Delft University of Technology

Emergency medical services and beyond
Addressing new challenges through a wide literature review

Aringhieri, Roberto; Bruni, Mariaelena E.; Khodaparasti, Sara; van Essen, Theresia

DOI
10.1016/j.cor.2016.09.016

Publication date
2017

Document Version
Accepted author manuscript

Published in
Computers & Operations Research

Citation (APA)
Aringhieri, R., Bruni, M. E., Khodaparasti, S., & van Essen, T. (2017). Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, *78*, 349-368. https://doi.org/10.1016/j.cor.2016.09.016

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Emergency Medical Services and beyond: Addressing new challenges through a wide literature review

R. Aringhieri*[1], M.E. Bruni†[2], S. Khodaparasti‡[3] and J.T. van Essen§[4]

[1]Department of Computer Science, University of Turin, Italy
[2]Department of Mechanical, Energy and Management Engineering, University of Calabria, Cosenza, Rende, Italy
[3]Department of Mathematics and Computer Science, University of Calabria, Cosenza, Rende, Italy
[4]Delft Institute of Applied Mathematics, TU Delft, The Netherlands and CWI, The Netherlands

September 21, 2016

Abstract

One of the most important health care services is emergency medical service as it plays a vital role in saving people's lives and reducing the rate of mortality and morbidity. Over the last years, many review papers have discussed emergency medical services (EMS) location problems, however, only few review papers consider the full range of EMS systems. This review paper tries to fill this gap. Our review introduces the concept of emergency care pathway following the current trend in health care systems, i.e., shifting the central role from health care providers to patients. Considering the emergency care pathway, we provide a broad literature review and analysis in order to identify emerging challenges for future research.
Keywords — Emergency Medical Service, Emergency Department, Operations Research, Data Mining, Health Technology Assessment, Clinical Pathway.

1 Introduction

Emergency medical service is one of the most important health care services as it plays a vital role in saving people's lives and reducing the rate of mortality and morbidity. The importance and sensitivity of decision making in the Emergency Medical Services (EMS) field have been recognized by operations research scientists, EMS planners, and health care practitioners who studied many problems arising in the management of EMS systems since the 1960s.

Locating EMS vehicles has been widely studied as is shown by numerous review papers that appeared over the last years. One of the first review papers that addressed EMS location models is the paper of ReVelle et al. [1]. Brotcorne et al. [2] presented a review paper on ambulance location and relocation problems, and classified the existing models in three main groups: static and deterministic models, probabilistic models, and dynamic models. Goldberg [3] surveyed operations research models for the deployment of EMS vehicles while focusing on modeling aspects and problems assumptions. In these

*roberto.aringhieri@unito.it
†mariaelena.bruni@unical.it
‡sara.khodaparasti@unical.it
§j.t.vanessen@tudelft.nl

review papers, challenges posed by real world applications, potentials gaps in the literature, and new trends have been addressed.

In a recent review paper, Li et al. [4] studied papers addressing different types of covering models for EMS planning (set covering, maximal covering, maximum expected covering, maximum availability, gradual coverage, and cooperative coverage), hypercube queuing models, and dynamic allocation and relocation models. They also discussed the solution methods used for solving the aforementioned models. Finally, Basar et al. [5] presented a taxonomic framework for the EMS location problem. Considering 84 papers published in major journals, they developed their classification based on three major features including problem type, modeling, and methodologies. For the problem type, they divided the models into four general groups based on the type of emergency, model structure (deterministic or stochastic), variation in time (static or dynamic), and the number of objectives. For the modeling, they considered five major sub-classes of models differing in the definition of the objective function, the parameters involved and the mathematical programming type (integer, non-linear, dynamic programming, etc.). They categorized the methodologies into exact solution methods, heuristics, metaheuristics, and simulation.

Amongst the review papers on EMS systems, there are only a few review papers that have a broader view on the management of an EMS system. Green and Kolesar [6] surveyed the role of operations research and management science in improving emergency responsiveness over time, which lead to new policies and practices. They also studied the relationship between historical events (terrorism attacks) and the evolution of emergency response models and methods. Henderson [7] presented a discussion on challenges in EMS, highlighting the role of system-status management (redeployment strategies) in improving EMS systems. Finally, Ingolfsson [8] surveyed research on planning and management for EMS, emphasizing four topics: (i) demand forecasting, response times, and workload; (ii) measuring performance; (iii) choosing station locations; and (iv) allocating ambulances to stations based on predictable and unpredictable changes in demand and travel times.

In EMS studies, an important question that should be addressed is whether the current direction of research is consistent with the main aims in EMS systems. As Sorensen and Church [9, page 9] mentioned "*the recent EMS location literature has diverged, at least to some extent, from the goals of many EMS agencies*". A possible explanation for this divergence could be the difficulty of modeling the increasingly complex system that EMS has become over time. In order to manage an EMS system, relevant data should be forecasted, complex situations should be modeled, efficient solution methods should be designed, and accurate dispatching policies should be implemented. The main purpose of this review is to account for such a complexity through a broad literature review in order to address new challenges. This will indicate relevant research opportunities for researchers in this field and will show practitioners how EMS systems can benefit from this research.

The structure of our review is based on the concept of Emergency Care Pathway (ECP). In doing so, we follow the current trend in health care systems, i.e., shifting the central role from health care providers to patients. The main concept of Clinical Pathways (CPs) in health care systems shifts the attention from single departments to the entire health care chain, which increases patient's safety and satisfaction, and optimizes the use of resources. Campbell et al. [10] defines the CPs as "health-care structured multidisciplinary plans that describe spatial and temporal sequences of activities to be performed, based on the scientific and technical knowledge and the organizational, professional and technological available resources".

In EMS management, many decision problems are connected to each other. Therefore, decisions taken in one step of the ECP can affect decisions in subsequent steps of the ECP. An example of such a decision is the relationship between Emergency Department (ED) overcrowding and ambulance diversion, as discussed in Section 5.1.

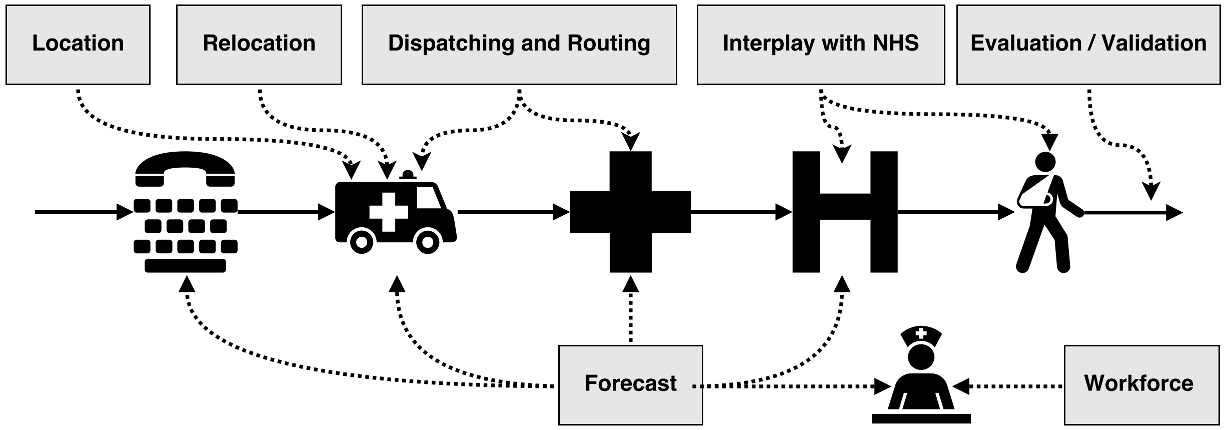

Figure 1: Emergency Care Pathway and related problems.

Figure 1 depicts a possible ECP composed of five main steps and its relationships with related management and organizational problems. The ECP starts when the EMS receives an emergency request. After determining the urgency of the incident, an ambulance is dispatched. The ambulance should reach the emergency scene as soon as possible to provide first-aid and to transport the patient to the ED of a hospital. Once the patient is discharged from the hospital, the ECP finishes.

Some of the problems highlighted in Figure 1 have already been introduced above, such as the ambulance location and relocation problems. In addition, ambulance fleet management should also decide upon a dispatching policy and routing of the vehicles. The next problem in line is the interplay with the National Health Service (NHS). EMS can be viewed as an entry point of NHS which contributes to ED overcrowding. In turn, overcrowding influences the throughput of the ECP and the outcome of patients in the ECP. Therefore, the interplay with the NHS system is a problem that should be addressed.

One of the prerequisites to guarantee an efficient and fair management of the entire ECP is a good forecast of, for example, emergency demand, travel time and workload. These forecasts are needed to ensure that enough resources are available to fulfil the emergency demand. These resources consist of, for example, EMS vehicles, paramedics and medical doctors. Composing sufficient workforces and determining proper rosters is essential to deliver high quality health care.

In order to determine whether the care provided for the entire ECP is *good* enough in terms of efficiency, effectiveness, and fairness, evaluation metrics are needed. For ambulance location problems, several metrics in terms of coverage and response times exists as described in Section 2, however, evaluation of the entire ECP is also needed. Related to this is the validation of the model outcomes resulting from simplifying assumptions in the modeling phase.

Following the lines depicted in Figure 1, the review is organized as follows. Section 2 is devoted to the analysis of ambulance location problems and classifies the existing literature on two key concepts with respect to the analysis of an ECP, namely *equity* and *uncertainty*. Section 3 provides an overview of ambulance relocation models. Section 4 is devoted to the analysis of dispatching and routing policies. In Section 5, we investigate the complex interplay between the EMS system and other components of the NHS, and in particular with other emergency health care delivery systems. The literature concerning the evaluation and validation of an EMS system is surveyed in Section 6. Forecasting techniques to determine demand, travel time and workload of an EMS are reviewed in Section 7.

Workforce management with respect to forecasted demand is reviewed in Section 8. Finally, Section 9 provides a discussion on the main challenges that EMS systems should face in the future.

2 Location of ambulances

The majority of operations research studies in the field of EMS focus on the issue of designing comprehensive EMS systems in which early response to emergency calls is provided. This critical challenge reveals the outstanding role of locational decision making in EMS. This has resulted in the provision of a rich and vast literature on location problems in EMS, providing a wide variety of models, solution methods, and real case studies.

Interestingly, as a by-product, the progress in location theory and models is also partially due to the efforts made by operations researchers in the EMS field. In fact, it is remarkable that most location models, which are widely applied in other areas, have been first inspired to address a specific aspect of the EMS.

Our viewpoint differs from those of other reviews [1–5] since it provides an overview of challenges, gaps and new trends that should be faced. We highlight some aspects of EMS location problems that have not been referred to in other review papers, classifying the existing literature based on two key concepts: *equity* and *uncertainty*.

2.1 Incorporating equity

Equity is one of the most challenging concerns in the healthcare sector and especially in EMS systems, since it evaluates the fairness of how resources (notably EMS vehicles) are allocated to patients.

Since the main aim of EMS planners is to provide early response, equity is usually expressed as a function of distance or Response Time (RT) traveled by EMS vehicles. Although there are some standard equity indicators in location literature (mostly distance-based deviation measures [11]), there exist only few studies that incorporate equity in EMS location problems (see also Li et al. [4]) and no general consensus exists on appropriate equity measures.

To evaluate the performance of different Response Time Thresholds (RTTs), McLay and Mayorga [12] design a location model which maximizes the overall marginal increase of the fraction of high priority calls that meet the performance standards. They also incorporate the concept of patient survival in their model by simplifying the survival function of Larsen et al. [13]. To address the equity concept in their model, they include a coefficient in the objective function which corresponds to the proportion of rural demand at each demand node. The model is tested on data of the Hanover County EMS system in Virginia. The results show that "*longer RTTs (*9 *and* 10 *min in the case of Hanover county) result in more equitable patterns of patient survival. In these solutions, patient lives were saved in the rural areas at the expense of losing patient lives in the urban areas, which reduced the disparity between patient survival rates in urban and rural areas. Since a* 9 *min response time threshold is the most common performance measure used by EMS systems in the United States, this suggests that many EMS systems implicitly consider patient equity or fairness*" [12, page 135].

