

# Logistics for Emergency Medical Service systems

December 16, 2016

## Abstract

Emergency Medical Service (EMS) systems worldwide are complex systems, characterised by significant variation in service providers, care pathways, patient case-mix and quality care indicators. Analysing and improving them is therefore challenging. Since EMS systems differ between countries, it is difficult to provide generic rules and approaches for EMS planning. Nevertheless, the common goal for all service providers is to offer medical assistance to patients with serious injuries or illnesses as quickly as possible. This paper presents an overview of logistical problems arising for EMS providers, demonstrating how some of these problems are related and intertwined. For each individual planning problem, a description as well as a concise literature overview of solution approaches considered is given. A summary table classifies the literature according to the problems addressed and connects it to the proposed taxonomy.

**Keywords:** EMS systems, emergency and non-emergency services, forecasting, location problems

## Introduction

Helping and rescuing patients in case of an emergency is an important and urgent service in every country in the world. Since ambulance response times can be a crucial factor in patient survival, ambulances are expected to arrive at the scene of reported incidents as quickly as possible, raising important questions of where ambulances should be located and how many should be deployed. As Emergency Medical Service (EMS) systems differ between countries, there might be different answers to these questions. Operational Research (OR) literature offers many different models and approaches to tackle the ambulance location problem. As most EMS systems can be grouped into one of two main systems, the Anglo-American and the Franco-German system (Dick, 2003), the approaches should be applicable to more countries than they were initially developed for, but maybe not to all. In addition to the ambulance location problem, several other planning problems and logistical tasks exist for EMS providers, such as shift scheduling for dispatchers and ambulance crews or the routing of patient transports. Besides a couple of reviews specifically on ambulance location and relocation, complete overviews over all problems including a classification of approaches into planning levels are still sparse. Aringhieri et al (2017) present a review of planning problems based on the Emergency Care Pathway, following the patient's trajectory through the system. This paper, on the other hand, classifies papers according to the planning problem they address and solution approach followed. It focuses on the dependencies between planning problems and planning levels and discusses the different organisational objectives.

Before presenting a derived taxonomy for EMS planning problems, we first describe the UK, Dutch and German EMS systems in detail, in order to highlight some of the main similarities and differences between European EMS systems. Such details become important when discussing specific forecasting and scheduling problems in the sections that follow.

## Short overview of EMS systems

Throughout the world, two main types of EMS systems exist, namely the Anglo-American system and Franco-German system (c.f. Dick (2003)). The main difference lies in the definition of staff. Whilst paramedics employed in the Anglo-American system are able to administer almost any treatment necessary, the German-Austrian system includes so-called emergency doctors. They are called in severe cases and are the only ones allowed to deliver certain kinds of treatments. Moreover, each country has its own set of laws and specific regulations. The main differences include the payment structure, response time targets and being either organised by private organisations or public institutions. An overview over the systems can be found in Dick (2003), for example.

In this paper, we will mainly consider three different countries which are the UK, Germany and the Netherlands when giving examples for the planning problems. Therefore, the next sections give short introductions into their EMS systems. Hoogeveen (2010) presents the results of a survey among ten European countries regarding the structure of ambulance care. In the description of the different systems, we give the terminology used for that particular system, so as to make it easy to find references. After overviewing the three systems in succession, we introduce a typical EMS system and the terminology for the remainder of this paper.

### UK system

EMS in the UK provide free urgent care services to patients suffering from illnesses ranging from acute ailments to minor injuries 24 hours a day, 7 days a week. There are 14 ambulance services within the UK (1 in Wales, 1 in Scotland, 1 in Northern Ireland and 11 in England that are separately managed by regional Clinical Commissioning Groups). Hence it is unsurprising that there are small discrepancies in the way that they operate (Lightfoot Solutions, 2009, NHS Choices, 2014, NHS Scotland, 2014, Workforce et al, 2014). Over and above the basic requirement to respond to emergency calls, each ambulance service is also responsible for responding to urgent admission requests from doctors and transporting patients requiring high dependency care between hospitals.

Ambulance crews are highly skilled professionals who are able to stabilise patients and administer medication during transportation. The emergency ambulances (EAs) used to transport patients are themselves state of the art, furnished with a wide range of technical apparatus and typically manned by two crew members - at least one of whom must be a fully trained paramedic. In addition to EAs, Rapid Response Vehicles (RRVs), air ambulances and emergency services co-responder (e.g Fire and Rescue Service) vehicles may also be dispatched to select incidents. RRVs are small vehicles which allow first responders to reach the scene of serious incidents quickly to assess the condition of the patient and offer immediate medical treatment. In the case of immediately life-threatening calls, RRVs should be automatically backed up by the nearest available double-staffed EA, but for all other calls, RRV clinicians must advise the type of back-up required as soon as they have made a primary assessment. Most ambulance trusts also encompass non-emergency patient transport service (PTS) vehicles to transport non-urgent scheduled patients to and from medical facilities. However, non-urgent patient transport work is becoming increasingly sub-contracted to private companies, voluntary organisations and community resource teams.

Patients will always be taken to hospital if there is a medical need, but several recent initiatives have been launched that support alternatives to conveyance wherever possible. The schemes have led to a successful reduction in conveyance rates in England from 98% to 58% over a 12-year period, which has helped to partially offset the rise in emergency calls (O’Cathain et al, 2015, Snooks et al, 2013).

Every time an ambulance service receives a call, emergency medical dispatchers use sophisticated software to prioritise it into one of the following categories of urgency:

- Immediately life-threatening: such calls should receive an emergency response within 8 minutes. The target compliance rate for this target varies locally, for example:
  - In England, 75% of critical ‘Category A Red 1 calls’ where patients are not breathing or do not have a pulse, should be responded to within 8 minutes. An extra 60 seconds is allowed for still serious, but less immediately time critical ‘Category A Red 2’ incidents e.g. strokes or fits. A secondary target also applies to ensure that 95% of all life-threatening calls are responded to within 19 minutes;
  - The monthly average performance targets for all life-threatening calls in Wales, Northern Ireland and Scotland are 65%, 72.5% and 75% respectively.
- All other calls: for conditions that are not life-threatening, response targets are set locally. A striking decision has recently been made by NHS Wales to drop time-based targets for all but immediately life-threatening calls - from October 2015, performance is instead being measured against clinical outcome indicators and patient experience information.

Unless prudence suggests otherwise, common practice is to send the closest available ambulance. Whilst ambulances are assigned to fixed ambulance stations, they are able to relocate to ‘hot-spots’ during the day, and the excessively high demand for ambulances often means that ambulance drivers are commonly routed to attend another incident directly from the hospital at which they previously handed over a patient for further treatment.

## German system

The German EMS system, as part of the German Emergency Health System (EHS), covers in general two main tasks - emergency rescues and patient transports. Patient transports, in contrast to emergency rescues, involve uncritical cases where medical attention is necessary while transporting the patient, for example, between hospitals or back home after a treatment. Besides, a number of further services are usually offered, as, for example, providing information on after-hour practices or on-duty pharmacies.

Each federal state in Germany - there are 16 in total - has its own EMS law organising the services within each state and defining rules and regulations like the response time target. This target only applies for emergency rescues and lies between 8 and 15 minutes. It is measured as the time between the call pick-up and the arrival at scene in most of the cases and has usually to be met in 95% of the cases. Keeping this response time target is, in general, the most important goal for the EMS planning in Germany.

Each federal state is again divided into a number of so-called EMS regions which are often urban or rural (political) districts. Overall, there are more than 250 EMS regions in Germany with more than 2000 base locations. The state of Baden-Württemberg, for example, is divided into 34 EMS regions. In some of the federal states, EMS regions were combined to form fewer and larger regions.

Per EMS region, in general, one coordination center is installed. In many of the regions, the EMS system is combined with the fire brigade. In some of the regions, the fire brigade is in charge of both, while in others relief organizations like the German Red Cross or the Workers’ Samaritan Federation Germany are organising the services.

For EMS, mainly three different vehicles are used - two kinds of ambulances for emergency rescues and patient transports with different levels of equipment and a special vehicle for the emergency doctors. Sometimes, multifunctional ambulances are used for both service types. For the ambulance staff two different levels exist. A basic training leads to becoming an Emergency Medical Technician (EMT). With additional education, a position basically comparable to a paramedic can be reached with the main difference that by law an emergency

doctor is needed for special types of treatments. An ambulance is always equipped by two staff members, for an emergency rescue at least one of the two is a paramedic whereas for a transport it can also be 2 EMTs. Usually, an emergency doctor works at a hospital or a private practice while being on duty as emergency doctor. In case of an emergency, he is either picked up by the ambulance if it is also stationed at the hospital or he is driving to the scene in a separate vehicle. In the latter case, often someone else is driving the car so that if necessary the emergency doctor can accompany the patient in the ambulance on the way to the hospital. This system is called “Rendez-vous-system” as ambulance staff and emergency doctor are meeting at the scene. Sometimes, the emergency doctor is called right away and sometimes he is requested by the ambulance staff when they arrive at the scene. In Baden-Württemberg, the response time target not only applies for the ambulances but also for the emergency doctors.

In addition to these vehicles, helicopters and aircraft are used in Germany for emergency rescues as well as for intensive care transports if needed, but only during daytime.

In general, one ambulance can only serve one patient at a time. Ambulances are typically stationed at hospitals or additional base locations throughout the EMS region. In the case of an emergency, the closest ambulance is usually dispatched. However, since not many ambulances have GPS, it is not always clear which ambulance this is.

The EMS provider that fulfils the emergency service only gets paid if the patient is transported from the scene to a hospital. The pure treatment at the scene is not sufficient.

## **Dutch system**

As in most systems, the ambulance providers in the Netherlands cover two main types of calls - emergency calls and patient transports. However, unlike the UK and German systems, patient transports may involve patients in critical conditions. Some ambulance providers further assist in acute home care during the nights. In total, 725 ambulances were in use in the Netherlands in 2012. These were divided over 207 base stations (Ambulancezorg Nederland, 2013). In the Netherlands, two types of emergency calls are distinguished: A1 and A2 calls. A1 calls are the most urgent, life-threatening calls. In this case, an ambulance is required to be at the scene within 15 minutes in 95% of the cases. For urgent, but not life-threatening calls, an ambulance should be there within 30 minutes. Within the patient transports, which are called B calls, sometimes a further two categories are distinguished: B1 and B2. Here, B1 calls involve the transportation of patients in critical conditions, for which a fully equipped ambulance is required. For B2 calls, less equipped ambulances suffice. In 2012, 1.1 million calls were served in the Netherlands, of which 500,000 were A1, 275,000 were A2, and 325,000 were patient transports.

The ambulance care in the Netherlands is divided into 24 more or less independently operating regions (RAVs), each of which is operated by a single organisation. The Dutch institute of public health and the environment (RIVM) computes the required capacity for each of the RAVs. These computations form the basis for the budgets of the RAVs. Although the budget is based on a number of ambulances and a number of bases, the RAVs are free to choose how to spend their budget. Every year, the branch organisation of ambulance care in the Netherlands (AZN) publishes the performance of the different regions. The AZN further organizes the training for the ambulance crew.

Originally, every ambulance region had his own call center from which calls were taken and ambulances were dispatched. However, in the past years, multiple call centers have merged in order to improve efficiency. Currently, there are 19 call centers in the Netherlands. This will further reduce to 10 in the coming years. Most call centers distinguish call takers and dispatchers. Call takers are responsible for triage, while dispatchers

instruct the ambulance crew. In principle, only the call taker requires medical training. However, in most call centers, call takers and dispatchers switch roles during the day, in which case both call takers and dispatchers are required to have a medical education.

In the Netherlands, all emergency calls that require transportation of a patient are served by an Advanced Life Support (ALS) ambulance. These ambulances are fully equipped and staffed by a paramedic and a driver. The paramedic is required to have completed a full nursing education and at least one follow-up studies in acute care. Additionally, specific training is given by the AZN at the moment of hiring. The driver, on the other hand, does not require a medical background. The driver is there to assist the paramedic at the scene, for this the driver gets training in providing medical assistance. Furthermore, training in driving an ambulance is required.

In patient transports, called B calls, two categories are distinguished depending on the medical conditions of the patient. In case life-threatening situations might occur during transport, an ALS ambulance is required. These calls are called B1. All other transports, B2 calls, might also be executed a Basic Life Support (BLS) ambulance. This is a lower equipped ambulance staffed with two regular nurses.

In particular cases, other vehicles might be used. For patients that need transportation between the intensive care units of two hospitals, Mobile Intensive Care Units (MICU) are used. Some regions additionally use rapid responders. This is a single paramedic that can provide care at the scene, but cannot transfer a patient to a hospital. The paramedic uses, for example, a car, motorbike, or even a normal bike to get to the scene.

As in Germany and the UK, an ambulance can serve only one patient at a time. For both A1 and A2 calls, regular practice is to always send the closest available ambulance. Since all ambulances are equipped with GPS, call center software shows the closest available ambulance to the dispatcher. If necessary, ambulance relocations are performed in order to maintain good coverage throughout the region. Often, these relocation decisions are based on so-called look-up tables. Some regions have specific locations where they can temporarily locate an ambulance as part of a relocation, for example, in the middle between two regular base stations.

