International Journal of Medical Informatics 133 (2020) 103975

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Using optimization to provide decision support for strategic emergency medical service planning – Three case studies

Henrik Andersson[a,*], Tobias Andersson Granberg[b], Marielle Christiansen[a], Eirik Skorge Aartun[a], Håkon Leknes[a]

[a] Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim, Norway
[b] Department of Science and Technology, Linköping University, Norrköping, Sweden

ARTICLE INFO

Keywords:
Ambulance station location
Ambulance allocation
Emergency response planning

ABSTRACT

To achieve high performing emergency medical services (EMS), planning is of vital importance. EMS planners face several challenges when managing ambulance stations and the fleet of ambulances. In this paper, three strategic cases for EMS planners are presented together with potential solutions. In the first case, the effects of closing down a local emergency room (ER) are analyzed together with how adding an ambulance station and an ambulance to the area affected by the closing of the ER can be used to mitigate the negative consequences from the closing. The second case investigates a change in the organization of EMS. Currently, many non-urgent transport assignments are performed by ambulances which make them unavailable for more urgent calls. The potential for a more effective utilization of the ambulances is explored through transferring these assignments to designated transport vehicles. The third case is more technical and challenges the common practice regarding how time dependent demand is handled. Looking at the busiest hour or the average daily demand, is compared with taking time varying demand into account. The cases and solutions are studied using a recently developed strategic ambulance station location and ambulance allocation model for the Maximum Expected Performance Location Problem with Heterogeneous Regions (MEPLP-HR). The model has been extended to also include multiple time periods. This article demonstrates an innovative use of the model and how it can be applied to find and evaluate solutions to real cases within the field of strategic planning of EMS. The model is found to be a useful decision support tool when analyzing the cases and the expected performance of potential solutions.

1. Introduction

The general challenge for emergency medical services (EMS) is to provide the best possible service to the public. Thus, a variety of planning problems arises. Within strategic planning, the main problem has been where to locate ambulances or ambulance stations. Tactical problems include dimensioning the ambulance fleet and allocating ambulances to stations. Within operational planning, there are problems such as which ambulance to dispatch to a call, if and where to relocate ambulances, and if the patient should be treated at the scene or brought to the emergency room (ER) of a hospital. To be able to make the best possible decisions for strategic, tactical and operational problems, operations researchers have been developing decision support tools for several decades. In recent years, computational power has increased and EMS has tracked more data. This has created tremendous opportunities for using Operations Research (OR) to provide decision support for EMS management.

This paper shows how the model for the Maximum Expected Performance Location Problem for Heterogeneous Regions (MEPLP-HR) from Leknes et al. [1] can be applied as a decision support tool. Three cases experienced by the county of Sør-Trøndelag are presented. In the first case, the consequences of closing down a local ER, and how this action can be partly mitigated by adding more ambulances, are analyzed. To do this, a new performance measure based on the time to ER is used. In the second case, the potential for a more effective utilization of ambulances is explored through transferring non-urgent transport assignments to designated transport vehicles. Finally, the importance of taking several time periods into account when locating stations is explored. Common approaches is to look at the busiest hour or the average daily demand when analyzing station locations, which might give rise to suboptimal locations for some other time periods [2]. For this analysis, the model has been extended to include time periods. Thus, the paper contributes to the literature and EMS practice by showing how one single optimization model can be used to do different

* Corresponding author.
E-mail address: henrik.andersson@ntnu.no (H. Andersson).

https://doi.org/10.1016/j.ijmedinf.2019.103975
Received 29 March 2019; Received in revised form 2 July 2019; Accepted 22 September 2019

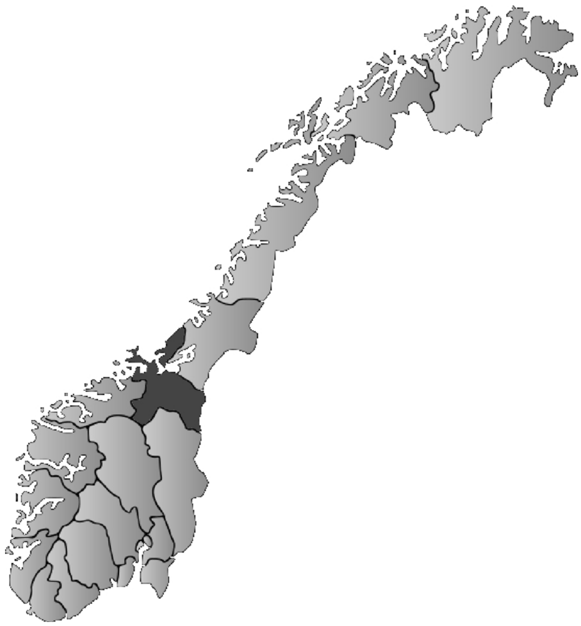

Fig. 1. Map of Norway in grey and the county of Sør-Trøndelag in black.

kinds of analyses for real problems.

The region of study is the county of Sør-Trøndelag in Norway. The county of Sør-Trøndelag is shown as the black area in Fig. 1. In Sør-Trøndelag there are approximately 300,000 inhabitants, with two thirds living in urban areas [3]. There are approximately 30,000 calls for EMS yearly, with one third being red, one third being yellow, and one third being green non-urgent transport calls. The red calls are the most urgent and time critical calls. The number of calls from a zone is also referred to as the demand in the zone. It is the EMS administrator of Sør-Trøndelag, the Emergency Medical Communication Central (AMK), that receives the calls. When the AMK receives a call, the general response process is as follows:

i. Call is received by trained personnel and then screened, classified and allocated to one or more available ambulance(s).
ii. Ambulance departs for incident scene.
iii. Ambulance arrives at scene and intervention by paramedics starts.
iv. Ambulance returns to hospital, station or is dispatched to new incident.

However, this is just an overview of the key operational EMS process. In addition, there are several other key processes for EMS, such as planning and training. All these processes are important for the AMK to be performing well. As the scope of work for the AMK contains several different processes, the performance of the AMK is divided into several performance objectives, listed below sorted by relative importance:

1. The patient should receive timely and correct treatment.
2. Partners and the public should have confidence in the organization.
3. The employees should have a good working environment and professional development.
4. The organization should appear transparent and be cost-effective.

All these performance objectives are important to achieve high performing EMS. However, OR has traditionally been used to optimize the system with respect to response time and survival. Response time is easy to measure and understand, and is often given as political targets. The National guidelines for Norway are that 90% of red calls should be responded to within 12 min in urban areas and 25 min in rural areas [4]. Nevertheless, these are just guidelines, and the local EMS planners are free to decide on other targets.

The rest of the paper is outlined as follows: Section 2 presents selected related research. Section 3 contains a brief overview of the model used. In Section 4 the data and the case region are presented, while Section 5 presents and discusses the three cases. Finally, Section 6 concludes on the results and proposes further research.

2. Related research

For more than four decades operations researchers have developed decision support for strategic, tactical and operational problems for EMS. Researchers have also put an effort in determining what should be measured to obtain the desired performance. In this section we review a selection of the literature within strategic decision support for EMS. Three recent review papers give a good overview of other OR, logistics and planning studies within EMS [5–7].

One of the earliest models, the maximal covering location problem (MCLP), was introduced by Church and ReVelle [8]. The MCLP maximizes the demand covered within a certain response time. This model with the covering performance measure has served as a basis for many strategic location models. Schilling et al. [9] developed a model that maximizes the demand covered by two different types of vehicles, while Hogan and ReVelle [10] created models that maximized the number of zones covered by two or more ambulance stations, i.e. provided double coverage.

The earliest models were pure strategic models that did not consider the operational aspects. One of the major challenges with locating ambulance stations is to create a model that in some way incorporates these. The most important one might be the risk that all ambulances allocated to a station are busy. Different locations will affect the workload of the ambulances, and hence the risk of an ambulance being busy when an incident occurs. To cope with this, Daskin [11] presented the maximum expected covering location problem (MEXCLP). In the MEXCLP there is a certain probability p that an ambulance is busy. This busy probability is in the earliest models set to a constant for all ambulances, while newer models such as the ones in [12] use iterative methods to find more realistic busy probabilities.