Interestingly, in contrast to [12], there is some evidence showing that EMS planners are not much concerned about equity in services. For instance, Erkut et al. [14] mention that the existence of different actual performance as well as the RT standards (especially in rural and urban areas) support the claim that EMS system managers are against the provision of equal access. They also state that "*although a policy of equal access seems difficult to criticize, such a policy implies that lives are valued differently in different areas, because the cost of saving a life can be much higher in sparsely populated rural areas than in urban centers*" [14, page 43]. The development of a consensus on the definition of equity in healthcare provision should rely on the composition of two different equity concepts referred to as horizontal equity and vertical equity [15]. Horizontal equity implies that all demand nodes are considered in an equal manner. It means that suburban nodes are treated without discrimination. Vertical equity is concerned with the distribution of the service between different groups. By this definition, EMS systems are equitable if they favor disadvantaged groups, compensating for overall inequities. Horizontal equity has been researched extensively in health economics, while vertical inequity is receiving growing attention in the last years [16].

To decrease the level of disparity for EMS systems in rural areas, Chanta et al. [17] present three different bi-objective location models solved by the ϵ-constraint method [18]. First, the demand nodes are classified in rural and urban nodes based on their geographical positions. Then, the authors investigate the trade-off between efficiency and equity criteria by maximizing the expected number of covered calls on one hand, and by choosing one the following three criteria on the other hand: the minimization of the number of uncovered rural demand zones, minimization of the maximum distance between uncovered demand zones and their closest open stations, and the minimization of the number of uncovered zones. It is worthwhile to note that maximizing the expected covered demand favors the location of ambulances in densely populated areas, which results in longer response times for patients in more rural areas. To compensate this drawback, the first two additional objective functions try to implement vertical equity, whereas the last one focuses on horizontal equity. As the authors acknowledge, comparing the three model solutions is very difficult given the lack of a unitary measure of vertical and horizontal equity.

In [19], Chanta et al. use the *Gini coefficient* [20] to measure horizontal equity. The Gini coefficient is a measure of statistical dispersion, which was firstly developed for evaluating inequity in economic and social welfare literature, and is the most commonly used measure of inequality. Extending the minimum-envy model proposed by Espejo et al. [21] in which the priority levels of serving stations are incorporated, Chanta et al. [19] develop a minimum p-envy location problem to provide equitable services in emergency response location problems. The contribution of this model is modeling customer dissatisfaction as a new distance-based metric denoted by the envy. The envy is a function of distance, which indicates the dissatisfaction level of a demand node with its serving station in comparison to another demand nodes for the same level of priority. The model then minimizes the overall envy value. There are several reasons supporting the use of satisfaction as a significant factor in health care. First and above all, satisfaction is an important health outcome in its own right, enhancing the compliance and the cooperation among customers. Second, evaluating satisfaction is important for continuous quality monitoring and improvement in healthcare delivery. To show the validity of their model, Chanta et al. compare their results with those obtained through the p-center model [22] and a model where the Gini coefficient is minimized [20], which are both accepted equity models in location literature. Following their previous work, Chanta et al. [23] present a modified version of the p-envy model in which the distance-based objective function is replaced with a survival function, dependent on the RT [12]. The model imposes lower bounds on the individual survival rates provided by first priority servers as well as on the system-wide survival rate. The model handles the unavailability of EMS vehicles by using the hypercube model (see the references in Section 2.2).

The Gini coefficient, together with the *squared coefficient of variation* [24], has also been used by Toro-Diaz et al. [25] to measure equity among servers. Here, it evaluates the relative differences among the workload of different stations obtained through the approximation procedure of Budge et al. [26]. Toro-Diaz et al. adopt an interesting fairness perspective that equalizes the performance of the system by reducing disparities in the mean response time of different demand zones, as well as in the workloads of the servers (EMS vehicles). To account for efficiency, the model also considers *the mean response time* and *the expected coverage*.

Besides the balance between efficiency and equity, an important issue which deserves thorough investigation is the trade-off between horizontal and vertical equity. Following this stream, Khodaparasti et al. [16] present an integrated framework for siting facilities in an equitable and efficient manner. The results of this research show that, especially for the applied equity measures, horizontal and vertical indices are consistent with each other.

Another way to address equity is by applying cooperative location games to EMS planning. Most of the previous research addresses non-cooperative location games while the incorporation of server cooperation, especially in heavy loaded systems, is inevitable. This issue is of interest in EMS planning due to the close relation of server cooperation with the busy fraction as we will show in Section 2.2.

In [27], Fragnelli and Gagliardo report a game theoretic approach for the EMS location problem with interaction between candidate locations for housing EMS vehicles. Their method orders the potential locations based on two location games, namely the *coverage game* and the *multi-coverage game*. The focus in the coverage game is on maximizing the coverage of the area while the multi-coverage game aims at reducing the overlap of covered zones. The model is able to deal with the equity criterion by maximizing the area covered by all ambulances without considering the number of calls received. This approach can be considered as a starting point for further investigation of applying game theory to the EMS location planning. Since one area can be covered by more than one EMS vehicle, the cooperative game should take the marginality issue into account, i.e., the contribution of each EMS vehicle to the total coverage. In cooperative game theory, the Shapley value (see, e.g., [28]) accounts for the total surplus generated by the coalition of all players and provides an index of the importance of each player concerning the overall cooperation. Fragnelli and Gagliardo define the Shapley value based on two fairness criteria: the coverage indifference and the demand indifference. Two algorithms are provided to compute the Shapley value of the games in polynomial time.

> Our literature review indicates that, even though some effort has been made in the last years to incorporate equity principles into EMS systems, there still exist important under-investigated areas on equity concepts that deserve attention. The existing literature on the incorporation of equity in EMS planning does not follow a straight direction and neglects important realistic and practical aspects of equity. One of the reasons for this might be the absence of a general consensus on equity measures for the EMS.
> Future research should focus on different aspects of equity in EMS planning and on the definition of a set of widely accepted evaluation metrics for developing equity based models in EMS planning. To this end, studying the equity concept from both a horizontal and vertical point of view might be a good starting point.
> The focus of equity in the EMS literature has mostly been on the geographical position of demand, which initially imposes the division of demand nodes into urban and suburban/rural zones. From this perspective, the efficiency of EMS systems has been evaluated by the quality of service (the provision of timely service) in urban areas and the fairness of service is, in general, evaluated by the coverage provided in suburban/rural zones. Even though spatial distribution is the most commonly used characteristic to differentiate between demand zones, also other characteristics can be used. Even demand zones located in the same urban zone can be different from each other in some characteristic, which calls for different treatment of these

demand zones. Finally, since EMS systems are subject to dynamic pressures and often the actual configuration does not comply with the original design of the system, it is important to embed the temporal aspect into the equity concepts and to carefully investigate this issue. In this way, we might also include equity considerations in ambulance relocation (see Section 3), which is very common in practice.

2.2 Incorporating uncertainty

Uncertainty plays a big role in the EMS analysis. There is uncertainty about the amount and location of demand, travel times, severity of incidents, availability of EMS vehicles, length of stay at the ED, etc. However, despite these uncertainties, costly decisions with long term implication need to be taken, which can be a challenging task. In addition, parameter estimates might be inaccurate due to poor measurements. Therefore, making strategic decisions under uncertainty requires tailored approach to exploit strategically relevant uncertain information. To address this issue of uncertainty in EMS literature, the existing literature adopts one of the following approaches: the *probabilistic paradigm*, the *stochastic programming approach*, the *robust counterpart*, and the *fuzzy framework*.

Probabilistic paradigm. EMS literature has focused on the uncertainty in three types of factors that can play a role in planning decisions: demand, availability of EMS vehicles, and response times.

The probabilistic paradigm mainly focuses on the last two sources of uncertainty. This results in a probability that a node is covered, which depends on the availability of EMS vehicles and uncertainty in RT. Two different streams in this framework can be distinguished. One of these streams is based on the results of the descriptive Hypercube Queueing Model (HQM). This model relies on queueing theory to estimate the busy fractions of EMS vehicles and other system performance measures. The other stream incorporates uncertainty directly into system parameters such as RT, service time, delay in response, busy fraction, etc., and is based on mixed-integer linear programming.

The descriptive HQM was developed by Larson [29] to evaluate steady-state busy fractions, loss probabilities, average RTs, and expected coverage, for any fixed configuration of facilities. Even though the hypercube model addresses cooperation between EMS vehicles as well as variation in workload for EMS vehicles, it relies on several assumptions for dispatching rules and it requires that the (exponentially distributed) service time is the same for all EMS vehicles, regardless the type of emergency call and the location of the vehicle. Moreover, HQM ignores many practical policies that most organizations employ regularly as, for instance, EMS vehicles being dispatched while on the road, diversions of EMS vehicles from low to high priority calls and many other features of EMS systems such as the presence of multiple depots. In spite of these disadvantages, HQM works well in practice and is considered to be a viable alternative to simulation. In fact, most researchers use it as a post-solution analysis tool for the evaluation of performance measures (see Erkut et al. [30]). Larson [31] presents an approximate HQM to avoid the computational difficulty of HQM. Other contributions to HQM are motivated by the need to overcome the restrictive assumptions of HQM.

Burwell et al. [32] present a modified version of Larson's approximate HQM in which the ties in preferences for EMS vehicles is captured. Jarvis [33] presents a generalized HQM in which service time distributions can be dependent both on base locations for EMS vehicles and on the type of emergency calls. Batta et al. [34] develop correction factors for the busy fractions in an embedded HQM. They include the fact that EMS vehicles do not operate independently, and thus, may have different busy fractions which depend on each other and on the location of the EMS vehicle. Takeda et al. [35] apply the hypercube approach to investigate the effect of the decentralization of EMS vehicles over the service area by applying the proposed model to the urban Emergency Medical Service of Campinas in Brazil. The results of this study show that by increasing the number of EMS vehicles that are partially decentralized, the performance measures and especially the RTs improve while variation in the workload of EMS vehicles is negligible. Further, the results of the model show that the total decentralization policy may not produce satisfactory results for the decision makers. Iannoni et al. [36] propose two hypercube queueing-based models to evaluate the EMS performance on highways. The model can deal with three priority levels for calls, different EMS vehicle types, partial backup, and a multiple dispatch policy.

In most previous works, the busy fraction is considered as an exogenous input parameter which is estimated through site-specific or area-specific formulations, overlooking the fact that a (site-specific) busy fraction is one of the model outputs and should be estimated after knowing the exact configuration of facilities (see, e.g., [2, 37]). In order to address this shortcoming, Cho et al. [37] propose an EMS location model for siting trauma centers and helicopters in which the busy fraction of each helicopter is endogenized as a variable depending on the number of patients transported by the helicopter per time unit. In view of the fact that vehicle specific busy fractions depend on the deployment of EMS vehicles, Shariat-Mohaymany et al. [38] present a strategic and tactical model for minimizing the costs of locating EMS vehicles and their base stations while imposing an upper bound on the busy fraction of each deployed EMS vehicle. This upper bound depends on the minimum number of EMS vehicles required to guarantee a desired reliability level as suggested in [39].

In the following, we consider probabilistic models based on mixed-integer linear programming. The apparent difficulty in this is that the stochastic elements of EMS systems result in non-linear expressions, which need to be approximated. In practice, this can compromise the accuracy of the model predictions and may lead to sub-optimal solutions for the original non-linear models.

The seminal contribution is undoubtedly the Maximal Expected Covering Location Problem proposed by Daskin [40]. The MEXCLP uses a system-wide busy fraction obtained by dividing the total workload of the system by the total workforce of EMS vehicles available. Assuming that EMS vehicles operate independently, it is possible to estimate the reliability of service at each node using probability rules. The assumption of EMS vehicles being uniformly busy throughout the area, regardless spatial variations of demand, is relaxed in the Maximum Availability Location Problem (MALP) formulation [39], where local busy fractions depend on both the number of EMS vehicles available and the aggregated level of demand within each local area. Sorensen and Church [9] present a probabilistic location model for EMS planning which integrates the area-specific busy fraction of MALP in MEXCLP. The simulation results show the superiority of the proposed model over MALP and MEXCLP in terms of the overall percentage of calls that receive coverage within the RTT.