## A typical EMS system

Based on the description of the three aforementioned systems, we extract what we call a typical EMS system. This allows us to define the terminology that we use in the remainder of this paper. In a typical EMS system, the EMS provider is responsible for two main tasks: emergency calls and patient transports. These emergency calls include emergency rescues in Germany and Category A calls in the UK and the Netherlands. Typically, a response time target is set for these calls. Patient transport calls are classed as PTS calls in the UK and B calls in the Netherlands. In most systems, two types of emergency calls are distinguished: life-threatening calls and urgent calls. A different target can be set for the two categories. For example, in the Netherlands, the first category must be reached within 15 minutes, whereas the second has a target of 30 minutes. Note that the target that is set by the regulator can significantly change the optimal system design. The rather stringent target of 8 minutes in the UK asks for the use of a large fleet of rapid responders. On the other hand, given the 15 minute threshold in the Netherlands, a configuration with only very few rapid responders might be preferred.

We define three main types of ambulances: Advanced life support (ALS), Basic life support (BLS), and Rapid responders (RR). ALS ambulances are fully equipped, staffed with at least one paramedic, and able to respond to all types of calls. BLS ambulances are lower equipped and typically not staffed with a paramedic. These ambulances are typically used for patient transports, although they could occasionally be dispatched to non life-threatening calls. Note that in the Dutch and German systems, BLS ambulance are strictly limited to serve patient transports. Rapid responders are defined as fully qualified paramedics without the ability of

transporting patients to a hospital. For coverage, a response of a RR within the target response time would suffice. However, in case the patient requires transportation, an additional ALS ambulance is required.

Although there are small differences between countries, the structure of the response process is similar. However, often different terminology is used, which can result in misunderstanding. To avoid this confusion, we introduce a typical response process with the terminology that we use in this paper. When a call arrives at the dispatch center, it takes some time for the dispatcher to assess the urgency of the call and assign an ambulance. This process is called triage and dispatch. The time between the assignment of an ambulance and the moment it starts driving is called the chute time. Together triage and dispatch, and chute time accumulate to the pre-trip delay. Adding the travel time to the pre-trip delay, gives us the response time, which is the main performance measure for EMS providers. In some countries, the definition of response time is slightly different. In England, for example, the clock starts ticking up to 60 seconds later for serious but less immediately time critical incidents, than for cases where patients are not breathing. After spending some time on the scene with the patient, the patient might require transport to a hospital. In that case, the ambulance becomes idle after dropping off the patient at the hospital. In many countries, congested Emergency Departments (EDs) result in increasing turnaround times (the time taken for ambulance crew to handover the patient and restock the vehicle so it is ready to attend another call), which can have an enormous impact on the performance of EMS systems (Channouf et al, 2007). Whenever an ambulance finishes a call, it is available for new calls. If no new calls are waiting, the ambulance returns to its base, or any other location where it waits for new calls. Figure 1 shows the different stages of the response process.

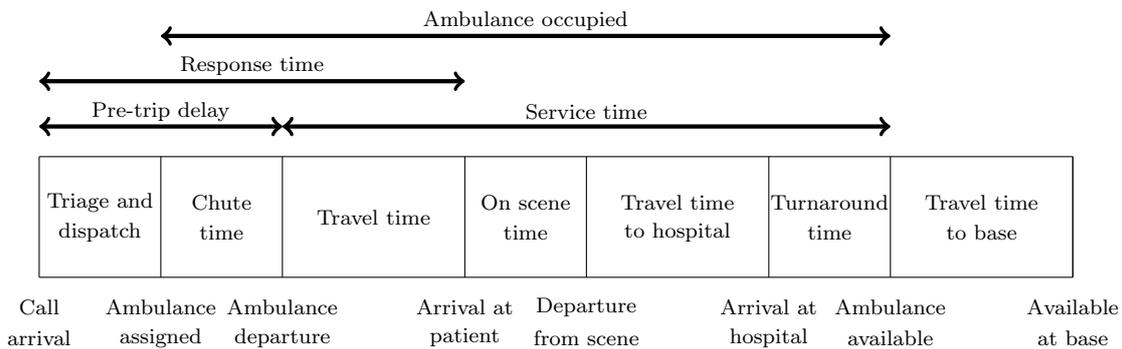


Figure 1: Response process

Even though there can be significant differences between countries, or even between EMS regions, we give some statistics to get insight in the duration of the components of the response time. In the Netherlands, the average time for triage and dispatch for the 24 different regions ranges from 1:18 minutes up to 2:48 minutes, with an average of 1:58 minutes. The average chute time is 0:58 minutes. This implies that the average pre-trip delay is almost three minutes, which allows for a travel time of at most 12 minutes. The average response time in 2013 was 9:36 minutes and in 92.6% of the cases an ambulance was available within the target response time of 15 minutes. The target that is set by law to reach 95% of the life-threatening calls within 15 minutes has recently been reached by 8 of the 24 EMS regions (Ambulancezorg Nederland, 2014).

## Organisational objectives

For EMS planning, as for most logistical planning, two aspects are important: (1) the length of the scheduling horizon and (2) the characteristics of the considered planning problem itself including the goal(s) when solving

the problem as well as possible constraints. In this section, we first look at the performance evaluation of EMS systems. Often a time-related performance measure is considered as the objective, which can result in undesired incentives. We distinguish the real objectives for EMS planning and the performance measures that are used to assess the performance. This often happens by means of simulation.

In addition, we present the different planning levels that are used to classify the time horizon a solution should be obtained for. For each planning level, we give exemplary planning problems. The concrete planning problems will be presented in the ‘Planning problems’ section.

## EMS performance evaluation

Often, improving the quality is named as the main objective when the (re)planning of an EMS system is considered in practice. The problem is that it is difficult to define the quality of an EMS system. While it is already much more difficult to define the quality of a medical system than for many other areas, this is especially true for emergency medicine, because of the extreme working conditions, the unexpected incidents and the often difficult circumstances. In addition, quality from the patient’s, the health insurance’s or the ambulance team’s perspective may differ significantly.

Even if a clear definition of quality could be provided, it is not always clear how the quality would be affected by certain changes in system. Due to the complexity of EMS systems, simulation is often required to estimate the performance in the new situation. We will first describe different definitions of quality of EMS systems. Then, we will discuss different studies that use simulation to estimate the impact of changes in the system.

### Quality definitions

A classical quality definition in healthcare is proposed by Donabedian (1980). According to his interpretation, there are three subcategories for a quality definition in healthcare: structure, process and outcome quality. Using this definition, quality can be defined as the degree of fulfilment of these characteristics. It can be said that the outcome quality (e.g., legal regulations) influences the process quality (e.g., handling of calls or transports, also response time). The process quality then determines the outcome quality (e.g., survival rate, patients’ satisfaction, cost-effectiveness). For most countries, in practice as well as in most of the OR approaches, the process quality (i.e., response time) is used as the objective for planning problems and to measure the performance of the EMS system while also considering the costs. Some mathematical approaches exist that take the survival rate into account, but it might also be appropriate to use the general improvement of the patient’s health status as the main quality measure to include all patients in the model and not only the ones in immediate life-threatening conditions. This seems reasonable at first sight, as this is of main interest for the patient. In general, this should also be the main goal for the EMS provider (while keeping the cost at a reasonable level) and additionally for the insurance companies and social security as they would have to pay for long-term care in case a patient does not fully recover. It would not be fair to only take the patient’s health status when arriving at the hospital or at the end of treatment into account, as the medical condition of the patient at the moment the ambulance is called highly influences the outcome. Scores like the NACA-Score (Schlechtriemen et al, 2005a,b) that is used in many of the German EMS regions express the health status of a patient as a set of integer values. They are often measured when the ambulance arrives at the scene, when the patient arrives at the hospital and if needed, during later stages of the recovering process. Unfortunately, these scores are very difficult to use in order to measure the quality of an emergency service. Consequently, the current practice of measuring performance by implicit measures as response times might be necessary. Future research might focus on how to incorporate the different quality types into an objective that maximizes the

overall performance for all participants.

With the aim of identifying common components of European EMS systems the European Emergency Data (EED) Project (1997-2002) was funded by the European Community under the Program on Health Monitoring within the Framework of Action in the Field of Public Health. The report of the project was published in Krafft et al (2006). The primary goal of this project was to create a common framework for monitoring and assessing EMS systems throughout Europe as an integrated part of a health monitoring strategy. Participants were representatives from Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Norway, Portugal, Slovenia, Spain, Sweden and the United Kingdom. The project team simultaneously sought to develop a comprehensive list of indicators based on the routine collection of EMS data that would enable the monitoring and evaluation of the respective activities of the member states in the area of pre-hospital emergency care, in addition to developing crucial indicators from evidence-based data to allow further comparisons among different member states. Five key indicators were ultimately recommended for monitoring and evaluating the pre-hospital emergency care of the EU countries. These five key indicators are: (1) unit hours (ALS + BLS) p. a. / 100 000 inhabitants, (2) response time (% within 480 sec) for highest priority p. a., (3) rate of highest priority responses p. a. / 100 000 inhabitants, (4) rate of First Hour Quintet incidences p. a. / 100 000 inhabitants and (5) rate of ALS interventions p. a. / 100 000 inhabitants (Fischer et al, 2011).

The unit hours (1) measures the availability of organised EMS resources (ALS and BLS) to the population. Here, one unit hour means a fully equipped response unit available for a call for 1 hour. The response time for highest priority response (2) measures the time until pre-hospital emergency care starts for patients who are presumed to have a life-threatening condition. It is measured by the percentage of patients that received first treatment within 480 seconds. The rate of highest priority response per 100,000 inhabitants (3) indicates the utilisation and demand of an EMS system. It counts the number of responses for which an EMS unit is dispatched to a perceived life-threatening emergency per year. The rate of First Hour Quintet incidents p. a. / 100,000 inhabitants (4) indicates the EMS demand for patients in critical conditions. The First Hour Quintet includes the five conditions cardiac arrest, severe respiratory failure, severe trauma, stroke and chest pain. For these conditions EMS can have a significant impact on the outcome as time is crucial and therefore, treatment needs to start as soon as possible (best chances for a patient to recover are within the first hour). Finally, the rate of ALS interventions p. a. / 100,000 inhabitants (5) measures the level of care provided by the organised EMS system. The number of ALS interventions such as drug application, infusion, tracheal intubation and assisted ventilation are counted for this indicator.

So overall the five indicators designed to monitor and evaluate the quality of pre-hospital emergency care across Europe are (1) availability, (2) reliable access, (3) demand/workload, (4) rate of critical conditions and (5) level of care. Not only are these metrics useful for benchmarking purposes, but since EMS planners seek to implement strategies that enable them to perform well against these metrics, they also significantly influence planning decisions. It is therefore important for all countries to share best practice internationally and seek to continually improve EMS provision within a European framework.

## **Performance estimation**

For EMS planning it is not only important to determine the current or past quality of the EMS system, but also to compute the expected service quality based on the current status and determined changes in order to prove the practicability or uncover the need for (additional) improvements. A typical way of determining the expected service quality is by simulation. The simulation can be used to provide estimates of the proportion of patients that can be reached within pre-defined response times targets, the average response times, the coverage

within a standard time  $T$  or the vehicle utilisation rates.

For the use of simulation it is crucial to have a realistic representation of the EMS system. For example, travel times should be incorporated in a realistic way. For some regions, Euclidean or Manhattan distances might yield good estimates, whereas for more irregular networks, other travel time models should be used. Another important step is the generation of calls. As it can be difficult to estimate the demand, an alternative could be to use trace-driven simulation where call streams are extracted from historical data. In this way, a particular period of time can be evaluated with the new configuration. A final example of modelling choices in simulation studies is the relocation policy. Since most EMS systems use at least some form of dynamic ambulance management, it is important to incorporate this in the simulation. However, often it is not clear under what circumstances relocations are executed. In order to realistically simulate the EMS system, some relocation rule should be implemented.

As a consequence of the high level of detail that can be incorporated in simulation models, most simulation studies focus on one specific ambulance region. The simulation tools are often not easily transferable to other regions. Two notable exceptions are Henderson & Mason (2004) and Kergosien et al (2015). The first introduces BartSim, a simulation tool that was originally developed for the region of Auckland, New Zealand. Later, BartSim formed the basis for a more general simulation tool commercialized by Optima (now belonging to the U.S. company Intermedix), which is the current market leader in EMS simulation software. Their software is applied to EMS systems in many different countries. The simulation tool “Optima Predict” and its application in Denmark is described in (Mason, 2013), for example. Kergosien et al (2015) have also proposed a generic discrete event simulation-based analysis model that can be adapted to a wide range of EMS facilities. In particular, it considers how to optimally serve emergency requests in addition to patient transports between their homes and other medical facilities. Other papers where simulation is used to evaluate scenarios are typically more region specific. The evaluated scenarios are proposed by decision makers (Aboueljine et al, 2014), ILP models (Aringhieri et al, 2013, De La Mota et al, 2015), or heuristics (Jain & McLean, 2003).

