clock = clock + clockIncrement
    createResultsReport()
```

6.7.1 Assumptions made in the simulation

Only "normal" emergencies

Ambulances will in reality have many assignments that are non-time-critical emergencies, including transport of patients between hospitals, from hospital to nursing homes, and other forms of transport assignments where the patients are bedridden or are in need for someway of medical care. These assignments are treated as green emergencies in real life. However, we have not included these kinds of emergencies. This is for two reasons:
- Firstly, the data we have is regarding this matter somewhat inconclusive, hence it is important to know its prevalence.
- Secondly, we are mostly interested in studying more time critical emergencies (red mainly, but also orange). Since these will have priority over green emergencies, they will not be affected very much by the green ones. As will be described in the next chapter, we have still decided to use the "full" number of green emergencies. This is to get more realistic availability times on ambulances.

Helicopters on red emergencies only

As shown while explaining the emergency handling, helicopters can only be dispatched on red emergencies. This is for the most part correct, but in real life there are some other occasions where helicopter can be used for orange and green emergencies. This is mainly when the area is unreachable for ambulances.

Hospital ferrying done with helicopters

As mentioned, thrombectomy-eligible patients that are diagnosed at a stroke unit, will be transferred to a stroke center by helicopter. Ambulances can not ferry patients between hospitals. Therefore, if the simulation is run without helicopters, thrombectomy-eligible patients will not be taken to a stroke center.

No upgrades/downgrades of priority

When an emergency is called in in real life, there is a tendency to rather treat the emergency as more important than what it really is, e.g. a green emergency can be categorized as an orange. This leads to a higher priority response. Usually, these emergencies are downgraded when the emergency personnel reach the patients. The opposite can also happen, the emergency operator can have a hard time realizing the gravity of the situation and therefore treating it as less important than what it really is. These events are measured, and are fortunately rare, and have therefore not been included.

Chapter 7

Input data and simulation validation

In this chapter we will start with describing the input data used for the area studied in the computational study chapter, and how the data is obtained or constructed. A lot of this data is common for both the optimization model and simulation program, while some data differs among the two. In the second section we will briefly present the schedule of today's ambulance location in Trøndelag as well as their time schedules. This current allocation will be used as one of the benchmarks later in the thesis. Finally, we will also use the current allocation to validate the input data. This part will also serve as a superficial verification of the simulation logic.

Appendix A.2.1 contains some more details regarding the area in question.

7.1 Describing the input data

The zones, their demand (population) and distances are static information that are used in both the optimization and simulation. Further, the optimization model must have some parameters and sets to run. Some of these will be briefly explained, while all parameters are listed in Appendix A.1. Since the simulation enables stochasticity and is more dynamic, we need to draw some random events. The distributions and parameters for these are listed in the end.

7.1.1 Common input data

Zones and demand

In order to capture the distribution of demand based on a patient's location, we have utilized public available information from The Norwegian Mapping Authority. They offer data sets including the borders of statistical units called "Grunnkretser". These are sections dividing a municipality into several smaller areas. Information about the number of people living in each "Grunnkrets" is available from Statistics Norway, the national statistical institute of Norway. The demand for emergency services is directly based on population.

However, it became clear that these divisions were too detailed considering the total area of land we want to cover in this thesis. In order to reduce the level of detail, we programmed a Python script and utilized GIS software to merge each "Grunnkrets" into larger divisions called "Delområder". One weakness discovered and discussed in Aarnseth and Hov (2017) is that the zones' geographical center points are used as the point location of the demand. This leads to some examples where the center location are far away from any populated area at all, as shown in Figure 7.1. Consequentially, the automatic gathering of distances we have developed recorded some very long travel times. In this thesis, we have manually moved the center points to represent the location where most people live. This will give *more* exact results, since these points are used to calculate travel times for both helicopters and ambulances.

All 139 zones, their population and coordinates are listed in Appendix A.2.2.

Figure 7.1: Some center points are placed on top of mountains far away from populated areas. Possible better alternative positions are indicated as red circles in the figure.

Lastly, we have manually categorized all zones as either urban or rural in order to implement differentiated response restrictions for the zones.

Road distances

Based on the zones' point locations mentioned above, a program was developed to gather and store driving distance and time from Google maps. From zones that contains hospitals, distance to all other zones were collected. From an ordinary zone without any hospital, the distances to all neighbouring zones within a 100 kilometer radius (by air) was collected. This limitation had to be done in order to not violate Google's terms regarding limitations on number of distances to collect. Also, in real life, responses longer than this for ambulances will only happen in special circumstances (typically moving patients between hospitals), and will not be included in this thesis. In total, approximately 20 000 distances and driving times were collected.

In Aarnseth and Hov (2017) the driving times are adjusted with a factor of 0.8, assuming that emergency vehicles will drive 20 % faster than normal traffic. As it turns out, the average speed of ambulances is pretty close to normal traffic on average.[1] Therefore, in this thesis we have not adjusted the travel times.

[1]According to the Ambulance Service at St. Olav Hospital

Air distances

To reduce memory usage and run time, we are interested in reducing the number of distances calculated and stored. If all distances between every zone had been calculated and store we would have ended up with a NxN matrix, where N is the set of all zones. Regarding air distances, in this thesis, we assume that a helicopter will never travel between two zones if neither zone has a hospital nor a helicopter base present. We create a subset, M, consisting of the zones with a helicopter base or a hospital and calculate the air distances between all zones, N, and the zones in the subset M. The air distance is approximated by using the Spherical law of cosines:

$$cosc = cosa \cdot cosb + sina \cdot sinb \cdot C \tag{7.1}$$

where C is the angle (difference in longitude) in radians between the two zones, a and b is the respective distances in radians from the pole to the zones, and c is the distance we need between the zones, used to calculate T^H_{ji}. The distance D_{ij} between zone i and j can be calculated:

$$D_{ij} = \arccos(cosc) = \arccos(cosa \cdot cosb + sina \cdot sinb \cdot C) \tag{7.2}$$

As with the road distances, we reduce the distances calculated and stored to those within a 500 kilometer radius. Note that in the simulation software, a helicopter can be diverted while being in the air, on the move from a hospital back to its base. This is a distance that most likely is not calculated in advance, but by using Equation 7.2 we are able to calculate distances on the fly. We assume an air-to-ground speed for helicopters on 200 kilometers per hour.

7.1.2 Optimization model only

Candidate lists - ambulance stations and helicopter bases

If not otherwise specified, today's ambulance station locations are used as possible zones for stations in our optimization model. These zones are listed in Table 7.1. As for helicopters, today's two bases, Tiller/Rosten and Ørland, are possible locations. We have also added the zones with hospitals in Namsos, Levanger and Orkanger as possible helicopter base candidate zones.

Ambulance and helicopter capacity

Since the optimization model has no availability concerns, we need some way of restricting their usage. This is described in Restrictions 5.16 and 5.17 for ambulances and CT ambulances respectively. There are many approaches to defining the demand and capacities in these constraints. For this instance, any ambulance is assumed to be available for assignment 50 % of its time. The rest of the time is used on maintenance, transport assignments, training etc.

In a year, there are 525 600 minutes. If we assume that a 24/7 operational ambulance has an efficiency of 50 %, this means that it is actually only available in 262 800 minutes/year. We therefore set:

$$K^A = K^{CTA} = 262800$$

According to figures from 2016, there are about 130 emergencies per 1000 inhabitant in Norway. (SSB, 2016b)

For a zone, the "total consumed time/emergency" for a mission is defined as: time to closest hospital (i.e. scramble time + transport time to patient + time at scene + time to hospital) + time to restock at hospital.

The consumed capacity for a zone, i, in a year if it is covered by an ambulance from zone j, is then:

$$D_{ij}^A = (T^{SA} + T_{ji}^A + T^{ID} + T_i^{CLST-A} + T^{RA}) * \frac{P_i}{1000} * 130$$

where T^{RA} is the average restock time needed for ambulances when they have delivered a patient to prepare the ambulance for new assignments, and P_i is the population on zone i. The total capacity for a station is then: ambulances * 262 800 minutes/year

For helicopters, the capacity per helicopter is set to 90 000 minutes/year, or 1500 hours/year. Scramble time, initial diagnosis or swap time, and restock time at the hospital will equal around 50 % of this capacity, which leaves about 750 hours in airtime. This is approximately the same amount of flight time the helicopter ambulances uses. (SNL, 2013)

Meeting places

Since the optimization model assumes that helicopters and ambulances are always available at their station/base, optimal meeting places can be determined in advance and specified as input data. For any combination of an ambulance station and a helicopter base, one can find the optimal place to meet such that the transportation time to hospital is the minimum. If several zones gives the same minimal transportation time, one is arbitrarily chosen.

7.1.3 Simulation input data

Most simulation input data is based on the data collected by Ambulansetjenesten in 2016. In order to avoid overfitting, most numbers, such as scramble time, time at patients location, time needed for restocking at hospital are based on approximated probability distributions. For helicopters, all times are set to constants.

Emergency generation

Emergencies are generated according to the explanation in Section 6.3, with priority probabilties: $P(\text{green}) = 28\%$, $P(\text{orange}) = 41\%$ and $P(\text{red}) = 31\%$. For red emergencies, the probability of a stroke is set to $P(\text{stroke}|\text{red}) = 10\%$. For strokes, the probability of it being an ischaemic is set to $P(\text{ischaemic}|\text{stroke}) = 90\%$, and the portion of these that are eligible to thrombectomy is $P(\text{trombectomy eligible}|\text{ishcaemic stroke}) = 10\%$

Scramble time

Scramble time is drawn from a separate truncated normal distribution function for each emergency priority, $t^{SC-A} \sim N(\mu, \sigma)$, with parameters and limits (truncation):

$$\begin{array}{ccc} \mu_{green} = 5 & \mu_{orange} = 2 & \mu_{red} = 0.7 \\ \sigma_{green} = 20 & \sigma_{orange} = 3 & \sigma_{red} = 1.5 \\ \text{Limits} = [1, 60] & \text{Limits} = [0, 30] & \text{Limits} = [0, 12] \end{array}$$

Initial diagnosis time

Time at patients location, or initial diagnosis time, for the different priorities are drawn from a truncated gamma distribution, $t^{ID} \sim \Gamma(\alpha, \theta)$:

$$\begin{array}{ccc} \alpha_{green} = 1.5 & \alpha_{orange} = 2 & \alpha_{red} = 3 \\ \theta_{green} = 13 & \theta_{orange} = 10 & \theta_{red} = 7 \\ \text{Limits} = [1, 70] & \text{Limits} = [1, 70] & \text{Limits} = [1, 70] \end{array}$$

Time at hospital

Time spent at the hospital for ambulances after the patient is handed over is also drawn from a truncated normal distribution function, $t^{SC-A} \sim N(\mu, \sigma)$, with parameters and limits (truncation):

$$\mu = 30$$
$$\sigma = 40$$
$$\text{Limits} = [10, 250]$$

This is the same for all emergencies.

7.2 Current case

Here, today's ambulance and helicopter distribution is presented. This allocation is run in the simulation, and this is used to validate the simulation program in the next section.

For the most part, all stations have one or two ambulances that are on duty 24/7. Some stations have extra capacity during daytime. The schedule for these are shown in the table. The stations in Trondheim (St. Olavs, Ranheim and Rosten) are in real life treated as one pool of ambulances and do not necessarily return to their station when available. We have therefore simply distributed them in this manner to get an "average" location.

Table 7.1: Current allocation of ambulances and helicopters. * indicates that this is a rescue helicopter. This was discussed in the Background Chapter.

Station/Base	Ambulances		
in zone	24/7	Extra	Helicopters
Melhus	1		
Støren	1		
Holtålen	1		
Røros	1		
Oppdal	1	1 (09:00-18:00)	
Rennebu	1		
Orkdal	2		
Hemne	1		
Hitra	1		
Frøya	1	1 (09:00-18:00)	
Ørland	2		1*
Rissa	1		
Åfjord	1		
Selbu	1		
Tydal	1		
St. Olavs	1	2 (07:00-16:00)	
Ranheim	1	1 (07:00-19:00)	
Rosten	1	1 (07:00-19:00)	1
Vikna	2	1 (09:00-18:00)	
Leka	1		
Namsskogan	1		
Lierne	1		
Grong	2		
Snåsa	1		
Flatanger	1		
Roan	1	1 (09:00-18:00)	
Namsos	2		
Verran	1		
Steinkjer	2		
Levanger	2	1 (09:00-18:00)	
Leksvik	1		
Meråker	1		
Stjørdal	2		

7.3 Validating the simulation program

The purpose of the simulation program is to test and evaluate the solutions given from the optimization. In order to do this, the simulation in itself must be verified and validated. In Table 7.2 the average numbers for the simulation is compared to the real data for 2017 for Trøndelag. The simulation has been run over 5 years, and the numbers are the average over these years. Meanwhile, the real data from the Ambulance does contain some incorrect reporting. Extreme values have been removed in order to minimize the impact of these.

Table 7.2: Simulation aggregated results compared to real data results.