The server independence assumption is relaxed in the Queueing based Probabilistic Location Set Covering Problem (Q-PLSCP) and the Queueing-based Maximum Availability Location Problem (Q-MALP) formulations of Marianov and ReVelle [41]. To evaluate the reliability of service at each demand node, Q-MALP makes use of an M/G/s-loss queueing model. Although Q-MALP relaxes the server independence assumption, it still relies on local service areas and on location-independent service times. Ingolfsson et al. [42] recognize that, beside ambulance availability, randomness has an impact on delays and travel times. This study emphasizes that by ignoring the randomness in delays, the coverage performance might be overestimated.

These findings are confirmed by Erkut et al. [30], who evaluated five probabilistic covering models, namely the Maximal Covering Location Problem (MCLP) proposed by Church and ReVelle [43] and a variation of it including probabilistic response times (MCLP+PR), the MEXCLP, and two non-linear formulations that include probabilistic response times and base location specific busy probabilities (MEXCLP+PR and MEXCLP+PR+SSBP). Erkut et al. compare the five models on three performance measures, namely expected coverage, loss probability and average RT, by applying them to the case of Edmonton, Canada. They conclude that the two non-linear models, MEXCLP+PR and MEXCLP+PR+SSBP, perform better in terms of coverage compared to MCLP+PR and MEXCLP, which confirms that the accuracy in modeling complex systems can lead to substantial improvements in performance. In particular, based on the obtained results as well as on their previous research in [14], they conclude that models that include uncertainty "*not only result in better coverage estimates, but also cause coverage to be a better proxy for lives saved*" [12, page 64].

> We should warn the reader that simplifying assumptions might come at a cost. This is, for example, the case for the MALP model where using a local busy fraction requires the assumption of a local service area. Which simplified assumption poses the main limitation on finding an optimal solution for the true problem is difficult to determine. Another limitation of mixed integer probabilistic problems is the needed computation time. To efficiently solve these optimization problems, tailored metaheuristic approaches have to be developed. However, it is not guaranteed that these solutions methods lead to an optimal solution. Therefore, simulation is needed to determine how well the found solutions perform in practice. Simulation might also shed some light on which simplified assumptions poses the main limitation on finding the optimal solution.

Stochastic programming paradigm. In the literature, different stochastic programming problems have been presented to explicitly consider the dependence of the system reliability on the randomness in demand. In the stochastic programming paradigm, the model parameters are assumed to be random variables with a known probability distribution function. This assumption can be justified by the availability of historical data (or outputs of a simulation model) from which the empirical probability distribution might be estimated. Under this assumption, two different strategies for making decisions can be considered. In the chance constrained paradigm, "here and now" decisions are taken under uncertainty and implemented whatever the future is. Since a probabilistic guarantee is imposed on the feasibility of the solution, there is a chance that the implemented solution is not feasible in practice. A different approach is represented by the stochastic programming paradigm with recourse. In this case, some of the decisions must be made under uncertainty, whereas other decisions, the recourse decisions, can be postponed until the realizations of the random variables become known. Hence, the decisions taken in the first stage can be corrected in other stages by the recourse decisions, which explicitly depend on the specific value (realization) assumed by the random variables. Both the probabilistic paradigm and the recourse paradigm have their advantages and disadvantages. The probabilistic paradigm is most appropriate when the reliability of the system (for example, in terms of coverage) is most important. However, the two-stage nature of facility location problems – where locations are chosen given uncertain future demand, but customers are assigned to facilities once the uncertainty has been resolved – advocates the use of the recourse paradigm.

Ball and Lin [44] develop a model with separate chance constraints, where the probability of failure (the inability to be serviced within the RTT) of each demand point should be less than a given threshold value. They consider the worst case busy fraction, which occurs when each server is attending all calls from its neighbourhoods. In [45], Beraldi et al. develop a location model based on joint probabilistic constraints to capture the inherent uncertainty in demand. The probabilistic constraints are jointly imposed on all demand points, which ensures that the reliability of the entire geographical area is kept above the prescribed level. In another work, Zhang and Li [46] develop another strategic and tactical location model in which the probabilistic chance constraints on the capacity of base locations are approximated by their equivalent second-order cone constraints, assuming that random demands can be modeled by continuous random variables. Even though the distribution of emergency calls is naturally discrete and follows a Poisson distribution, the considered approximation can be justified by the computational tractability of the resulting model.

To capture the inherent two-stage nature of location decisions, Beraldi and Bruni [47] present a two-stage stochastic model with embedded joint probabilistic constraints. The paper relaxes the independence assumption among servers at the expense of discretizing the uncertainty into a finite set of scenarios. Following the same structure, Noyan [48] proposes a two-stage stochastic location model for EMS planning in which the uncertainty in demand is taken into consideration by integrated chance constraints. These integrated chance constraints can be considered to be a relaxed version of separate chance constraints and allows a convex approximation of the non-convex feasible set defined by the probabilistic constraints.

Since full discretization of uncertainty in scenarios results in solving huge mixed integer problems, often a limited number of scenarios is generated by using proper scenario generation techniques. These techniques aim at finding a good trade-off between 1) including enough scenarios to guarantee an accurate representation of the underlying stochastic process and 2) including not too many scenarios such that the computational tractability of the resulting scenario-based mixed integer problem is preserved.

> Regardless the tractability of the model, we should first investigate whether it is worthwhile to include randomness into EMS models. Even though uncertainty plays a big role in EMS models, it is not evident that this should indeed be included as is done in the stochastic programming paradigm. Although well known measures can be used to evaluate the advantages of using a stochastic model over a deterministic model (namely the *Value of the Stochastic Solution* and the *Expected Value of Perfect Information*), these quantities are not useful to decide whether a stochastic model should be used in practice. The validation of potential advantages of stochastic models over deterministic ones should be critically assessed (see also Section 6.2). In addition, stochastic models usually require information on distributions and other values that might not be known. As a conclusive remark, the art of modeling amounts to describing the crucial aspects of a problem and approximating the less relevant ones. Including randomness might turn out to be not the most important issue to address.

Robust optimization and fuzzy paradigms. A more recent approach for optimization under uncertainty is "robust optimization". While there are many high level similarities with the stochastic paradigm, robust optimization is a distinct field that seeks a solution that is feasible for any realization in a given uncertainty set instead of immunizing the solution against stochastic uncertainty. This paradigm is successful in various application areas because of its computational tractability. Moreover, this approach is the only reasonable alternative when information about the probability distribution is not readily available.

Adopting a robust counterpart approach, Zhang and Jiang [49] propose a bi-objective location model for strategic and tactical EMS planning in which uncertain demand is addressed. The uncertainty in the number of emergency calls is captured by an ellipsoidal uncertainty set and the weighting method is applied to make a trade-off between costs and responsiveness.

Another computational tractable way of addressing uncertainty is the fuzzy paradigm. This framework is mostly applied when the probabilistic paradigm or stochastic framework cannot be used. In the absence of historical data, qualitative information extracted from expert's observations is a crucial aspect in the decision making process. As a matter of fact, some service level concepts in EMS decision making such as patient satisfaction and staff performance are inherently known as qualitative information. The fuzzy paradigm facilitates the use of qualitative data as well as expert-based knowledge by characterizing them as linguistic terms. There are only a few papers in this stream. Some papers adopt the fuzzy set theory concept to deal with uncertainty, others apply the fuzzy programming approach to solve multi-objective location models.

Araz et al. [50] present a three-objective location model for strategic EMS planning in which different Fuzzy Goal Programming (FGP) solution methods are applied. The first and the second criteria in the objective function maximize the total number of demand nodes covered at least once and twice, respectively. The third criterion minimizes the total traveled distance for the uncovered demand nodes. The authors use the fuzzy goal programming approach to express the aspiration levels as imprecise values (fuzzy numbers). The efficiency of the proposed approach, in terms of first and backup coverage, is shown by comparing the results of different proposed FGP models with the lexicographic multi-objective technique. Although, the proposed FGP models provide better results with respect to the backup criterion, the performance of such techniques highly depends on the choice of the weights and on the priority levels assigned to different criteria. To allow the decision maker to find near-optimal solutions for different problem inputs in a short period of time, tailored heuristics could be developed. Following this idea, Uno et al. [51] present an interactive FGP approach combined with the particle swarm optimization method to solve an emergency location problem.

In [52], a fuzzy location model for EMS planning based on the Double Standard Model (DSM) [53] is presented. Koc and Bostancioglu use linguistic variables to capture the uncertainty in both travel times and RTTs. A triangular fuzzy number is associated with each linguistic term and a probability-possibility transformation is applied to find the probability distribution of each fuzzy variable. A Monte Carlo simulation shows the efficiency of the proposed approach for reproducing unknown data.

A clear advantage of the fuzzy paradigm is that it does not require much information on the probability distribution of uncertain parameters. However, relying only on expert knowledge or qualitative data may lead to severe loss of information about the behavior of the system, which in turn may lead to inaccurate and unrealistic results. Therefore, it is best to only use the fuzzy approach in cases where the incorporation of qualitative data as well as expert-based opinions is unavoidable.

Remarks. In summary, the type of uncertainty as well as the type of available information should be considered when choosing an appropriate solution approach. Applying a combination of approaches that includes different types of uncertainty, could help EMS managers to come up with a more realistic model in which the complexity of the system is fully captured and a better overview of the whole system is provided.

The EMS models reviewed so far, focus either on minimizing the expected costs/response time or on maximizing the expected coverage. This focus is appropriate for risk-neutral decision makers as these models are insensitive to costs/coverage variations. However, arriving at an incident location only a few seconds earlier can already save a human's life. Thus, instead of minimizing expected values, risk averse decision makers might accept higher costs in return for higher protection against losing lives. Unfortunately, traditional location models fail to meet the needs of risk-averse planners as only a few papers suggest mechanisms to reduce the chance of unfavorable large RTs or to increase the probability of arriving at the scene on time.

> It is important to incorporate the notion of risk aversion in EMS location models, for example, by providing risk-based measures especially designed for EMS systems. This is one of the most interesting directions for future research within the stochastic framework. Furthermore, there is an increasing need for developing models that simultaneously incorporate realistic information and several uncertainty sources. Most existing studies in the literature focus on only one or two stochastic aspects and there is a lack of location models handling all sources of uncertainty simultaneously. One reason for this deficiency is the complexity of the resulting models, which necessitates the design of tailored solution methods with high computational efficiency. In addition, future research within the robust stream could enrich the literature in the EMS field and could also provide a comparative framework in which the distinguishing features of stochastic and robust models can be investigated.

3 Relocation of ambulances

Most existing models in the EMS literature belong to the class of static strategic problems in which long-term and mid-term decisions are taken for establishing base stations, assigning EMS vehicles to base stations, and determining the fleet size. Unlike other applications of location models, EMS location models must include relocating EMS vehicles to deal with variations in demand. In this setting, the repositioning of idle EMS vehicles to back-up busy EMS vehicles (also called redeployment) represents the core problem [54]. Clearly, this entails the incorporation of dynamic aspects into EMS location problems. Two alternative strategies are found in the literature: multi-period mixed integer problems or truly dynamic problems. These problems do not focus on finding an optimal final configuration of EMS vehicles, but on the transformation from a given configuration to a reconfiguration at the end of some planning horizon. Although redeployment is a real-time decision making process that captures dynamic fluctuations in the system, mainly multi-period problems have been studied which provide decision makers with EMS vehicle locations for a given set of time periods within a given planning horizon.

A common redeployment strategy used in practice, is the use of a compliance table for real-time relocation of EMS vehicles. A compliance table shows where EMS vehicles should be located when calls come in and when vehicles become available. As such a compliance table can be determined in advance, real-time optimization is not required. However, a weakness of compliance tables is that dispatchers have to relocate the EMS vehicles frequently to maintain a satisfying compliance level. This can cause many problems for drivers and paramedic staff [55]. Furthermore, when the number of idle ambulances changes rapidly, the system will not be in compliance for some time.