Although the patient survival rate reflects the ability of an EMS system to meet its primary objective of saving lives, very few simulation models have considered it until recently, due to the difficulty of linking quantitative measure of survival rate to feasible changes in the rescue process. Inoue et al (2006) initially proposed overcoming this difficulty by approximating the survival rate for severely injured patients using a sigmoid curve and both Sacco et al (2005) and Wang et al (2012) have more recently considered estimating the degradation rate, which is the declining probability of patient’s survival rate as he waits for appropriate care to be administered, through the Delphi method. A simulation model for locating EMS by incorporating survival functions for capturing multiple classes of heterogeneous patients has further been proposed by Knight et al (2012), which aims to maximise the overall expected survival probability of multiple classes of patients with different medical conditions with corresponding survival functions. The model is demonstrated using data from the ambulance service in Wales, which is now beginning to issue reports on clinical based outcome performance measures alongside response times.

Aboueljine et al (2013) provides a more extensive overview of published papers on the topic of EMS computer simulation.

## Planning levels

In general, three different levels can be defined for EMS planning that differ in the time horizon the decision is made for: (1) strategic, (2) tactical and (3) operational.

Planning level	Time horizon	Planning problems
Strategic	Yearly or longer	Locating bases Number of ambulances Hiring crew Long-term demand forecast
Tactical	Monthly or weekly	Locating ambulances Staff scheduling (dispatchers, paramedics/EMTs) Schedules for emergency doctors
Operational	Daily	Relocating ambulances Assigning ambulances to calls Daily forecast Patient transport scheduling Handling unavailability of crew and vehicles

Table 1: Planning problems at the different planning levels

**Strategic level** Decisions at the strategic level are usually made for several years or even decades. This includes, for example, the construction of new buildings for a dispatching center or base stations. In addition, also the design of the EMS regions within a country or a federal state is a strategic decision. Hiring crew can also be seen as part of this phase as contracts are usually envisioned for several years.

**Tactical level** Decisions at the tactical level often hold for periods of one month up to one year. Typical problems are the number of ambulances (per base and overall) and the crew scheduling and rostering, including the number of dispatchers per shift. Often, the ambulance location problem is solved simultaneously for the strategic and the tactical level.

**Operational level** Decisions at the operational level are made on a daily bases or even in real time. Examples are the dynamic relocation of ambulances throughout the day which have to be determined in real time or the assignment of crew to ambulances. Also the assignment of ambulances to emergency calls can be a problem at that level, if dispatchers are allowed to deviate from the closest-idle dispatch policy. In addition, it is necessary to deal with the unavailability of crew and / or vehicles on a daily bases.

Table 1 gives an overview over the EMS planning problems at the different levels that are discussed in this paper.

## EMS planning

EMS planning can be divided into three main parts: the general design of the services, the logistics to fulfil the services and the analytics of the services. Figure 2 gives an overview over these parts. The design of the services is mainly determined by existing laws and regulations and additionally by the decisions which services are offered for the corresponding EMS region. In addition, general decisions have to be made, for example, whether the closest ambulance must always be assigned. The logistics shall assure that the designed services can be offered as intended while fulfilling existing laws and regulations. The main services are of course emergency rescue and patient transport, but also the hotline of the dispatch center can be considered a service, especially if they also inform about on-duty doctors/pharmacies or serve calls from in-house emergency call systems. Analytics mainly includes two aspects: (1) forecasts for the logistical planning problems and (2) data analyses of historic data to control the fulfilment of laws and regulations and the provided service levels. Statistics for last year's data are often used to check whether the laws and regulations were fulfilled. OR covers the logistics as well as

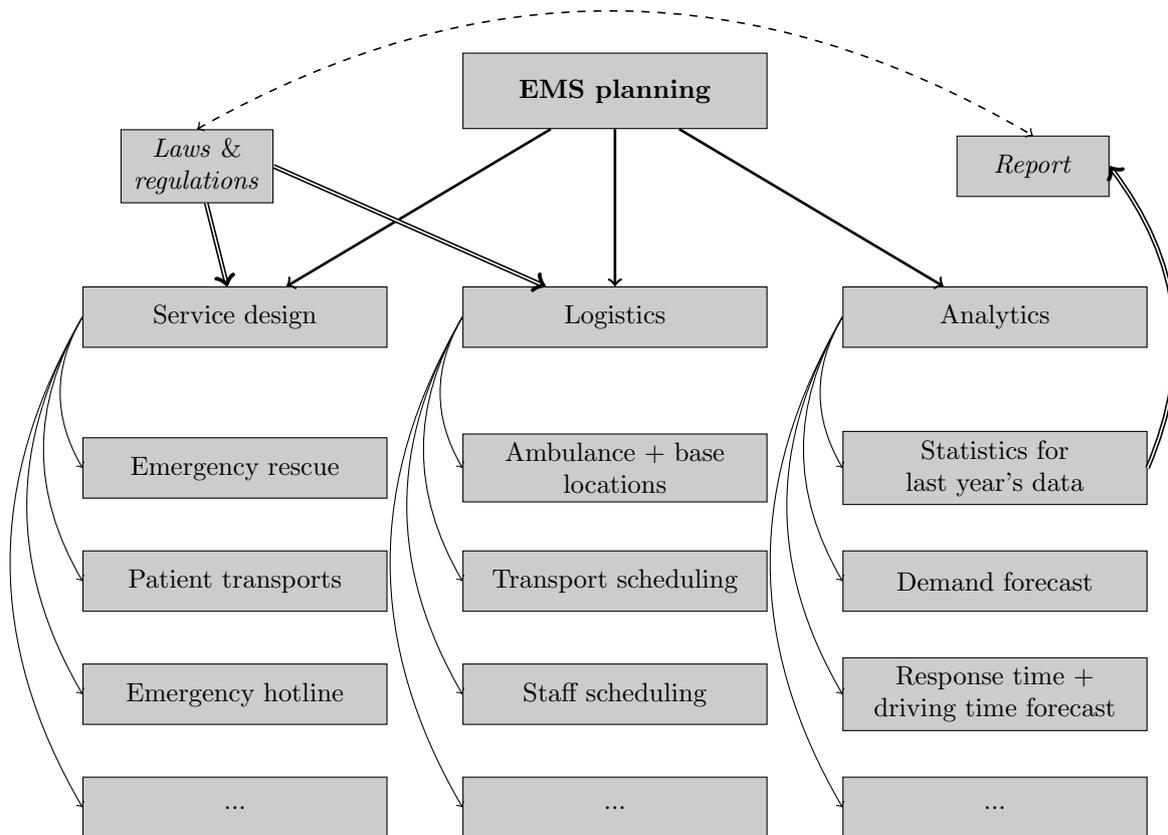


Figure 2: EMS Planning - different involved disciplines, assigned planning problems and additional aspects EMS planning includes Service design, Logistics and Analytics. The three disciplines themselves contain different aspects. Service design as well as the logistics are mainly defined by laws and regulations. These should match the (yearly) EMS report that is based on the statistics for last year's data.

the analytics. Therefore, this will be the focus of this paper.

Figure 3 shows the interdependencies that exist between the planning problems arising for EMS systems in a compact version, while Figure 4 does so in a more detailed manner. Both figures include the planning levels as the vertical axis. On top one can find the strategic level, then the tactical and at the bottom the operational level. Obviously, all the EMS planning problems need some form of data input. At the strategic and tactical level corresponding forecasts are used to determine and model the demand for emergency and transport services. At the operational level, the actual demand in real-time is the main input. Often, also operational forecasts are necessary, for example the expected emergency demand for the upcoming hours is important when deciding about ambulance relocations or possible return transports are included as dummy jobs for patient transport planning. At all three levels ambulance planning problems exist.

Besides the relations between planning problems at different levels (e.g. the number, location and relocation of ambulances) important dependencies exist between the number of ambulances and the need for staff. The resulting shifts have an influence on tactical ambulance locations and relocations as well as on the planning of patient transports.

Which methods need to be used for tackling the planning problems is often defined by the applied strategies. At the strategic level, dispatching strategies and assignment strategies can influence the planning problems.

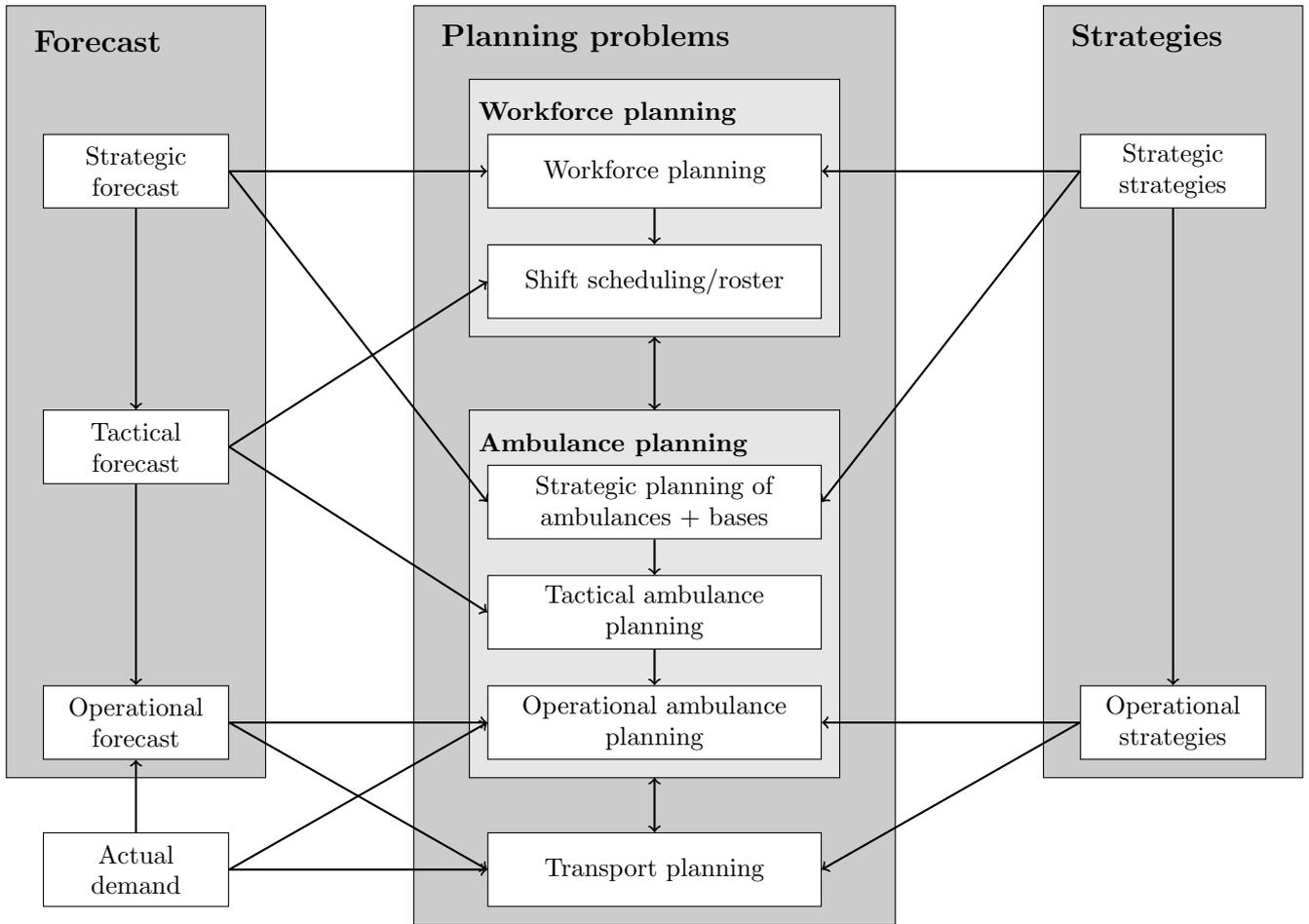


Figure 3: Compact representation of dependencies between EMS planning problems

Here, dispatching strategies mainly include: (1) if the tasks of call taking and dispatching are fulfilled simultaneously by the same staff members or if there are divided and (2) if the call taker stays on the phone until the ambulance arrives at the scene as both may influence the time that a staff member is occupied (when answering a call). Here, strategies for the assignment of ambulances to calls is of primary interest. Examples are: (1) is always the closest ambulance sent to an emergency or (2) is there a strict separation between ALS (emergency) and BLS (patient transports) or do ALS also do patient transports if necessary - this also correlates with the decision on different types of ambulances installed. Of course, ambulance assignment strategies applied at the operational level should go along with the ones used at the strategic level to assure the desired performance. At the operational level, relocation strategies determine (1) whether relocations are fulfilled at all (2) if relocations are determined beforehand or in real-time (3) where and when relocations are applied (to a fixed set of locations; only when an ambulance returns to a base or when also when an ambulance starts serving a call).