Emergencies	Metric	Simulation 5 years averages	Real data 2017	Deviation
All emergencies	Number of records	51504	51696	-0.37 %
	Scramble time	6.087	5.61	8.44 %
	Response time	15.43	16.64	-7.27 %
	Initial diagnosis time	19.82	19.80	0.10 %
	Time to hospital	77.73	84.72	-8.25 %
	Time at hospital	51.41	51.11	0.59 %
	Total time	129.14	128.40	0.58 %
RED	Number of records	16005.4	16031	-0.16 %
	Scramble time	1.59	1.56	2.19 %
	Response time	11.836	10.55	12.19 %
	Initial diagnosis time	20.844	21.91	-4.87 %
	Time to hospital	60.71	64.75	-6.24 %
	Time at hospital	51.47	52.49	-1.94 %
	Total time	112.18	106.10	5.73 %
ORANGE	Number of records	20989.4	21119	-0.61 %
	Scramble time	2.96	3.11	-4.91 %
	Response time	15.033	14.77	1.78 %
	Initial diagnosis time	19.646	18.23	7.79 %
	Time to hospital	71.57	80.47	-11.06 %
	Time at hospital	51.31	53.85	-4.71 %
	Total time	122.88	127.60	-3.70 %
GREEN	Number of records	14509.2	14546	-0.25 %
	Scramble time	15.19	16.25	-6.52 %
	Response time	19.674	27.26	-27.83 %
	Initial diagnosis time	18.884	20.03	-5.73 %
	Time to hospital	103.97	109.49	-5.04 %
	Time at hospital	51.48	46.32	11.14 %
	Total time	155.45	154.20	0.81 %

The simulation software has been validated in greater detail, but only these aggregated results are presented. These numbers in Table 7.2 show that the simulation program gives almost all these aggregated averages below 10 % deviation from the real data. Furthermore, since the real data probably still has some incorrectly reported data registered, any more adjustments to the simulation could lead to overfitting. Furthermore, Figure 7.2 shows that the number of emergencies generated are close to the real data.

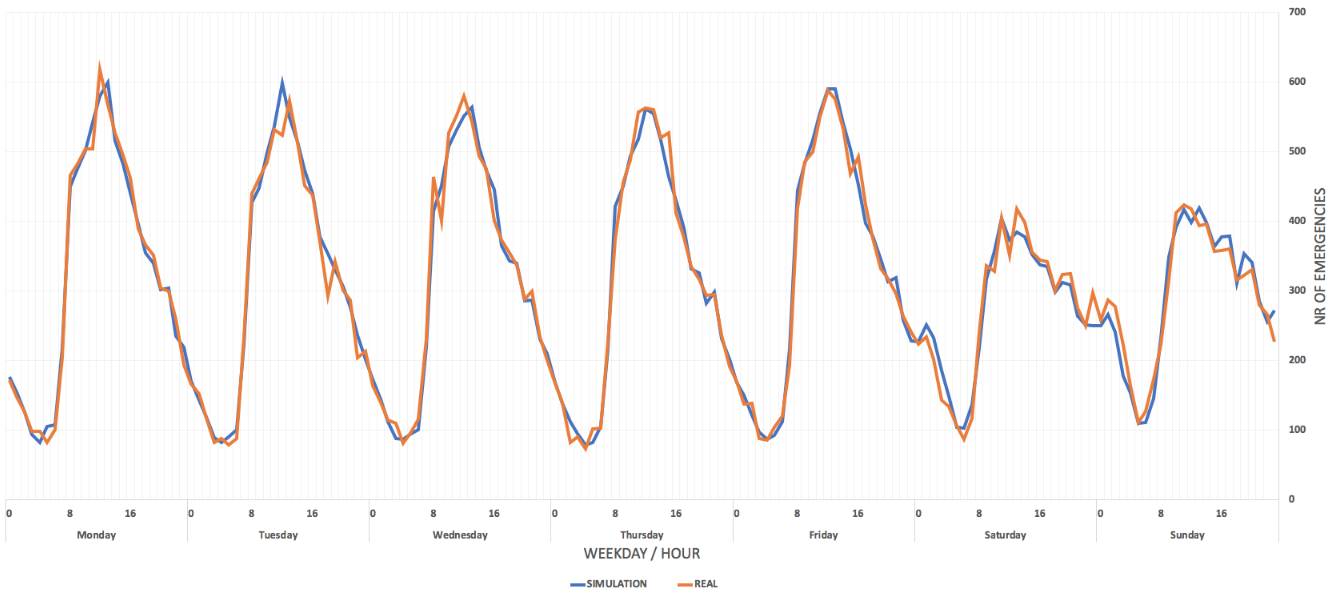


Figure 7.2: A comparison of real and simulated emergency per weekday and hour.

For this thesis, we will mainly use the simulation to evaluate the different results from the optimization models. Any potential mistakes in the simulation will most likely have somewhat the same effect on the different solutions.

Chapter 8

Computational study

This chapter starts with determining how the model should prioritize the stroke and general parts of the objective function. In Section 8.2, different solutions produced by the optimization when varying the number of ambulances and helicopters, are evaluated using simulation. After this evaluation is presented, the different alternative cost functions described in Subsection 5.4.2 are evaluated. In Section 8.4 the effects of introducing CT ambulances on both the optimization model and the simulation are presented. This section is followed by a study on how establishment of new ambulance stations affect the times to different treatments. Section 8.6 ends this chapter by evaluating how new technologies have a potential to significantly decrease the time to stroke treatment. In this section a method for solving computationally hard and time consuming problems by running multiple instances in parallel is described. This method has been utilized in many instances presented in this chapter. Scenarios that have been solved using this method are marked with * in the solution time column in the result tables. Software and hardware information is described in Appendix B.1.

When the simulation validation was presented above, each emergency unit had its own schedule based on real data. Some of them were active 24 hours a day, while others were active only during the daytime. In order to enable schedules and different day/night capacities the optimization must be run twice. This was discussed in Subsection 5.4.4. As a simplification, a full capacity run of the current scenario is solved and used as a benchmark instead of the current allocation. This means that all ambulances are set to be on duty 24/7 using the current allocation. The results of the two compared to each other are shown in Table 8.1.

Table 8.1: Simulation results for current allocation with (C) and without schedule (C24). Average response times for all emergencies show the biggest discrepancy among the two. For red emergencies (including stroke) there are only minor differences.

	Response time		Time to hospital		Stroke			
	All	Red	All	Red	$\tilde{t}^T$	$\tilde{t}^E$	$\tilde{t}^{RSP}$ <12 min	$\tilde{t}^{RSP}$ <25 min
C	35.36	14.28	87.02	59.77	90.49	110.9	57.19 %	87.36 %
C24	30.78	13.78	82.54	59.14	89.48	110.1	59.80 %	88.21 %

In order to evaluate the performance of different solutions, several performance measures are discussed. The measures are divided into three categories: response time, time to hospital and time to stroke treatment. All times are given in minutes. The times produced by the optimization are a weighted arithmetic mean based on the population in each zone. These are indicated with the same variables used in the optimization model. The times given by the simulation are an arithmetic mean of all emergencies. These are sometimes indicated with variables marked with tilde. Since the emergencies in the simulation are generated based on population, the arithmetic mean is comparable with the weighted arithmetic mean from the optimization.

8.1 Evaluating weighting between stroke and general cost function in the objective function

In this section, we study how the weighting among the two terms in the objective function affects the results. It must be pointed out that the terms are not independent of each other. Both stroke patients and other patients benefit from quicker response, and minimal time to treatment/hospital. There are some particular cases when these terms can be somewhat contradictory. This happens when we introduce CT ambulances. If CT ambulances are available and there is an emphasis on the stroke part of the objective function, response time and time to hospital might increase. This is because some areas will rather be covered by a CT ambulance that in return will provide faster treatment for stroke patients, but will not respond to non-stroke patients. Nonetheless, this evaluation is based on using different weights on the optimization of ordinary ambulances only. The effects of having CT ambulances in the fleet are discussed in Section 8.4.

Before evaluating the weights, the contribution to the objective value that each term generate needs to be balanced so that each term contribute with 50 % if the stroke weight, α, is set to 0.5. This balancing is done by forcing one solution (here, the C-24 solution is used) and solving with stroke weights 0 and 1. This removes the cost of stroke treatment and the cost of general response respectively, from the objective function. The two objective values can then be compared. The difference among the two is then used as a constant for one of the terms to balance the objective. This leads to equal contribution to the objective from both terms when $\alpha = 0.5$. That way different weights can be evaluated, knowing that the objective function is balanced.

Table 8.2: Results from the optimization model with different stroke weights. When running the same number of ambulances and helicopters, 50 and 2 respectively, there are only minor differences in solutions and results when changing the stroke weight.

Stroke weight α	MIP solution	Solution time [s]	t^T	t^E	t^{RSP}	t^{CLST}
0	36820398	29	93.539	110.961	14.090	63.539
0.2	36830808	32	93.494	110.916	14.136	63.494
0.4	36840475	38	93.494	110.916	14.136	63.494
0.5	36845308	35	93.494	110.916	14.136	63.494
0.6	36850142	55	93.494	110.916	14.136	63.494
0.8	36859809	69	93.494	110.916	14.136	63.494
1	36865448	473	93.408	110.870	14.265	63.408

As seen in Table 8.2, although the MIP solution value differs slightly for different weights, the practical differences in average response time and time to treatments/hospital are minimal. The figures also show a strong link between the stroke weight and the solution time. This can be explained by the concave cost function of costs related to stroke patients. This type of cost function makes the optimization much harder than the linear dependency for general response costs.

When the stroke weight α is set to 0, the model only considers the cost of general response time and time to hospital. In this case, all but one of today's existing stations are utilized. This makes sense, as having ambulances located in rural areas will reduce response time for the patients living there. However, when the model only considers the cost of stroke treatment ($\alpha = 1$) two stations are closed in zones where the population is rather sparse (in Tydal and Leka). Instead the ambulances are moved to Vikna and Verran. These are higher populated zones that lie far away from the closest stroke center, zones that would benefit from having ambulances ready to commence the stroke treatment fast.

Weighting the individual parts of the objective is in itself a difficult task, and the effects are dependent on the input data, especially the cost function parameters. For the rest of this chapter, 0.5 is used and kept constant for practical purposes. That way, response time and time to stroke treatment are prioritized equally in the objective function.

8.2 Using simulation to evaluate the optimization model

In chapter 7, it was showed that the simulation provides results that are reasonably similar to real data when the real ambulance allocation is run. Now, the allocations given by the optimization model will be simulated. Several different scenarios are put into the optimization and the optimal allocation is then used as input in the simulation. As mentioned above the allocation produced describe units operating around the clock, i.e. there are no shifts. Since these solutions are only compared to each other, the premises are equal for all scenarios.

The scenarios tested are all combinations of 40, 45, and 50 ambulances, with 0, 1, and 2 helicopters. The results are shown in Table 8.3.

For the optimization model, the objective value is used to evaluate the results. Meanwhile, the simulation program gives no aggregated number that can be used to measure the performance. Therefore, several numbers are used to compare the performance, and each of them are discussed.

Table 8.3: Comparing optimization and simulation results. X: Time to thrombectomy in the simulation will be extremely high when there are no helicopters, since only those transported directly to a stroke center will receive thrombectomy.

Settings		**Optimiziation**							**Simulation**					
Nr of units		Solution			Variables weighted averages				Response time		Time to closest hospital		Stroke specific	
A	H	LP sol.	MIP sol.	Time [s]	t^T	t^E	t^{CLST}	t^{RSP}	All	Red	All	Red	$\tilde{t}^T$	$\tilde{t}^E$
50	2	21425716	36845308	57	93.49	110.92	63.49	14.14	24.73	13.55	76.57	59.21	90.15	111.1
45	2	21425716	36928218	199	93.61	111.04	63.61	14.28	27.05	14.04	78.77	59.64	90.27	109.5
40	2	21524891	37143927	64	94.12	111.69	64.12	14.54	29.36	14.66	81.05	59.92	91.01	108.6
50	1	22101835	37726105	33	95.57	113.14	65.57	14.31	25.06	13.78	77.51	61.15	91.8	118.2
45	1	22101835	37826818	82	95.73	113.31	65.73	14.47	27.41	14.43	79.84	61.76	93.29	118.2
40	1	22276706	38190191	228	96.37	113.89	66.37	15.07	30.55	15.26	82.68	62.33	92.84	119.9
50	0	22980581	38902869	23	98.45	116.13	68.45	14.31	25.9	14.57	79.61	66.13	95.01	X
45	0	22980581	39092719	49	98.76	116.44	68.76	14.62	28.29	15.25	81.82	66.82	95.83	X
40	0	23204920	39487841	58	99.43	117.10	69.43	15.29	31.47	16.51	85.08	68.34	97.81	X

The MIP solutions show that the optimization model significantly favors helicopters more than ambulances, as sacrificing 10 ambulances for one helicopter gives improved objective value. Meanwhile, others will not be transported after thrombolysis is given.

In order to provide a measurement on the similiarity on the objective value (MIP solution) and the numbers from the simulation, correlations are calculated. The correlations are shown in Table 8.4.

Table 8.4: Correlation MIP solution and simulation results. ** $\tilde{t}^E$ correlation are only based on those instances with helicopters.

Simulation metric (i)	Correlation(MIP, i)
Avg. response time	0.4682
Avg. response time (red)	0.8009
Avg. time to hospital	0.7153
Avg. time to hospital (red)	0.9856
Avg. time-to-thrombolysis, $\tilde{t}^T$	0.9801
Avg. time-to-thrombectomy, $\tilde{t}^E$	0.9334**

All four averages from the simulation have a positive correlation with the objective value, as one would hope. Moreover, the averages for red emergencies show significantly higher correlations with the objective values than the averages for all emergencies regardless of priority. These results can be explained by considering the assumptions for the optimization model and how red emergencies are prioritized: in the optimization model, all units are assumed to be available at all times since we optimize with regards to red emergencies. For all other emergencies, this represent a best-case outcome. In the simulation, red emergencies will in most cases be responded to immediately and their expected scramble times will be lower. However, all other emergencies will have a longer response time, and can be delayed.