While compliance table policies are easy to implement in real-time, it is a challenge to determine the best one. Alanis et al. [56] propose a Markov chain model to evaluate performance measures (such as RT distribution, the distribution of the number of busy servers, expected travel time, and average server utilization) for different compliance tables. A recent paper by Sudtachata et al. [57] uses a mixed integer program to define the best nested compliance table in order to maximize expected coverage. A limitation of this approach is that it considers only nested policies (only one EMS vehicle already on the move can be relocated), which restricts the solution space. Considering a single type of EMS vehicle and a single type of call priority, the optimal nested compliance table is determined using steady state probabilities of a Markov chain model. Differently from [56], these parameters are approximately evaluated regardless the exact compliance table used. The model validation is performed by using a discrete event simulation model applied on a real case study.

The first multi-period model on ambulance relocation is [58], in which the dynamic DSM at time t (DDSM^t) – a multi-period extension of the static DSM [53] – is presented. Gendreau et al. include a penalty term in the objective function to limit repeated relocations of the same EMS vehicle, round trips, and long trips between two location sites. In the model, the input data are updated whenever a new call arrives. Based on this, a new deployment pattern is computed. To solve the resulting mixed integer program (MIP) model in real-time, the authors implement a parallel tabu search heuristic in a parallel computing system. The simulation on real data of the Montréal EMS system, shows that, when two successive emergency calls occur at the same time or within a very short time frame, the computation of the redeployment plan can be time-consuming or even infeasible.

Moeini et al. [59] apply a slightly modified version of DDSM^t to the EMS planning of the Val-de-Marne county in France, in which each demand node has two different demand values, namely the total average number of calls and the average number of concurrent calls. They claim that differentiating between the number of concurrent calls and the total number of calls can be useful to deal with uncertainty in the EMS planning. In contrast to DDSM^t, which maximizes the total number of demand nodes covered by two EMS vehicles, the aforementioned model maximizes the total number of demand nodes covered once as well as the total number of concurrent demand nodes covered twice. Basar et al. [60] propose another variant of the DDSM^t in which two different coverage thresholds are considered. The objective function maximizes the total number of demand nodes that have been covered twice by two different EMS vehicles.

Schmid and Doerner [61] present a multi-period version of DSM with time dependent travel times in which the fraction of demand nodes assigned to EMS vehicles is taken into consideration. The model can capture variations in travel speed and travel time over the multi-period planning horizon. To control the number of relocations over the planning horizon, a penalty term is included in the objective function. A metaheuristic method based on variable neighborhood search is used to solve the model. The results show that it is essential to consider time-dependent variations in travel times and coverage.

Gendreau et al. [62] propose an extension of the MEXCLP to dynamically relocate EMS vehicles between stations. Differently from their previous model, they impose some constraints to limit the number of vehicle relocations. Their model considers a system with n EMS vehicles, which results in $n+1$ system states that correspond to the number of idle EMS vehicles. The model precomputes a series of location scenarios for different system states. However, the model is only tractable for systems with a small numbers of vehicles. Gendreau et al. present results for n up to 6.

Jagtenberg et al. [63] develop a heuristic algorithm for real-time EMS vehicle redeployment based on the idea of MEXCLP. The dispatching policy for idle EMS vehicles is based on maximizing the marginal expected coverage of the corresponding EMS vehicle. In [64], van den Berg and Aardal present a time-dependent version of MEXCLP addressing fluctuations in travel times, service times, demand, and the availability of servers. The model also restricts the costs of establishing EMS base stations as well as the number of EMS vehicle relocations over the planning horizon.

The models presented above are powerful tools that account for variations in the number of EMS vehicles and travel times throughout the planning horizon. However, they do not explicitly consider real-time system changes due to EMS vehicles becoming idle or new calls arriving. To deal with this issue, the dynamic programming paradigm can be used. In [54], Maxwell et al. present an Approximate

Dynamic Programming (ADP) model that repositions idle EMS vehicles such that the coverage is maximized. In contrast to the models that are based on integer programs, the dynamic programming formulation captures the real-time evolution of the system while solutions can be computed quickly. This also holds for realistically sized instances, since only a simple optimization problem has to be solved. This work is a good starting point for future research on redeployment of EMS vehicles.

Uncertainty has also been incorporated into multi-period models that follow different paradigms. For example, Naoum-Sawaya and Elhedhli [65] present a two-stage stochastic model for EMS vehicle redeployment. The scenario-based approach captures the uncertainty in the location of emergency calls during each time period. The model minimizes the number of relocations as well as the number of calls not being responded to within the RTT.

Another multi-period covering location model for EMS vehicle redeployment is developed by Rajagopalan et al. [66]. The model minimizes the total number of EMS vehicles and base stations over all time periods while meeting coverage requirements with a predetermined reliability. Their model extends the queueing probabilistic location model of Marianov and ReVelle [67] in which ambulance specific busy fractions are incorporated. They also adopt the Jarvis hypercube approximation algorithm [33] to take the server cooperation probability into account. To solve the model, a reactive Tabu Search algorithm was designed.

Degel et al. [68] develop a multi-period model for EMS location and relocation planning at a tactical level in which the temporal and spatial variations in demand and travel times are captured. Three different criteria are considered in the objective function of the model: the first criterion maximizes the total demand covered with respect to the flexible required coverage level, the second criterion minimizes the total number of relocations over the planning horizon, and the third criterion minimizes the total number of flexible EMS vehicle sites (volunteer fire departments and hospitals) occupied by EMS vehicles during all time periods. The model is to some extent similar to MALP, but it uses empirical data instead of busy fractions to estimate the required coverage regarding the dynamic fluctuations in demand as well as travel times.

A stochastic dynamic model combining dynamic EMS vehicle relocation and dispatching decisions has been proposed by Schmid [69]. The model addresses both the fluctuations in call arrival rates and travel times. Further, it is based on the EMS law in Austria prohibiting the relocation of idle EMS vehicles between waiting sites and only allowing the relocation of EMS vehicles after an emergency call is served. The model outputs are: (1) a dispatching plan determining which EMS vehicle should respond to a potential emergency call and (2) a relocation pattern indicating where the EMS vehicle should be send after serving a call. The objective function minimizes the expected RT over all periods while variations in demand and travel times are explicitly taken into account. An ADP method was used to solve the model. In addition, Schmid criticizes the closest dispatching policy by comparing the results obtained from a real case study of the Vienna EMS system.

> The variety of methods available to determine EMS vehicle redeployment policies poses the challenge to determine which redeployment policy performs best in practice. All redeployment policies have their own advantages and disadvantages, and there is no evidence on the superiority of one method over the other. The paper by Bélanger et al. [70] presents the first attempts to evaluate redeployment strategies in real-time EMS fleet management, however, determining a trade-off between different redeployment policies and providing bounds on their performance measures are still interesting topics for future research. Both studies reported in [54] and [69] share the idea of applying ADP as a tool to handle a high-dimensional and

uncountable state space in dynamic programs. These studies emphasize the need of incorporating realistic features such as call priority and multiple EMS vehicle dispatching. On the other hand, the computational challenges posed by dynamic programming approaches should be addressed to foster the design of novel redeployment models.

4 Dispatching and routing policies

Important real-time operational problems in EMS management are dispatching and routing. Dispatching is the act of choosing appropriate EMS vehicles to respond to emergency calls based on the nature and location of calls. Routing decisions are concerned with defining the exact route that a dispatched ambulance should follow to reach a patient.

4.1 Dispatching strategies

An EMS vehicle can be dispatched after an emergency call is received or after the vehicle becomes idle after serving a call. Based on this fact, Lee [71] addresses dispatching from a *call-initiated* and an *EMS vehicle-initiated* perspective. In the call-initiated framework, the dispatcher should select one of the current idle EMS vehicles to be dispatched after the arrival of an emergency call. In the EMS vehicle-initiated framework, the dispatcher should choose which of the emergency calls in the waiting queue will be served whenever an EMS vehicle becomes idle. It is clear that for EMS systems in which the arrival rate of calls in comparison with the number of idle EMS vehicles is low (low-load system), call-initiated dispatching decisions should be made, while in EMS systems with a high call rate, EMS vehicle-initiated dispatching policies should be applied.

In practice, many EMS systems apply simple dispatching policies, such as the closest-idle policy, which always sends the closest available EMS vehicle to each call, or the first-in first-out policy, which is based on the priority of incoming emergency calls. The closest-idle policy has gained much attention in EMS planning due to its simplicity. However, it has been recently criticized for resulting in suboptimal solutions [55, 69, 72–75], as well as increasing the waiting time for subsequent calls. To overcome the drawbacks of the closest-idle policy, many researchers suggest to take the priority of calls into consideration.

Andersson and Varbrand [73] present a dynamic relocation model combined with a dispatching policy based on the preparedness measure. The preparedness measure enables the dispatcher to choose the most appropriate EMS vehicle to respond to a call given that the system should be sufficiently prepared to respond to subsequent calls. The integer program relocates EMS vehicles through the system to maintain a minimum preparedness threshold and, more importantly, to minimize the maximum travel time. Then, they use the preparedness measure to construct a dispatching policy based on the call priority by considering three priority levels. They also apply the concept of *pseudo priority* in which the priority of calls can be changed according to their waiting times. This will guarantee that the RT for priority 2 and priority 3 calls will not be too long.

McLay and Mayorga [75] present a Markov Decision Process (MDP) to dispatch distinguishable EMS vehicles to prioritized calls while considering the fact that errors in the classification of patient priorities might occur. The model determines the dispatching policy that maximizes the expected coverage of true high-risk calls. To show the efficiency of their model, they consider three different priorities for incoming calls as well as three different policies including the closest-idle policy while only considering priority 1 calls to be high-risk (under-responding) and the closest-idle policy while considering priority 1 and 2 calls to be high-risk. The results of their study show that over-responding (under-responding) is preferable only when there is a high (low) rate of classification errors.

To complete their previous work, McLay and Mayorga [76] propose an equity- and efficiency- based dispatching MDP model in which distinguishable EMS vehicles as well as emergency call priorities are addressed. The efficiency-based objective function of the model focuses on the fraction of covered calls. In addition, four equity constraints are considered to reflect the fairness from both the patient's and vehicle's points of view. The patient's equity measures, imposed through linear constraints in the MDP model, are related to the equity of process (*ex ante* equity) and the equity of outcomes (*ex post* equity). The first equity constraint ensures for each priority level and demand point that the fraction of patients served by their absolute closest (and not necessarily the closest available) EMS vehicle is greater than a given threshold. The second equity constraint ensures that the fraction of high priority surviving patients is greater than a given lower bound for each demand point. To deal with fairness for EMS vehicles, two different sets of constraints are considered. The first one guarantees that all EMS vehicles have approximately the same busy fraction by restricting the busy fractions to lie between a predetermined lower and upper bound. The second equity constraint makes sure that for each EMS vehicle, the dispatching rate to high-priority calls lies above a given lower bound. This will help to improve the paramedics' skills. The study also investigates how the incorporation of different equity measures can affect the dispatching policy. However, the authors also mention that incorporating equity might lead to a lower service or negative outcomes.

Bandara et al. [55] present a priority-based heuristic dispatching rule based on static dispatching rules and fixed deployment. It always sends the closest available EMS vehicle to priority 1 calls and sends a less busy EMS vehicle to priority 2 calls. In other words, this rule provides an ordered list (contingency table) for priority 2 calls based on the busy fraction of EMS vehicles. The results of applying this dispatching policy indicate that the patient survivability, the average RT, and the percentage of covered calls corresponding to priority 1 are improved in comparison with the closest-idle policy. The average RT for priority 2 calls slightly increased to ensure shorter response times for priority 1 calls. However, this slight increase does not affect the survivability of priority 2 calls.

Kanchala et al. [77] present a priority-based dispatching policy for multiple-unit EMS systems in which two types of vehicles, namely Advanced Life Support (ALS) and Basic Life Support (BLS) vehicles, as well as three priority levels are considered. Such systems are referred to as two-tiered EMS systems. They apply a simulation method combined with a heuristic algorithm to find a good dispatching policy. Based on the results of this study, the two closest ALS and BLS vehicles are dispatched to priority 1 calls while the closest BLS vehicle is dispatched to priority 2 calls. This maximizes the survival rate for priority 1 calls. The dispatching policy for responding to priority 3 calls is based on sending a farther EMS vehicle to balance the workloads among servers.