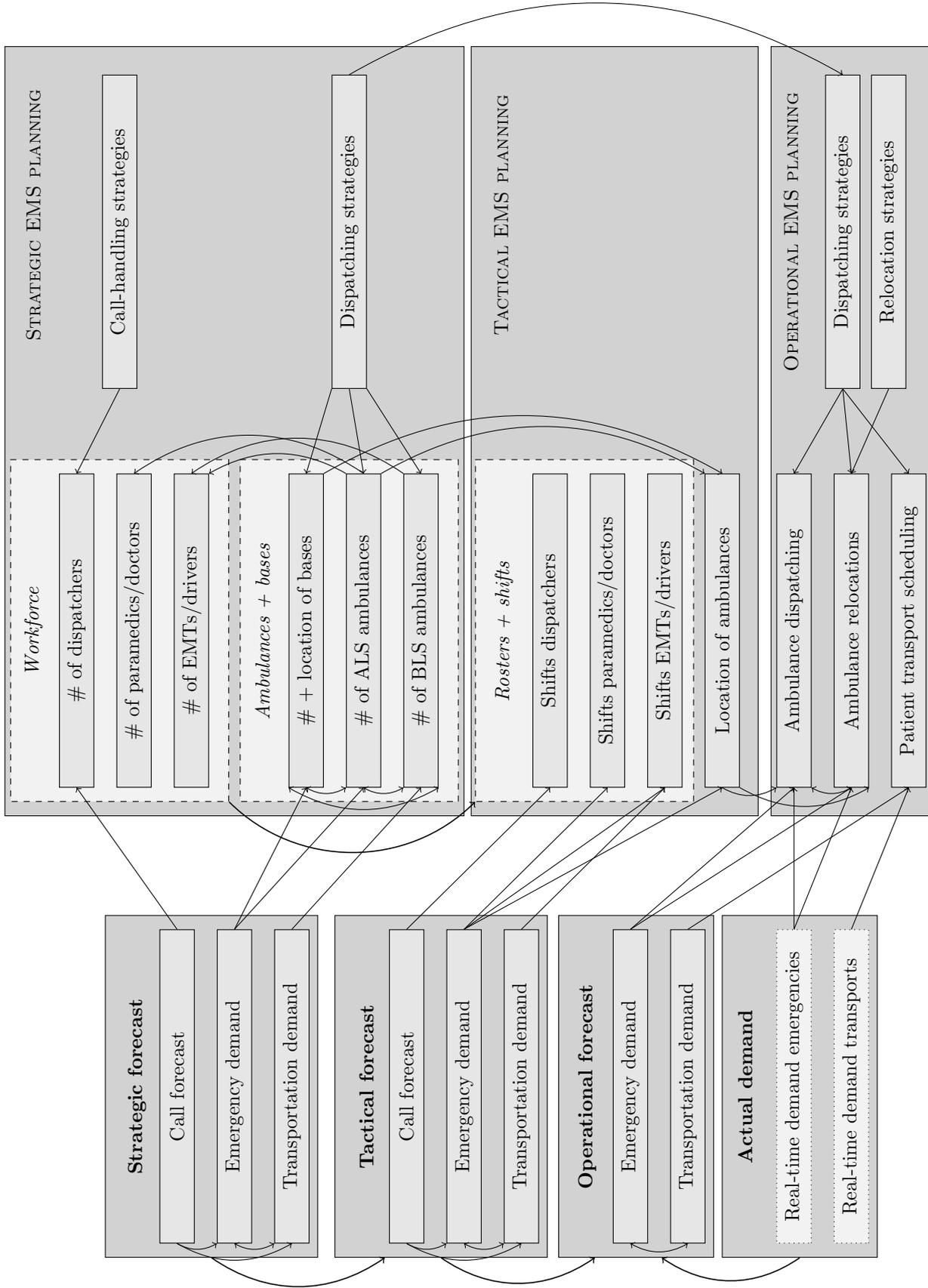


Figure 4: EMS Planning - interdependencies between planning problems

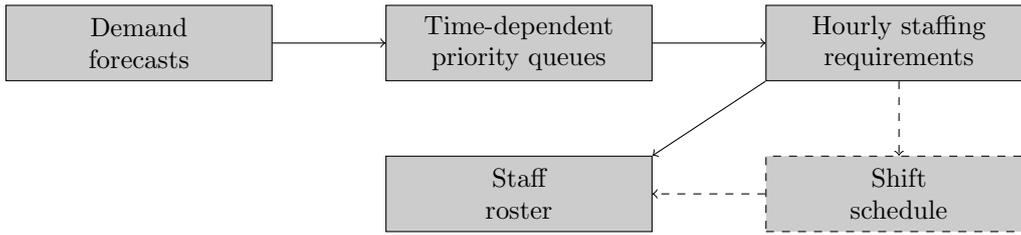


Figure 5: Staff Scheduling (c.f. Vile et al (2016))

To end this section we want to give an example that shows the interdependencies for one specific problem arising in EMS planning. For the probabilistic staffing problem, Vile et al (2016) presented the dependencies of related planning problems as shown in Figure 5. It can be seen that when considering one specific problem more detailed information about the included steps is of interest. For certain ambulance location problems, for example, the predicted emergency demand is used to determine a so-called busy fraction for the ambulances. In addition, the number of available ambulances needs to be known, either by solving a strategic problem beforehand or by just taking the actually existing vehicles into account. Unfortunately, a figure including all the possibilities and “side-problems” would not be helpful any more. Therefore these must be defined for the specific planning problems whenever necessary, as addressed by Vile et al (2016), for example.

## Planning problems

In this section, the different planning problems arising for emergency medical services are described in more detail and emerging topics for future research are presented. For each problem, a literature review is given showing the existing approaches and methods. The structure of this section follows Figure 3. First, we describe the forecasting process, which corresponds to the left hand side of this figure. Second, the strategies from the right hand side of the figure are discussed. Then, the middle part of the figure is discussed in three parts: workforce planning, ambulance planning, and patient transports. Finally, Table 2 summarizes the content at the end of this section.

### Forecasting

While accurate forecasts are important for EMS logistics on all three planning levels, existing literature mainly focuses on the strategic level. Several authors have extracted hourly, daily and monthly trends in ambulance demand; and forecasted forward assuming that future demand will behave in a similar way (Brown et al, 2007, Kamentzky et al, 1982, Setzler et al, 2009). Whilst several regression models have been published that successfully explain demand over large regions over long periods of time, they are only useful for strategic planning and budgeting. In order to provide real-time decision support for dynamic ambulance deployment and hourly operational deployment plans, attention has only more recently shifted to short-term forecasts. These models suggest that factors such as the weather and the total volume of calls in the previous day can significantly improve the 1- to 7-day forecasts of EMS demand (Channouf et al, 2007, Matteson et al, 2011, Wong & Lai, 2014).

Mainly two areas are distinguished in literature, demand forecast and workload / service time forecast. We have structured this section accordingly. Demand forecasts on the strategic and tactical level are the main input to locate bases and determine the number of necessary ambulances. On the operational level demand forecasts

are essential for efficient relocations and assignments of ambulances. The other type of forecast considers the workload and service times including driving times. For those forecasts, accuracy on the operational level is important in order to determine the expected availability of an ambulance, while predictions on the strategic and tactical level are usually more basic.

## **Demand forecasting**

In order to develop effective EMS deployment strategies at the strategic and tactical level, it is essential to first assimilate accurate predictions of demand per time period (e.g. hourly or at the required granularity); yet despite the potential of advanced statistical models to offer accurate demand forecasts, most ambulance service providers still use rudimentary prediction methods when developing strategic plans. Typically, these methods involve dividing the week into 168 one hour increments, accumulating historical records of service requests and evaluating the number of calls received during each hour of the week (Matteson et al, 2011). In Wales, for example, the highest value for each hour of each day in each 10 week period in the previous 50 weeks is observed and the average of these is selected as the ‘average peak demand value’. The number of ambulances then deployed for this hour in future weeks is based on the concept that there must be a sufficient number to cope with such demand. In Germany however, a more complex risk-based method by Behrendt & Schmiedel (2002) is often used to determine the maximum number of ambulances needed. They do not consider the average demand but the more unlikely event of several emergencies happening simultaneously. This is modelled by a Poisson distribution. The number of needed ambulances is determined in such a way that the probability of having to serve more emergencies simultaneously than ambulances exist is below a certain threshold.

The demand for EMS has the characteristics of the essentially random occurrence of individual calls with historically discernible seasonal patterns (Vile et al, 2016) and an underlying sustained increase over the past 20 years (Lowthian et al, 2011b). Several research papers have suggested different methods to account for such fluctuations including linear, sinusoidal and support vector regression (Chen et al, 2015), simple moving averages and more complex time series approaches that allow inclusion of neighbouring hours in the forecast (Baker & Fitzpatrick, 1986, Matteson et al, 2011). Integrated solutions have also been presented that both estimate ambulance demand and recommend deployment plans using queueing theory, simulation models and theoretical distributions (Bell & Allen, 1969, Larson, 1974, Rajagopalan et al, 2011).

Since the late 1980’s, classical time series models such as Autoregressive Integrated Moving Average (ARIMA) and Holt-Winters methods have been used extensively to forecast call volumes (Andrews & Cunningham, 1995, Bianci et al, 1993, Holcomb & Sharpe, 2007) and specifically applied to ambulance demand in Channouf et al (2007). These models however require restrictive data assumptions.

As ambulance deployment strategies at the operational level are ever increasingly becoming more flexible and dynamic in nature, the past decade has seen the development of more responsive model-free methods, which not only offer more accurate short-term call volume predictions, but can also integrate with location and deployment models. In conjunction with evaluating the potential of conventional time series methods to predict future demand levels, Vile et al (2012) have recently considered the potential of Singular Spectrum Analysis (SSA) to produce accurate forecasts whilst adequately accounting for non-stationarities. They show that it considerably outperforms traditional methods for long-term forecasts and offers at least comparable forecasts for a short term planning horizon. Artificial Neural Networks (ANNs) have also been demonstrated to be capable of producing accurate forecasts for small areas by Setzler et al (2009).

## Workload and service time forecast

Whilst a large number of models have been developed to better predict demand for EMS, the most comprehensive mathematical models of EMS systems also take account of how response times and workload are expected to fluctuate over time (Ingolfsson, 2013). The relationship between these components is extremely complex, but the detailed call logs now standardly collected by most modern day EMS providers have supplied operational researchers with a wealth of historical data to analyse and interpret, in order to better understand the effect of behavioural and temporal issues. Response and service time forecasts are mainly important for strategic EMS planning

Each EMS call has an associated response and service time (see Figure 1), which are important for different reasons: the response time is often looked to as an indicator of the quality of service provided by each EMS body and the service times determine the workload on the EMS system. The travel time is usually the largest component of the response time (Ingolfsson, 2013) and has thus unsurprisingly been found to be one of the main factors to influence overall system quality. Hence, by prudently distributing ambulances to bases, ambulance planners are able to somewhat improve their performance (Takedaa et al, 2007). Most statistical analysis of EMS travel times has focussed on either predicting travel time based on the characteristics of distinct road types and travel conditions encountered when travelling from the dispatch location to the scene of the incident (Harewood, 2002, Henderson & Mason, 2004), or based simply on the birds eye distance between both the two points, scaled by correction factors (Aringhieri et al, 2007, Fujiwara et al, 1987). Beyond the standard techniques, operational researchers have also considered modelling this component of service time using graphical analysis, factor scaling (comparing travel time data to Google Maps travel times and distances), cluster analysis to group demand locations and find factors, and cluster analysis to group demand locations so that a significant distributional fit might be found to individual groups.

Fewer studies have focussed on modelling the on scene time, which might be attributable, at least in part, to data limitations (Aboueljineane et al, 2013). Despite the significant time spent on triage and dispatch, some researchers neglect this time in their modelling efforts. Others, such as Maxwell et al (2010), considered deterministic preparation times depending on the initial location of vehicles. Where only partial information is known about the service time, processing times have been modelled in an aggregated form, for example, as in Goldberg et al (1990). More recently, Knight & Harper (2012) have studied the effect of individual components of the ambulance service cycle using Coxian phase-type distributions. By fitting distributions to both the overall cycle time for different classes of patient priorities, as well as to sub-cycles, they were able to identify expected gains from adjusting specific aspects of the response process on the overall efficiency of an ambulance service. Ultimately, the insight they offered on the benefit of reducing turnaround times pointed towards the need for an entire systems approach given that the congestion in the hospital impacts on the ED and in turn on EMS turnaround times.

Further work in this area could involve studying how load-dependent average service times could be incorporated into mathematical models on EMS systems. Ingolfsson (2013) has already shown that chute time appears to decrease with load, whilst hospital time increases, but such observations have not yet been incorporated into time-varying models of the system.

## Strategies

For EMS systems in practice as well as EMS planning strategic decisions on dispatching, deployment and relocation strategies must be made. These decisions on the strategies then influence planning problems on the strategic and the operational levels.

## **Call-handling strategies**

When optimising the ambulance distribution, most research focusses on adjusting the travel time component with a fixed or ‘known’ pre-trip delay time. However, the pre-trip delay time can greatly influence the overall travel time. Hence by adequately managing the EMS call center, some reduction could be achieved. Typically, incoming calls first receive complete triage before an ambulance is dispatched. Another approach could be to dispatch an ambulance even before triage is completed. Potentially, quicker response times could be achieved because the pre-trip delay would be reduced. On the other hand, inappropriate ambulance assignment as a result of the incomplete triage could lead to a higher workload. It would be interesting to investigate the overall impact on the performance. Some EMS systems differ between call takers and dispatchers, e.g., the Dutch system as described above. Studies are missing that prove or disprove the efficiency increase when dividing the tasks. This might also depend on the size of the dispatch center, i.e., on the call volume and the number and type of tasks the staff at the dispatch center needs to fulfil.