It is also worth noticing that the simulated response time and time to hospital for red emergencies, as well as the time to thrombolysis/thrombectomy are very similar to those in the optimal solution in the optimization model.

These results show that improved objective values give improved simulation performance, which strengthens the optimization model's reliability and credibility.

Practical implications of changing the number of units

When the number of ambulances is reduced, the optimization model keeps the number of ambulances in densely populated areas the same, at the expense of ambulances in remote/rural zones, where most of the areas loses ambulances. This indicates that ambulances in densely populated areas are more pressed on capacity, while ambulances are placed in remote areas to reduce response time rather than to satisfy capacity constraints.

Table 8.5: Helicopter improvements on response time and time to hospital. 1 helicopter gives at least 0.79 minutes improvement vs no helicopters, and at best 1.25 minutes faster response time for red emergencies. Meanwhile, 1 helicopter yields at least 4.98, and at best 6.01, minutes improved time to hospital for red emergencies.

	Improvement on response times (red)	Improvement on time to hospital (red)
1 helicopter vs None	[0.79, 1.25] min	[4.98, 6.01] min
2 helicopters vs None	[1.02, 1.85] min	[6.92, 8.42] min
2 helicopters vs 1 helicopter	[0.23, 0.60] min	[1.94, 2.41] min

It can also be observed that helicopters have little effect on the average response times. This is first of all because helicopters are only dispatched to red emergencies. Secondly, only in rare occasions are helicopters the closest available unit to an emergency. However, when considering the average time to hospital for red emergencies, the helicopters have a significant effect. These effects are depicted in Table 8.5. It must be emphasized that helicopters come in addition to the ambulances, effectively increasing the number of units. In addition to increasing capacity, this gives less availability conflicts, hence it might be the indirect effects of one or two helicopters that leads to some of the improvements.

8.3 Different shapes of the cost function

In Subsection 5.4.2, different shapes of the cost of stroke treatment were presented. Although the concave shaped cost function is chosen as the default, this section will evaluate the effects of the other cost functions as well.

Table 8.6: Optimization effects of different objective functions. B indicates the base case (50 ambulances and 2 helicopters) with the concave function.

	Settings			Optimization							
Scenario	Nr of units			Solution				Variables			
Name	A	CTA	H	LP-solution	MIP sol	Time [s]	Gap	t^T	t^E	t^{CLST}	t^{RSP}
C24	50	0	2	34715968	40592873	107	0 %	99.47	116.95	69.47	18.74
B	50	0	2	21425716	36845308	57	0 %	93.49	110.92	63.49	14.14
B-Convex	50	0	2	10859466	10884782	27	0 %	93.34	110.86	63.34	14.31
B-Onedim	50	0	2	21581098	37117384	63	0 %	93.49	110.92	63.49	14.14
B-S-shaped	50	0	2	46799563	62595689	23222*	0.14%	93.35	110.97	63.35	14.31

As seen in Table 8.6 there are very few differences between the solutions for the different cost functions. The MIP solutions do vary between the different cost functions. This is because the parameters of the stroke costs functions (C^T and C^E) are different for each shape. When changing from the base case (concave cost function) to the convex cost function, one ambulance is moved from Trondheim to Rennebu. Using the one-dimensional cost function results in one ambulance being moved from Trondheim to Åfjord. The exact same solution is given by the S-shaped cost function, although the solution time for the S-shaped model is significantly higher.

Changing the shape of the stroke treatment cost function does not modify the cost of general response. The stroke weight is set to 0.5, so the model considers both parts of the objective function equally much. However, the similarity between the solutions suggests that minimizing the general response time is important for obtaining an optimal solution. Intuitively this makes sense, as it sounds reasonable that the EMS arrives to the patient, the faster the patient will get stroke treatment.

8.4 Effects of CT Ambulances

In this part, CT ambulances are added in several different ways in order to estimate their effects. To begin with, two naive approaches are implemented based on the current ambulance allocation. Then some scenarios where the model can freely place all units are studied. The effects on the surrounding areas of the stations holding CT ambulances are also studied in this section. The table below holds a summary of all results that are used for this section, and will be referenced trough out the section.

Table 8.7: Results of CT ambulances. Solution times marked with * indicates that the scenario was solved by running several model instances in parallel as described in section 8.6.1.

	Settings			Optimization							Simulation					
Scenario	Nr of units			Solution			Variables				Response time		Time to hospital		Stroke	
Name	A	CTA	H	LP-solution	MIP sol	Time	t^T	t^E	t^{CLST}	t^{RSP}	All	Red	All	Red	$\tilde{t}^T$	$\tilde{t}^E$
C-24	50	0	2	34715968	40592873	107	99.475	116.952	69.475	18.737	31.32	13.76	83.13	59.11	89.82	110
C+1CT	50	1	2	25180396	38591904	172*	96.327	113.904	66.327	16.137	30.74	13.82	82.88	60.27	86.9	109.6
C-R	49	1	2	26804282	40300695	150*	97.073	119.548	69.559	18.737	31.19	13.86	83.07	59.66	88.89	116
B	50	0	2	21425716	36845308	57	93.494	110.916	63.494	14.136	24.73	13,55	76,57	59,21	90,14	110,9
B-R	49	1	2	23425746	36531520	155*	90.853	114.274	63.789	14.146	25.16	13.64	77.21	60.04	87.58	112.7
B+1CT	50	1	2	23425746	36529209	107*	90.850	114.271	63.785	14.143	24.89	13.54	77.04	60.05	88.34	115.5
B+2CT	50	2	2	23577294	36236774	1173*	88.272	114.031	64.000	14.264	25.05	13.55	77.4	60.56	85.6	111.8
B+3CT	50	3	2	23475609	35944597	2491*	85.870	116.627	64.084	14.264	25.01	13.46	77.21	60.43	84.18	118.3

8.4.1 Using current allocations to study effects of a CT ambulance

The current ambulance allocation was discussed in Chapter 7, and the results of this allocation (Scenario C-24) without any schedules are now used as a comparison. Now, two different scenarios based on the current allocation will be studied. In one, an extra CT ambulance is granted to the current allocation, we call this scenario for C+1CT. In the alternative scenario, C-R, the number of ambulances is kept the same, and it is simply decided which ambulance to replace with a CT ambulance.

In the first scenario, the optimization model allocates the CT ambulance to Trondheim (St. Olav) station. According to the optimization results, this would give an improvement of approximately 3 minutes on time-to-thrombolysis and -thrombectomy. Furthermore, the optimization averages indicate that this also will give an improvement of approximately 3 minutes on general response time and time to closest hospital. A 5 year simulation supports the indicated improvement on time to thrombolysis, also with 3 minutes compared to C-24. However, the additional CT ambulance gives no significant changes to general response times or time to hospital.

In the second scenario (C-R), the model chose to replace one ordinary ambulance with a CT ambulance in Vikna. This has limited effect on the overall average time to thrombolysis, which was one minute lower. The average response time for red emergencies was about the same as without a CT ambulance. The simulation does not allow the CT ambulance to respond to red emergencies that are not strokes. So in the case of Vikna, non-stroke patients have in reality lost one ambulance.

The optimization model focuses heavily on reducing the time to thrombolysis, and less on time to endovascular treatment, since only 10 % of ischaemic strokes are eligible for thrombectomy. Moreover, the number of generated thrombectomy eligible strokes in the simulation are low relative to the total amount of red emergencies, causing time to thrombectomy to vary notable between simulations of the same scenario. This may explain why the model does not seem to have any predictable effect regarding time to endovascular treatment between different scenarios.

In Subsection 8.4.3 we look into whether boundary effects may contribute to the decision to place the CT ambulance in Vikna.

8.4.2 Effects of more CT ambulances

Now, all units can be placed freely, i.e. current allocation is not enforced. In Table 8.7 the base case (50 ambulances and 2 helicopters) is included as a benchmark for the different results with CT ambulances. Again, we will start with replacing one of the ordinary ambulances with a CT ambulance. Then, we will examine the effects of adding one to three CT ambulances.

When replacing one ordinary ambulance with a CT ambulance, the model was able to allocate

49 ordinary ambulances, two helicopters and one CT ambulance. The CT ambulance was positioned in Steinkjer. By comparing the simulation results for this scenario (B-R) to the base case (B), we can see that average time-to-thrombolysis has been reduced with 2.56 minutes. This reduction is almost the same in the optimization model, 2.64 minutes. Compared to the above mentioned current scenarios, more ambulances are placed in densely populated areas like Trondheim, Stjørdal and Steinkjer at the expense of areas that are sparsely populated. Based on the simulation, this results in a minimal improvement of average time to thrombolysis (about 1 minute). On the other hand, the average response time and time to hospital are improved significantly. This might imply a somewhat low stroke weight.

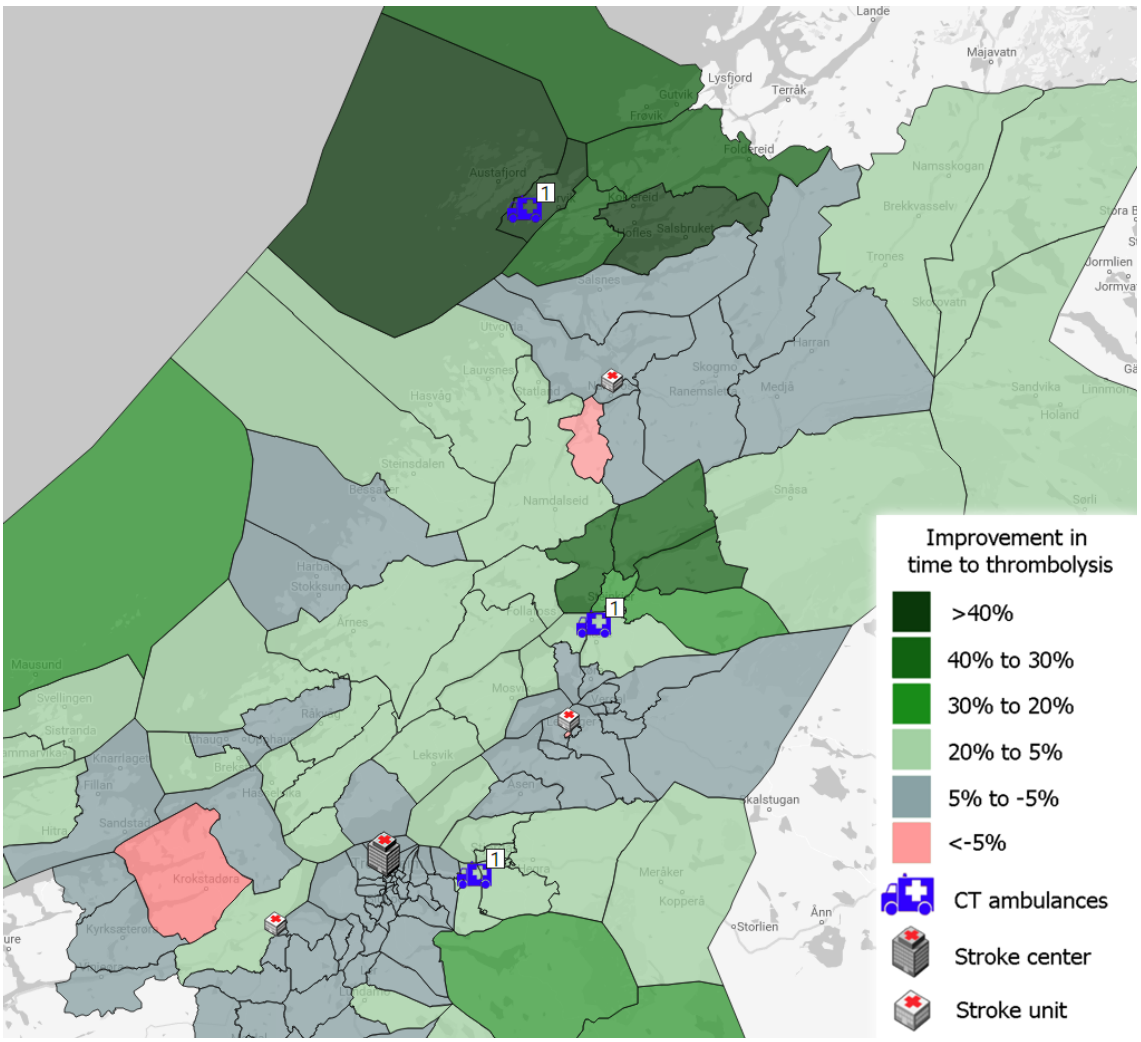


Figure 8.1: Improvement in time to thrombolysis with three CT ambulances. All numbers can be found in Appendix B.2. Map data ©2018 Google.

When adding one CT ambulance to the base case (B+1CT), the model positioned the CT ambulance in Steinkjer. Additionally, one ordinary ambulance was moved from Trondheim to

Steinkjer. For two CT ambulances, the model suggests that these should be placed in Steinkjer and Stjørdal. In this case, two ordinary ambulances are moved out of Trondheim and distributed to Steinkjer and Stjørdal. This clustering of ordinary ambulances along with a CT ambulance can be explained from the synergistic effect in Restriction (5.17) in Chapter 5.