Knight et al. [2] extended the model by Erkut et al. [12] with multiple performance measures, and used an iterative procedure to calculate and update the busy probabilities. Instead of using different busy probabilities, van den Berg [13] developed an integer linear model allowing for fractional coverage, which can be interpreted as the probability of reaching a call within the specified time standard. Another approach is to use stochastic programming, like, e.g. Boujemaa et al. [14] who developed a two-stage stochastic programming ambulance location-allocation model and showed that it outperforms a deterministic model. Sung and Lee [15] also used stochastic programming to solve a model with stochastic call arrivals, taking into account that the frequency of ambulance calls typically varies both spatially and temporally during the day and night.

To evaluate a certain location and allocation of ambulances, both simulation and stochastic models can be used. Simulation is applied by Davis [16], Goldberg et al. [17] and Ünlüyurt and Tunçer [18] among others, while the stochastic hypercube queuing model (HQM) was introduced by Larson [19]. Both simulation and stochastic models have their pros and cons, but as argued by Ingolfsson [20], a primary advantage of stochastic models is that they can be solved analytically. HQM is used by Erkut et al. [12], Knight et al. [2] and Ansari et al. [21] among others to find the busy probabilities (or equivalent factors) for the ambulances at different stations.

In addition to having a realistic simulation or stochastic model, it is important to know what characterizes a good solution to be able to evaluate a certain location and allocation. The earliest models, such as the MCLP, evaluated locations based on the covering performance measure. A problem with this performance measure is however that as long as the origin of the call is outside the cover threshold, it does not contribute to the objective function. Hence, it does not matter how far

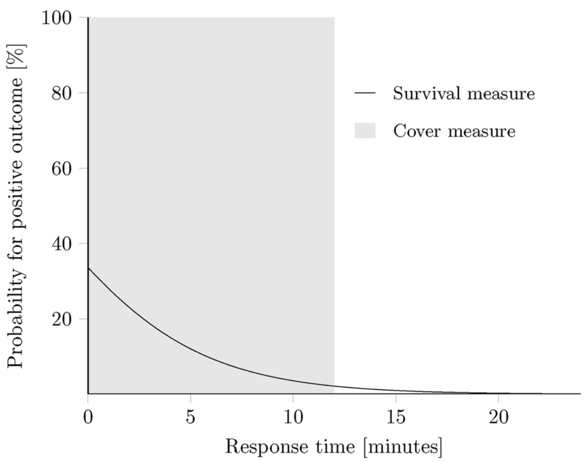

Fig. 2. Comparison of survival and cover performance measure.

the demand zone is away from a station as long as it is outside the cover threshold. The covering performance measure does not differentiate between response times within the threshold either. This is a challenge as the outcome of some calls is highly dependent on a short response time.

As a response to these challenges, Erkut et al. [12] introduced the maximum survival location problem (MSLP). The MSLP maximizes the probability of survival for cardiac arrest patients. The objective is based on the probability of survival given a specific response time. Fig. 2 illustrates the difference between the survival and cover measure. The 1/0 cover measure is seen as the grey square. For all demand within 12 min, the probability of positive outcome is 100%. For demand outside 12 min, the probability for a positive outcome is 0%. The black line is the survival function from De Maio et al. [22]. The probability for positive outcome from cardiac arrest is about 35% at the time the cardiac arrest occurs. It is assumed that there is no interaction from bystanders. The function decreases with the response time, and after 12 min the probability for a positive outcome is about 3%. Erkut et al. [12] investigated different survival functions. However, the functions were found to give approximately the same locations of the ambulance stations. The conclusion was that the important characteristic is the exponential slope of the function. Knight et al. [2] and McCormack and Coates [23] built on the work of Erkut et al. [12], but also included standard cover performance measures in the objective function. Knight et al. [2] show that heterogeneous outcome measures can provide additional decision support. He et al. [24] also incorporate multiple objectives, when studying how to locate rural EMS stations for managing traffic accidents in South Dakota, US. They have one coverage component, and then minimize the average response time for the calls that are not covered.

For all these models, response time is considered the main parameter to use when evaluating a potential location configuration. The validity of response time as a parameter for patient outcomes has been the background for several articles. Weiss et al. [25] and Pons and Markovchick [26] found that response time did not play an important role for patient survival after traumatic injuries. However, by using distance from ambulance station to patient as a proxy for response time, Wilde [27] showed that response time significantly affects mortality of patients in need of emergency services. By studying patient data, Peyravi et al. [28] show that using temporary locations for ambulances will decrease the response times and also the pre-hospital mortality rate, compared to using permanent stations. In some research, e.g. in [29], the cost/benefit of response time has been investigated. However, most of the existing related research has focused on determining the minimum number of ambulances needed to fulfill some requirements, or how to best use the existing resources [30].

Alongside with the progress on model development and solution methods, numerous case studies have been performed, where OR is utilized when analyzing specific real world problems. Su and Shih [31] used computer simulation to study the EMS system in Taiwan. By varying input parameters such as staffing levels and number of rescue units assigned to hospitals, they were able to analyze the effectiveness, care quality and cost-efficiency for many different ways of organizing the operations. Su et al. [32] used a refined double coverage model to produce a new plan for ambulance deployment in Shanghai that could potentially lower the operation cost with 10 million RMB per year. Andrade and Cunha [33] suggested moveable ambulance stations for São Paulo in Brazil, and use an artificial bee colony algorithm to get allocation and reposition suggestions. Based on the results, the ambulance managers in São Paulo decided to increase the number of bases, using the model to get the appropriate decision support. Both Amorim et al. [34] and He et al. [24] focused on traffic accidents, but Amorim et al. [34] for urban areas (Porto city) and He et al. [24] for rural. Liu et al. [35] used a double coverage model to locate stations and allocate ambulances in the Songjiang District, Shanghai, China. Dibene et al. [36] calculated the optimal locations for ambulances in Tijuana, Mexico using a double standard model, where all demand should be covered within one time standard, and a fraction of the demand also within another, shorter time standard. They show that the coverage can be significantly improved without adding new resources, compared with the current real locations. Fritze et al. [37] use a variant of the MCLP to determine locations for EMS stations in Lower Austria. Their study was motivated by a call for bids to restructure the locations, having very specific instructions for the analysis, e.g. that 95% of the population should be reachable by a physician within 20 min, that an ambulance speed of 1.25 times the maximum allowed speed should be used, and that no double coverage had to be considered. The results showed that it was possible to improve the coverage substantially, even while reducing the resources. However, the authors also report that the final decision by the government, which was made before the study was finished, was to keep all the present locations.

3. Mathematical model

To analyze the cases in this paper, the model for the Maximum Expected Performance Location Problem for Heterogeous Regions (MEPLP-HR) is used. Given a set of possible locations for ambulance stations and a set of zones with demand for EMS, the model locates ambulance stations and allocates ambulances to these stations based on a set of given performance measures. Each performance measure has a certain weight, and the model maximizes the total performance with a limited number of ambulances and stations at disposal (see Section 4). The model incorporates operational aspects by calculating the probability that there is an available ambulance at a station. The model is explained in depth in Leknes et al. [1].

The problem analyzed in this paper combines both strategic (locating ambulance stations), and tactical (allocating ambulances to the stations) decisions. To reflect that the tactical decisions can be adopted to different demands, the model in Leknes et al. [1] is extended with multiple scenarios representing different demand situations. These scenarios represent different time periods of a week and are henceforth called time periods. The problem thus resembles a two-stage problem [38]. In the first stage, the stations are located, and in the second stage the available ambulances are allocated to the stations for each time period. In the following subsection, the model is presented and briefly explained.