## **Dispatching strategies**

Many EMS systems demand for an assignment of the closest available ambulance for all emergencies. In others, dispatching decisions can be made based on the severity of the emergency as well as the overall coverage of the system. In any case, the dispatching policy influences the decision on ambulance and base locations on the strategic level as well as ambulance relocations and the actual assignments of ambulances on the operational level. In addition, patient transports must be assigned to ambulances, if they are served in the EMS system. It must be decided when ambulances are assigned to the transport tasks (right before the pick-up or earlier) and which ambulances can be assigned, if there is not a dedicated part of the fleet.

## **Relocation strategies**

While in some countries relocations to basically every location in the region at every point in time are possible (e.g., in the US), in other countries relocations are not wanted at all (e.g., in many German EMS regions). In between these two rather “extreme” policies, relocations might only be allowed to a defined set of locations (e.g., the set of bases) or at specific points in time, for example when an ambulance becomes available again after having finished a service. The strategy that is chosen for an EMS region can be used as input in a simulation to determine the expected service quality and determines the set of applicable relocation approaches that can be used in practice.

## **Workforce planning**

For EMS planning mainly two different sets of staff exist: (1) the staff working in the ambulances (and as emergency doctors) and (2) the staff working in the dispatching center. In smaller German regions it is preferred that dispatchers also work at an ambulance from time to time to not lose contact to practice. In most other countries, the two staff sets are completely disjoint and therefore, the corresponding planning problems can be considered independently.

In general, workforce planning problems have been extensively discussed in literature. A review over existing approaches, methods and application areas can, for example, be found in Van den Bergh et al (2013).

## Ambulance crew

While some of the general approaches might be used for fixing shifts for paramedics and assigning them to ambulances, several papers have already been published that explicitly consider the ambulance rostering problem. Bradbeer et al (2000) discuss the ambulance roster problem and present three approaches that build upon each other and are based on genetic algorithms. They assume that the number and locations of ambulances is given. Defraeye and Van Nieuwenhuyse (2012) analyse the probability that patients wait longer than maximal acceptable times under varying staffing levels using an extension of the Iterative Staffing Algorithm (ISA). The ISA is a simulation-based approach to determine staffing requirements under time-varying arrivals, which targets a stable delay probability throughout the day.

In contrast, several researchers such as Erdoğan et al (2010), McCormack & Coates (2015) and Vile et al (2016) combine the crew rostering and ambulance location problems. In particular, Erdoğan et al (2010) schedule ambulance crews in order to maximize the coverage throughout a planning horizon by first running a tabu search to locate ambulances and then using the output to solve the crew rostering problem. For that they present two integer programming models. Li & Kozan (2009) define two stages for solving the problem. First, shift start times and the necessary number of ambulance staff to be assigned to each shift are determined using a deterministic model. Then, an allocation model assigns all ambulance staff to shifts resulting in a schedule for four weeks. Also Rajagopalan et al (2011) present a two-stage approach for crew rostering and ambulance location planning. In the first stage, they solve a dynamic expected coverage model using tabu search. For the second stage an integer programming model is presented. Jasim (2002) present a set partitioning approach to solve the staff scheduling problem for the New Zealand EMS provider St John. In addition, a “fatigue model” is applied to the optimal solution to predict the impact of working hours on the fatigue that a staff member experiences during a working day.

## Dispatching center

Apart from the potential response time reduction, an efficiency gain could be obtained by improving the call center staffing. For other applications, significant research has been performed on the optimal staffing of call centers (c.f. Koole & Mandelbaum (2002)). Research specific for EMS call centers is limited. One of the few peer-reviewed papers by Kozan & Mesken (2005) introduces a simulation tool that can be used for what-if scenarios to improve the staffing levels in EMS call centers. Dwars (2013) introduces a simulation tool that contains both specialised call takers and dispatchers and generalists that can do both. The tool is designed to find good configurations of call center crews. The main application is to investigate the potential gain of merging call centers. On the one hand, economies of scale can lead to significant savings, whereas the loss of region-specific knowledge can result in longer call durations and lower efficiency. Incorporating this regional knowledge, Dwars (2013) shows that significant efficiency gains can be obtained by merging call centers. In the Netherlands, for example, one can observe a trend of merging call centers. From the 24 call centers that were there some years ago, only ten will remain in the upcoming years.

Despite the mentioned results, it is fair to say that the call center domain of EMS systems is not as well-studied as other domains and contains some good areas for future research.

## Ambulance planning

In this section, we describe different models that can be used for an efficient planning of emergency rescues. First, we describe the problems that arise at a strategic and tactical level. These include models to determine

good base locations and a good distribution of ambulances over the bases. Often, these two problems are solved simultaneously. Most models assume a fixed capacity and try to maximize the performance with the available resources. However, it is also interesting to consider the problem of deciding on an optimal capacity level so as to obtain a minimum performance. Secondly, we give an overview of some models that consider problems at the operational level. This includes, for example, dispatch rules and real-time relocation. The decisions made on the strategic and tactical level are typically considered as input at the operational level. Finally, we highlight some related problems that might have a significant impact on the performance of an EMS system.

### **Strategic and tactical level**

The most important problems on the strategic and tactical level are to determine good base locations and a good distribution of ambulances over these bases. Although the first problem is more a strategic decision and the second more a tactical decision, these problems are often solved simultaneously. In Van Essen et al (2013), approaches for solving the two problems simultaneously or subsequently are presented and compared. When fixing the set of bases, we can use the same models to solve the problem of distributing the ambulances separately. The models are typically formulated in a way to maximize the performance given a fixed set of resources. However, with slight modifications, most of the models can also be used to determine the required capacity to satisfy a minimum performance requirement. The vast majority of models use coverage-based performance measures. These models maximize the fraction of calls that can be reached within a given target response time. This is mainly due to the fact that in almost all countries, EMS providers are assessed on these kind of measures. Nevertheless, there are models that use different objectives. For example, Dzator & Dzator (2013) minimize the average response time by applying the  $p$ -median model (ReVelle & Swain, 1970) to ambulance location.

Two of the first ambulance location models did not incorporate the ambulance distribution. In 1971, Toregas et al (1971) introduced the Location Set Covering Model (LSCM) to determine the minimum required number of bases to cover the entire region within a fixed time threshold. Later, the Maximum Coverage Location Problem (Church & ReVelle, 1974) was introduced to maximize the coverage given a limited number of bases. Inspired by these two models, much research was done to include the ambulance distribution in the models. At first, it was assumed that a fixed number of ambulances was required to obtain full coverage. Examples of models of this type are DSM (Gendreau et al, 1997), BACOP (Hogan & ReVelle, 1986), MALP (ReVelle & Hogan, 1988) and multi-objective MALP (Harewood, 2002) . After that, the concept of marginal coverage was introduced by Daskin (1983). Here, each additional ambulance covering some area provides some coverage to that region. This model uses expected coverage as opposed to the all-or-nothing coverage of the previous models. Many models were introduced that extended on Daskin's MEXCLP by incorporating time-dependent demand (Repede & Bernardo, 1994, Van den Berg & Aardal, 2015), multiple vehicle types (Chong et al, 2015), stochastic response times (Ingolfsson et al, 2008, Van den Berg et al, 2015), or survival probabilities (Erkut et al, 2008). Alternatively, Beraldi et al (2004), Beraldi & Bruni (2009), and Nickel et al (2015) use Stochastic Programming techniques to obtain robust solutions.

As it is typically necessary to highly simplify the EMS system in order to obtain tractable solutions, optimization is sometimes combined with simulation to ensure solutions that perform well in practice. Approaches in which there is interaction between the simulation and the optimization are called simulation-optimization approaches. Lee et al (2012) iteratively use simulation to estimate busy fractions in a static ambulance location model. With the new busy fraction, a new solution is found for which the busy fraction is estimated by the simulation. Mason (2013) also presents a simulation-optimization algorithm for determining improved base

locations. Finally, in McCormack & Coates (2015) simulation is integrated in a genetic algorithm to give the fitness of the current solution.

A more extensive overview of the literature on ambulance location models can be found in Brotcorne et al (2003), Li et al (2011), and Basar et al (2012). Recently, Ahmadi-Javid et al (2016) presented a review of more general health care facility location problems. In this paper, also non-emergency facilities are discussed. A computational comparison of some of the basic ambulance location models is given by Erkut et al (2009) and Van den Berg et al (2016).

### **Operational level**

At the operational level of the planning, real-time decision should be made, such as which ambulance to send to a call and how to relocate the remaining vehicles.

**Real-time locations** Gendreau et al (2001) were one of the first to address the real-time ambulance location problem. The proposed model was a dynamic version of the static Double Standard Model. It incorporates the current state of the system in finding good relocations. Whenever a redeployment decision must be made, the adapted version of DSM is solved. A similar approach is used by Gendreau et al (2006), where MEXCLP is solved in stead of DSM. Over the last ten years, many models were introduced that were specifically designed to capture the dynamics of an EMS system. Zhang et al (2008) solve the real-time relocation problem for a small number of ambulances by Dynamic Programming. For larger instances, their model suffers from the curse of dimensionality. Bjarnason et al (2009) evaluate policies using a simulation tool. Based on the results of the simulation, an optimization tool is used to find better policies. As opposed to most models that significantly simplify the system, Maxwell et al (2010) include as many details of the real system as possible and apply approximate dynamic programming (ADP) to find good relocation policies. A simulation model is used to assess the performance of a given EMS allocation policy. ADP is further used by Schmid (2012) to find dispatch and relocation policies in case travel times and call rates fluctuate over time. However, redeployment decision are limited to the moment an ambulance becomes available after finishing a call. Yue et al (2012) embed a simulation in a greedy algorithm to obtain the objective value of solutions. In the optimization, the simulation is called every time the value of a solution is requested. Alanis et al (2013) pose a two-dimensional Markov chain to evaluate the system given a compliance table. A simulation-optimization framework for allocating ambulances to bases and relocating them throughout the day was proposed by Zhen et al (2014). In the simulation a stochastic environment is implemented, in which the arrival time of requests, travel time and on scene time are uncertain. For the optimization a genetic algorithm is used. Jagtenberg et al (2015) introduce a heuristic in which an ambulance is sent to the base where it provides the highest marginal coverage according to the MEXCLP objective function.

**Dispatching** Besides redeployment decisions, real-time decisions must be made on which ambulance to dispatch to a call. Even though Carter et al (1972) has already shown that it is not always optimal to dispatch the closest idle ambulance, it is still by far the most common dispatch rule (as highlighted in our overviews of the UK, German and Dutch systems). This assumes knowledge about the locations of the available ambulances. As observed by Dean (2008) and mentioned in the description of the German system, this information is not always present. One notable exception of the closest idle dispatch rule is Andersson & Värbrand (2007), which adopt alternative dispatch rules for low priority calls. However, they do not try to find optimal dispatch rules. Schmid (2012) uses approximate dynamic programming to find dispatch policies, and find that deviating from the closest idle dispatch rule for non life-threatening calls can improve the overall performance. Ibri et al (2012)

were one of the first to consider the coverage while dispatching ambulances. They propose a decentralized distributed solution approach as a multi-agent system. Lee (2011, 2012a,b, 2013, 2014) studies the dispatching for the U.S. EMS system where only part of the patients are actually transported and ambulances therefore often drive from one demand location to the next. First, the preparedness for future demand is considered when deciding which ambulance to assign to an emergency (Lee, 2011). As a next step, the concept of centrality is introduced (Lee, 2012a,b, 2013). The centrality is determined for each demand node based on the distances and distribution of the current calls. For the dispatching closeness and centrality are combined. To test the approaches generated instances based on a 5 by 5 demand grid with 25 nodes are used. In another extension, not only idle but also busy ambulances are included that reduces the response time again for the generated test instance (Lee, 2014). Bandara et al (2014) include the severity of calls into the dispatching decision with the aim of increasing the survival probability of the patient. The authors use a function proposed by Larsen et al (1993) to determine the survival probability. They propose a heuristic to apply the dispatching strategy to large-scale EMS systems. Sudtachat et al (2014) also study multiple priorities, but additionally different types of ambulances that can be assigned. The authors specifically consider three priorities and two types of vehicles, ALS and BLS, and assume that it can also happen that one ambulance of both types is assigned to one call. The objective is to maximize the overall expected survival probability of patients with a “life-threatening” emergency. Haghani et al (2003) use a simulation to compare different dispatching strategies. They focus on the position of the call in the waiting queue in case all ambulances are busy or no ambulance can reach the scene in the response time. In general, the closest available ambulance is assigned. Also Van Buuren et al (2012) present a simulation tool for EMS dispatchers in the Netherlands to evaluate different dispatch strategies. A review of dynamic ambulance relocation models from the perspective of dispatch policies is given by Lim et al (2011). The authors then present an adapted version of the MCLP for the ambulance location problem and assign ambulances to calls in a simulation study. In the simulation they dispatch ambulances considering the priority and the closeness, as these they identified these two policies as the most common. A Markov Decision Process is proposed by McLay & Mayorga (2013b). The model determines how to optimally dispatch ambulances to patients in order to maximize the the expected coverage of high-risk patients. The optimal policies are tested with real-world data. The authors show that it is not always best for the system to dispatch the closest ambulance to a patient. The authors extend the model in a second paper and consider the impact of equity constraints by using linear programming models and algorithms (McLay & Mayorga, 2013a) and four potential equity measures, two for patient and two for server equity. Zarkeshzadeh et al (2016) propose a hybrid method with a linear combined metric that is based on network centrality measures, the nearest neighbor method and the first-in first-out (FIFO) policy. Additional parameters such as the operating environment, rate of incoming emergency calls, available resources, hospitalization probability of the patients as well as distances and locations of units are also considered in the approach. A simulation is used to test the proposed metric.