If the model is to place three CT ambulances, the third CT ambulance is placed in Vikna. These are all densly populated zones which have a distance to the closest hospital that is relatively long. In driving distance, Stjørdal is 35 kilometers away from St. Olavs Hospital in Trondheim. Steinkjer is located 41 kilometers away from Levanger Hospital while Vikna is 85 kilometers away from Namsos. Allowing patients in these zones early thrombolysis in a CT ambulance would be beneficial for their time to treatment.

The map in Figure 8.1 shows the improve in time-to-thrombolysis from scenario B to B+3CT. The CT ambulances will have significant effects for patients suffering from stroke in the surrounding zones, especially in zones far from a hospital, such as Vikna in north-west. However, from a cost-benefit perspective it is important to point out that the CT ambulances were only used 206, 834 and 909 times each over the course of a 5-year simulation for Vikna, Steinkjer and Stjørdal respectively. In other words, if the CT ambulances only are allowed to be used for stroke patients, they will on average be called out approximately every 9th day (Vikna) and every other day (Stjørdal).

8.4.3 Boundary effects

Vikna is located in the northern part of the geographical area of the model, on the boundary between Trøndelag and Nordland. Areas located close to the boundary may have seemingly fewer ambulances and helicopters closer to them because emergency units outside the boundary are not available to our model. In the case of Vikna, the ambulance helicopter located in Brønnøysund could potentially respond to emergencies in Vikna. According to SNL (2013), the helicopter does indeed respond to patients in northern parts of Trøndelag today. If the boundary effect were avoided, the model's choice of ambulance allocation would most likely be affected.

8.5 Establishment of new stations

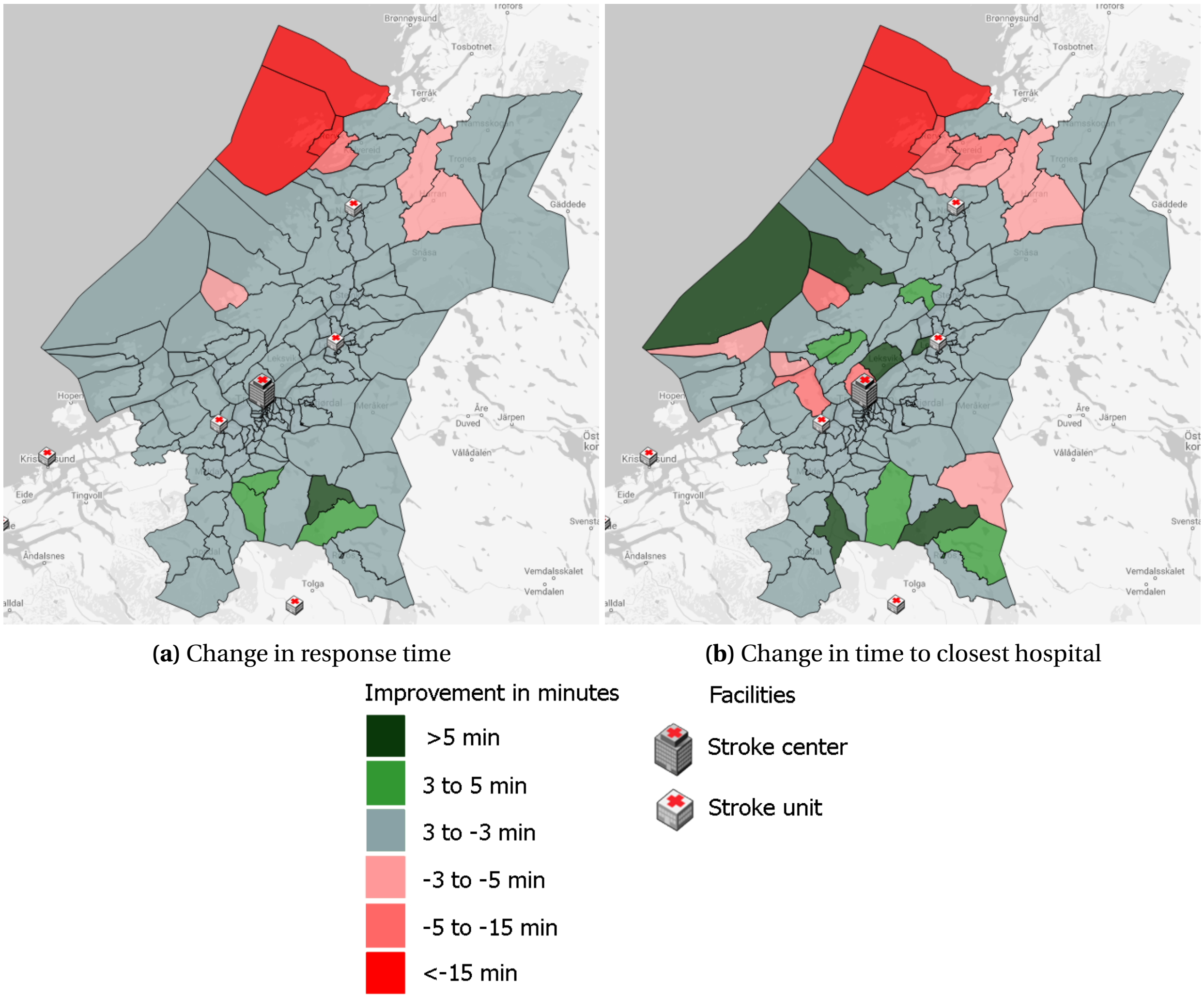
In this section, the model can choose to open new stations in zones that has no stations today. The model is also free to close any of today's existing stations. The results are compared with the base case solution.

Table 8.8: Results when allowing new stations. W indicates the number of possible new stations. Solution times marked with * indicates that the scenario was solved by running several model instances in parallel as described in Subsection 8.6.1.

	Settings				Optimization							Simulation					
Scenario	Number of units				Solution			Variables				Response time		Time to hospital		Stroke	
Name	A	CTA	H	W	LP	MIP	Sol time [s]	t^T	t^E	t^{CLST}	t^{RSP}	All	Red	All	Red	$\tilde{t}^T$	$\tilde{t}^E$
C24	50	0	2	0	34715968	40592873	107s	99.47	116.95	69.47	18.74	31.32	13.76	83.13	59.11	89.82	110
B	50	0	2	0	21425716	36845308	57	93.49	110.91	63.49	14.14	24.73	13.55	76.57	59.21	90.14	110.9
B-W1	50	0	2	1	22785877	36282014	90*	91.96	109.42	61.96	14.35	25.25	13.5	77.04	59.38	90.07	110.1
B-W2	50	0	2	2	22788147	35831357	1819*	90.71	108.17	60.71	14.51	26.01	13.95	77.89	59.64	89.87	112.1
B+1CT	50	1	2	0	23425746	36529209	107*	90.84	114.27	63.78	14.14	24.65	13.49	76.68	59.79	87.59	114.1
B+1CT-W1	50	1	2	1	26833615	35976605	1935*	89.27	112.83	62.20	14.50	26.13	13.72	78.03	59.56	87.17	107.2

New stations with ordinary ambulances

When allowing one new station (B-W1), the model chooses to close the station in Ørland and open a new one in the neighboring zone to the east. Two ambulances are moved out of Trondheim and to the stations in Støren and Vikna. When allowing two new stations (B-W2), the station in Ørland is again moved to the same zone, while the stations in Leka and Vikna are closed down in favour for a new station in Kolvereid. Two ambulances are moved out of Trondheim to Haltdalen and Støren.



(a) Change in response time

(b) Change in time to closest hospital

Figure 8.2: Effects of allowing two new station. Map data ©2018 Google.

In Figure 8.2, improvements in response time and time to closest hospital with two new stations are presented. The results are based on a 5-year simulation. Although a new station is opened in Kolvereid, the total number of ambulances is reduced in that region as stations in Vikna and Leka are closed. This has some serious implications. In Leka, the average time to closest hospital is 42 minutes longer. In Vikna, 30 minutes longer. The number of ambulances are however increased in southern parts of Trøndelag around Støren and Røros (see green cluster in the south), improving both response time and time to closest hospital in that area. The population in this area is more dense than in the northern parts, which may explain why the model chooses to relocate resources here at the expense of Vikna and Leka. This result may also be affected by boundary effects. The hospital in Tynest, just south of Røros, is not available to our model. If the hospital had been taken into consideration, the solution to relocate more ambulances to the south might be affected.

New stations with one CT ambulance

The model was run with 50 ambulances, two helicopters and one CT ambulance at its disposal. When allowing one new station to be build, the station at Ørland is closed and reopened further east as in the cases above. The CT ambulance is positioned in Steinkjer. Other main observations are that two ambulances were moved out of Trondheim and one helicopter was positioned in Namsos.

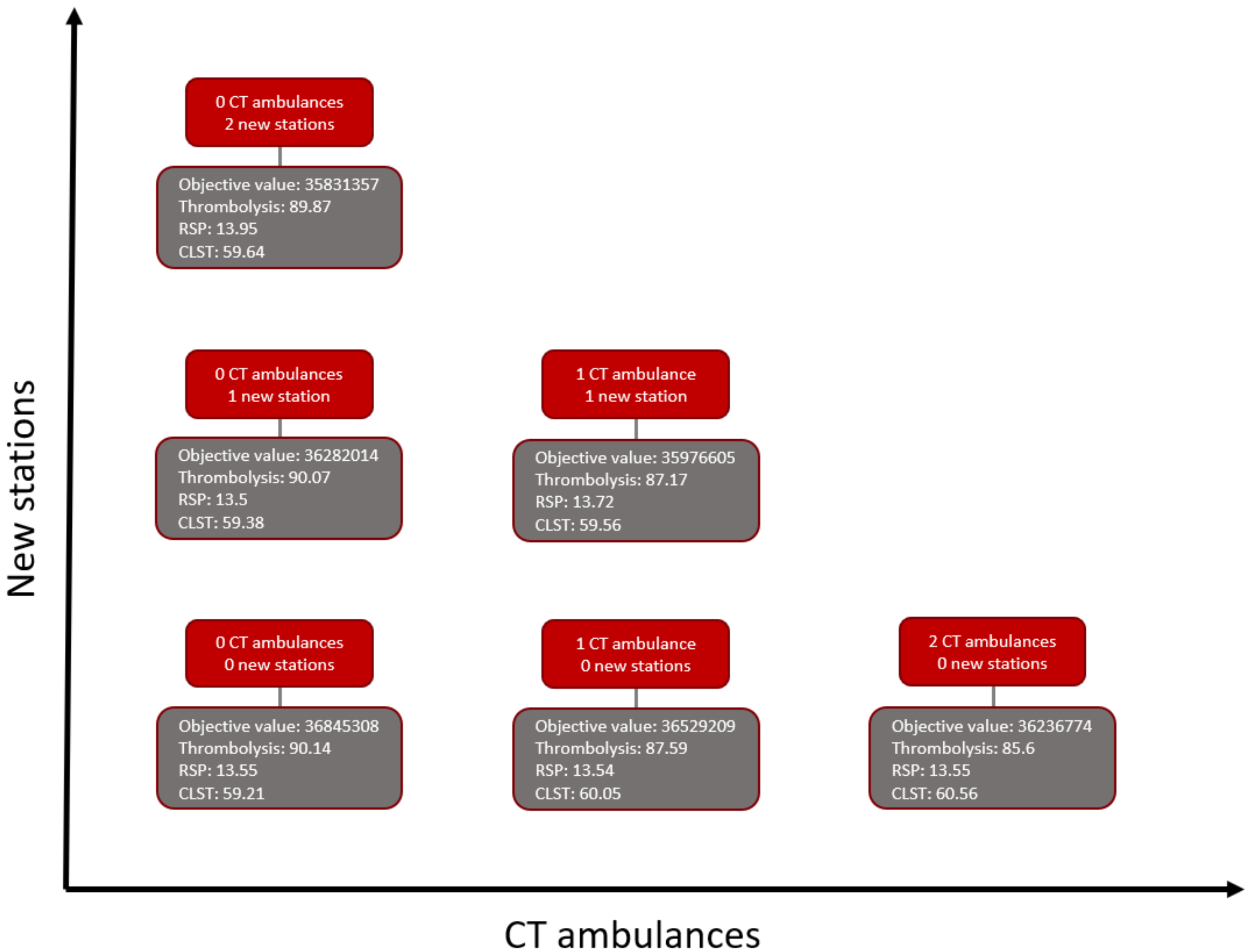


Figure 8.3: Comparison between different cases with CT ambulances and new stations.

A comparison between some of the cases with CT ambulances and new stations is presented in Figure 8.3. Objective values are provided by the mathematical optimization, while time-to-thrombolysis, response time (RSP) and time to closest hospital (CLST) are provided by the simulation. Increasing the number of CT ambulances yield the best improvement in time-to-thrombolysis. The objective values shows that the optimization model would prefer 0 CT ambulances and 2 new stations over 2 CT ambulances and 0 new stations. However, the simulation shows that 2 CT ambulances and 0 new stations is superior in average time-to-thrombolysis by 4 minutes. The optimization model evaluates the combination of 1 CT ambulance and 1 new station to be better than having 1 new station or 1 CT ambulance respectively. The simulation confirms that 1 CT ambulance and 1 new station results in a reasonable good time-to-thrombolysis, but its not that much better than having 1 CT ambulance and 0 new stations. This indicates that having CT ambulances is more beneficial for the time-to-thrombolysis than opening new stations.