3.1. Model formulation

Indices and sets$j \in \mathcal{J}$zones where ambulance stations can be located$i \in \mathcal{I}$zones with a demand for EMS$q \in \mathcal{Q}$ranking of stations$l \in \mathcal{L}$performance measures of the EMS provider$t \in \mathcal{T}$time periodsParameters$W_l$weight of performance measure $l$$D_{ilt}$number of calls in zone i relevant for performance measure l in time period $t$$H_{ijl}$performance value per call of zone i being covered by a station in zone j, given performance measure $l$$A_t$number of available ambulances in time period $t$$S$number of stations that can be located$\bar{A}$maximum number of ambulances at a stationλ_{it}rate of calls from zone i in time period $t$$T_{ij}$average time for an ambulance at a station in zone j to serve calls from zone iVariablesz_j1 if a station is located in zone j, 0 otherwisex_{jt}number of ambulances allocated to a station in zone j in time period $t$$y_{ijqt}$proportion of demand in zone i at time period t covered by a station in zone j ranked as the qth station for zone $i$$\rho_{ijt}$1 if station in zone j is the primary station for zone i in time period t, 0 otherwiseθ_{jt}arrival rate of calls to the station in zone j in time period $t$$\mu_{jt}$service rate of the station in zone j in time period t

$$\max \sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{q\in Q}\sum_{t\in\mathcal{T}} W_l D_{\mathrm{ilt}} H_{\mathrm{ijl}} y_{\mathrm{ijqt}} \tag{1}$$

$$\sum_{j\in\mathcal{J}} z_j \le S \tag{2}$$

$$\sum_{j\in\mathcal{J}} x_{\mathrm{jt}} \le A_t \quad t\in\mathcal{T} \tag{3}$$

$$x_{\mathrm{jt}} \le \bar{A} z_j \quad j\in J,\ t\in\mathcal{T} \tag{4}$$

$$\sum_{j\in\mathcal{J}}\sum_{q\in Q} y_{\mathrm{ijqt}} = 1 \quad i\in\mathcal{I},\ t\in\mathcal{T} \tag{5}$$

$$\rho_{\mathrm{ijt}} \ge y_{\mathrm{ij}1t} \quad i\in\mathcal{I}, j\in\mathcal{J},\ t\in\mathcal{T} \tag{6}$$

$$1-\rho_{\mathrm{ijt}} \ge y_{\mathrm{ij}2t}, \quad i\in\mathcal{I},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{7}$$

$$\sum_{j\in\mathcal{J}} \rho_{\mathrm{ijt}} = 1 \quad i\in\mathcal{I},\ t\in\mathcal{T} \tag{8}$$

$$\sum_{j\in\mathcal{J}} y_{\mathrm{ij}1t} \ge \sum_{j\in\mathcal{J}} y_{\mathrm{ij}2t}, \quad i\in\mathcal{I},\ t\in\mathcal{T} \tag{9}$$

$$\theta_{\mathrm{jt}} = \sum_{i\in\mathcal{I}} (\lambda_{\mathrm{it}}\rho_{\mathrm{ijt}} + \lambda_{\mathrm{it}} y_{\mathrm{ij}2t}) \quad j\in\mathcal{J},\ t\in\mathcal{T} \tag{10}$$

$$\mu_{\mathrm{jt}} = \frac{\sum_{i\in\mathcal{I}}\sum_{q\in Q}\lambda_{\mathrm{it}} y_{\mathrm{ijqt}}}{\sum_{i\in\mathcal{I}}\sum_{q\in Q}\lambda_{\mathrm{it}} T_{\mathrm{ij}} y_{\mathrm{ijqt}}} \quad j\in\mathcal{J},\ t\in\mathcal{T} \tag{11}$$

$$y_{\mathrm{ijqt}} \le f(\theta_{\mathrm{jt}}, \mu_{\mathrm{jt}}, x_{\mathrm{jt}}) \quad i\in\mathcal{I},\ j\in\mathcal{J},\ q\in Q,\ t\in\mathcal{T} \tag{12}$$

$$z_j \in \{0, 1\} \quad j\in\mathcal{J} \tag{13}$$

$$x_{\mathrm{jt}} \in \{0, 1, 2, ..., A_t\} \quad j\in\mathcal{J},\ t\in\mathcal{T} \tag{14}$$

$$y_{\mathrm{ijqt}} \ge 0 \quad i\in\mathcal{I},\ j\in\mathcal{J},\ q\in Q,\ t\in\mathcal{T} \tag{15}$$

$$\rho_{\mathrm{ijt}} \in \{0, 1\}, \quad i\in\mathcal{I},\ j\in\mathcal{J},\ t\in\mathcal{T} \tag{16}$$

$$\theta_{\mathrm{jt}} \ge 0 \quad j\in\mathcal{J},\ t\in\mathcal{T} \tag{17}$$

$$\mu_{\mathrm{jt}} \ge 0 \quad j\in\mathcal{J},\ t\in\mathcal{T} \tag{18}$$

The model includes variables for where a station can be located (z_j), and how many ambulances that are located in each station at each time period (x_{jt}). y_{ijqt} is the required minimum probability that an ambulance from a station in zone j will be able to respond to a call in zone i at time period t, given that the station is qth ranked station for zone i. A station can be ranked as primary or secondary station, i.e. q can be 1 or 2, and each zone has to have exactly one primary station. ρ_{ijt} is 1 if the station in zone j is the primary station for zone i. The arrival rate of calls to a station, θ_{jt}, is variable, depending on how much of the demand that is handled by the surrounding stations, while the average service rate for a station, μ_{jt}, depends on the number of calls served and the average service time.

The objective function (1) calculates the total performance of the

Table 1
Performance measures used in the cases.

Performance measure	Function	W_l	D_{ilt}
Survival	$H(t^R) = \frac{1}{1+e^{-0.679+0.262t^R}}$	20	Red calls
Cover urban	$H(t^R) = \begin{cases} 1 & \text{for } 0 \le t^R \le 12 \\ 0 & \text{for } t^R > 12 \end{cases}$	1	Yellow calls
Cover rural	$H(t^R) = \begin{cases} 1 & \text{for } 0 \le t^R \le 25 \\ 0 & \text{for } t^R > 25 \end{cases}$	1	Yellow calls

locations and allocations. Depending on the type of call (red or yellow), different performance measures, $l \in \mathcal{L}$ are used, as specified in Table 1, where each measure is weighed by W_l. Thus, as an example, for yellow calls, if zone i is an urban zone, a station in zone j covers 70% of the demand and the response time from j to i is less than 12 min, the objective function contribution is $W_l D_{ilt} H_{ijl} y_{ijqt} = 1 \cdot D_{ilt} \cdot 1 \cdot 0.7$.

The deployment constraints are given by constraints (2)–(4). Constraints (2) and (3) make sure that no more than the number of available stations and ambulances are located and allocated, respectively. The logical restriction that an ambulance cannot be allocated to a zone without a station is handled by constraints (4). The covering constraints (5)–(9) keep track of which zones the different stations cover, as well as the primary station for each zone. All calls from each zone have to be covered by a station. This is taken care of by constraints (5). For each zone there is one primary station and one or more secondary station(s). The secondary station(s) cannot be the same as the primary station. These properties are handled through constraints (6)–(8). In addition, constraints (9) ensure that the primary station receives a higher proportion of calls than the secondary station(s).

A station receives all the calls from a zone that has the station as its primary station, as well as the respective proportion of calls it covers from a zone that has it as secondary station. This constitute the arrival rate and is given by constraints (10). It should be noted that a station typically will not serve all arriving calls, as all ambulances at one station at times might be busy. In those cases, the calls are served by the secondary stations. The average service rate of ambulances at a station is given by constraints (11). The numerator is the number of calls served by the station, and the denominator is the time it takes to serve all calls. This expression is nonlinear and therefore linearized as described in Leknes et al. [1]. The proportion of calls covered by a station has to be less than or equal to the long time probability that there is an ambulance available at the station, expressed through the function $f(\theta_{jt}, \mu_{jt}, x_{jt})$. This is given by constraints (12). The long time probability that there is an ambulance at a station depends on the arrival rate of calls to the station, the service rate of the ambulances at the station, as well as the number of ambulances at the station. This expression is nonlinear and based on queuing models. The full explanation and linearization of this expression is given in Leknes et al. [1]. Finally, constraints (13)–(18) are the variable definitions.