Haghani et al. [78] present a simulation model to evaluate the performance of three different dispatching policies, namely the first called first served strategy, the nearest origin assignment policy, and the flexible assignment dispatching strategy. Their main goal is minimizing average RTs while considering several call priorities. By applying a dynamic shortest path algorithm based on real-time travel time information, the model is able to guide EMS vehicles through non-congested routes. The results of their study show the superiority of the flexible assignment strategy when it comes to reducing average RTs.

Henderson and Mason [79] present an integrated priority-based dispatching model based on simulation and a Geographical Information System (GIS). They solve the routing problem by computing the shortest paths based on time-dependent travel times.

In contrast to the papers highlighting the role of call priority, there are other dispatching policies that follow a different stream. These works mostly integrate the closest assignment policy with some notions borrowed from other fields.

Lee [74], extending the work of Andersson and Varbrand [73], proposes a composite policy in which the preparedness as well as the RT are considered simultaneously, whereas the priority of calls is neglected. Moreover, the effect of applying different aggregation functions used in social welfare literature is studied to evaluate preparedness. The simulation shows that the results of a composite dispatching policy are superior to those obtained by the closest-idle policy.

Lee [71, 80] introduce the centrality-based dispatching policy, which combines the centrality notion, used in complex networks, and the closeness notion. The main idea behind incorporating centrality into dispatching decisions is that by responding to the most central calls, especially when the probability of transferring patients to a hospital is low, the chance of in-time response to the next calls will increase. Moreover, the dispatcher is able to cover the highly populated areas appropriately and prevents the EMS vehicles from remaining in rural areas. As the author mentions, the applicability of the centrality-based policy highly depends on the probability of transferring the patients to the hospital (hospital probability) which in turn depends on many factors such as crew expertise, the nature and severity of accidents, and resource scarcity. In EMS systems with a lower hospital probability, the role of the centrality concept is less evident. The computational experiments reveal the superiority of the centrality-based dispatching policy over the closest-idle policy with respect to the average RT as well as the variation in RTs. The papers suggest that the simultaneous incorporation of priority-based and centrality-based policies is able to provide more efficient dispatching decisions. Another hybrid model, combining network centrality measures and a first-in first out policy, is presented by Zarkeshzadeh et al. [81]. To overcome the shortcoming of a pure centrality dispatching policy, Lee [82] adopted, in a recent paper, the notion of parallelism to develop a parallelized centrality-based dispatching policy in which both idle and busy vehicles are considered simultaneously. The idea originates from the fact that a currently busy server might respond sooner to a call after completing the service, than idle vehicles located farther from the call.

Apart from the design of efficient dispatching policies, the evaluation of dispatching performance measures should also be investigated. In [26], Budge et al. present a methodology to evaluate the dispatch probability by considering four characteristics. These characteristics are variations in location-specific workload, multiple vehicles at stations, call- and vehicle (station)-dependent service time, and vehicle cooperation which is related to the dependency between EMS vehicles. In [83], Lim et al. consider three components, namely the queueing of calls, the assignment of EMS vehicles, and add-on dispatches. Using these components, the relative performance of various dispatch policies is examined and the advantages and disadvantages of different policies is discussed.

> Incorporating different EMS vehicles and several emergency priority levels is important for future research on EMS redeployment and dispatching policies. In addition, priority-based dispatching systems should also take into account the equity of their solutions. [73] made a first attempt to include equity by developing a pseudo priority approach for determining dispatching policies.
> Using communication technologies when implementing novel dispatching policies will enable dispatchers to dynamically update the priority of assigned calls and to reroute EMS vehicles if necessary. Even though practitioners and scientists acknowledge the importance of this

approach, it has not yet been applied in practice (see also the discussion in Section 9). The seminal paper of [84], which proposes an optimization model that uses real-time traffic information to assist dispatchers in assigning multiple EMS vehicles to incidents, seems to be an isolated example.

The recent popularity of emergency response decisions in disaster management emphasizes the importance of vehicle-initiated dispatching policies. In disaster management, decision makers are often challenged to respond to a high number of calls in the waiting list which results in drastic fluctuations in vehicle availability. Therefore, the incorporation of other vehicle-initiated dispatching policies (such as priority-based dispatching policies) designed especially for disaster response is more than welcome.

Finally, the knowledge obtained by dispatchers in practice can be useful when developing reliable dispatching policies. To this end, knowledge management methods such as spatial analysis combined with data mining techniques can be used as suggested in Section 7.1.

4.2 Ambulance routing problem

The ambulance routing problem is one of the other challenges in EMS management which has not been surveyed in detail. Routing decisions can directly affect RTs and, as a consequence, other performance measures.

In [85], Oran et al. propose an integrated location-routing model for emergency response which considers the priority of calls. The model ensures that high-priority calls are prioritized. Panahi and Delavar [86] design a dynamic routing system in which real-time variations in traffic congestion (dynamic travel times) are incorporated through the use of geographical information systems. The results of the study indicate up to 20% improvement in RTs as well as the superiority of dynamic routing strategies over static routing policies.

In [87], Jotshi et al. develop an integrated EMS routing and dispatching system for disaster management based on simulation and data fusion. The dispatching decisions are divided into the *patient pickup problem* and the *patient delivery problem*, i.e., sending an EMS vehicle to the incident location and sending an EMS vehicle from the incident location to the hospital. The patient pickup problem considers the priority of calls and clustering criteria. The dispatching policy gives a higher priority to responding to a cluster of demand. In the absence of demand clusters, an available EMS vehicle is dispatched to the nearest high-priority incident. The routes of the EMS vehicles are calculated based on real-world road networks, existing road damage, and congestion. Data fusion is used to provide a reliable estimate for parameters (such as casualties, road status, and demand clusters) for which data is obtained from multiple sources.

Beaudry et al. [88] propose a two-phase algorithm to solve a patient transportation problem arising in large hospitals. The aim is to provide an efficient and timely patient transportation service between several locations on a hospital campus. The main difficulty is that requests arrive in a dynamic fashion. The solution methodology must therefore be capable to quickly insert new requests in the current routes. Different from standard dial-a-ride problems, the problem under study includes several complicating constraints which are specific for the hospital context.

In another paper, Ardekani et al. [89] present two vehicle routing heuristics for inter-facility patient transfer in which both scheduled and unscheduled transfers are considered. The objective of the proposed heuristic algorithms is to minimize travel times and to minimize the violation of time windows and working hours. The results of the simulations show improvements on all performance measures when compared to the existing schedule.

In a recent paper, Talarico et al. [90] develop two ambulance routing models for disaster response management where calls are divided into two groups: low-priority calls, which require only on-scene care, and high-priority calls, which should be transferred to the hospital. The first model minimizes the sum of the weighted maximum service completion time for low-priority and high-priority calls. Due to the large size of the model and the necessity to solve the model in real-time, a modified version of the former model is presented which considers all EMS vehicles to be identical but located at different hospitals.

> It is essential that efficient solution methods are developed that can determine optimal or near-optimal routes in real-time. To incorporate more realistic travel times, stochastic aspects should be included into the developed models. In addition, by improving the communication between the drivers and the call center, better routing decisions can be made.

5 Interplay with other emergency health care delivery systems

The EMS can be viewed as an entry point of the National Health Service (NHS), which plays a fundamental role in the delivery of emergency care. As a consequence, it is important to investigate the interplay of the EMS system with other components of the emergency care system since the performance of the EMS system can be affected by any of them. In this section, we consider three components that are important in the ECP, namely the EDs, the location of static emergency devices, and the care systems dealing with *First Hour Quintet* care delivery.

5.1 Emergency departments

The ED plays a crucial role in delivering on-time care to emergency patients. After a patient is transported to a hospital by an ambulance, he or she needs to be transferred to the hospital such that appropriate care can be given. However, because of ED overcrowding or ambulance offload delay, this care might be postponed. In addition, the ambulance is not available for serving other emergency patients since it has to wait until the patient is taken over by the hospital. One of the solutions to avoid or to limit this negative effect determined by the ED overcrowding is the ambulance diversion. This means that ambulances are redirected to other EDs where there is no overcrowding yet. However, the EDs the ambulances are redirected to are often further away, and in addition, because of the ambulance diversion, less central EDs may also become overcrowded.

Hoot and Aronsky [91] describe the causes, effects and solutions to the ED overcrowding in a systematic review. The causes described are non-urgent visits of patients, influenza seasons, hospital closures, ambulance diversion, inadequate staffing, delay in diagnostics and hospital bed shortages. As described before, ED overcrowding leads to delayed patient care. This results in an increased risk of mortality, patient elopement and also financial losses. The relationships between ED overcrowding, ambulance diversion and its causes and effects are depicted in Figure 2 from Pham et al. [92].

In a search for solutions to ED overcrowding, Hoot and Aronsky distinguish three remedies: increased resources, demand management and operations research. Increasing resources means that there is more capacity to treat patients. These resources can include personnel, beds or diagnostic services.

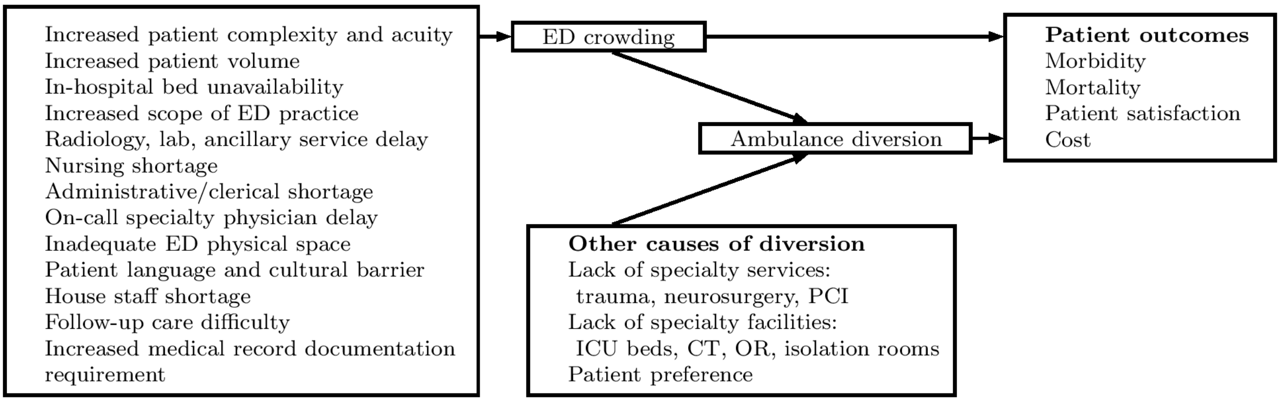

Figure 2: Relationship between ED overcrowding and ambulance diversion (source [92])

Demand management considers methods to redistribute patients or to improve the utilization of resources. These methods include non-urgent referrals, ambulance diversion and destination control. The latter means that ambulances are distributed over the EDs by an information system as described by Sprivulis and Gerrard [93]. Their information system is an internet portal which displays near-real-time information on the ED workload conditions. Based on these conditions, it is determined to which ED a patient is transported. For the most severely injured patients this will be the nearest hospital, but the other patients are evenly distributed over the other EDs. Ambulance diversion and destination control are not the preferred solutions from the ambulance perspective. These solutions often mean that an ambulance has to transport the patient to a hospital further away, which reduces the ambulance availability for other patients. In addition, this leads to delayed patient care. However, when the additional travel time is less than the offload delay at an overcrowded ED, these options should be preferred.

The operations research solutions described by Hoot and Aronsky [91] consist in simulation and queueing theory. Delgado et al. [94] provide a systematic review of insights from simulation models. All studies described by Delgado et al. found that reducing the time admitted patients were awaiting placement into inpatient beds at the ED and the length of stay at the ED were expected to reduce ambulance diversion. In addition, Delgado et al. report that Nafarrate et al. [95] found that the number of patients in the waiting room is a better trigger for ambulance diversion than inpatient bed availability, as it provides the best balance between accessibility and waiting times.

Ramirez et al. [96] show that adding a fast track unit for low acuity patients can significantly reduce the ambulance diversion time. Storrow et al. [97] find that decreasing the lab turnaround time decreases the ED length of stay which in turn decreases the time spent on ambulance diversion. Kolker [98] concludes that canceling some elective admissions could reduce ambulance diversions. In addition, smoothing the elective surgery schedule allowing no more than five intensive care admissions per day can reduce ambulance diversion even more. Hagtvedt et al. [99] and Ramirez-Nafarrate et al. [100] provide quantitative evidence that cooperative strategies between hospitals to reduce ambulance diversion have more impact than non-cooperative strategies. Deo and Gurvich [101] also showed this with the use of a queueing game.