8.6 Imaging equipment on all ambulances: standard in the future?

As mentioned in the background, in addition to the stroke ambulance, another research project is evaluating an alternative imaging technique, capable of diagnosing patients more easily than what can be done with a CT. This makes it a more practical piece of equipment, that potentially can be availabile on any ambulance. In this section it is examined what the effects would be if the *Strokefinder* (or similar equipment) would be standard equipment on all ambulances, just like EKG (electrocardiography) is today for diagnosing the heart. In other words: what happens if all ambulances are "stroke ambulances"? It is assumed that the diagnosis time is equal to the one with CT ambulances.

In order to conduct this test, the simplest approach is to change the x^{CTA} variable to an integer variable. Moreover, the CT scramble time is set equal to the scramble time for ordinary ambulances.

Table 8.9: Effects of new imaging technology on all ambulances.

	Settings			Optimization								Simulation					
Scenario	Number of units			Solution				Variables				Response time		Time to hospital		Stroke	
Name	A	CTA	H	LP-solution	MIP sol	Gap	Solution time	t^T	t^E	t^{CLST}	t^{RSP}	All	Red	All	Red	$\tilde{t}^T$	$\tilde{t}^E$
B	50	0	2	21425716	36845308	0.00%	57	93.49	110.92	63.49	14.14	24.73	13.55	76.57	59.21	90.15	111.1
H	0	50	0	24571691	33228844	0.00 %	28	64.31	124.73	68.45	14.31	25.2	15.05	79.33	71.67	65.75	117.7
H1	0	50	1	22634738	32420899	0.00 %	526	64.27	121.67	65.63	14.27	24.15	14.19	76.66	65.46	64.34	111.9
H2	0	50	2	22147561	31766118	0.00%	192*	64.14	119.18	63.54	14.09	24.68	14.08	76.79	63.84	64.75	120.5
H2	0	50	2	22147561	31766118	0.22%	81969	64.14	119.18	63.54	14.09	24.68	14.08	76.79	63.84	64.75	120.5

As seen in Table 8.9, the time to thrombolysis is reduced by approximately 25 minutes when all ambulances have diagnosis equipment. Since thrombolysis can be initiated immediately after the diagnosis is determined, the average time to thombolysis equals approximately the average response time for red emergencies + diagnosis time. Increasing the number of helicopters does not have a big impact on time-to-thrombolysis, since most of the patients initiate the thrombolysis in an ambulance, as these are usually the first responders. In the case where two helicopters are available, only Frosta and Ytterøy are covered by helicopter only. However, several zones are responded to by an ambulance, and then met by a helicopter after the diagnosis is done. When a helicopter meets an ambulance, the response time for that zone is not affected, but the time to closest hospital is reduced. The simulation confirms that the average time to hospital is reduced by 8 minutes when the model goes from zero to two ambulances.

8.6.1 Solving several model instances in parallel

Solving the case above with two helicopters is computationally hard, and after 81969 seconds the solution was still not proven optimal using a single Mosel optimization instance. The following section presents a method for utilizing built-in functions in FICO's Xpress Mosel to solve several instances in parallel. Using this approach significantly reduces the total solution time. The case with two helicopters is solved to optimality after 192 seconds using this approach. In this instance, we divided the solution space by fixing the helicopters locations in the different submodels.

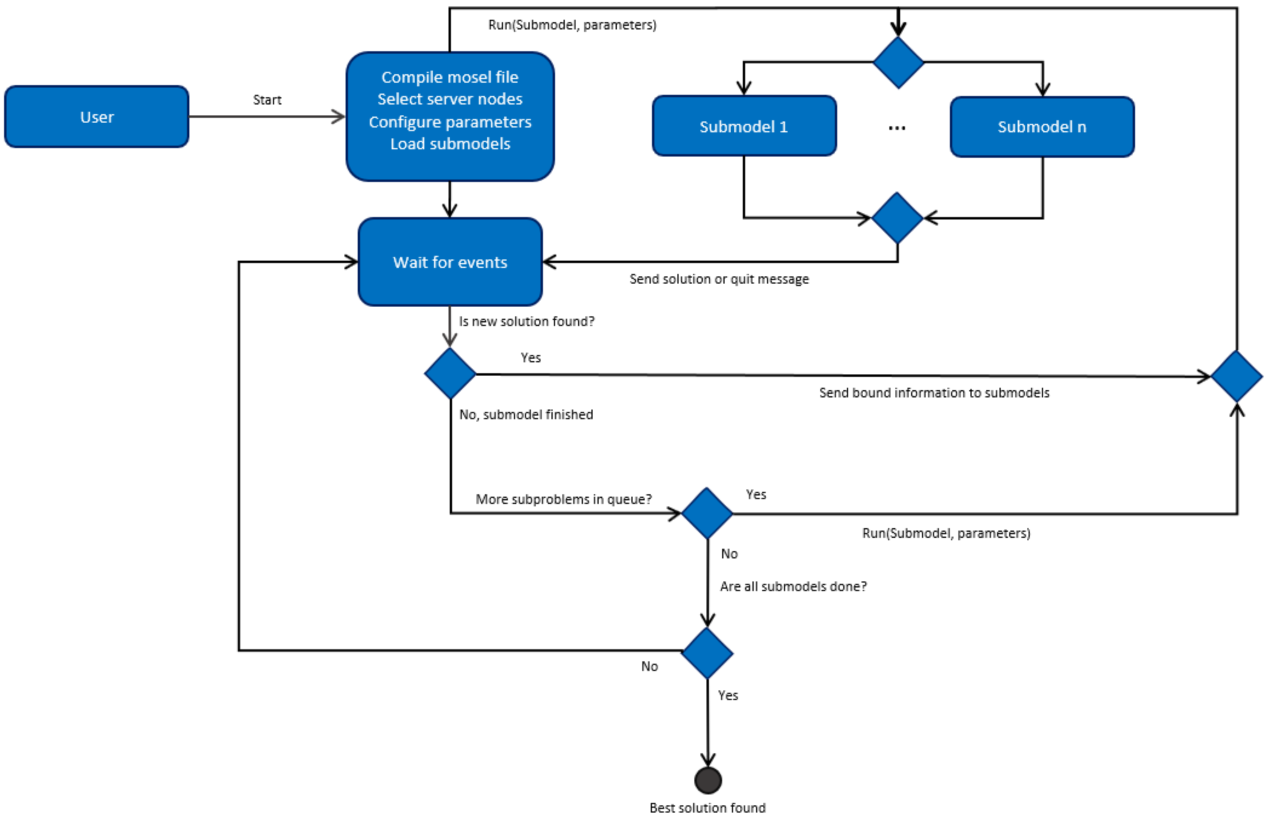


Figure 8.4: Flowchart describing the basic logic of the implementation of parallel solving

The Mosel framework supports parallel solving such that a division of solution space mentioned in Subsection 5.4.3 is possible. In order to load several models and run them concurrently, a Mosel module called *mmjobs* is utilized as described in FICO (2018).

The Mosel Remote Launcher *xprmsrv* is used to run model instances on several computers across a network. Before executing the optimization, *xprmsrv* must be started at each remote computer and configured with appropriate connection and security settings.

A master Mosel code is then run to perform several tasks: It compiles Mosel code into an executable file. It then connects to every remote *xprmsrv* instance, loads the executable file onto the remote instance along with input files and start parameters. Then the master code starts the optimization at each remote instance by calling the *run* command as seen in Figure 8.4. The parameters sent to each instance describe which part of the total solution space that the instance should solve for, such that each instance handles only a subproblem and not the original problem.

When all submodels are running, the master Mosel code enters a state where it listens for events.

The *mmjobs* module has functionality for sending and receiving events between Mosel instances. Whenever a submodel finishes or finds a solution, this is sent as an event to the master model. If a new solution is sent, the master problem will check if this is the best solution yet found, and in that case send the objective value to the other submodels. The submodels check regularly, by the callback function *XPRS_CB_PRENODE*, if a new solution is received. If a new solution is received, this will be used as a new cutoff bound by setting the *XPRS_MIPABSCUTOFF* parameter.

When a submodel finishes, the master problem will check in a queue whether parts of the solution space remains to be checked for optimal solution. If the queue is not empty, the master model will reload the submodel with parameters containing the new solution space and the current best objective value. When all submodels have finished, and the queue of remaining solution space is empty, the best solution found by the subproblems is the optimal solution to the original problem. This method is also used to solve the cases with CT Ambulances in Section 8.4 as well as establishment of new stations in Section 8.5.

Chapter 9

Concluding remarks

This thesis presents a mixed integer linear program for EMS allocation of multiple vehicle types, including specialized CT ambulances built for diagnosing and treating stroke patients. The model uses a survival function based approach, minimizing the costs related to both stroke and general emergencies. The problem opens up for differentiated treatment paths for stroke patients if they receive prehospital diagnosis. This means that patients eligible for thrombectomy treatment can be transported directly to a stroke center. By including helicopters in the model formulation, the results given on real life applications will be much more realistic for areas dependent of such resources.

The optimization is complemented with a discrete event simulation software, specifically created for this problem. This program is based on real data. The simulation is validated by comparing its results to this real data, and most aggregated numbers are relatively close. The logic of the simulation is also reviewed with the personnel at the Ambulance Services at St. Olavs Hospital. This has given it face validity, although all eventualities in real life and the human factor cannot be programmed. The simulation also provides some dynamics that the linear program is missing, since ambulances are "always" on the move. This makes the results even more realistic.

The results for Trøndelag, which is the application case study of this thesis, shows that CT ambulances have a limited overall effect, since the prevalence of stroke in rural areas are small. For patients in urban areas who often are closer to hospitals, the time saved is limited. However, for those stroke patients who receives care from stroke ambulances, the time-to-treatment saved can have a significant effect on expected life quality.

As an extra part of this thesis, we have described a method to divide large problems into several subproblems which are then solved in parallel. Problems that are at first deemed unsolvable (in reasonable time, or due to memory issues) actually can be solved quite effectively if the solution space is divided in a clever manner.

9.1 Recommendations for future research

While working with this thesis we have discovered some areas which can extend and improve our models.

Relocating ambulances for alertness

A possible extension of both the optimization and simulation would be to implement a way of relocating emergency resources in order to give higher preparedness. From testing the simulation, we noticed that areas with few ambulances in some cases are left "uncovered" while other stations have several available ambulances. Sending those to fill in for empty stations could result in notably improved response times.

CT scanners at emergency clinics

In our model, CT scanners can only be in CT ambulances and in hospitals. While hospitals can be far away, and the benefit of a few CT ambulances is arguable, another possibility could be to invest in CT scanners at emergency clinics. This was mentioned in the Background chapter. Measuring the effects of a shared stationary CT scanner at local emergency clinics in rural areas compared to CT ambulance will be of interest. Especially if equipment that can enable ordinary ambulances to determine diagnosis, such as the Strokefinder, proves to have low value.

Finding alternative ways of implementing demand and capacity

The implementations of capacity and demand in this thesis are directly based on population. The capacity used considers the driving distances related to covering a zone, but some results indicates that our implementation slightly regards urban areas as more important than rural areas. This might also have to do with our hard constraint implementation of capacities. This forces ambulances into urban areas to provide enough capacity, while response times in rural areas are neglected. By implementing a form of shared capacity between stations (like those in Trondheim), this problem can be reduced. An alternative approach would be to implement share responsibility for the zones, giving reduced demand for each station.

Appendix A

Appendix - Input data

A.1 Medical parameters

$P^T = 0.9$: Probability that a patient can only be treated with thrombolysis
$P^E = 0.1$: Probability that a patient is eligible for the endovascular procedure
$T^{CALL} = 3$ minutes : Emergency call time (EC response time)
$T^{ID} = 20$ minutes : Time needed for initial diagnosis when arriving at patient location
$T^{D-TRNS} = 30$ minutes : Time needed for CT and diagnosis in ambulance with CT
$T^{D-HOSP} = 39$ minutes : Time needed for admission, CT and diagnosis at hospital
$T^{PREP1} = 2$ minutes : Preparation time for thrombectomy if a patient has been diagnosed in ambulance
$T^{PREP2} = 10$ minutes : Preparation time for thrombectomy if a patient has not been diagnosed prehospitally
$T^{LIMIT} = 30$ minutes : Minimum improved transport time needed to justify helicopter transport
$T_i^{MAX} = 12/25$ minutes (urban/rural) : Maximum response time for general emergencies for zone i
$T^{SA} = 1$ minute : Scramble time for ambulances (unit reaction time)
$T^{SCTA} = 5$ minutes : Scramble time for CT ambulances (unit reaction time)
$T^{SH} = 5$ minutes : Scramble time for helicopters (unit reaction time)

A.2 Area of study

A.2.1 Trøndelag

In Trøndelag, there are four hospitals. Levanger, Namsos and Orkdal are all emergency hospitals, and are considered as stroke units.[1] St. Olav Hospital in Trondheim is a stroke center, and performs endovascular treatment for stroke patients. It is the only stroke center in the region.