4. Data

The basis for the case studies is AMK data from a time period of four years. The dataset contains the time, date, location and severity (red, yellow and green) of each call. For Sør-Trøndelag today, there are no formalized performance measures for the different types of calls. However, AMK's objective is to give the best possible service to the public. For response time, this can be summarized by performance objective 1 and 2 from Section 1. As the criticality of time is different for red, yellow and green calls, the performance measures should be different for these calls. For the time critical red calls, the survival measure from De Maio et al. [22] is used. For yellow calls, the response time is not critical for the condition of the patient, but still important for the service level and for the population to maintain a confidence in

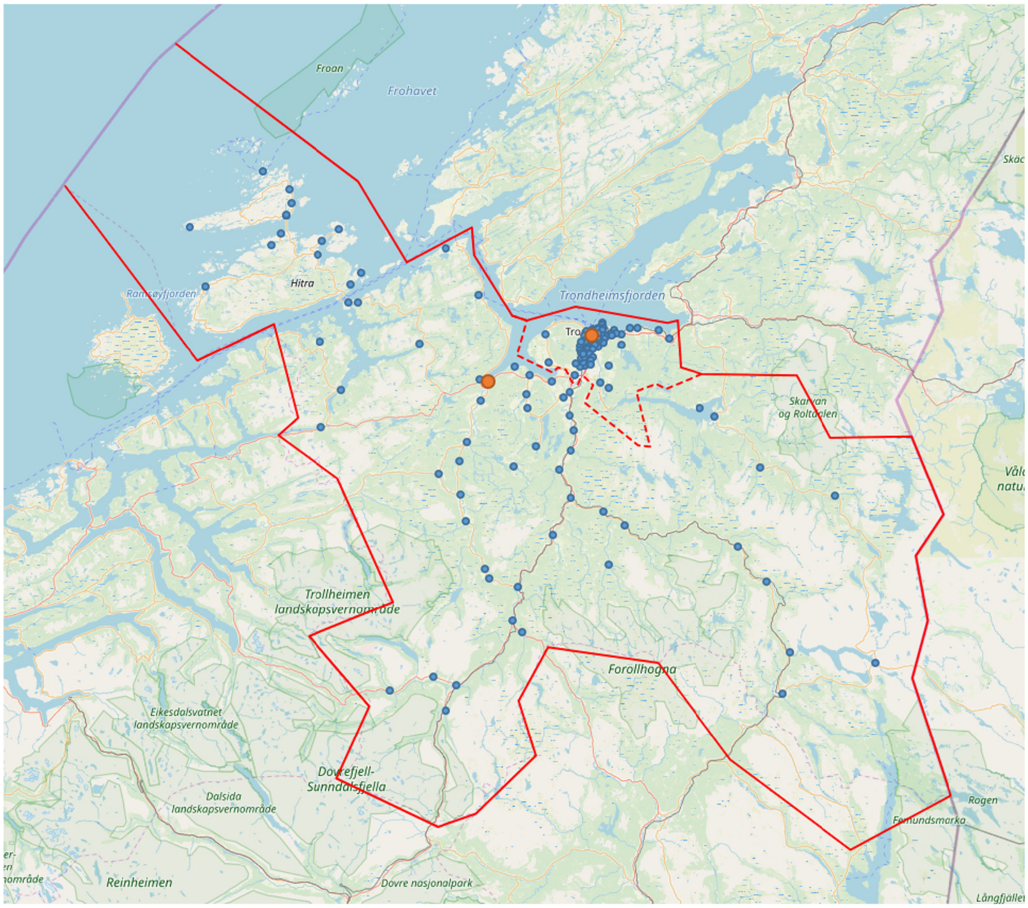

Fig. 3. The county of Sør-Trøndelag, the blue dots represent the population center in each zone and the orange dots indicate where the hospitals with ER are located today. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the organization. Still, it is sufficient that the ambulance arrives within a given threshold, and thus, it is reasonable to use a cover measure for these calls. There is no performance measure for green calls as these are mostly non-urgent transport assignments. However, as ambulances are used also for these, they indirectly affect the other performance measures, and have to be taken into account in the analysis. The weights for the different performance measures are based on the weights in Knight et al. [2]. The summarized performance measures are given in Table 1, where t^R is the response time in minutes.

The region contains 139 zones with demand for EMS and 76 of these are potential locations for ambulance stations. The zones are postcode areas, and thus vary somewhat both in size, resident population and number of historical calls, with rural zones typically being larger with a sparser population distribution. Partitioning the county into just 139 zones will of course induce some aggregation errors [39]. However, the strategic nature of the problem, along with the fact that it is impossible to exactly forecast where and when future calls will occur, makes it a reasonable trade-off between model accuracy and the amount of effort required to collect data and solve the model. The region can be seen in Fig. 3, where the blue dots represent the population center in each zone and the orange dots indicate where the hospitals with ER are located today. The driving distance from the south east to the north west of the county is approximately 350 km with an estimated driving time of 5 h. The hospital located to the west is Orkdal hospital and the easternmost hospital is the regional hospital of Sør-Trøndelag. The area within the dashed line is the urban area of Trondheim and Malvik. The Trondheim and Malvik area consists of 67 zones out of which 44 are potential locations for stations. This smaller area has also been used for some of the tests.

At the time of the study, there were 24 ambulances allocated to 16 stations in the region. The travel times between the zones were found using a tool developed in the Python language that gather the travel times between each node pair from Google Maps. The average service times T_{ij} are calculated using the travel times between the zones, stations and the nearest ER, as well as adding a constant that represents the time on the scene. For Sør-Trøndelag, 43% of all calls end at an ER, and the average time spent on the scene is 16 min. Hence, the formula for T_{ij} is given by Eq. (19), where T^T_{ji} is the travel time from zone j to i, T^T_{iE} is the travel time from zone i to the nearest ER, and T^T_{Ej} is the travel time from the ER to zone j. In some specific cases, an ambulance may take a patient to another ER than the closest one, e.g. for specialized care. However, it is common practice in Norway to choose the closest ER, and for a majority of cases this is done. Thus, in the model we assume that this always happen:

$$T_{\mathrm{ij}} = T^T_{\mathrm{ji}} + 16 + 0.43(T^T_{\mathrm{iE}} + T^T_{\mathrm{Ej}}) + 0.57T^T_{\mathrm{ij}} \tag{19}$$

5. Case studies

The cases we present show different analyses that can be done using the proposed model. In the first case we evaluate the effects of closing down a local emergency room (ER). We also analyze how adding an ambulances station and an ambulance to the area affected by the closing of the ER can be used to mitiage the bad consequences from the closing. The second case investigates a change in the organization of EMS. Currently, many non-urgent transport assignments are performed by ambulances which make them unavailable for more urgent calls. The potential for a more effective utilization of the ambulances is explored through transferring these assignments to designated transport vehicles. The third case is more technical and challenges the common practice regarding how time dependent demand is handled. The cases are studied using the model for the MEPLP-HR. The model is written in Mosel and solved by Xpress-Optimizer Version 8.5.6. Each case begins with a description of the problem and proceeds to show how the model is used to analyze the case.

5.1. Case 1: Closing down a local emergency room

There are several small local ERs in Norway today. These are controversial as they are expensive, and there are discussions about the possibility to uphold proper competence in such small facilities compared with the regional hospitals. However, there are substantial local political forces that want to keep these facilities open, as they fear that the emergency medical services for their local area will be weakened if the facility is closed down. For Sør-Trøndelag, the local ER under discussion is the one located in Orkdal. The ER in Orkdal is approximately 35 min from the regional hospital of Sør-Trøndelag. 40 of the 139 zones has this as its nearest ER, and the accumulated demand in these 40 zones counts for 14.9% of the red and yellow calls in Sør-Trøndelag.