In addition, there exist some more recent papers describing operations research solutions for ED crowding and ambulance diversion. Ramirez-Nafarrate et al. [102] use a simulation model to determine the effect of several ambulance diversion policies on the patient's waiting time. A policy based on an MDP performs the best, followed by a policy based on the number of beds occupied. The worst performing policy is the no diversion policy. Almehdawe et al. [103] use queueing theory to model the interface between an ambulance provider and several EDs. They investigate the effect of several scenarios on offload delays. The results show that by giving ambulance patients a higher priority, the offload delays decrease at the cost of longer waiting times for walk-in patients. The model can also be used to determine the effect of routing choices and increasing certain resource capacities on the offload delay. Ramirez-Nafarrate et al. [104] use an MDP formulation to study optimal ambulance diversion control policies. Their focus is to minimize the average time that patients wait beyond their recommended safety time threshold. In addition, they analyze the value of information on the time to start treatment in one of the other hospitals.

> The main challenge here it to develop a model that takes all the above mentioned aspects into account, i.e., a model that combines the existing research trends. This would shed light on the system as a whole and could provide solutions acceptable for all stakeholders.
> In addition, due to budget costs, ED networks are re-organized. Because of this, many EDs will only house a few medical specialties, whereas only a few EDs will house all medical specialties. This will lead to a hierachical network with many logistically challenging aspects.

5.2 First Hour Quintet care systems

First Hour Quintet is related to critical conditions for which the EMS system can have a significant impact on the outcome. Examples are cardiac arrests, acute coronary syndromes, strokes, respiratory failures and severe trauma. Cardiac arrest and stroke are the most relevant from a social point of view: cardiac arrest is responsible for approximately 400,000 deaths per year in North America [105] while the total number of deaths caused by strokes is currently estimated at 1,239,000 per year for the 48 European countries combined (508,000 per year for the 27 European Union members [106]). Furthermore, Di Carlo [106] reports that "*About half of stroke survivors are left with some degree of physical or cognitive impairment. The need of support for common daily activities directly impacts quality of life of patients and their relatives, frequently taking the role of caregivers. Although often neglected, informal care is of paramount relevance to maintain stroke survivors in the community, and a valuable economic resource for health care systems*".

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 [107]. Based on this, Fischer et al. [108] confirm the relevance of First Hour Quintet diseases for human life and introduce the rate of First Hour Quintet incidences per year over 100,000 inhabitants. The organization of emergency care systems for First Hour Quintet diseases directly involves the use of EMS resources as reported by Churilov and Donnan [109]. Furthermore, Churilov and Donnan reported "*the complete lack of operations research studies regarding the EMS component*" of a stroke care system [109, Table 2, page 12].

Cardiac arrest occurs when the heart stops pumping blood in a coordinated fashion due to abnormal heart rhythms. The likelihood of survival from cardiac arrest decreases by 7 to 10% for every minute of delay in treatment. In order to support EMS intervention, trained people can use Automated External Defibrillators (AEDs). An AED is a portable device that automatically diagnoses cardiac rhythms and delivers a shock to correct abnormal activity in the heart if needed.

Usually, AEDs are located near public places such as subways, train stations, airports, universities, and so on. Locating AEDs is quite different from classic location problems. The main difference is that the AED location is not necessarily known by AED users. As a consequence, demand is not always satisfied by the closest idle "facility" as is usually the case in classic location problems. This implies that the AED location problem resembles a multiple-responder model that maximizes coverage in which more than one AED can contribute to the coverage of a cardiac arrest. Siddiq et al. [110] showed that "*optimization models can be a policy evaluation tool, to measure the potential impact of technological innovations that can increase the effective radius of public AEDs*".

Only a limited amount of research has been dedicated to AED location problem. Mandell and Becker [111] focus on the equitable distribution of AEDs over EMS vehicles by using a multi-objective integer programming model. Rauner and Bajmoczy [112] developed a decision model to evaluate the cost-effectiveness of placing AEDs in EMS vehicles. Myers and Mohite [113] use the MCLP of Church and ReVelle [43] to determine locations for AEDs on a university campus. Dao et al. [114] optimize the locations of AEDs in an indoor environment. Chan et al. [115] show that an MCLP-driven approach to AED deployment outperforms an intuitive population-based method. Chan et al. [116] develop three mixed integer non-linear models that guide the deployment of public AEDs for bystander use in case of an emergency. These models generalize existing location models and incorporate differences in bystander behavior. The authors derive equivalent integer linear reformulations or easily computable bounds. Using data from Toronto, Canada, they show that optimizing AED deployment outperforms the existing approach by 40% in coverage. In addition, substantial gains can be achieved by relocating existing AEDs.

> The main challenge here is the development of quantitative models to evaluate the interplay between the EMS system and First Hour Quintet care systems as proposed in [109]. For example, the availability of dedicated care units (such as stroke units or cardiac intensive care units) should be taken into account when deploying an EMS vehicle. This will guarantee that the patients receive appropriate care in a timely fashion. Furthermore, it is important to optimize the availability of these dedicated care units as these units may not be available 24 hours per day.

6 Evaluation and validation

The evaluation of the performance of EMS systems should be based on practical performance measures such as health outcomes. By developing a unitary evaluation framework, different ECPs can be compared. Such an evaluation framework can also provide insight in the effect of simplifying assumptions on the accuracy of the model when applied in practice [6]. In this section, we consider ways to evaluate EMS systems, to validate models and to determine the accuracy of solutions obtained by the models.

6.1 Evaluation

Most of the existing models in the literature apply coverage or time based performance measures as a proxy for health outcomes. In the following, a review of papers addressing survivability in EMS literature, one way of evaluating the health outcome, is provided.

The survival function, which models the survival probability for a patient served by an EMS vehicle, monotonically decreases depending on many factors, notably the response time. Survival-based models can be viewed as the probabilistic version of gradual coverage location models. To overcome the drawback of covering models, Erkut et al. [14] develop different survival-maximizing models in which the concept of coverage is replaced with a survival function. In their research, they refer to four existing survival functions in literature which have been especially computed for cardiac arrest calls. By using these survival functions, it is emphasized that EMS systems should be designed to minimize response times, and not to maximize the number of calls served within a given time limit. By incorporating the idea of a survival function into covering models (MCLP, MEXCLP, and MEXCLP+PR), they present a class of survival-based location models in which the concept of coverage in replaced with a survival function. The results of the study indicate that survival-based models will provide more realistic solutions compared to standard covering models, whose main shortcoming is that they cannot differentiate between the response times of covered demand nodes.

McLay and Mayorga [12] present a method to evaluate the performance of RTTs in terms of resulting patient survival rates. In [117], McLay and Mayorga surveyed the relation of different RTTs and patient survival rates. The results of this study show that RTTs can be a good proxy for survival rate when appropriate RTTs are chosen. Knight et al. [118] propose a simulation model for locating EMS vehicles that incorporates survival functions for multiple classes of heterogeneous patients. The model aims to maximize the overall expected survival probability of multiple classes of patients with different medical conditions and corresponding survival functions. The model is validated using data from the EMS system in Wales.

One of the main challenges in EMS evaluation is the development of a unitary framework to compare different EMS systems. This is not a new topic in the EMS practitioners' community. The main objective of the EED, ended in 2004, was to identify common components of EMS systems and to create a common data set for monitoring and assessing these systems. One of the main outcomes of this project is a list of five key indicators to assess the performance of EMS systems. Two of these indicators are the response time (% within 480 sec) for calls with the highest priority and the ratio unit hours (ELS + BLS + ALS) p.a. over 100,000 inhabitants. ELS stands for Emergency Life Support and such a vehicle is used to provide first aid until a BLS vehicle arrives at the scene. Unfortunately, the proposed indicators only consider the efficiency of EMS systems and omits other measures such as fairness.

To overcome this limitation, the Health Technology Assessment (HTA) can be used. HTA systematically evaluates the properties and impacts of new technology. This approach is particularly appropriate for EMS systems since it does not only consider cost issues, but also a related wide range of information such as safety, clinical effectiveness, patients' needs and benefits, costs of therapy, as well as social, organizational, legal, and fairness implications. Performance is indeed a multidimensional concept that should take the following into account: *efficiency*, measured by comparing input and output, *effectiveness*, assessed by comparing output and outcomes, and *fairness*, measured in terms of equal access and concerning both horizontal and vertical equity, i.e., providing the same treatments for the same needs and different treatments for different needs (see also Section 2.1). The mentioned dimensions must be evaluated and assessed since performance cannot be limited to only the efficiency dimension (the manager's point of view) or to the effectiveness dimension (the doctor's point of view) or the fairness dimension (the policy maker's point of view).

> In our opinion, the main challenge here is to develop a multidimensional system of indicators that is able to evaluate and compare new solutions by assessing the trade-offs that exists

among different objectives.
Further, it should be noted that the relation between the survival function and response time for other types of emergency calls (except cardiac arrest calls) has not been investigated in detail yet, which could help investigators to present more realistic models.

6.2 Validation

Simplifying assumptions are unavoidable in modeling. This usually results in less accurate models, however, the most important issue is the effect of these simplified assumptions on the performance of the system in practice [6]. The close scrutiny of these simplifying assumptions is of paramount importance to control possible modeling failures. The use of simulation can be effective in identifying the sources of errors that invalidate the models, and in the assessment of the validity of alternative approaches. Aboueljinane et al. [119] present a review of the many simulation models that have been developed over the years. Most of the available simulation approaches are based on a Discrete Event Simulation (DES) (see, e.g., [79, 120–126]).

One of the most critical issues to be addressed when developing a simulation model for an EMS system is how to model the movement of EMS vehicles. Within a DES framework, the movement of an EMS vehicle is usually modeled by an event which represents the fact that a vehicle has reached its destination. Such an event will occur after a given time interval modeling the travel time of the vehicle from its origin to its destination. This travel time can be computed by using a travel time model [79] or exploiting route planning software [127] based on an accurate speed estimation. Basically, the actual movement of an EMS vehicle is not explicitly modeled in simulation models.

Emerging simulation methodologies can be useful in addressing the vehicle movement issue. Agent Based Simulation (ABS) allows to track the behavior of each individual acting in the simulated environment [128]. The ABS model is based on the Environment-Rules-Agents framework proposed by Gilbert and Terna [129]. Such a framework requires an environment on which the agents can interact and move. Macal and North [130] report several examples of environment topologies. Among them, the authors list the euclidean space model and GIS topology. Therefore, we can affirm that the ABS methodology naturally models the vehicle movement since it requires an agent to move in its environment while interacting with other agents. This ABS characteristic makes the model more flexible when testing different vehicle management policies because it allows to track where the vehicle is located at the time of the decision. For instance, it allows to reroute a vehicle while it is moving if a more serious incident occurs in the proximity of the vehicle as discussed in [131, 132]. Recently, Anagnostou et al. [133] reported the development of a distributed hybrid ABS and DES model within the context of EMS systems. The aim of this research is to demonstrate the feasibility of using distributed simulation technology to implement hybrid simulation models for EMS systems.

GIS is a relatively new tool in health care services and organizations. However, health care professionals, who know how to use GIS and other spatial tools, are provided with a powerful decision support tool [134]. Using GIS, Peleg and Pliskin [135] performed a retrospective review of emergency calls and dispatch logs in two different Israeli regional districts. The precise locations of all calls provided by GIS were included and the data was stratified by weekday and by daily shifts. Geographic areas (polygons) of, at most, 8 minutes response time were then simulated to maximize the timely response of calls. When vehicles were positioned within the modeled polygons, more than 94% of calls met the 8-minute criterion. Huang et al. [136] propose geospatial analysis to assess the current EMS system of New Taipei City. The study allows to discover regions that cannot be reached within 10 minutes under pre-assumed conditions.