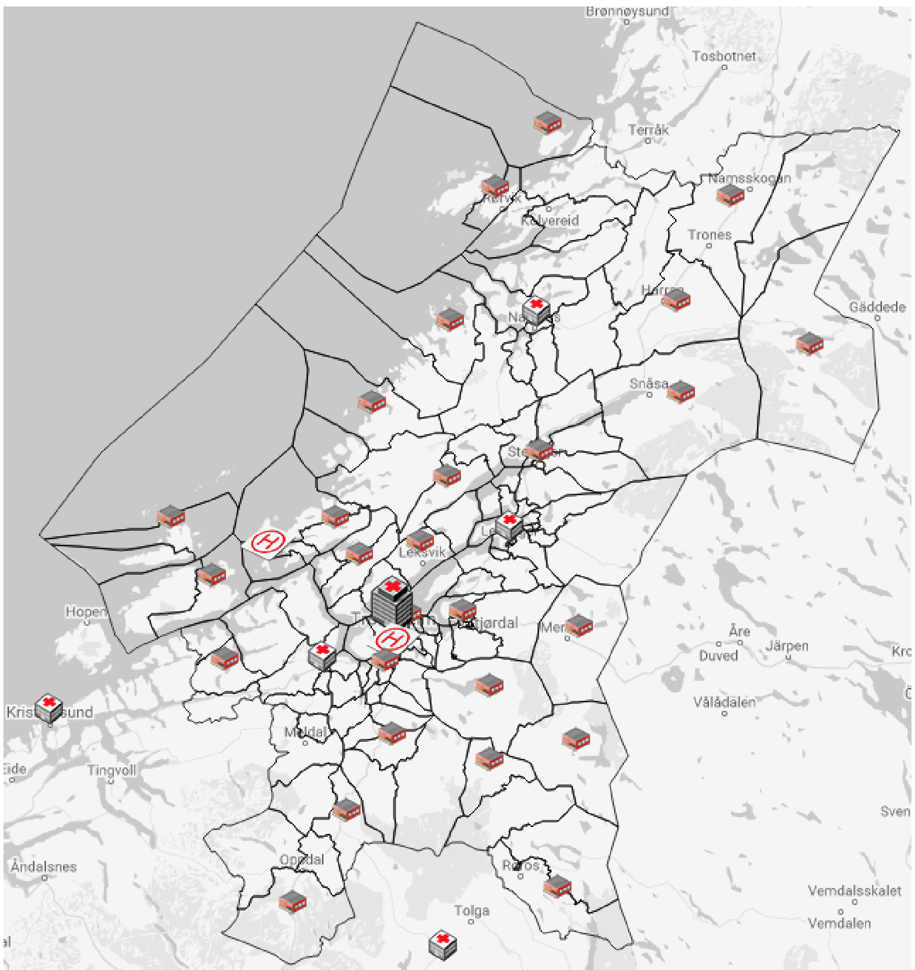


Figure A.1: A map overview of all hospitals, ambulance stations and helicopter bases in Trøndelag.

Trøndelag has a total of 139 zones. 33 of the zones hold ambulance stations today, and 2 of them have helicopter bases. The helicopter bases are also ambulance stations (this is not always the

[1]It has recently been decided that Orkdal Hospital no longer be a stroke unit, and that patients belonging to its coverage area will be transported directly to St. Olav Hospital in Trondheim if they have stroke symptoms (Skjesol, 2017). This change came after we had started our work and is therefore not considered in this report.

case). All zones with hospitals has an ambulance station. The number of ambulances vary with the time of day. We use 48 ambulances as our base case. There is a rescue helicopter based on Ørlandet (the most Western on the map). In this thesis this helicopter is included in the base case.

As shown in Figure A.1 there are two hospitals outside Trøndelag (Kristiansund and Tynset). Many patients will in reality belong to these hospitals. In addition, there is a helicopter base in Brønnøysund (the base is not shown on the map) North of our focus area. This helicopter cover many of the most northern zones in Trøndelag. These boundary effects can lead to "wrong" resource allocation.

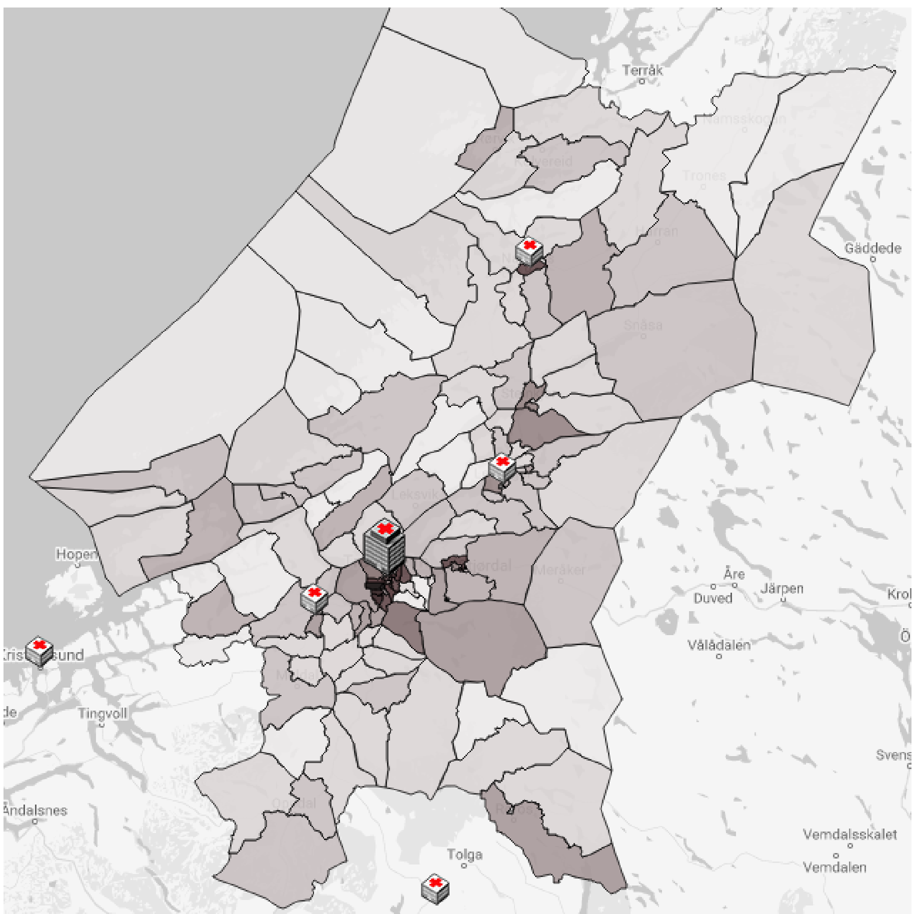


Figure A.2: A map overview of the population in all zones. Population is strongly correlated with demand for stroke treatment. Less demand is colorized with a lighter color. The population in the zones varies from below 1000 people up to 17000 people. Map data ©2018 Google.

A.2.2 Zones

Zone ID	Name	Latitude	Longitude	Population
16011100	Ila-Trolla	63.43369391533816	10.350278303919595	6521
16011200	Midtbyen	63.42940329592425	10.392924899637137	4441
16011300	Øya-Singsaker	63.42156746635383	10.388475710128887	7226
16011400	Rosenborg-Møllenberg	63.431062699	10.4194575739	9009
16011500	Lademoen	63.4395022815	10.4271014928	5634
16011600	Lade	63.4487214772116	10.447382037908369	5525
16012100	Strindheim	63.4300465009	10.4558776736	9105
16012200	Charlottenlund-Jakobsli	63.42185197691861	10.495169033352795	11214
16012300	Ranheim	63.428066036048264	10.559158665048471	8986
16013100	Berg-Tyholt	63.417773485	10.42989931	8902
16013200	Åsvang-Stokkan	63.398604584	10.4705321534	8653
16014100	Nardo	63.4077448502	10.4069624524	5028
16014200	Nidarvoll-Leira	63.3767088422	10.4314605478	9830
16014300	Risvollan-Othilienborg	63.397324873	10.4360195535	7198
16014500	Jonsvatnet	63.3828707215324	10.597254413343308	1065
16014600	Bratsberg	63.34949721679002	10.496863240539028	861
16015100	Sverresborg	63.41733863346363	10.353571856856092	9571
16015200	Byåsen	63.398229086553904	10.334514910304733	17039
16015300	Hallset	63.3914728874	10.3704586915	4588
16016200	Sjetne-Okstad	63.37040258855522	10.392855843921211	4839
16017200	Flatåsen-Saupstad	63.37481768690517	10.33580198681966	14186
16017400	Tiller-Hårstad	63.36242980446893	10.380503788593842	8432
16017500	Heimdal	63.35124094615263	10.334031671623848	13949
16018100	Byneset-Leinstrand	63.369950991012374	10.177237857357113	4928
16120100	Heim	63.41890851704344	9.077748563249997	391
16120200	Hemne	63.2945696508	9.08063327241	3445
16120300	Vinje	63.21149677311361	9.019819890712483	415
16130100	Snillfjord	63.40261692733106	9.502707333836838	977
16170100	Vestre Hitra	63.53166995817958	8.38488772330993	1411
16170200	Østre Hitra	63.609086970003034	8.953030055937234	3179
16200100	Sør-Frøya	63.705453127625645	8.564266338067227	1816
16200200	Nord-Frøya	63.77645810378017	8.788996642372808	2590
16200300	Øyrekken	63.8638522312	8.64785018381	335
16210100	Ørland øst	63.70957073501399	9.674933786617203	2477
16210200	Ørland vest	63.68069619285014	9.59493412124175	2711
16220100	Agdenes	63.61822199716428	9.640002906594646	1716
16240100	Stjørna	63.752602482100066	10.132605404931269	761
16240200	Hasselvika/Fevåg	63.655422222883445	9.843246125865335	631
16240300	Rissa	63.56381608557983	9.923661668001614	3319
16240400	Stadsbygd	63.49764672228582	10.007972813175797	1875
16270100	Bjugn/Stjørna	63.7826042873	9.94531689485	2901
16270200	Nes/Jøssund	63.85125763445151	9.776830169094865	1866
16300100	Åfjord	63.9649031133876	10.219432590371866	2615
16300200	Stoksund	64.03735348309519	10.054785903964785	656
16320100	Roan	64.18350299030025	10.430573005357928	956
16330100	Osen	64.29522543289313	10.514502405192161	976
16340100	Olbu/Vognill	62.6638099813	9.34021566641	1207
16340200	Oppdal nord/Fagerhaug	62.6272102762	9.84183369055	2806
16340300	Oppdal syd/Drivdalen	62.4296577593	9.61010923365	2861
16350100	Øvre Rennebu	62.7620554819	10.032116635	1664
16350200	Nedre Rennebu	62.8555609698	9.72110570558	895
16360100	Løkken	63.124493251277606	9.70375657420368	1553
16360200	Bygda	63.0297350099	9.662101756	2395
16380100	Orkanger	63.30678091987356	9.850327794302757	3590
16380200	Orkdal østre	63.26530279818667	9.813717171610278	4254
16380300	Orkland	63.1891604149	9.70259569624	1641
16380400	Geitastrand og Orkdal vestre	63.3142424432	9.70024961427	2282
16400100	Røros	62.57916646895283	11.380520270803117	4482
16400200	Brekken/Glåmos	62.646497868680434	11.865541607807813	1147
16440100	Haltdalen	62.926492643909505	11.13875910402271	556
16440200	Ålen	62.84270854818596	11.298880952198488	1473
16480100	Støren	63.03431525939707	10.292959879431237	2976
16480200	Singsås	62.955892348625554	10.731973719431267	1526
16480300	Soknedal	62.949855102160726	10.190095675037469	1773
16530100	Melhus vest	63.2822389241	10.2262389705	4432
16530200	Melhus øst	63.284140329290985	10.282268042892156	4037
16530300	Kvål	63.22796675088258	10.28591078782813	1382
16530400	Ler	63.19740082911765	10.305774239217158	1282
16530500	Lundamo	63.1338785261	10.4236546975	1835
16530600	Hovin	63.10901113199569	10.233541042056231	1512
16530700	Hølonda	63.11297358711993	10.011807330685997	1568
16570100	Skaun	63.2370029813	10.0145981279	1408
16570200	Buvik	63.2912221694	10.1654635885	3363
16570300	Børsa	63.32607308403534	10.068828978832812	2974
16620100	Klæbu	63.29787909660028	10.483393553843712	6058
16630100	Hundhamaren-Saksvik bygdelag	63.43837909204476	10.602849170195327	3795
16630200	Vikhamar Malvik-Torp Midtsandan	63.43521256217446	10.672355964891267	3045
16630300	Hommelvik-Mostadmark	63.41152412068411	10.800859789806282	4417
16630400	Malvik utmarksområder	63.39379741083556	10.661205405428063	653
16630500	Sveberg	63.42348536431069	10.751702639835344	1789
16640100	Selbu	63.2001143849	11.137160261	4121
16650100	Tydal	63.044491789636794	11.652728082134445	849
17020100	Sparbu/Henning	63.91854337223989	11.431680449849978	5379
17020200	Steinkjer sentrum	64.01512044831435	11.495591786980071	5552
17020300	Egge	64.06914887570072	11.50994352959924	5192
17020400	Beitstad	64.08506474125056	11.362281285490553	2148
17020500	Kvam	64.14113725462411	11.70036664473048	1013
17020600	Stod	64.08232403436286	11.702265633921911	1096
17020700	Ogndal	64.02993351995835	11.651365541810264	1327
17030100	Spillum	64.348693531	11.5724362548	2080
17030200	Bangsund	64.393673549128	11.400612297187536	1088
17030300	Otterøy/Vemundvik	64.53502480012634	11.27977907447655	1401
17030400	Namsos	64.46832077657702	11.508888241592217	8424
17110100	Meråker	63.416212456541146	11.743540344400003	2517
17140100	Hegra	63.464684259543844	11.114225940184383	3494
17140200	Lånke	63.39378635648877	11.063734736669517	3468
17140300	Skatval	63.5176798074	10.8573870664	2590
17140400	Stjørdal	63.4832280135403	10.96689920190579	8574
17140500	Stjørdalshalsen	63.47047586826663	10.923603549672407	5144
17170100	Frosta	63.56492915297449	10.69933038786803	2621
17180100	Leksvik	63.669357781463184	10.614850637109384	2233
17180200	Stranda	63.57493955134386	10.362831195582771	1297
17190100	Nesset	63.75327380663823	11.261013757320711	5875
17190200	Skogn	63.70318744102392	11.194593893707179	2522
17190300	Midt-Skogn	63.680620930429946	11.126109076464672	2128
17190400	Åsen	63.60797560191839	11.043133468259498	1764
17190500	Frol	63.73317605296473	11.407966035403547	2170
17190600	Bruborg med omland	63.745767574394925	11.325650076449165	3051
17190700	Levanger	63.744309818107894	11.296369930111268	1497
17190800	Ytterøy	63.774676246818395	11.058953466988214	548
17210100	Ørmelen	63.7920221457	11.4381902551	2223
17210200	Verdalsøra	63.7968463149	11.4986843874	4298
17210300	Sjøbygda	63.8220139913	11.4458655613	1530
17210400	Stiklestad	63.798211841364186	11.565809517514026	2516
17210500	Vinne og Ness	63.767669890807355	11.561526978024972	2560
17210600	Øvre Verdal	63.71640917419102	11.866036823237437	1673
17240100	Malm	64.07062601942987	11.216364155164001	1823
17240200	Verran	63.868256742449084	10.783909787864104	664
17250100	Namdalseid	64.22199984799803	11.22611470673121	1622
17360100	Snåsa	64.24288767111584	12.36677231623139	2137
17380100	Lierne	64.44178484170573	13.627853866900068	1374
17390100	Røyrvik	64.8833004076437	13.563581367017832	468
17400100	Namsskogan	64.85029409189875	13.030141399034392	866
17420100	Grong	64.53028196776981	12.400906250359412	2462
17430100	Høylandet	64.7443544758	12.3565738883	1250
17440100	Overhalla	64.48082210984937	11.890943709251587	3815
17480100	Fosnes	64.65551947551177	11.446944538688285	632
17490100	Flatanger	64.49421806412352	10.898099910441374	1101
17500100	Vikna indre	64.89749566946048	11.252719310389011	3920
17500200	Vikna ytre	64.9138261699524	10.932404896570347	460
17510100	Nærøy	64.796022424	11.2976830256	1621
17510200	Kolvereid	64.86480082002383	11.60077872867032	2410
17510300	Gravvik/Foldereid	64.98658017531233	11.760149194382848	1092
17550100	Leka	65.08864712673133	11.711525520251598	562
17560100	Røra	63.85405216151494	11.398997961067153	1219
17560200	Straumen	63.8735781047	11.2879647548	2326
17560300	Inderøy vest	63.869267699287256	11.160752833618744	1311
17560400	Sandvollan	63.958606889290444	11.336243958754721	1101
17560500	Mosvik	63.82057411557572	11.007391081220248	804

Table A.1: Zones and population.