A proposed mitigating action for closing local ERs is to procure additional ambulances and/or stations for the area affected by the closing. In this manner, the extra stations and ambulances should weigh up for the longer distance to the ER. A share of the savings from the closed ER can finance these additional resources. However, it is important to emphasize that the closing of local ERs is not solely based on cost cutting.

The traditional performance measures of the ambulance station location are only based on response time. However, if the ambulances should weigh up for closing down a local ER, it is the time from a call is received until the patient arrives at the ER that is of greatest interest. This makes sense when considering, e.g. stroke, where the time until a CT-scan is performed is of great importance [40]. It is also important for local politicians, as the time until a person arrives at the ER affects the perceived safety and convenience for the population.

To analyze the effect of closing the local ER, a new performance measure based on the time from a call arrives until the patient is at the ER is added to the model. The idea is that people far from the ER will be compensated by having an ambulance closer to reduce the time to ER. A cover measure is used because the objective is to get as many as possible to the ER within a reasonable time, not to minimize the average time to ER. However, there are no official guidelines to what should be defined as a reasonable time to the ER. For Sør-Trøndelag, some interest groups claim that 60 min is reasonable, while others claim that it would be acceptable to have up to 120 min. Based on this, the performance measure is implemented as being 1 if an ambulance from a specific station can get a person from a specific zone to the ER within 90 min, and 0 otherwise. The weight is set to be the same as for the response time cover measure, and the calls relevant for this measure are the red and yellow. The summarized performance measures used in this case are given in Table 2. t^{ER} is the time to ER and defined as the response time plus the travel time from the zone to the closest ER.

To analyze the mitigating actions, one extra ambulance and one extra station have been made available for the zones that are affected by the closing of the local ER and four different scenarios are analyzed. At first, the model is run with the existing ERs and current location and allocation to get a current situation. We have solved the problem for the time period with the highest demand, workdays from 08:00 to 16:00, and used the corresponding number of ambulances. In the second scenario, the location of ambulance stations from the current state is fixed and the proposed closed ER is removed from the data. The ambulances located at the ER are allocated to the stations that had Orkdal as the closest ER. In the third scenario, the stations are fixed as in the second scenario and one extra ambulance is added to the stations that had Orkdal as the closest ER. In the fourth scenario, the stations are fixed as in the second scenario and one extra station is located within the area that had Orkdal as the closest ER. One extra ambulance is also added to the area. The objective values for the different performance measures and scenarios are shown in Table 3. The scenarios are named *Current*, $X00$, $X10$ and $X11$, and refer to the current situation, the scenario without the ER, the scenario without ER and an extra ambulance, and the scenario without the ER and an extra ambulance and station, respectively. We treat Cover urban and Cover rural as one measure, called Cover, and have separated the three performance measures to better study the effects on each of them. The objective value is the sum of the values of the performance measures, and thus, a high value is preferable. For instance, the scenario with the highest value for Cover ER is *Current*, which means that the largest amount of calls can reach an ER within 90 min in that scenario.

Table 2
Performance measures for the first case.

Performance measure	Function	W_l	D_{ilt}
Survival	$H(t^R) = \frac{1}{1+e^{-0.679+0.262t^R}}$	20	Red calls
Cover urban	$H(t^R) = \begin{cases} 1 & \text{for } 0 \le t^R \le 12 \\ 0 & \text{for } t^R > 12 \end{cases}$	1	Yellow calls
Cover rural	$H(t^R) = \begin{cases} 1 & \text{for } 0 \le t^R \le 25 \\ 0 & \text{for } t^R > 25 \end{cases}$	1	Yellow calls
Cover ER	$H(t^{ER}) = \begin{cases} 1 & \text{for } 0 \le t^{ER} \le 90 \\ 0 & \text{for } t^{ER} > 90 \end{cases}$	1	Red and yellow calls

Table 3
Performance values and optimality gap for the scenarios tested in the first case.

	Survival	Cover	Cover ER	Objective value
Current	101.8	17.2	29.3	148.3
$X00$	95.2	16.1	28.4	139.8
$X10$	96.1	16.2	28.4	140.7
$X11$	102.5	17.2	28.4	148.0

The results in Table 3 show that there is little value in adding an additional ambulance without any additional stations. When the ER is removed, the possibility of allocating ambulances to that zone is also lost. Therefore, the response time to the population in Orkdal increases and the performance (especially with respect to the survival measure) deteriorates. With an extra ambulance and ambulance station ($X11$), the negative consequences of closing the ER are almost completely mitigated. A possible explanation for the increase in the survival measure between the current situation and $X11$ is that this measures the time to site, which naturally decreases with extra ambulances. The Cover ER measure is very marginally affected by the addition of ambulance stations and ambulances. The decrease from the current situation can largely be described by zones that in the current situation can reach the ER in Orkdal within 90 min but not the regional hospital of Sør-Trøndelag. For most of these zones, an extra ambulance in the area affected by the closing does not change this. Improved response time could therefore be seen as a compensation for longer time to ER.

The results in Table 3 are for the entire Sør-Trøndelag. As the affected zones only account for 14.9% of the red and yellow calls in Sør-Trøndelag, the consequences of the closing do not appear drastic for the county as a whole. However, for the affected area, the consequences are significant. To see the effect for the affected area, the cumulative distribution of the time to ER and the cumulative distribution of the response time for each of the tests are calculated. The cumulative distribution of the time to ER is shown in Fig. 4, while the cumulative distribution of the response time is seen in Fig. 5. Note that there are some minor inconsistencies due to that the model was not solved to optimality. However, the main trends are clear.

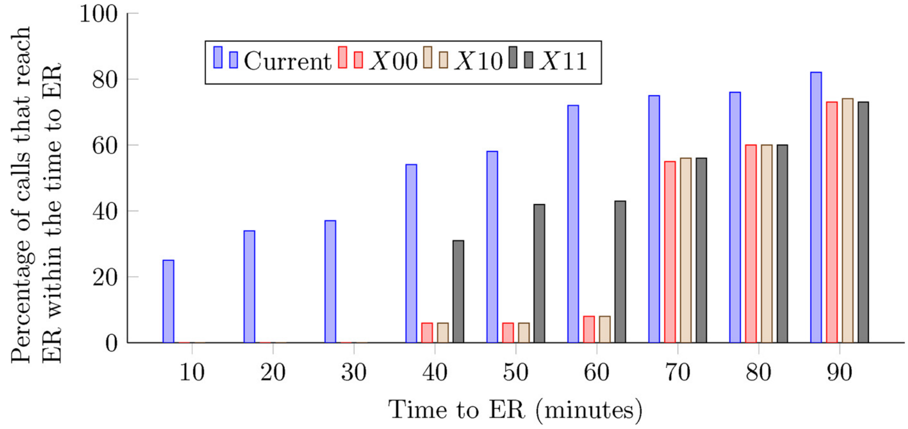

Fig. 4. Cumulative distribution of the time to ER for the affected area.

As seen in Fig. 4, 25% of the calls in the current situation are able to get to the ER in 10 min or less. 72% of the calls is within 60 min or less, and 82% are able to get to the ER within 90 min. Closing the local ER significantly affects the cumulative distribution. For $X00$, $X10$ and $X11$, none of the calls in the affected area are able to get to the ER within 30 min. With no extra station, less than 8% of the calls can get to the ER within 60 min, a drastic decrease in the quality of service. The same measure with an extra station is 43%. Even at 90 min, there is a difference of almost 10 percentage points, and the extra station does not help.

As expected, adding an additional ambulance and station cannot mitigate the effect that removing a local ER has on the time to ER. The model however gives valuable information about how severe the consequences will be, and a quantitative evaluation of the benefits of adding an ambulance to compensate. From a patient perspective, the time to ER is one thing that might affect the health outcome. Another factor is the time from the start of the emergency, until the ambulance arrives, i.e. the response time, which is expected to decrease when an ambulance is added.