### **Other services**

Besides the problems that arise in any EMS system, there some aspects of EMS systems that are not always present. Depending on the system, these can have a significant impact on the performance. We discuss three here: emergency doctors, helicopters, and hospital drop-offs.

**Emergency doctors** In Franco-German EMS systems, determining the number and locations of emergency doctors might be of interest. This is of particular interest in Baden-Württemberg as they also have defined a response time target of 15 minutes in 95% of the cases for emergency doctors. Therefore, their locations are crucial. To the best of our knowledge, there are no publications on explicitly locating emergency doctors using

OR to be found in the literature. A main reason is probably that emergency doctors usually work in hospitals or practices while being on duty and these locations cannot be changed. In addition, it is not possible to simply assign further emergency doctors as a specific training is a prerequisite for working as an emergency doctor that is time consuming as well as expensive. Nevertheless, the assignment of shifts is an important task, especially if in total more doctors and locations are available to choose from than necessary for each shift. This can result in a combination of a simple maximum coverage problem to make sure that the considered region is covered (as good as possible) with a shift scheduling problem. If, for example, there are only two emergency doctors in an area, these two should not have overlapping duties.

**Helicopters** In many countries, helicopters are used in the most severe cases. Typically, the helicopter is not the first responder, but is used to provide more specialized medical assistance. This is mainly due to the high start-up times of helicopters. A land ambulance is used for the first response and if necessary, a specialized doctor arrives by helicopter. The helicopter can then also be used for transport to a hospital or trauma center. In many cases, the patient is transported by the land ambulance to a suitable place for the helicopter to land. One notable exception is the region of Ontario, Canada, where aircrafts and helicopters are also used for non-urgent patient transports. For this region, Carnes et al (2013) develop a model to better schedule the aircrafts that are used for these transports. For the case where helicopters are only used for trauma patients, Erdemir et al (2010) introduce a MCLP-based model for the simultaneous optimization of land and air ambulances. Here, a patient can be served by a land ambulance, a helicopter or both. Cho et al (2014) optimize the location of the helicopters as well as the location of the trauma centers. The trauma centers can only be located at specific existing hospitals. Furuta & Tanaka (2014) consider the case where land transport is also necessary and the ambulance and helicopter meet at a rendez-vous point. The goal is to reduce the access time for specialized care compared to the case where only land ambulances are used. Given the land ambulance distribution, the best location for helicopters and rendez-vous points is determined.

**Drop-off at hospitals** It is in the interest of all ambulance trusts, patients and health care workers for a patient to experience a swift handover of care, since a long handover not only wastes valuable resources but can also be harmful to patients, whose condition might deteriorate whilst waiting in the ambulance bay. This is, however, not always possible and conflicting targets for the ambulance service and the emergency departments are not always conducive to small handover times. In order to promote swift handovers, some European countries issue target times for ambulance personnel to transfer patient care to the ED (for example, this target is fifteen minutes in the UK). Some countries additionally issue ‘turnaround’ targets, which are typically a few minutes longer since these also incorporate time for the crew to fully replenish the ambulance so it is ready to attend another emergency call. Whilst the exact formulation of the targets varies from country to country, their aims are consistent, in as much to reduce costs, and improve both patient outcomes and the equitable use of resources (of vehicles and crews).

In the majority of European countries, it seems that the turnaround targets are attained at a reasonable level and therefore, patient handover is currently not a major area of concern. However, it is a notorious problem in the UK, as well as for some areas of the USA and Canada. In fact, the number of ‘lost’ ambulance hours due to long handovers has been estimated to have cost the UK NHS millions of pounds a year. In Wales alone, there has been a five-fold increase in ‘lost’ ambulance hours in the last few years (from around 8,000 in 2008 to 40,000 in 2014); hence it is not surprising that this has been closely monitored by the media (Clarke, 2015, Hughes, 2009, Jones, 2011). Not only is this money wasted that could be better used within the service, distressed patients spend long periods of time waiting for transfer of care, resulting in potential deterioration

in their condition and effectiveness of subsequent treatment. Furthermore, whilst waiting to handover patients, the crew's vehicles are blocked, which results in decreasing coverage (Lowthian et al, 2011a) and puts lives at risk.

The effect of reducing turnaround time on performance has been investigated in several simulation studies, which have in the main focussed on its impact on response time as this is the common measure for EMS systems and comparable across the countries (Knight et al, 2012). However, since some ambulance services are moving to clinical outcome based measures (e.g. the Welsh Ambulance Service Trust, WAST), it is also of interest to see if survival between different scenarios alters. Investigations into the effect of reducing the turnaround time have been undertaken in several studies; notably with Knight & Harper (2012) showing that if turnaround times were reduced to the extent that the government targets were met in Wales, this would be equivalent to 15% extra capacity on the ground.

As mentioned before, patient handover is also a critical issue in Canada. Therefore, Carter et al (2015) propose the introduction of offload zones in hospitals to shorten the drop-off time while also controlling the workload in the ED. They describe it as an additional area next to the ED where a nurse looks after patients that were taken to hospital by an ambulance.

## Patient transports

When patients need to be transported to, from or between hospitals this is often organized by the EMS provider, sometimes even regulated by (EMS) law, as, for example, in Germany. If medical attention is needed during the transportation phase, an ambulance is sent. In Germany, this is called a qualified patient transport. Sometimes, intensive care is needed which result in the use of specially equipped vehicles and trained personnel. If medical attention is not necessary, in Germany and other countries also taxis or private transport companies can fulfill the tasks. This is then called unqualified patient transport. Usually, taxis as well as private transport companies organize themselves. Therefore, they are excluded in this section. Nevertheless, they also have to determine routes and schedule their transports. Therefore, similar methods to those described in this section can be used to tackle their planning problems.

A transport task involves picking up a patient at one location and dropping them off at a second location. Often, for one of the two actions a time window is given. Depending on the regulations in the country, these time windows are hard and must be fulfilled or soft and might be violated. In the latter case, minimizing the violations is often (part of) the objective function. In general, a distinct set of ambulances is reserved to fulfil the transport tasks. In that case, there are mainly two decisions to be made: (1) tasks must be assigned to the ambulances and (2) the routes for the ambulances must be constructed.

If not all patients can be served by the set of ambulances, some tasks need to be assigned to ALS ambulances. This results in higher costs and coverage reduction and should therefore be prevented, if possible. The underlying problem can be expressed with a dial-a-ride (DARP) formulation.

The DARP itself is already quite well studied, see, for example, Cordeau & Laporte (2007) and Parragh (2009). There are only a few publications that study models and approaches for planning the patient transport problem using OR. Ritzinger & Puchinger (2012), Schilde et al (2011) and Parragh et al (2009) study patient transports in the Austrian EMS system, focusing on the two relief organizations Arbeiter Samariter Bund (ASB) in Vienna and the Austrian Red Cross. They provide different DARP formulation and solution approaches. Parragh et al (2009) include two different types of vehicles having different capacities for transport tasks and they additionally assign drivers to vehicles. In the system presented by Schilde et al (2011), only transports between patients' home locations and hospitals are considered with patients, instead of hospitals and doctors,

ordering the transports. Ritzinger & Puchinger (2012) assume about 1,000 transports per day. Therefore, an approach is needed that is explicitly fast. Unfortunately, this is usually going along with some compromises on solution quality.

Of course, solving a DARP formulation to optimality can only be applied in practice if all patient transport tasks are known in advance, e.g., the night before, and if the problem size can be solved during the time available. If this is not the case, heuristics might be more reasonable for a use in practice. If none or only part of the tasks are known in advance and the rest becomes known throughout the day, scheduling the tasks resembles an online problem. For this case, Kergosien et al (2011) introduce a Tabu Search heuristic which is called every time a new call arrives.

Kergosien et al (2015) have proposed a generic discrete event simulation-based analysis model that simultaneously decides how to schedule patient transports between their homes and other medical facilities as well as how to optimally serve emergency requests.

Especially in Germany, having a coordination platform for planning the routes can be expected to be highly beneficial for the EMS system. A dispatcher would get decision support for assigning the tasks while the tool determines possible solutions in real-time if needed. The outline of such a coordination platform for the use in Germany is described in Reuter-Oppermann et al (2015), for example.

## Summarising table

The following Table 2 summarises the literature presented in this paper and matches it to the taxonomy presented in Figure 4.

	Forecasting			EMS Planning											
				Strategic				Tactical				Operational			
	Strategic	Tactical	Operational	Staff		Ambulances			Shifts			Dispatching	Relocation	Transports	
				Dispatchers	Paramedics	EMTs	Bases	# ALS	# BLS	Dispatchers	Paramedics				EMTs
Aboueljane et al (2014)				x				x	x	x			x		
Alanis et al (2013)		x												x	
Andersson & Värbrand (2007)													x	x	
Aringhieri et al (2007)							x	x							
Aringhieri et al (2013)							x							x	
Baker & Fitzpatrick (1986)	x	x													
Bandara et al (2014)														x	
Behrendt & Schmiedel (2002)	x						x	x					x		
Bell & Allen (1969)	x	x	x												x
Beraldi et al (2004)							x						x		
Beraldi & Bruni (2009)							x						x		
Bjarnason et al (2009)															x
Bradbeer et al (2000)					x	x						x	x		
Carter et al (1972)														x	
Channouf et al (2007)	x	x													
Chen et al (2015)	x	x													
Cho et al (2014)							x								
Chong et al (2015)								x	x						
Church & ReVelle (1974)							x	x					x		

	Forecasting			EMS Planning												
				Strategic					Tactical			Operational				
	Staff		Ambulances			Shifts										
	Strategic	Tactical	Operational	Dispatchers	Paramedics	EMTs	Bases	# ALS	# BLS	Dispatchers	Paramedics	EMTs	Distribution	Dispatching	Relocation	Transports
Daskin (1983)							x	x					x			
De La Mota et al (2015)							x									
Dean (2008)														x		
Dwars (2013)				x						x						
Dzator & Dzator (2013)							x	x					x			
Erdemir et al (2010)							x									
Erdoğan et al (2010)				x	x		x	x			x	x	x			
Erkut et al (2008)							x	x					x			
Fujiwara et al (1987)													x			
Furuta & Tanaka (2014)							x									
Gendreau et al (1997)							x	x					x			
Gendreau et al (2001)															x	
Gendreau et al (2006)															x	
Haghani et al (2003)														x		
Harewood (2002)							x	x								
Henderson & Mason (2004)							x	x					x			
Hogan & ReVelle (1986)							x						x			
Ibri et al (2012)														x	x	
Ingolfsson (2013)							x	x					x			
Ingolfsson et al (2008)													x			
Jagtenberg et al (2015)															x	
Jasim (2002)				x	x						x	x				
Kergosien et al (2011)																x
Kergosien et al (2015)														x		x
Knight et al (2012)													x			
Kozan & Mesken (2005)				x						x						
Larson (1974)	x	x	x												x	
Lee (2011)														x		
Lee (2012b)														x		
Lee (2012a)														x		
Lee (2013)														x		
Lee (2014)														x		
Lee et al (2012)								x								
Li & Kozan (2009)				x	x						x	x				
Lim et al (2011)							x	x						x		
Lowthian et al (2011b)	x	x														
Mason (2013)								x								
Matteson et al (2011)	x	x														
Maxwell et al (2010)															x	
McCormack & Coates (2015)								x					x			
McLay & Mayorga (2013b)														x		
McLay & Mayorga (2013a)														x		
Nickel et al (2015)								x	x				x			
Parragh et al (2009)																x
Rajagopalan et al (2011)	x	x	x		x	x	x	x			x	x	x		x	

	Forecasting			EMS Planning												
				Strategic					Tactical			Operational				
	Staff		Ambulances			Shifts										
	Strategic	Tactical	Operational	Dispatchers	Paramedics	EMT's	Bases	# ALS	# BLS	Dispatchers	Paramedics	EMT's	Distribution	Dispatching	Relocation	Transports
Repede & Bernardo (1994)							x	x					x			
Reuter-Oppermann et al (2015)																x
ReVelle & Hogan (1988)							x	x					x			
ReVelle & Swain (1970)							x									
Ritzinger & Puchinger (2012)																x
Schilde et al (2011)																x
Schmid (2012)													x	x	x	
Setzler et al (2009)	x	x	x													
Sudtachat et al (2014)														x		
Takeda et al (2007)							x	x					x			
Toregas et al (1971)							x	x								
Van Buuren et al (2012)														x		
Van den Berg & Aardal (2015)							x	x					x			
Van den Berg et al (2015)							x	x					x			
Van Essen et al (2013)							x	x					x			
Vile et al (2016)	x	x														
Vile et al (2012)	x	x														
Yue et al (2012)															x	
Carter et al (1972)														x		
Zarkeshzadeh et al (2016)															x	
Zhen et al (2014)							x									
<b>Reviews</b>																
Ahmadi-Javid et al (2016)							x						x			
Aringhieri et al (2017)	x	x	x		x		x	x			x		x	x	x	
Basar et al (2012)							x						x			
Brotcorne et al (2003)							x	x					x		x	
Li et al (2011)							x						x			

Table 2: Literature overview

## Conclusions and recommendations for further research

In this paper, we have overviewed several logistical problems arising for EMS providers. Although demand, response time and workload are often considered separately in the literature, we have highlighted the high degree of interconnection between each of these components. Consequently solving only one problem at a time might not be the best option. Thus, we have advocated papers that have addressed these issues simultaneously. At the very least, hierarchical planning with feedback loops is advantageous although simultaneous planning is preferable. Until recently, most papers have additionally focussed on one particular EMS provider but over the last decade there has been an increase in the emergence of generic models, acclaimed as straightforward to adapt to country specifics. In detailing a “typical” EMS system, we hope that our review will support further research of a generic nature and hence minimise any repetition of research works.