A.3 Simulation probability data

Data provided by Ambulansetjenesten, based on stats from 2017.

Based on all emergencies where ambulance belongs to Trøndelag region where dates/times are recorded. This means that emergencies served by units from outside Trøndelag are not included. Data is also filtered to remove inconclusive records, however, not all are removed. This explains the discrepancies in the sums. The small variations will most likely have minor impact on the results.

Table A.2: Distribution of emergencies based on priority.

Priority	# records	Probability
Red	16035	0.3101
Orange	21119	0.4085
Green	14550	0.2814
SUM	51704	1

Table A.3: Distribution of emergencies based on weekday.

Weekday	# records	Probability
mandag	7805	0.1510
tirsdag	7565	0.1464
onsdag	7420	0.1436
torsdag	7469	0.1445
fredag	7886	0.1526
lørdag	6526	0.1263
søndag	7004	0.1355
SUM	51675	1

Table A.4: Distribution of emergencies based on hour of weekday/weekend day.

	All week	Monday - Friday		Saturday - Sunday	
Hour	# records	# records	Probability	# records	Probability
0	1 319	837	0.0219	482	0.0356
1	1 233	714	0.0187	519	0.0384
2	1 055	576	0.0151	479	0.0354
3	832	467	0.0122	365	0.0270
4	720	423	0.0111	297	0.0220
5	679	462	0.0121	217	0.0160
6	737	525	0.0138	212	0.0157
7	1 373	1 085	0.0284	288	0.0213
8	2 620	2 159	0.0566	461	0.0341
9	2 934	2 283	0.0599	651	0.0481
10	3 240	2 500	0.0655	740	0.0547
11	3 517	2 691	0.0705	826	0.0611
12	3 636	2 868	0.0752	768	0.0568
13	3 629	2 819	0.0739	810	0.0599
14	3 377	2 584	0.0677	793	0.0586
15	3 126	2 415	0.0633	711	0.0526
16	2 904	2 202	0.0577	702	0.0519
17	2 632	1 930	0.0506	702	0.0519
18	2 336	1 721	0.0451	615	0.0455
19	2 322	1 677	0.0440	645	0.0477
20	2 155	1 501	0.0393	654	0.0483
21	2 029	1 475	0.0387	554	0.0409
22	1 710	1 194	0.0313	516	0.0381
23	1 559	1 037	0.0272	522	0.0386
SUM	51 674	38 145	1	13 529	1

Appendix B

Appendix - Computational study

B.1 Hardware and software configuration

All instances were run using Fico Xpress 8.4.4 on CentOS Linux 7 (Core) operating system.

Lenovo NeXtScale M5 with 2 x Intel E5-2643v3 CPUs, each with 3,4 GHz. Memory available: 512 GB RAM. 1 x 120GB SSD harddrive.

The cluster configuration was run on different hardware setups. This is not included in this thesis.

B.2 Comparing stroke treatment times: 3 CT ambulances and base case

Here are the numbers behind the calculations in Subsection B.2 presented.

Table B.1: Effects of 3 CT ambulances versus regular base case with no CT ambulances.

	3 CT Ambulances		Base case		Differences	
Patient Zone	Avg t^T	Avg t^E	Avg. t^T	Avg. t^E	t^T	t^E
Agdenes	129,30		126,15	177,00	2,49 %	
Bangsund	84,56	163,00	74,24	182,25	13,91 %	-10,56 %
Beitstad	83,68	197,00	121,90	175,00	-31,36 %	12,57 %
Berg-Tyholt	72,58	75,86	71,63	70,67	1,33 %	7,35 %
Bjugn/Stjørna	126,54	97,00	132,98	128,25	-4,84 %	-24,37 %
Bratsberg	85,50		88,82	94,00	-3,74 %	
Brekken/Glåmos	172,00	266,00	185,25	224,00	-7,15 %	18,75 %
Bruborg med omland	64,76	127,00	64,48	117,25	0,43 %	8,32 %
Buvik	82,91	122,50	83,02	121,50	-0,13 %	0,82 %
Bygda	113,97	155,00	115,33	170,33	-1,18 %	-9,00 %

Byneset-Leinstrand	77,45	79,33	81,37	81,14	-4,81 %	-2,23 %
Byåsen	70,57	69,04	71,32	69,61	-1,06 %	-0,83 %
Børsa	80,02	116,83	78,37	119,00	2,10 %	-1,82 %
Charlottenlund-Jakobsli	73,44	77,90	69,97	69,81	4,96 %	11,58 %
Egge	82,42	165,00	105,37	162,50	-21,78 %	1,54 %
Flatanger	111,05		120,07	239,67	-7,51 %	
Flatåsen-Saupstad	70,70	69,13	70,61	72,05	0,13 %	-4,06 %
Fosnes	108,89	205,00	109,58	231,00	-0,63 %	-11,26 %
Frol	69,83	135,25	67,64	128,83	3,24 %	4,98 %
Frosta	89,54	123,83	105,94	95,00	-15,48 %	30,35 %
Geitastrand og Orkdal vestre	87,05	141,00	93,23	124,00	-6,63 %	13,71 %
Gravvik/Foldereid	110,86	397,50	181,32	304,00	-38,86 %	30,76 %
Grong	109,00	209,40	110,50	217,29	-1,36 %	-3,63 %
Hallset	70,90	68,50	71,11	72,11	-0,30 %	-5,01 %
Haltdalen	106,29		122,00	137,00	-12,88 %	
Hasselvika/Fevåg	105,80	63,00	116,43	118,50	-9,13 %	-46,84 %
Hegra	87,16	106,29	97,76	101,00	-10,84 %	5,23 %
Heim	127,30		133,71		-4,80 %	
Heimdal	71,29	74,71	71,68	67,88	-0,55 %	10,08 %
Hemne	108,30	142,75	106,26	142,43	1,92 %	0,23 %
Hommelvik-Mostadmark	80,11	85,00	89,51	91,55	-10,50 %	-7,15 %
Hovin	96,14		93,00		3,38 %	
Hundhamaren-Saksvik bygdelag	79,76	74,80	77,06	75,67	3,49 %	-1,15 %
Hølonda	84,67		82,36	142,00	2,80 %	
Høylandet	130,79	334,00	132,33	234,50	-1,16 %	42,43 %
Ila-Trolla	63,25	60,40	64,44	64,17	-1,85 %	-5,87 %
Inderøy vest	88,00		103,24		-14,76 %	
Jonsvatnet	79,59	80,00	83,47	69,00	-4,65 %	15,94 %
Klæbu	81,06	96,67	79,85	72,71	1,52 %	32,94 %
Kolvereid	90,92	365,00	152,67	257,00	-40,44 %	42,02 %
Kvam	88,27	197,00	126,21	179,00	-30,07 %	10,06 %
Kvål	84,30	76,00	85,83		-1,77 %	
Lade	79,73	84,00	76,91	70,22	3,67 %	19,62 %
Lademoen	70,24	70,58	72,90	77,63	-3,65 %	-9,07 %
Leka	129,18	387,00	186,40		-30,70 %	
Leksvik	103,47	166,40	119,57	146,80	-13,47 %	13,35 %
Ler	89,38	103,50	90,54	87,00	-1,29 %	18,97 %
Levanger	61,92	124,40	55,52	100,00	11,51 %	24,40 %
Lierne	149,46	290,00	160,38		-6,80 %	
Lundamo	91,18	62,00	100,63	117,50	-9,39 %	-47,23 %
Løkken	102,00	142,00	96,00	134,00	6,25 %	5,97 %
Lånke	84,39	109,40	95,73	89,22	-11,85 %	22,62 %
Malm	99,35	224,00	123,96	229,00	-19,85 %	-2,18 %
Malvik utmarksområder	78,17	90,00	80,14	75,00	-2,47 %	20,00 %
Melhus vest	81,88	78,17	79,64	71,00	2,82 %	10,09 %
Melhus øst	75,68	81,50	78,48	76,45	-3,56 %	6,60 %
Meråker	90,83	159,00	112,54	144,50	-19,29 %	10,03 %
Midt-Skogn	81,74	137,50	83,06	142,00	-1,58 %	-3,17 %
Midtbyen	61,18	65,67	60,96	54,82	0,36 %	19,79 %
Mosvik	88,29	216,00	94,40		-6,48 %	
Namdalseid	112,84	206,00	119,44	225,00	-5,52 %	-8,44 %
Namsos	65,47	169,76	64,43	158,27	1,61 %	7,26 %
Namsskogan	120,18	307,00	145,23	233,00	-17,25 %	31,76 %
Nardo	65,80	67,60	66,53	63,43	-1,09 %	6,58 %
Nedre Rennebu	122,57		117,10	179,00	4,67 %	-100,00 %
Nes/Jøssund	128,64	84,50	156,49	199,20	-17,80 %	-57,58 %
Nesset	65,81	120,63	63,84	122,71	3,08 %	-1,70 %
Nidarvoll-Leira	73,45	70,50	72,09	75,62	1,89 %	-6,77 %
Nord-Frøya	136,51	262,50	144,65	225,50	-5,63 %	16,41 %
Nærøy	91,46	308,00	152,28	310,00	-39,94 %	-0,65 %
Ogndal	87,38	166,50	114,42	152,50	-23,63 %	9,18 %
Olbu/Vognill	137,38		149,47	243,50	-8,09 %	-100,00 %
Oppdal nord/Fagerhaug	136,32	263,00	145,05		-6,02 %	
Oppdal syd/Drivdalen	147,46	254,60	153,87	205,67	-4,17 %	23,79 %
Orkanger	61,97	96,17	63,44	111,50	-2,33 %	-13,75 %
Orkdal østre	64,26	106,75	66,78	115,67	-3,77 %	-7,71 %
Orkland	99,00	124,00	95,92	141,33	3,21 %	-12,26 %
Osen	133,63	277,33	141,18		-5,34 %	
Otterøy/Vemundvik	99,55		96,00	202,50	3,69 %	
Overhalla	93,47	194,00	91,67	186,25	1,97 %	4,16 %

Ranheim	69,54	71,10	69,00	70,86	0,79 %	0,34 %
Rissa	100,71	120,67	112,19	122,67	-10,23 %	-1,63 %
Risvollan-Othilienborg	76,17	74,67	78,83	81,60	-3,38 %	-8,50 %
Roan	131,50	232,00	137,43		-4,31 %	
Rosenborg-Møllenberg	70,47	73,71	71,79	69,67	-1,84 %	5,80 %
Røra	80,71	158,00	83,21		-3,00 %	
Røros	135,13	172,14	164,76	167,57	-17,99 %	2,73 %
Røyrvik	160,00		186,71		-14,31 %	
Sandvollan	83,94	163,00	98,50	147,50	-14,78 %	10,51 %
Selbu	90,31	137,71	113,30	111,36	-20,29 %	23,66 %
Singsås	103,74	147,00	121,32	82,00	-14,49 %	79,27 %
Sjetne-Okstad	66,26	58,60	68,74	68,25	-3,60 %	-14,14 %
Sjøbygda	82,36		84,76	132,00	-2,83 %	
Skatval	78,46	115,20	89,88	129,60	-12,70 %	-11,11 %
Skaun	83,88		87,16	148,00	-3,76 %	
Skogn	87,78	145,00	89,48	143,00	-1,90 %	1,40 %
Snillfjord	101,44	154,50	86,00		17,95 %	
Snåsa	100,76	252,00	125,58	217,33	-19,76 %	15,95 %
Soknedal	100,58		113,45	109,00	-11,34 %	
Sparbu/Henning	81,27	156,57	98,04	168,80	-17,10 %	-7,24 %
Spillum	100,35	236,50	99,05		1,31 %	
Stadsbygd	109,17	130,25	122,16	109,50	-10,63 %	18,95 %
Steinkjer sentrum	76,10	164,78	93,90	150,25	-18,96 %	9,67 %
Stiklestad	96,02	173,40	93,93	144,25	2,23 %	20,21 %
Stjørdal	86,24	107,47	96,43	94,82	-10,57 %	13,34 %
Stjørdalshalsen	75,52	84,00	90,23	89,33	-16,31 %	-5,97 %
Stjørna	128,75		154,00	165,00	-16,40 %	
Stod	79,39		119,25	172,40	-33,43 %	
Stoksund	138,00		138,33	207,00	-0,24 %	
Stranda	105,30		99,91	132,50	5,39 %	
Straumen	94,46	176,67	97,24	164,33	-2,87 %	7,51 %
Strindheim	71,14	65,08	73,73	79,83	-3,51 %	-18,48 %
Støren	94,86	159,00	99,36	106,00	-4,54 %	50,00 %
Sveberg	85,62	82,75	90,19		-5,07 %	
Sverresborg	65,84	65,29	67,28	66,12	-2,13 %	-1,26 %
Sør-Frøya	144,05	199,00	160,44	200,50	-10,22 %	-0,75 %
Tiller-Hårstad	66,81	65,18	65,21	65,46	2,46 %	-0,43 %
Tydal	101,44	114,67	140,92		-28,01 %	
Verdalsøra	84,69	125,00	87,50	148,00	-3,21 %	-15,54 %
Verran	102,17	169,00	115,23	210,00	-11,34 %	-19,52 %
Vestre Hitra	157,05		167,72	229,67	-6,36 %	
Vikhamar Malvik-Torp Midtsandan	81,91	86,80	82,08	80,00	-0,20 %	8,50 %
Vikna indre	77,09	301,00	143,19	293,75	-46,17 %	2,47 %