Another aspect is the increase in workload for the surrounding ERs, when an ER is closed down. Here, it is assumed that they can handle the increased demand, but in reality, this may be a problem.

Looking at the cumulative distribution of the response times (see Fig. 5), we see that a new station is crucial for maintaining the response times. With an extra station and an extra ambulance, the response times are actually improved compared with the current situation. The new station is placed in the zone where the ER was and the extra ambulance then helps to improve the response times. Without the new station, the response times increase considerably and about 50% of the calls are reached within 30 min, compared with 75% in the current situation and the scenario with an extra station.

For the affected area, the consequences of closing the local ER and adding an ambulance at a new ambulance station is a significantly increase in the time to ER while the response time are only marginally affected. To fully analyze the value of the proposed solution, there needs to be a proper weighting between the response time and the time to ER. However, as there is a stronger focus on treating patients on scene [41], the solution of introducing extra ambulances and stations as mitigating actions is interesting.

5.2. Case 2: Designated non-urgent transport vehicles

Ambulances in Norway are used for almost every type of transport to and from hospitals. Ambulances are expensive vehicles, with trained staff that are specialists in handling emergencies. However, for Sør-Trøndelag in the busiest period, 57% of the calls are categorized as green calls, mainly consisting of normal transport assignments without any need for advanced medical treatment. These assignments may be planned or unplanned, but they are not urgent.

To cut cost and utilize the resources effectively, a proposed solution is to transfer the green calls from the ambulances to specialized transport vehicles. These vehicles may be administered by the emergency medical communication central or a designated transport organization. The main idea is that it is inefficient to use specially trained paramedics

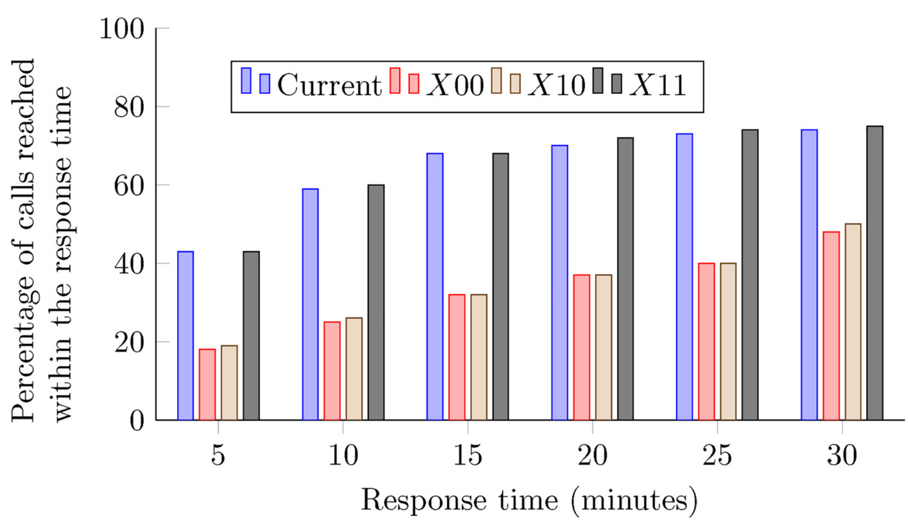

Fig. 5. Cumulative distribution of the response time for the affected area.

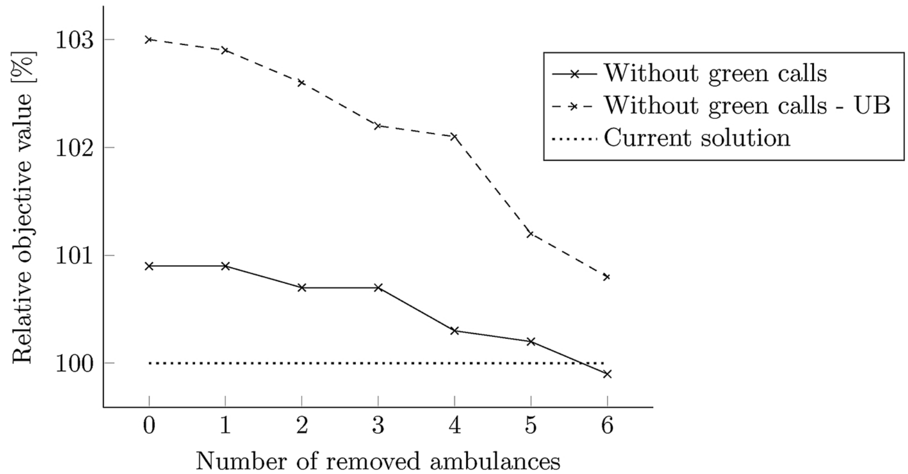

Fig. 6. Objective function value without green calls relative to the situation with green calls and the current number of ambulances when ambulances are removed.

with expensive EMS equipment for normal transport assignments. To efficiently utilize the resources, expensive ambulances could be replaced by cost effective transport vehicles. This means that fewer vehicles would be used for the urgent calls, but that these vehicles also would have less calls to handle. To be able to handle the urgent calls without any decrease in mean response time, the number of ambulances that are removed must not be too large.

To analyze the benefit of introducing designated non-urgent transport vehicles, we investigate how many ambulances that can be removed while still keeping the same performance level. All the green calls are removed from the dataset, as they are assumed to be handled by the specialized transport vehicles. The analysis is carried out on the busiest period, workdays from 08:00 to 16:00. The problem is solved for the busiest time period and the current number of ambulances. The station locations found are then fixed and used when the number of ambulances is decreased.

The relative impact on the objective value from removing ambulances is presented in Fig. 6. The solid line marked with $\times$ is the best objective function value without the green calls included and the dashed line marked with $\times$ is the corresponding upper bound since the instances are not solved to proven optimality. The dotted line marks the current situation, i.e. including the green calls and with no ambulances removed. As seen in Fig. 6, five ambulances can be removed while still keeping the same total expected performance level as with the green calls. The relatively low level of improvement in objective function value when removing the green calls indicates that with the current fleet, the best (closest or close enough) station will almost always have available ambulances. In fact, with the green calls included, the primary stations cover approximately 92% of all calls, and without, this number increases to around 97%, which is a relatively small increase. Thus, it is not possible to improve the situation much by removing demand.

By removing 57% of the calls, 5 out of 24 ambulances can be removed while still keeping the same performance level. This seems a bit low, but can be explained by the fact that each station requires at least one ambulance to contribute to the performance measures. For the busiest period, 57% of the calls represents 22 calls each day. Hence, for designated non-urgent transport vehicles to be an interesting option, these vehicles need to be able to handle at least 22 calls each day and cost less than five ambulances. However, the analysis presented here is just an indication of what is possible. To fully explore the potential of designated non-urgent transport vehicles, more research on the green calls as well as the specialized vehicles is needed.

5.3. *Case 3: Time varying demand and resources*

For EMS districts, both the demand for EMS and the available EMS resources vary throughout the day. The demand for EMS is typically highest in the daytime of workdays and lowest during the night of the workdays. The areas with demand may also change as most people are at work during the day and at home at night. This presents a challenge for EMS managers when they are locating resources. Some of the resources, such as the ambulances, can be moved during the day, but the station locations are of course fixed. For the county of Sør-Trøndelag, the week is divided in 6 time periods based on the demand. The periods are 00–08, 08–16 and 16–24 for workdays and weekends. These periods have different resources at disposal, as shown in Table 4. The demand is also significantly different, where workdays 08–16 is the busiest period and accounts for 44.6% of the total number of calls.

Common approaches when locating ambulance stations are to focus on the busiest period of the week, use an average demand, or base the demand on where people have their homes. This gives a simpler problem to solve, than when taking all periods into account. However, it is not certain that the solutions obtained using this strategy give locations that are good for all time periods, especially if the demand is shifting spatially. As explained in Section 2, one strategy to handle this problem is to use dynamic relocations, or to have time dependent temporary locations for the ambulances. Another, and the one used here, is to take all time periods, and their varying demand, into account when solving the strategic model.