For the use of planning approaches in practice, a good forecast is important. Incorporating the resulting uncertainty is still a research focus when determining the locations of ambulances and bases. In the future, models that also incorporate the impact of borders between neighbouring regions or even countries will be of interest.

The growing number of complex data sources, together with the ever increasing availability of data mining tools, present a number of directions for future EMS studies. For example, in line with the weakening in the perceived importance of hard response time targets when evaluating EMS quality, future studies might treat the coverage requirement as a goal constraint while assessing the impact of changing resource team's positions on clinical based outcome measures. Furthermore, the call centre is a fertile domain for future research, as is the study of load-dependent service times. Finally, it would be interesting to compare the performance of hand generated or commercially developed deployment plans against those generated by complex mathematical models. Since most commercially available software lacks the capability to efficiently combine demand forecasting, deployment and crew scheduling algorithms, it would be interesting and useful in future research to develop integrated decision support tools necessary for practice that aid the dispatcher to make better or faster decisions by offering them different methods and algorithms to choose from.

## Acknowledgements

This research was financed in part by EPSRC grant EP/F033338/1 as part of the LANCS initiative and the Dutch Technology Foundation STW under contract 11986, which we gratefully acknowledge.

The authors would like to thank the organisers of the EURO Summer Institute XXXI conference, at which the stimulus of this paper arose, for providing an excellent forum for the authors to discuss and debate new modelling and solution techniques to aid EMS decision makers. They would also like to express their gratitude to each of their University supervisors for their helpful comments and support, as well as each of the ambulance trusts that have actively provided data, comments and advice for each of their related doctoral projects.

## References

- ABOUELJINANE L, SAHIN E, and JEMAI Z (2013) A review on simulation models applied to emergency medical service operations. *Computers and Industrial Engineering* 66:734–750.
- ABOUELJINANE L, SAHIN E, JEMAI Z, and MARTY J (2014) A simulation study to improve the performance of an emergency medical service: Application to the French Val-de-Marne department. *Simulation Modelling Practice and Theory* 47:46–59.
- AHMADI-JAVID A, SEYEDI P, and SYAM S S (2016) A survey of healthcare facility location. *Computers & Operations Research* pp –.
- ALANIS R, INGOLFSSON A, and KOLFAL B (2013) A Markov Chain Model for an EMS System with Repositioning. *Production and Operations Management* 22(1):216–231.
- Ambulancezorg Nederland (2013) Ambulances in-zicht 2012. Technical report National Institute for Public Health and the Environment.
- Ambulancezorg Nederland (2014) Ambulances in-zicht 2013. Technical report National Institute for Public Health and the Environment.

- ANDERSSON T and VÄRBRAND P (2007) Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society* 58:195–201.
- ANDREWS B and CUNNINGHAM S (1995) L. L. Bean improves call-center forecasting. *Interfaces* 25(6):1–13.
- ARINGHIERI R, BRUNI M, KHODAPARASTI 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.
- ARINGHIERI R, CARELLO G, and MORALE D (2007) Ambulance location through optimization and simulation: the case of milano urban area. In: *The 38th annual conference of the Italian operations research society optimization and decision sciences*, Università degli Studi di Milano, Polo didattico e di ricerca di Crema.
- ARINGHIERI R, CARELLO G, and MORALE D (2013) Supporting decision making to improve the performance of an Italian Emergency Medical Service. *Annals of Operations Research* pp 1–18.
- BAKER J and FITZPATRICK K (1986) Determination of an optimal forecast model for ambulance demand using goal programming. *Journal of Operational Research Society* 37(11):1047–1059.
- BANDARA D, MAYORGA E M, and MCLAY A L (2014) Priority dispatching strategies for ems systems. *Journal of the Operational Research Society* 65(4):572–587.
- BASAR A, ÇATAY B, and ÜNLÜYURT T (2012) A taxonomy for emergency service station location problem. *Optimization Letters* 6:1147–1160.
- BEHRENDT H and SCHMIEDEL R (2002) Ermittlung der bedarfsgerechten fahrzeugvorhaltung im rettungsdienst. *Notfall und Rettungsmedizin* 5(3):190–203.
- BELL C and ALLEN D (1969) Optimal planning of an emergency ambulance service. *Socio-Economic Planning Science* 3:95–101.
- BERALDI P and BRUNI M (2009) A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research* 196(1):323–331.
- BERALDI P, BRUNI M, and CONFORTI D (2004) Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research* 158(1):183–193.
- BIANCI L, JARRETT J, and HANUMARA C (1993) Forecasting incoming calls to telemarketing centers. *Journal of Business Forecasting* 12(2):3–11.
- BJARNASON R, TADEPALLI P, FERN A, and NIEDNER C (2009) Simulation-based Optimization of Resource Placement and Emergency Response. In: *Proceedings of the Twenty-First Innovative Application of Artificial Intelligence Conference* pp 47–53.
- BRADBEER P, FINDLAY C, and FOGARTY T C (2000) An ambulance crew rostering system. In *Real-World Applications of Evolutionary Computing* pp 267–279, Springer.
- BROTCORNE L, LAPORTE G, and SEMET F (2003) Ambulance location and relocation models. *European Journal of Operational Research* 147(3):451–463.

- BROWN L H, LERNER E B, LARMON B, LEGASSICK T, and TAIGMAN M (2007) Are ems call volume predictions based on demand pattern analysis accurate? *Prehospital Emergency Care* 11(2):199–203.
- CARNES T A, HENDERSON S G, SHMOYS D B, AHGHARI M, RUSSELL D, HENDERSON S G, and SHMOYS D B (2013) Air-Ambulance Routing at Ornge. *Interfaces* 43(3):232–239.
- CARTER A J, GOULD J B, VANBERKEL P, JENSEN J L, COOK J, CARRIGAN S, WHEATLEY M R, and TRAVERS A H (2015) Offload zones to mitigate emergency medical services (ems) offload delay in the emergency department: a process map and hazard analysis. *CJEM* pp 1–9.
- CARTER G M, CHAIKEN J M, IGNALL E, and JUN N M (1972) Response Areas for Two Emergency Units. *Operations Research* 20(3):571–594.
- CHANNOUF N, L'ECUYER P, INGOLFSSON A, and AVRAMIDIS A (2007) The application of forecasting techniques to modelling Emergency Medical System calls in Calgary, Alberta. *Health Care Manage Science* 10(1):25–45.
- CHEN A, LU T, MA M, and SUN W (2015) Demand forecast using data analytics for the pre-allocation of ambulances. *IEEE Journal of Biomedical and Health Informatics*.
- CHO S H, JANG H, LEE T, TURNER J, and CHO S H (2014) Simultaneous Location of Trauma Centers and Helicopters for Emergency Medical Service Planning Simultaneous Location of Trauma Centers and Helicopters for Emergency Medical Service Planning. *Operations Research* 62(4):751–771.
- CHONG K C, HENDERSON S G, and LEWIS M E (2015) The vehicle mix decision in Emergency Medical Service systems. *Manufacturing & Service Operations Management* pp 1–45.
- CHURCH R and REVELLE C (1974) The maximal covering location problem. *Papers in Regional Science* 32(1):101–118.
- CLARKE O (2015) 40,000 hours of ambulance delays at Welsh hospital A&Es. *BBC News Wales*.
- CORDEAU J F and LAPORTE G (2007) The dial-a-ride problem: models and algorithms. *Annals of Operations Research* 153(1):29–46.
- DASKIN M S (1983) A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science* 17(1):48–70.
- DE LA MOTA I F, GARDUÑO A V, and PÉREZ E S (2015) Simulation and Optimization of the Pre-hospital Care System of the National University of Mexico. In *Applied Simulation and Optimization* (MUJICA MOTA M, DE LA MOTA I F, and GUIMARANS SERRANO D, eds), pp 233–276, Springer International Publishing Cham.
- DEAN S F (2008) Why the Closest Ambulance Cannot be Dispatched in an Urban Emergency Medical Services System. *Prehospital and Disaster Medicine* 23(02):161–165.
- DICK W F (2003) Anglo-american vs. franco-german emergency medical services system. *Prehospital and Disaster Medicine* 18:29–37.
- DONABEDIAN A (1980) *Definition of Quality and Approaches to its Assessment. Explorations in Quality Assessment and Monitoring*. Health Administration.

- DWARS R P (2013) Capacity planning of emergency call centers. Master's thesis VU University Amsterdam.
- DZATOR M and DZATOR J (2013) An effective heuristic for the P-median problem with application to ambulance location. *Opsearch* 50(1):60–74.
- ERDEMIR E T, BATTÀ R, ROGERSON P A, BLATT A, and FLANIGAN M (2010) Joint ground and air emergency medical services coverage models: A greedy heuristic solution approach. *European Journal of Operational Research* 207(2):736–749.
- ERDOĞAN G, ERKUT E, INGOLFSSON A, and LAPORTE G (2010) Scheduling ambulance crews for maximum coverage. *Journal of the Operational Research Society* 61(4):543–550.
- ERKUT E, INGOLFSSON A, and ERDOĞAN G (2008) Ambulance location for maximum survival. *Naval Research Logistics (NRL)* 55(1):42–58.
- ERKUT E, INGOLFSSON A, SIM T, and ERDOĞAN G (2009) Computational comparison of five maximal covering models for locating ambulances. *Geographical Analysis* 41:43–65.
- FISCHER M, KAMP J, RIESGO L G C, ROBERTSON-STEEL I, OVERTON J, ZIEMANN A, KRAFFT T, GROUP E, and OTHERS (2011) Comparing emergency medical service systems - a project of the european emergency data (eed) project. *Resuscitation* 82(3):285–293.
- FUJIWARA O, MAKJAMROEN T, and GUPTA K K (1987) Ambulance deployment analysis: A case study of bangkok. *European Journal of Operational Research* 31(1):9–18.
- FURUTA T and TANAKA K I (2014) Maximal Covering Location Model for Doctor-Helicopter Systems with Two Types of Coverage Criteria. *Urban and Regional Planning Review* 1:39–58.
- GENDREAU M, LAPORTE G, and SEMET F (1997) Solving an ambulance location model by tabu search. *Location Science* 5(2):75–88.
- GENDREAU M, LAPORTE G, and SEMET F (2001) A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing* 27(12):1641–1653.
- GENDREAU M, LAPORTE G, and SEMET F (2006) The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society* 57(1):22–28.
- GOLDBERG J, DIETRICH R, CHEN J M, MITWASI M, VALENZUELA T, and CRISS E (1990) A simulation model for evaluating a set of emergency vehicle base locations: Development, validation, and usage. *Socio-Economic Planning Sciences* 24(2):125–141.
- HAGHANI A, TIAN Q, and HU H (2003) A simulation model for real-time emergency vehicle dispatching and routing. In: *CD-ROM), Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, DC*.
- HAREWOOD S (2002) Emergency ambulance deployment in Barbados: A multiobjective approach. *Journal of the Operational Research Society* 53(2):185–192.
- HENDERSON S and MASON A (2004) Ambulance service planning: Simulation and data visualization. In *“Handbook of Operations Research and Health Care Methods and Applications”, Sainfort, F., Brandeau, M.L., Pierskalla, W.P. (Eds.), International series in Operations Research and Management Science* 70:77–102.