To the best of our knowledge, Henderson and Mason [79] were the first to use GIS information within an EMS simulation model. Other works using GIS information are Aringhieri [131] and Pinciroli et al. [137] within the research project “Decembria”, which was aimed at optimizing the vehicle fleet usage of the EMS system in the province of Milan, Italy. Pinciroli et al. [137] propose an interactive software simulator that allows for exact geocoding of request points, correct reproduction of vehicle routes, and computation of shortest paths according to graph distances while also taking forbidden manoeuvres and streets accessibility constraints into account. The simulator uses historical data obtained from the database of the EMS system in Milan. It also includes on-line optimization algorithms that provide the user with possible decisions within real time. Azizan et al. [138] present a prototype of a simulation framework to study the performance of the EMS delivery in Johor Bahru. Several real-life dispatching policies are simulated to evaluate the efficiency of local EMS delivery.

Pinto et al. [139] introduce the concept of reusable models for EMS systems. Usually, EMS simulation models are developed in an ad hoc fashion to evaluate a particular EMS system. Even if EMS systems are quite similar, simulators can often not be used for other EMS systems. The authors report the construction of a reusable model for EMS systems. Further, they describe the associated parameters, data sources, and performance measures, and report on the collection of information.

> The main challenge in this area is the development of a general purpose and reusable simulation model, as in Kergosien et al. [140], which exploits GIS and real time traffic information to compute shortest paths (usually time dependent) on a real network and based on the ABS methodology. On a large network, such a simulation model can require a large amount of computational resources. To overcome this problem, new advances in parallel and distributed programming can be really useful as discussed in [133].

As discussed in Section 2.2, several approaches have been developed to deal with the stochasticity of the EMS vehicle location problem. Combining optimization and simulation is a promising methodology as discussed in [141, 142] and it could be an alternative methodology to incorporate stochasticity within EMS vehicle location problems.

A first attempt to combine simulation and optimization within the field of EMS vehicle location problems can be contributed to Aringhieri et al. [132], who extend the 2008’s work presented in [143]. In their work, the authors propose the ABS-EMS simulation model and a low-priority calls coverage (LPCC) deterministic optimization model to locate EMS vehicles in the urban area of Milan. In the LPCC model, the authors introduce the concept of *EMS vehicle capacity*, i.e., a way to take the limited number of emergency requests that a vehicle can serve during a time interval into account. The main advantage of introducing this capacity parameter is the possibility to take the vehicle availability into account in a deterministic optimization model. The authors then propose an iterative greedy procedure to compute an alternative set of EMS vehicle locations for the urban area of Milan. The procedure starts from the optimal solution obtained by the LPCC model. This solution provides a lower bound on the number of vehicles needed to cover the area under the deterministic assumption. The solution is then evaluated via ABS-EMS by determining a ranking for the vehicles with respect to their utilization. The iterative procedure ends when a location is determined for all available EMS vehicles. The described procedure can also be used to determine the location of EMS vehicles such that a given overall system performance is guaranteed: instead of ending the procedure when a location has been determined for all available vehicles, the procedure continues by adding EMS vehicles until the performance of the system reaches a given threshold.

In 2009, Maxwell et al. [144] formulated a simulation model to evaluate the performance of a given allocation policy and to use this model in an ADP context to compute high-quality redeployment policies. Maxwell et al. find that the resulting ADP policies perform much better than sub-optimal static policies and marginally better than near-optimal static policies. Representative computational results for Edmonton, Alberta are included. This work has been extended and published in [54]. In his work, Zhang [145] deals with the dynamic vehicle redeployment problem by proposing optimal move-up policies based on three small-scale Markov models. Then, Zhang exploit simulation-based numerical optimization to tune the Markov model parameters. Mason [146] discusses operations research models and methods for simulating and optimizing EMS vehicle operations. The author describes a new simulation-optimization algorithm for determining EMS vehicle base locations, which was tested during a major reorganisation of ambulance operations in Copenhagen, Denmark. The author also reviews the literature in this area.

More recently, Zhen et al. [147] propose a simulation-optimization method to evaluate the operational performance of deployment plans through a detailed simulation model. A simulator takes a potential solution for the considered optimization problem (i.e., the vehicle deployment problem) and evaluates the solution's performance in a stochastic environment, in which the arrival time of requests, traveling time, and servicing time are uncertain. To guide the search process of the simulation-optimization method, the authors propose a genetic algorithm where the simulation model evaluates each individual. The authors also propose a mathematical model for vehicle relocations which is adapted to dynamically changing environments. Finally, a demo-example is reported based on the case of Shanghai. McCormack and Coates [148] introduce a similar approach where simulation is combined with a genetic algorithm to maximize the overall expected survival probability across different patient classes.

In contrast, Ingolfsson [8] highlights the use of stochastic models to predict how the EMS system performance changes as the deployment of EMS vehicles changes. He focuses on analytically solvable stochastic models rather than simulation models and states the following: "*a primary advantage of analytical models is their short computation time, which is important when using such a model as a component in a procedure to search for optimal or near-optimal deployment plans or as a component in a decision support system that allows EMS planners to experiment with deployment policies and to (almost) immediately see the likely consequences for system performance*" [8, page 117].

> The field of combining optimization with other quantitative methodologies requires further research to gain more insights in these possible hybridizations and to understand when simulation should be preferred over analytical models. From this perspective, the idea of combining optimization techniques with (generalized) stochastic Petri nets, which can be solved more efficiently than a standard DES simulation model, seems completely unexplored.

7 Forecasting

Usually, EMS systems can provide huge amounts of data to be used for forecasting since they are obliged to collect data for each call and EMS vehicle. The collected data usually contains information on the time a call is received, answered and finished, the location of the incident, the outcome of the triage evaluation, etc. For the EMS vehicles, the collected data contains information on the time of dispatching, arrival at the scene, departure from the scene, arrival at the hospital, etc. The availability of such data enables the development of forecasting models to predict – in some way – the expected *emergency demand*. This is a topic which is usually not considered in literature reviews concerning EMS management except for the review by Ingolfsson [8].

In our opinion, a good forecasting model is a key feature to provide an efficient and effective EMS management. A counter-intuitive example – which supports our remark – is discussed in Aringhieri et al. [132]. In this paper, the authors show that demands that heavily fluctuate in time and space can lead to managerial solutions that worsen the EMS performance even if more resources are used. The literature mostly deals with demand forecasting, i.e., how call volumes vary over time and space. However, only few contributions are available on forecasting travel times and workload.

7.1 Forecasting the demand

One of the earlier models, proposed by Hall [149], uses basic statistics in order to determine daily demand. Its main shortcoming is the lack of accounting for daily or weekly trend data, or other causal factors. Aldrich et al. [150] develop a model capable to predict the total demand using a large set of independent variables reflecting socio-demographic characteristics using least squares regression. Other regression models were developed by Siler [151] and Kvalseth and Deems [152]. Recently, McConnel and Wilson [153] developed a regression model that takes the age distribution of the population into account.

Another stream of forecasting models are those based on time series, which are able to overcome some weaknesses of regression techniques. Baker and Fitzpatrick [154] propose a Winters exponential smoothing model whose optimal parameters are defined through a goal programming model. The model is able to combine the forecast of emergency calls with *routine calls*, i.e., non-urgent calls. Channouf et al. [155] develop and compare several time series models to generate daily and hourly forecasts of EMS calls in Calgary, Alberta. The authors' objective is to provide a simple and effective model that can be used to develop effective simulation models.

Setzler et al. [156] are the first to recognize that EMS demand depends on the day of the week and the time of day. They develop an artificial neural network to forecast demand volume for specific areas during different times of the day. The forecasted demand is compared with current practice to determine the accuracy of the prediction.

The paper by Vile et al. [157] aims at exploring new methods to produce accurate forecasts. The authors propose a non-parametric technique for time series analysis known as singular spectrum analysis. The model – tested on data provided by the Welsh Ambulance Service Trust – produces superior long-term forecasts and comparable short-term forecasts to well established methods.

An alternative method is presented by Micheletti et al. [158]. The authors assume that the time and location of an emergency call are outcomes of a space-time marked point process. They estimate the process intensity by exploiting a rich database covering three years, from 2005 to 2007. Further, they relate the daily number of emergency call to various exogenous variables.

Since the problems discussed in Section 2 are NP-hard, a common practice is to aggregate demand points in order to reduce the computational burden. Aggregation makes the size of the problem more manageable, but may also reduce the quality of the solution when applied in practice. Identifying and controlling this quality loss is the subject of the study reported in [159]. Considering the importance of covering models in emergency location problems, Emir-Farinas and Francis [159] develop aggregation methods supported with a priori error bounds, which facilitate the identification and control of errors in covering location models. A survey of different aggregation approaches has been reported in Francis et al. [160] where Francis et al. provide a framework for different aggregation approaches applied to various location models, including covering, p-median, and p-center models. This study also discusses the efficiency and inefficiency of different aggregation error measures via a comparative study.

EMS data is a classical example of a spatio-temporal dataset. In these kind of datasets, it is possible to define a path in its embedded spatio-temporal framework. Examples of such paths are the sequence of the calls served by an EMS vehicle or the sequence of emergency calls in a given area. The analysis of such paths could lead to the detection of interesting sub-paths in spatio-temporal data [161], where not only the spatial features are considered but also time. As a result, the most frequently used sub-paths could provide evidence for a lack of coverage and, as a consequence, the need to restructure the EMS vehicle deployment. Further, regional co-location patterns (representing collections of feature types frequently located together in certain localities or regions [162]) could constitute a characteristic rule that is only valid in the detected regions and not in the overall dataset. These patterns might form the basis of a classification rule that can be used to identify a class of interesting situations (such as for the prediction of emergency situations with a high demand of healthcare services). Finally, the combination of spatial analysis with data mining techniques can identify regions or districts with a high level of demand, as in [163]. This can be determined by analyzing the districts with respect to their geographical neighborhood and the uniformity of the geographical features distribution in the space. Based on this spatial model, data mining might allow the development of a prediction system for emergency demand, i.e., to identify the most likely region from where the next emergency request could arrive.

> We remark that, in most EMS relocation and dispatching models, demand is considered as discrete points in space. However, it is more realistic to consider the problem space as a network in which the emergency requests arrive both on the links and the nodes [78]. Therefore, an extension of demand forecasting models dealing with this issue will be welcomed. Although some researchers have addressed the demand aggregation problem by bounding it or removing some of the factors causing aggregation errors, this problem is worth to be studied in more detail (see Goldberg [3]).
> Combining spatial analysis and data mining in EMS modeling could help to identify the most likely region from where the next emergency request could arrive. Such a forecasting tool could improve real-time EMS fleet management. To the best of our knowledge, there is no literature on the application of such techniques to EMS systems.

7.2 Forecasting travel times and workloads

The first model aimed at forecasting travel times was developed by Kolesar et al. [164]. They propose a model for fire engines, which was validated by a field experiment. Their main finding is that regional traffic conditions and hourly variations have only minor effects on average travel times.

Alanis et al. [56] estimate the dependence of travel times on distance by using data of high-priority calls in Calgary, Alberta. The EMS system is represented as a two-dimensional Markov chain model and EMS vehicle redeployment is included by using a compliance table policy as discussed in Section 3. First, Alanis et al. prove that the mean fire engine travel times reported in [164] are a valid and useful description of median ambulance travel times through a non-parametric estimate of the median and coefficient of variation. Then, they propose a new specification for the coefficient of variation, which decreases with distance. The validated model is capable to provide accurate estimates of the travel time distribution, the distribution of the number of busy ambulances, and other system performance measures. The results of the model largely depends on the compliance table used, and therefore, the authors emphasize the importance of a good redeployment policy.

Westgate et al. [165] propose an innovative approach to evaluate travel times based on a Bayesian model that uses global positioning system (GPS) data from the Toronto EMS system. The paths, travel times, and parameters of each road segment travel time distribution are estimated simultaneously by using Bayesian data augmentation. They also consider two simpler estimation methods based on GPS speed data. The proposed approaches outperform estimates from alternative methods and their robustness with respect to GPS location errors. Finally, as in [166], the authors construct probability-of-coverage maps for ambulances.

Workload is concerned with how long an ambulance and its crew are occupied with a call. As discussed in [8], workload is a service time that could largely influence the performance of an EMS system. For instance, ambulances with non-urgent patients can wait longer than average when an ED is overcrowded which has an impact on the EMS system as discussed in [56].