To evaluate the value of including all demand information, we have tested three different strategies for solving the problem:

– Solve the full model and then solve all single time period problems with the station locations fixed to the locations found in the full model. Solving the full model means including all six time periods, with their varying demand and number of resources. Ambulance allocations can vary between time periods, but not the station locations. This is called *Full*.

Table 4
Time periods, ambulances at disposal and demand.

Period	Ambulances	Calls/h
Workday 00:00–08:00	17	1.06
Workday 08:00–16:00	24	4.97
Workday 16:00–24:00	19	2.37
Weekend 00:00–08:00	17	1.52
Weekend 08:00–16:00	22	3.00
Weekend 16:00–24:00	19	2.33

Table 5
Bounds and gaps for the strategies tested in the third case.

Area	Strategy	LB†	LB	UB†	Gap
ST	*Full*	56.894	57.453	58.479	1.8%
ST	*Max*		56.907		2.8%
ST	*Total*		57.458		1.8%
TM	*Full*	30.652	30.685	30.843	0.5%
TM	*Max*		30.649		0.6%
TM	*Total*		30.649		0.6%

– Solve the model only for the busiest time period and then use the station locations of this solution for all single time period problems. This is called *Max*.
– Aggregate the demand from all time periods, solve this problem with the number of ambulances given by the busiest period and then use the station locations of this solution for all single time period problems. This is called *Total*.

The location problem in each strategy is solved for 24 h and the single time period problems are solved for 10 h each. In a single period problem, the station locations are fixed, and the problem is to find the allocation of ambulances, the covers and the ranks given the demand for that period.

Table 5 summarizes the results. *ST* in the first column refers to the large area (Sør-Trøndelag), while *TM* is the smaller (Trondheim-Malvik). The column **LB**† shows the lower bound from the full model. The column **LB** is the lower bound from the solutions of the single time period problems and **UB**† is the upper bound from the full model. **Gap** is calculated as $\mathbf{Gap} = (\mathbf{UB}^{\dagger} - \mathbf{LB})/\mathbf{LB}$.

Some of the single period models of *ST* are not solved to proven optimality, but the difference between the lower and upper bounds in these cases is less than 0.1%. Table 5 shows that taking all demand information into consideration, either through the full model or by aggregating the demand, gives better station locations than only considering the busiest period. The differences are however rather small. In Fig. 7, the locations of the ambulance stations in the three cases are shown. The left figure shows the whole county, the green dots represent locations common in all three strategies and the blue dots show locations appearing in at most two strategies. The right figure is zoomed in on Trondheim, the largest city in the county. The green dots still represent locations common in all three strategies. To highlight the difference between the max strategy (only taking the busiest period into account) and the other strategies, we have used blue squares for the locations suggested by the max strategy (and possibly the other strategies) and blue dots for locations only suggested by the other strategies. From the figure it is clear that there is consensus about the locations in the rural areas, while there are differences in the urban areas. One reason for the differences might be that the location problems are not solved to optimality. The *ST Max* location problem ended with a gap of 1.2% after 24 h (the gap was 0.9% after 100 hours with no change in the solution). For *TM*, all strategies give the same station locations and they are therefore considered as equal.

6. Conclusions

In this work, three managerial cases of EMS are studied using a strategic ambulance station location and ambulance allocation model. In particular, the effect and mitigating actions of closing down a local emergency room, the benefit of introducing designated non-urgent transport vehicles instead of ambulances, as well as the value of taking multiple time periods into account when planning, are studied.

The case studies are performed on the county of Sør-Trøndelag in Norway. For the first case, the results show that to close the local ER will significantly increase the time to ER for the affected area. However, adding an extra ambulance station and ambulances can at a certain degree mitigate the consequences, especially by ensuring that the affected population will get a quicker first response and treatment. The analysis in the second case shows that there is a potential to reduce the number of ambulances by 20% if designated transport vehicles handle the non-urgent assignments. For the third case, the key finding is that taking all demand information into consideration, either through the full model or by aggregating the demand, gives better station locations than only considering the busiest period. However, as in most large strategic cases where OR is used, more analyses are needed for each case to make evidence based decisions. Despite this, the model can be used to provide insight when analyzing real cases and potential solutions experienced by EMS planners.

As future research, it would be interesting to formalize what defines high performing EMS. Then it would be possible to point out where OR has its greatest potential. There is also a need for new performance measures in the models that are not solely based on response time. To build on this, it could be interesting to find a monetary value on the different levels of performance. Then the decision makers could calculate if extra investments to reduce, e.g. response time are beneficial from a cost-benefit point of view.

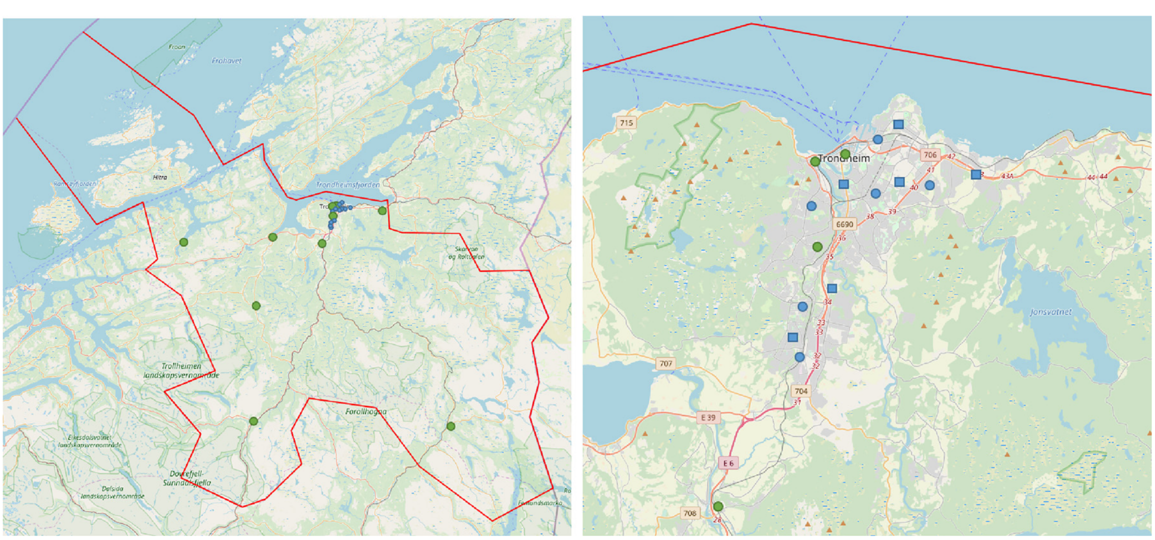

Fig. 7. Location of the ambulance stations using the three strategies tested in Case 3 for the whole county (left figure) and the urban area around Trondheim (right figure).

Authors' contribution

Mr. Aartun and Mr. Leknes developed a first version of the mathematical model used in the paper under the supervision of Prof. Andersson, Prof. Andersson Granberg and Prof. Christiansen. The model has after this been further developed by Prof. Andersson, Prof. Andersson Granberg and Prof. Christiansen. This model has been published in [1].

The data has mainly been collected by Mr. Aartun and Mr. Leknes. Prof. Andersson has been involved in the data processing together with Mr. Aartun and Mr. Leknes. The data is collected from publicly available sources as well as from the Emergency Medical Communication Central (AMK) in Sør-Trøndelag. No sensitive data has been used.

Mr. Aartun and Mr. Leknes wrote a first version of the paper as part of their Master's thesis under the supervision of Prof. Andersson, Prof. Andersson Granberg and Prof. Christiansen. The submitted version of the paper has been prepared by Prof. Andersson, Prof. Andersson Granberg and Prof. Christiansen.

Conflict of interest

There are no conflicts of interest.