- HOGAN K and REVELLE C (1986) Concepts and applications of backup coverage. *Management Science* pp 1434–1444.
- HOLCOMB J and SHARPE N (2007) Forecasting police calls during peak times for the city of Cleveland. *CS-BIGS* 1(1):47–53.
- HOOGEVEEN M (2010) Ambulance care in Europe. Technical report Ambulancezorg Nederland.
- HUGHES O (2009) Crews waste 30,000 hours at A&E. *Daily Post. Wales* p 8.
- IBRI S, NOURELFATH M, and DRIAS H (2012) A multi-agent approach for integrated emergency vehicle dispatching and covering problem. *Engineering Applications of Artificial Intelligence* 25(3):554 – 565.
- INGOLFSSON A (2013) *Operations Research and Health Care Policy*. EMS Planning and Management, Springer New York.
- INGOLFSSON A, BUDGE S, and ERKUT E (2008) Optimal ambulance location with random delays and travel times. *Health Care Management Science* 11(3):262–274.
- INOUE H, YANAGISAWA S, and KAMAE I (2006) Computer-simulated assessment of methods of transporting severely injured individuals in disaster - case study of an airport accident. *Computer Methods Programs in Biomedicine* 81(3):256–265.
- JAGTENBERG C, BHULAI S, and VAN DER MEI R (2015) An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care* 4:27–35.
- JAIN S and MCLEAN C (2003) In: CHICK S, SÁNCHEZ P, FERRIN D, and MORRICE D (Eds). *Proceedings of the 2003 Winter Simulation Conference*. pp 1068–1076.
- JASIM H (2002) Relief staff rostering for the st john ambulance service.
- JONES A (2011) Slow ambulance turnarounds cost NHS more than £10m. *BBC News Wales*.
- KAMENTZKY R, SHUMAN L, and WOLFE H (1982) Estimating need and demand for prehospital care. *Operations Research* 30(6):1148–1167.
- KERGOSIEN Y, BÉLANGER V, SORIANO P, GENDREAU M, and RUIZ A (2015) A generic and flexible simulation-based analysis tool for ems management. *International Journal of Production Research* 53(24):7299–7316.
- KERGOSIEN Y, LENTÉ C, PITON D, and BILLAUT J C (2011) A tabu search heuristic for the dynamic transportation of patients between care units. *European Journal of Operational Research* 214(2):442–452.
- KNIGHT V, HARPER P, and SMITH L (2012) Ambulance allocation for maximal survival with heterogeneous outcome measures. *Omega* 40:918–926.
- KNIGHT V A and HARPER P R (2012) Modelling Emergency Medical Services with phase-type distributions. *Health Systems* 1:58–68.
- KOOLE G and MANDELBAUM A (2002) Queueing models of call centers: an introduction. *Annals of Operations Research* 113:41–59.

- KOZAN E and MESKEN N (2005) A Simulation Model for Emergency Centres. In: ZERGER A and ARGENT R (Eds). *Proceedings of the International Congress on Modelling and Simulation. Advances and Applications for Management and Decision Making*. pp 2602–2608.
- KRAFFT T, GARCIA-CASTRILLO RIESGO L, FISCHER M, LIPPERT F, OVERTON J, and ROBERTSON-STEEL I (2006) Health monitoring and benchmarking of european ems systems: Components, indicators, recommendations. Technical report.
- LARSEN M P, EISENBERG M S, CUMMINS R O, and HALLSTROM A P (1993) Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Annals of emergency medicine* 22(11):1652–1658.
- LARSON R C (1974) A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and Operations Research* 1(1):67–95.
- LEE S (2011) The role of preparedness in ambulance dispatching. *Journal of the Operational Research Society* 62(10):1888–1897.
- LEE S (2012a) Ambulance-initiated dispatching by centrality principle in emergency medical service. In: *IIE Annual Conference. Proceedings*, Institute of Industrial Engineers-Publisher.
- LEE S (2012b) The role of centrality in ambulance dispatching. *Decision Support Systems* 54(1):282 – 291.
- LEE S (2013) Centrality-based ambulance dispatching for demanding emergency situations. *Journal of the Operational Research Society* 64(4):611–618.
- LEE S (2014) Role of parallelism in ambulance dispatching. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44(8):1113–1122.
- LEE T, JANG H, CHO S H, and TURNER J G (2012) A simulation-based iterative method for trauma center - air ambulance location problem. In: *Proceedings of the 2012 Winter Simulation Conference* pp 955–966.
- LI X, ZHAO Z, ZHU X, and WYATT T (2011) Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research* 74(3):281–310.
- LI Y and KOZAN E (2009) Rostering ambulance services. *Industrial engineering and management society* pp 795–801.
- Lightfoot Solutions (2009) Time to make a difference: Transforming ambulance services in Wales. A modernisation plan for ambulance services and NHS Direct Wales. Technical report.
- LIM C S, MAMAT R, and BRAUNL T (2011) Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems* 12(2):624–632.
- LOWTHIAN J, CAMERON P, STOELWINDER J, CURTIS A, CURRELL A, COOKE M, and MCNEIL J (2011a) Increasing utilisation of emergency ambulances. *Australian Health Review* 35(1):63–69.
- LOWTHIAN J, JOLLEY D, CURTIS A, CURRELL A, CAMERON P, STOELWINDER J, and MCNEIL J (2011b) The challenges of population ageing: accelerating demand for emergency ambulance services by older patients, 1995-2015. *The Medical Journal of Australia* 194(11):574–578.

- MASON A J (2013) *Simulation and Real-Time Optimised Relocation for Improving Ambulance Operations*. pp 289–317 Springer New York New York, NY.
- MATTESON D, MCLEAN M, WOODARD D, and HENDERSON S (2011) Forecasting emergency medical service call arrival rates. *The Annals of Applied Statistics* 5(2B):1379–1406.
- MAXWELL M, RESTREPO M, HENDERSON S, and TOPALOGLU H (2010) Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing* 22(2):266–281.
- 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* 247:294–309.
- MCLAY L A and MAYORGA M E (2013a) A dispatching model for server-to-customer systems that balances efficiency and equity. *Manufacturing & Service Operations Management* 15(2):205–220.
- MCLAY L A and MAYORGA M E (2013b) A model for optimally dispatching ambulances to emergency calls with classification errors in patient priorities. *IIE Transactions* 45(1):1–24.
- NHS Choices (2014) Emergency and urgent care services. Technical report.
- NHS Scotland (2014) Annual Reports and Accounts 2013/14. Technical report.
- NICKEL S, REUTER-OPPERMANN M, and SALDANHA-DA GAMA F (2015) Ambulance location under stochastic demand: A sampling approach. *Operations Research for Health Care*.
- O’CATHAIN A, KNOWLES E, TURNER J, HIRST E, GOODACRE S, and NICHOLL J (2015) Variation in avoidable emergency admissions: multiple case studies of emergency and urgent care systems. *Journal of Health Services Research and Policy*.
- PARRAGH S (2009) *Ambulance routing problems with rich constraints and multiple objectives*. Dissertation, Fakultät fuer Wirtschaftswissenschaften Universitaet Wien, Vienna.
- PARRAGH S N, DOERNER K F, HARTL R F, and GANDIBLEUX X (2009) A heuristic two-phase solution approach for the multi-objective dial-a-ride problem. *Networks* 54(4):227–242.
- RAJAGOPALAN H K, SAYDAM C, SETZLER H, and SHARER E (2011) Ambulance deployment and shift scheduling: An integrated approach. *Journal of Service Science and Management* 4(01):66.
- REPEDE J and BERNARDO J (1994) Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky. *European Journal of Operational Research* 75(3):567–581.
- REUTER-OPPERMANN M, KUNZE VON BISCHHOFFSHAUSEN J, and HOTTUM P (2015) Towards an it-based coordination platform for the german emergency medical service system. In *Exploring Services Science* (NOVOA H and DRAGOICEA M, eds), volume 201 of *Lecture Notes in Business Information Processing* pp 253–263, Springer International Publishing.
- REVELLE C and HOGAN K (1988) A reliability-constrained siting model with local estimates of busy fractions. *Environment and Planning B: Planning and Design* 15(2):143–152.
- REVELLE C and SWAIN R (1970) Central facilities location. *Geographical Analysis* 2(1):30–42.

- RITZINGER U and PUCHINGER J (2012) Real-world patient transportation. In: *19th ITS World Congress*.
- SACCO W, NAVIN D, FIEDLER K, WADDELL R, LONG W, and BUCKMAN R (2005) Precise formulation and evidence-based application of resourceconstrained triage. *Academic Emergency Medicine* 12(8):759–770.
- SCHILDE M, DOERNER K F, and HARTL R F (2011) Metaheuristics for the dynamic stochastic dial-a-ride problem with expected return transports. *Computers & operations research* 38(12):1719–1730.
- SCHLECHTRIEMEN T, BURGHOFFER K, LACKNER C K, and ALTEMEYER K H (2005a) Validation of the naca score based on objectifiable parameters: Analysis of 104,962 primary air rescue missions in 1999–2003. *Notfall & Rettungsmedizin* 8(2):96–108.
- SCHLECHTRIEMEN T, BURGHOFFER K, STOLPE E, ALTEMEYER K H, and LACKNER C K (2005b) The munich naca score: Modification of the naca score for preclinical emergency medicine. *Notfall & Rettungsmedizin* 8(2):109–111.
- SCHMID V (2012) Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research* 219(3):611–621.
- SETZLER H, PARK S, and SAYDAM C (2009) EMS call volume predictions: A comparative study. *Computers & Operations Research* 36:1843–1851.
- SNOOKS H, KINGSTON M, ANTHONY R, and RUSSELL I (2013) New models of emergency prehospital care that avoid unnecessary conveyance to Emergency Department: Translation of research evidence into practice? *The Scientific World Journal* pp 1–6.
- SUDTACHAT K, MAYORGA M E, and MCLAY L A (2014) Recommendations for dispatching emergency vehicles under multitiered response via simulation. *International Transactions in Operational Research* 21(4):581–617.
- TAKEDA R A, WIDMER J A, and MORABITO R (2007) Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. *Computers and Operations Research* 34(3):727–741.
- TAKEDAA R, WIDMERA J, and MORABITOB R (2007) Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. *Computers & Operations Research* 34:727–741.
- TOREGAS C, SWAIN R, REVELLE C, and BERGMAN L (1971) The location of emergency service facilities. *Operations Research* 19(6):1363–1373.
- VAN BUUREN M, VAN DER MEI R, AARDAL K, and POST H (2012) Evaluating dynamic dispatch strategies for emergency medical services: Tifar simulation tool. In: C. LAROQUE, J. HIMMELSPACH R P O R and UHRMACHER A M (Eds). *Proceedings of the Winter Simulation Conference*. p 46.
- VAN DEN BERG P L and AARDAL K (2015) Time-dependent MEXCLP with start-up and relocation cost. *European Journal of Operational Research* 242(2):383–389.
- VAN DEN BERG P L, KOMMER G J, and ZUZÁKOVÁ B (2015) Linear formulation for the maximum expected coverage location model with fractional coverage. *Operations Research for Health Care*.

- VAN DEN BERG P L, VAN ESSEN J T, and HARDERWIJK E J (2016) Comparison of static ambulance location models. In: *International Conference on Logistics and Operations Management (GOL)*.
- VAN DEN BERGH J, BELIËN J, DE BRUECKER P, DEMEULEMEESTER E, and DE BOECK L (2013) Personnel scheduling: A literature review. *European Journal of Operational Research* 226(3):367–385.
- VAN ESSEN J T, HURINK J L, NICKEL S, and REUTER M (2013) Models for ambulance planning on the strategic and tactical level. Technical report Beta Research School for Operations Management and Logistics.
- VILE J, GILLARD J, HARPER P, and KNIGHT V (2012) Predicting ambulance demand using singular spectrum analysis. *Journal of the Operational Research Society* 63(11):1556–1565.
- VILE J, GILLARD J, HARPER P, and KNIGHT V (2016) Time-dependent stochastic methods for managing and scheduling emergency medical services. *Operations Research for Health Care*.
- WANG Y, LUANGKESORN K, and SHUMAN L (2012) Modeling emergency medical response to a mass casualty incident using agent based simulation. *Socio-Economic Planning Sciences* 46(4):281–290.
- WONG H T and LAI P C (2014) Weather factors in the short-term forecasting of daily ambulance calls. *International journal of biometeorology* 58(5):669–678.
- WORKFORCE, FACILITIES TEAM H, and CENTRE S C I (2014) Ambulance Services, England 2013-14. Technical report.
- YUE Y, MARLA L, KRISHNAN R, HEINZ H J, and COLLEGE I I I (2012) An Efficient Simulation-based Approach to Ambulance Fleet Allocation and Dynamic Redeployment. In: *AAAI Conference on Artificial Intelligence*.
- ZARKESHZADEH M, ZARE H, HESHMATI Z, and TEIMOURI M (2016) A novel hybrid method for improving ambulance dispatching response time through a simulation study. *Simulation Modelling Practice and Theory* 60:170–184.
- ZHANG O, MASON A, and PHILPOTT A (2008) Simulation and optimisation for ambulance logistics and relocation. In: *Presentation at the INFORMS 2008 Conference*.
- ZHEN L, WANG K, HU H, and CHANG D (2014) A simulation optimization framework for ambulance deployment and relocation problems. *Computers and Industrial Engineering* 72:12 – 23.