Vikna ytre	94,91	314,50	163,80	291,00	-42,06 %	8,08 %
Vinje	97,00					
Vinne og Ness	82,39	140,00	85,58	132,88	-3,73 %	5,36 %
Ytterøy	123,00		112,29	236,00	9,54 %	
Ørland vest	120,97	218,00	134,27	205,00	-9,91 %	6,34 %
Ørland øst	126,18	106,80	132,31	138,83	-4,64 %	-23,07 %
Ørmelen	83,32	152,40	85,14	144,50	-2,13 %	5,47 %
Østre Hitra	121,56	199,00	118,46	208,00	2,62 %	-4,33 %
Øvre Rennebu	124,00	191,50	116,76	195,33	6,20 %	-1,96 %
Øvre Verdal	104,06	180,33	106,85	164,50	-2,61 %	9,63 %
Øya-Singsaker	56,64	61,55	54,49	50,43	3,94 %	22,04 %
Øyrekken	152,25		211,40		-27,98 %	
Åfjord	115,05	209,00	133,45		-13,79 %	
Ålen	128,86	150,67	142,87	161,75	-9,80 %	-6,85 %
Åsen	89,37	148,00	89,41	151,00	-0,04 %	-1,99 %
Åsvang-Stokkan	71,85	74,06	71,80	74,37	0,07 %	-0,42 %

Bibliography

Aarnseth, A. B. and Hov, E. M. (2017). A deterministic optimization model for differentiated stroke treatment chains. Project report in TIØ4500.

Andersson, T., Värbrand, P., and Mustafee, N. (2016). Operational research for emergency planning in healthcare: Volume 1.

Aringhieri, R., Brunib, M., Khodaparastic, S., and van Essen, J. (2017). Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers Operations Research*, 78:349–368.

Banks, J. L. and Marotta, C. A. (2007). Outcomes validity and reliability of the modified rankin scale: Implications for stroke clinical trials. *Stroke. American Heart Association.*, 38:1091–1096.

Bordvik, M. (2014). Hjelm kan diagnostisere slag. URL: https://www.dagensmedisin.no/artikler/2014/06/17/hjelm-kan-diagnostisere-slag/, Accessed: 2018-06-01.

Brooks, R. J. and Tobias, A. M. (1996). Choosing the best model: Level of detail, complexity and model perfomance. *Mathematical Computer Modelling*, 24:1–14.

Budge, S., Erkut, E., and Ingolfsson, A. (2008). Optimal ambulance location with random delays and travel times. *Health Care Manage Sci*, 11:262–274.

Churilov, L. and Donnan, G. A. (2012). Operations research for stroke care systems: An opportunity for the science of better to do much better. *Operations Research for Health Care*, 1:6–15.

Daskin, M. S. (1983). A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science*, 17:48–70.

Daskin, M. S. and Haghani, A. E. (1987). A combined model of train routing, makeup, and empty car distribution. *Logistics and Transportation Review*, 23:173–177.

Daskin, M. S. and Maass, K. L. (2015). A combined model of train routing, makeup, and empty car distribution. *Logistics and Transportation Review*, 23:173–177.

Diaz, R. and Behr, J. G. (2010). *Modeling and Simulation Fundamentals: Theoretical underpinnings and practical domains. Discrete-Event Simulation chapter.* McGraw Hill Higher Education.

Erkut, E., Ingolfsson, A., and Erdogan, G. (2008). Ambulance location for maximum survival. *Naval Research Logistics*, 55:42–58.

FICO (2018). Solving several model instances in parallel. URL: http://www.fico.com/fico-xpress-optimization/docs/latest/mosel/mosel$_parallel/dhtml/secsingleparmod.html, Accessed: 2018-06-05$.

Folkestad, E. H., Gilbert, M., and Steen-Hansen, J. E. (2004). Når det haster – prehospitale responstider i vestfold og troms i 2001. *Tidsskriftet Den Norske Legeforening*, 3(124):324–8.

Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., Roberts, K., Coyle, E., Bevan, G., and Palmer, S. (2003). Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health Medicine*, 25:225–335.

Garrido, J. (1998). *Practical Process Simulation Using Object-Oriented Techniques and C++*. Artech House.

Gjesdal, E. (2017). Kan revolusjonere behandlingen av slag. URL: https://www.nrk.no/rogaland/kan-revolusjonere-slagbehandlingen-1.13459706, Accessed: 2018-06-01.

Gumbinger, C., Reuter, B., Stock, C., Sauer, T., Wiethölter, H., Bruder, I., Rode, S., Kern, R., Ringleb, P., Hennerici, M., and Hacke, W. (2014). Time to treatment with recombinant tissue plasminogen activator and outcome of stroke in clinical practice: retrospective analysis of hospital quality assurance data with comparison with results from randomised clinical trials. *British Medical Journal*, 348(7961).

Hakimi, S. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, 13:462–475.

Holodinsky, J. K., Williamson, T. S., Kamal, N., Mayank, D., Hill, M., and Mayank, G. (2017). Drip and ship versus direct to comprehensive stroke center: Conditional probability modeling. *Stroke*, 48:233–238.

Law, A. M. and Kelton, W. D. (2000). *Simulation Modelling and Analysis*. McGraw Hill Higher Education.

Lyden, P. D., Barreto, A. D., and Grotta, J. C. (2015). *Thrombolytic Therapy for Acute Stroke*. Springer International Publishing Switzerland.

Mazighi, A., M., Chaudhry, D., S., Ribo, H., M., Khatri, A., P., Skoloudik, I., D., Mokin, I., M., Labreuche, I., J., Meseguer, I., E., Yeatts, I., S., Siddiqui, I., A., Broderick, I., J., Molina, I., C., Qureshi, I., A., and Amarenco, I., P. (2013). Impact of onset-to-reperfusion time on stroke mortality: A collaborative pooled analysis. *Circulation*, 127(19):1980–1985.

Mccormack, R. and Coates, G. (2015). A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *European Journal of Operational Research*, 247(1):294–309.

Nahin, P. J. (2002). *Duelling Idiots and Other Probability Puzzlers*. Princeton University Press.

Nilsen, L. (2018). Dagens medisin: Langt flere slagpasienter kan reddes. URL: https://www.dagensmedisin.no/artikler/2018/01/30/-langt-flere-slagpasienter-kan-reddes/, Accessed: 2018-06-01.

Norsk Hjerneslagregister (2017). Nasjonal servicemiljø for medisinske kvalitetsregistre - norsk hjerneslagregister - resultater publisert i 2017. URL: https://www.kvalitetsregistre.no/registers/353/resultater, Accessed: 2017-12-11.

Pidd, M. (1994). An introduction to computer simulation. *Proceedings of Winter Simulation Conference*, 0:7–14.

Robinson, S. (2004). *Simulation: The Practice of Model Development and Use*. John Wiley Sons, Ltd.

Saver, J. L., Goyal, M., van der Lugt, A., Menon, B. K., Majoie, C. B. L. M., and more. (2016). Time to treatment with endovascular thrombectomy. *Journal of American Medical Association*, 316(12):1279–1288.

Schellinger, P. D., Köhrmann, M., and Nogueira, R. G. (2016). The location of emergency service facilities. *International Journal of Stroke*, 11(5):502–508.

Skjesol, H. (2017). St. olav blir størst i landet på hjerneslagpasienter. *Adresseavisen*, Published 01.10.2017.

SNL, S. N. L. (2013). Kapasitet og basestruktur. URL: https://norskluftambulanse.no/wp-content/uploads/2013/09/SNLA-Kapasitet-og-basestruktur-rapport-sept2013.pdf, Accessed: 2018-06-03.

SNL, S. N. L. (2017). Slagambulansen - veien til ct i helikopter. URL: https://norskluftambulanse.no/vart-arbeid/forskning-2/slagambulansen/, Accessed: 2017-12-13.

SSB (2016a). Statistisk sentralbyrå, statistikkbanken, spesialisthelsetjenesten. URL: https://www.ssb.no/statbank/table/09556/?rxid=30a52a4c-776e-40eb-a140-c5ad45647ac9, Accessed: 2017-12-12.

SSB (2016b). Statistisk sentralbyrå, statistikkbanken, spesialisthelsetjenesten. URL: https://www.ssb.no/statbank/table/09556/tableViewLayout1/?rxid=f769e41f-e6d3-4ea2-a577-e582952464f3, Accessed: 2018-06-05.

SSB (2017a). Statistisk sentralbyrå, ambulansetjenester koster mest i nord. URL: https://www.ssb.no/helse/artikler-og-publikasjoner/ambulansetjenester-koster-mest-i-nord, Accessed: 2017-12-12.

SSB (2017b). Statistisk sentralbyrå, statistikkbanken, spesialisthelsetjenesten. URL: https://www.ssb.no/statistikkbanken/selectout/ShowTable.asp?FileformatId=2Queryfile=20171213141104016
Accessed: 2017-12-13.

Steiger, N. and Cifu, A. (2016). Primary prevention of stroke. *JAMA*, 316(6):658–659.

Stoinska-Schneider, A., RObberstad, B., and Fure, B. (2016). Mekanisk trombektomi ved akutt hjerneinfarkt del2. helseøkonomisk evaluering.

Strbian, D., Piironen, K., Meretoja, A., Sairanen, T., Putaala, J., Tiainen, M., Artto, V., Rantanen, K., Häppölä, O., Kaste, M., and Lindsberg, P. J. (2012). Intravenous thrombolysis for acute ischemic stroke patients presenting with mild symptoms. *International Journal of Stroke*.

Tawil, S. E., Cheripelli, B., Huang, X., Moreton, F., Kalladka, D., MacDougal, N. J., McVerry, F., and Muir, K. W. (2016). How many stroke patients might be eligible for mechanical thrombectomy? *European Stroke Journal*, 1(4):264–271.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, 19:1363–1373.

Uleberg, O., Zakariassen, E., Grytten, S., Normann, E., and Sellevold, S. (2013). Kapasitet og basestruktur, en utredning om luftambulansetjenesten i norge.

Waaler, H. (1999). Scenario 2030. sykdomsutvikling for eldre frem til 2030. *Statens helsetilsyn*, 19:0.