Acknowledgements

We thank Erik Solligård and Lars Vesterhus at AMK Sør-Trøndelag for data, as well as input on challenges and performance measures. We also thank Jørgen Einerkjær at AMK Vestfold & Telemark and Håkon Gammelsæter at Ambulanse Midt-Norge for insightful information. Finally, we thank the two anonymous reviewers for their valuable and constructive suggestions.

References

[1] H. Leknes, E.S. Aartun, H. Andersson, M. Christiansen, T. Andersson Granberg, Strategic ambulance location for heterogeneous regions, Eur. J. Oper. Res. 260 (2017) 122–133.
[2] V.A. Knight, P.R. Harper, L. Smith, Ambulance allocation for maximal survival with heterogeneous outcome measures, Omega 40 (2012) 918–926.
[3] S.T. Fylkeskommune, Facts and Numbers, (2018) http://www.stfk.no/no/Fylket_vart/Fakta_og_tall/.
[4] Stortinget, Stortingsmelding nr.43 1999–2000, 4.5 responstid for ambulansebiler må klargjøres, (2000) http://www.regjeringen.no/nb/dep/hod/dok/regpubl/stmeld/19992000/stmeld-nr-43-1999-2000-/4/5.html?id=193572.
[5] R. Aringhieri, M.E. Bruni, S. Khodaparasti, J.T. van Essen, Emergency medical services and beyond: addressing new challenges through a wide literature review, Comput. Oper. Res. 78 (2017) 349–368.
[6] V. Bélanger, A. Ruiz, P. Soriano, Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles, Eur. J. Oper. Res. 272 (2019) 1–23.
[7] M. Reuter-Oppermann, P.L. van den Berg, J.L. Vile, Logistics for emergency medical service systems, Health Syst. 6 (2017) 187–208.
[8] R. Church, C. ReVelle, The maximal covering location problem, Papers Reg. Sci. 32 (1974) 101–118.
[9] D. Schilling, D.J. Elzinga, J. Cohon, R. Church, C. ReVelle, The team/fleet models for simultaneous facility and equipment siting, Transp. Sci. 13 (1979) 163–175.
[10] K. Hogan, C. ReVelle, Concepts and applications of backup coverage, Manag. Sci. 32 (1986) 1434–1444.
[11] M.S. Daskin, A maximum expected covering location model: formulation, properties and heuristic solution, Transp. Sci. 17 (1983) 48–70.
[12] E. Erkut, A. Ingolfsson, G. Erdoğan, Ambulance location for maximum survival, Naval Res. Logist. 55 (2008) 42–58.
[13] P. van den Berg, G. Kommer, B. Zuzáková, Linear formulation for the maximum expected coverage location model with fractional coverage, Oper. Res. Health Care 8 (2016) 33–41.
[14] R. Boujemaa, A. Jebali, S. Hammami, A. Ruiz, H. Bouchriha, A stochastic approach for designing two-tiered emergency medical service systems, Flex. Serv. Manuf. J. 30 (2018) 123–152.
[15] I. Sung, T. Lee, Scenario-based approach for the ambulance location problem with stochastic call arrivals under a dispatching policy, Flex. Serv. Manuf. J. 30 (2018) 153–170.
[16] S.G. Davis, Analysis of the deployment of emergency medical services, Omega 9 (1981) 655–657.
[17] J. Goldberg, R. Dietrich, J.M. Chen, M.G. Mitwasi, T. Valenzuela, E. Criss, Validating and applying a model for locating emergency medical vehicles in Tuczon, AZ, Eur. J. Oper. Res. 49 (1990) 308–324.
[18] T. Ünlüyurt, Y. Tunçer, Estimating the performance of emergency medical service location models via discrete event simulation, Comput. Ind. Eng. 102 (2016) 467–475.
[19] R.C. Larson, A hypercube queuing model for facility location and redistricting in urban emergency services, Comput. Oper. Res. 1 (1974) 67–95.
[20] A. Ingolfsson, EMS planning and management, in: G.S. Zaric (Ed.), Operations Research and Health Care Policy, Springer, 2013, pp. 105–128.
[21] S. Ansari, L.A. McLay, M.E. Mayorga, A maximum expected covering problem for district design, Transp. Sci. 51 (2017) 376–390.
[22] V.J. De Maio, I.G. Stiell, G.A. Wells, D.W. Spaite, Optimal defibrillation response intervals for maximum out-of-hospital cardiac arrest survival rates, Ann. Emerg. Med. 42 (2003) 242–250.
[23] R. McCormack, G. Coates, A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival, Eur. J. Oper. Res. 247 (2015) 294–309.
[24] Z. He, X. Qin, Y. Xie, J. Guo, Service location optimization model for improving rural emergency medical services, Transp. Res. Rec. 2672 (2018) 83–93.
[25] S. Weiss, L. Fullerton, S. Oglesbee, B. Duerden, P. Froman, Does ambulance response time influence patient condition among patients with specific medical and trauma emergencies? South. Med. J. 106 (2013) 230–235.
[26] P.T. Pons, V.J. Markovchick, Eight minutes or less: does the ambulance response time guideline impact trauma patient outcome? J. Emerg. Med. 23 (2002) 43–48.
[27] E.T. Wilde, Do emergency medical system response times matter for health outcomes? Health Econ. 22 (2013) 790–806.
[28] M. Peyravi, S. Khodakarim, P. ”Ortenwall, A. Khorram-Manesh, Does temporary location of ambulances ("fluid deployment") affect response times and patient outcome? Int. J. Emerg. Med. 8 (2015) 37.
[29] G. Nichol, A. Laupacis, I.G. Stiell, K. O'Rourke, A. Anis, H. Bolley, A.S. Detsky, Cost-effectiveness analysis of potential improvements to emergency medical services for victims of out-of-hospital cardiac arrest, Ann. Emerg. Med. 27 (1996) 711–720.
[30] L. Brotcorne, G. Laporte, F. Semet, Ambulance location and relocation models, Eur. J. Oper. Res. 147 (2003) 451–463.
[31] S. Su, C.L. Shih, Modeling an emergency medical services system using computer simulation, Int. J. Med. Inform. 72 (2003) 57–72.
[32] Q. Su, Q. Luo, S.H. Huang, Cost-effective analyses for emergency medical services deployment: a case study in Shanghai, Int. J. Prod. Econ. 163 (2015) 112–123.
[33] L.A.C.G. Andrade, C.B. Cunha, An ABC heuristic for optimizing moveable ambulance station location and vehicle repositioning for the city of S ao Paulo, Int. Trans. Oper. Res. 22 (2015) 473–501.
[34] M. Amorim, S. Ferreira, A. Couto, Road safety and the urban emergency medical service (UEMS): strategy station location, J. Transp. Health 6 (2017) 60–72.
[35] M. Liu, D. Yang, F. Hao, Optimization for the locations of ambulances under two-stage life rescue in the emergency medical service: a case study in Shanghai, China, Math. Probl. Eng. (2017) 14 p..
[36] J.C. Dibene, Y. Maldonado, C. Vera, M. de Oliveira, L. Trujillo, O. Schütze, Optimizing the location of ambulances in Tijuana, Mexico, Comput. Biol. Med. 80 (2017) 107–115.
[37] R. Fritze, A. Graser, M. Sinnl, Combining spatial information and optimization for locating emergency medical service stations: a case study for lower Austria, Int. J. Med. Inform. 111 (2018) 24–36.
[38] J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer, 2011.
[39] R.L. Francis, T.J. Lowe, M.B. Rayco, A. Tamir, Aggregation error for location models: survey and analysis, Ann. Oper. Res. 167 (2009) 171–208.
[40] J.L. Saver, S.R. Levine, Alteplase for ischaemic stroke – much sooner is much better, Lancet 375 (2010) 1667–1668.
[41] H.A. Snooks, J. Dale, C. Hartley-Sharpe, M. Halter, On-scene alternatives for emergency ambulance crews attending patients who do not need to travel to the accident and emergency department: a review of the literature, Emerg. Med. J. 21 (2004) 212–215.