> To the best of our knowledge, workload forecasting has not yet been discussed in the EMS literature. This research could take inspiration from similar work in the hospital setting as the one discussed by Kc and Terwiesch [167].

8 Workforce Management

Workforce management is closely related to personnel scheduling and rostering (see, e.g., Ernst et al. [168, 169]). In healthcare, the nurse scheduling problem is probably the most studied problem in the field of workforce management (see review of Cheang et al. [170]). Even though many models have been published in this field, the application of these models in practice is limited as argued by [171]. However, there are only a few papers that consider workforce management in the field of EMS systems.

Bradbeer et al. [172] use an evolutionary algorithm approach to determine an acceptable roster for the EMS vehicle crew duties while assuming that the number and locations of the EMS vehicles are given. In contrast, other authors deal with both the crew rostering problem and EMS vehicle location problem. Erdogan et al. [173] apply a neighbourhood search to locate EMS vehicles and use the solution found as input for two integer programming models to solve the crew rostering problem. Vile et al. [174] propose interrelated advanced statistical and operational research methods to recommend minimum staffing requirements and generate low-cost rosters. Rajagopalan et al. [175] also present a two-stage approach for crew rostering and EMS vehicle location planning. In the first stage, they solve a dynamic expected coverage model using tabu search while an integer programming model is presented in the second stage to solve the crew rostering problem. Finally, Li and Kozan [176] propose two-stage models that use non-linear integer programming techniques. In the first stage, shift start times and the number of staff required to work during each shift are determined. The results of the first stage are used as input for the second stage, which determines a balanced scheduled for the EMS vehicle crew.

In this section, we consider a new approach for workforce management in healthcare and its possible application to the management of personnel employed along the ECP by using demand forecast.

A new way of thinking and of organizing health care delivery is to focus on the patient instead of on facilities. A patient-centered approach to health care means to deliver a service which is "*closely congruent with and responsive to patients' wants, needs, and preferences*" [177]. In May 2004, the International Alliance of Patients' Organizations (IAPO) conducted a consultation with its member patients' organizations in order to investigate which health care policy issues were most important to them. The final report showed that 74% of the respondents indicated that "*defining patient-centered healthcare was very relevant to their organization*" [178]. Along the lines of patient-centered health care delivery, approaches for medical workforce management that guarantee efficiency and fairness of the delivered services becomes crucial as pointed out by Aringhieri [179].

Usually, health care is delivered by a team of individuals who work together and who share knowledge, experiences and skills. As reported in [180], higher patient volumes lead to superior outcomes for hospitals, teams of physicians and single physicians. Similar insights are discussed by Grantcharov et al. [181] who reports the learning rate for laparoscopic skills when trained on a virtual reality system. McIntosh and Sheppy [182] argue that skill maximization (e.g., increasing the responsibilities of healthcare practitioners) is the key to increase productivity and quality of care. They also argue that an improvement in the output (number of cases treated) and quality of care is not just necessary, but essential. Because of this, McIntosh and Sheppy conclude that maximizing the use of human resources is key in the future of health care.

In order to increase the responsibilities of healthcare practitioners, we need to measure the efficiency of individuals and teams with respect to the service demand. Various metrics can be used to measure the efficiency of one individual (such as the number of patients visited per hour, the volume of surgical patients per year, the probability of making an incorrect decision, etc.). These measures can be used to evaluate the overall team performance with respect to the service demand forecast. Actually, a good demand prevision plays a fundamental role when health care managers want to guarantee the efficiency and fairness of the provided service. Demand forecast is crucial in the work of Feyter [183, 184], where the author focuses on the long-term supply of employees in a company based on forecasted needs of recruitments, lay-offs and retraining of the current workforce.

In [185], Addis et al. deal with managing the personnel working at the operation center of the EMS system of Milano by taking demand forecast into account. The staff members work together in teams according to predefined shift patterns. The considered problem consists in assigning individuals to teams and teams to shifts while providing the best service to citizens by guaranteeing the needed number of operators with respect to the demand forecast. The authors provide mixed integer linear programming models solved by a general purpose solver.

Outside the field of health care, there are a few papers that consider, in some way, demand forecast in their planning. Billionnet [186] discusses scheduling a hierarchical workforce with variable demands under the restriction that a higher qualified worker can substitute a lower qualified one, but not vice versa. An ILP model has been proposed which is solved using a general purpose solver. In [187, 188], Atlason et al. consider the problem of minimizing staffing costs in an inbound call center while maintaining an acceptable level of service in multiple time periods. The staffing level in one time period can affect the service levels in subsequent periods. The authors present an iterative cutting plane method for minimizing staffing costs in a service system which must meet service level requirements over multiple time periods. Furthermore, it is assumed that the service level cannot be computed a-priori, and thus, it is evaluated using simulation. More recently, Gurvich et al. [189] address the problem of staffing call centers with multiple customer classes, agent types operating under probabilistic quality-of-service constraints, and uncertainty in demand rate. The proposed formulation of the problem is solved by a two-step solution method. In the first step, a random static planning problem is solved which provides an approximation of the optimal staffing levels and a staffing frontier. In the second step, a finite number of staffing problems is solved with arrival rates provided by the staffing frontier of the first step. The output of the procedure is a solution that is feasible with respect to the chance constraints. For large call centers, the solution method provides near-optimal solutions.

> Including demand forecast in workforce management poses several challenges. In many real-life settings, teams have to be composed in a fair manner (see [185]) to guarantee a (quasi) constant quality of health service. This opens the debate (already introduced in Section 2.1 for the location problems) on how fairness should be incorporated when assigning shifts to medical crews. From a methodological point of view, the problem of composing teams has an intrinsic stochastic nature that should be addressed by adopting unconventional approaches as discussed in Aringhieri et al. [190].

9 Big EMS, big data, big challenge

In the previous sections, many challenges were highlighted such as those arising when dealing with the problem of incorporating equity and uncertainty aspects, the need for reliable forecasts and new methodological hybridizations.

In our opinion, the biggest challenge is to adopt a holistic outcome-based approach for the ECP, which should be conceived as a methodology that details all decisions, treatments, and reports related to a patient. From this point of view, one of the main difficulties is the collection of information regarding all events involving the patient before, during and after the EMS intervention. Usually, EMS systems collect a large amount of data, but this does not include data on what happens before and after the involvement of an EMS vehicle. Therefore, the main challenge is to develop new reliable models (probably hybrids and supported by new ICT solutions) that are capable to represent the inherent complexity behind the definition of an ECP.

Focusing on the ECP means to shift the attention from the EMS system to the whole Emergency Care Delivery System (ECDS) in order to enhance the quality of care, which will improve patient outcomes, promote patient safety, increase patient satisfaction, and optimize the use of resources. In this respect, the interplay between the EMS system and other stakeholders of the ECDS plays a crucial role. Due to budget cuts, this challenge is faced by reorganizing the EMS system to better exploit the availability of other resources or by taking advantage of economies of scale. In the first case, especially in North America, the idea is to involve firemen services, which usually have low utilization rates: after a training, it will be possible to use firemen to provide ELS until a BLS provider arrives at the scene. In the second case, especially in Europe, the idea is to merge several EMS systems into one organization. Apart from economies of scale, another benefit of this is that it solves the problem of dealing with emergency requests arising from contiguous EMS systems.

One of the main consequences of taking the whole ECDS into account could be the need to analyze large amounts of data from different data sources. Smart cities are more and more connected, which provides interesting information such as real-time traffic conditions. Another example is related to the ED workload: to deal with ED overcrowding, ICT solutions are used that provide real-time information on the ED workload. An experimental project on this subject is ongoing in the Piedmont region in Italy. Therefore, the *big challenge* is to connect the obtained data with the EMS systems to deliver more effective and efficient emergency care.

As an example, consider the development of a new dispatching decision support system. Such a system should at least deal with the following decisions. The first decision is to dispatch an EMS vehicle that can reach the patient in time and that reduces the overall coverage of the region the least. To support this decision, the spatial analysis discussed in Section 7 could be fruitfully used. The quality of this analysis is expected to be high, because accurate data is available. The second decision is to select the most appropriate ED to transport the patient to. For instance, in case of a non-urgent request, the dispatcher can select the ED with the lowest workload among the nearest EDs to evenly distribute the workload. The possibility of selecting non-overcrowded EDs may be crucial in the case of urgent requests. Finally, the third decision encompasses the positioning of the EMS vehicle after serving an emergency request. Again, a good forecast, in particular in combination with a spatial analysis, could be crucial in indicating the least covered area given the expected emergency demand. To some extent, this decision is connected with the second decision, i.e., the dispatcher can select a specific ED to cover the area around this ED after the patient is transported to the hospital. It is evident that the exploitation of real-time traffic information is crucial when determining the set of the "nearest" EMS vehicles or EDs.

Glossary. In this paper, we have used several acronyms. To help the reader, we list them here, in alphabetical order: Advanced Life Support (ALS), Agent-Based Simulation (ABS), Approximate Dynamic Programming (ADP), Automated External Defibrillator (AED), and Basic Life Support (BLS), Clinical Pathway (CP), Discrete Event Simulation (DES), Double Standard Model (DSM), Dynamic DSM at time t (DDSM^t), Emergency Care Delivery System (ECDS), Emergency Care Pathway (ECP), Emergency Department (ED), Emergency Life Support (ELS), Emergency Medical Service (EMS), European Emergency Data Project (EED), Fuzzy Goal Programming (FGP), Geographical Information System (GIS), Global Positioning System (GPS), Health Technology Assessment (HTA), Hypercube Queueing Model (HQM), Low-Priority Calls Coverage (LPCC), Markov Decision Process (MDP), Maximal Covering Location Problem (MCLP), Maximal Covering Location Problem including Probabilistic Response times (MCLP+PR) Maximal Expected Coverage Relocation Problem (MECRP), Maximum Availability Location Problem (MALP), Maximum Expected Covering Location Problem (MEXCLP), Maximum Expected Covering Location Problem including Probabilistic Response times(MEXCLP+PR), Maximum Expected Covering Location Problem Probabilistic Response times and Station Specific Busy Probabilities (MEXCLP+PR+SSBP), Mixed Integer Program (MIP), National Health Service (NHS), Queueing Maximum Availability Location Problem (Q-MALP), Queueing based Probabilistic Location Set Covering Problem (Q-PLSCP), Response Time (RT), Response Time Threshold (RTT).

Acknowledgements. The authors would like to thank the numerous colleagues who suggested useful references. In particular, we would like to thank Rosa Meo to help us in identifying the link between data mining techniques and EMS forecasts, Vito "Franco" Fragnelli for his useful suggestions in the field of game theory, Derya Demirtas for introducing the authors to the AED location problem and Angela Testi and Elena Tànfani for introducing the authors to the Health Technology Assessment. Further, the authors would like to thank the anonymous reviewers for their really accurate work and for providing such useful suggestions, which really improved the quality of this review.

References

[1] C. ReVelle, D. Bigman, D. Schilling, J. Cohon, and R. Church. Facility location: A review of context-free and EMS models. *Health Services Research*, 12(2):129–146, 1977.

[2] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147(3):451–463, 2003.

[3] J.B. Goldberg. Operations research models for the deployment of emergency services vehicles. *EMS Management Journal*, 1(1):20–39, 2004.

[4] X. Li, Z. Zhao, X. Zhu, and T. Wyatt. Covering models and optimization techniques for emergency response facility location and planning: A review. *Mathematical Methods of Operations Research*, 74(3):281–310, 2011.

[5] A. Basar, B. Catay, and T. Unluyurt. A taxonomy for emergency service station location problem. *Optimization Letters*, 6(6):1147–1160, 2012.

[6] L.V. Green and P.J. Kolesar. Improving emergency responsiveness with management science. *Management Science*, 50(8):1001–1014, 2004.

[7] S.G. Henderson. *Wiley Encyclopedia of Operations Research and Management Science*, journal article Operations Research Tools for Addressing Current Challenges in Emergency Medical Services. Wiley, New York, 2011.

[8] A. Ingolfsson. EMS planning and management. In *Operations Research and Health Care Policy*, volume 190 of *International Series in Operations Research & Management Science*, pages 105–128. Springer, New York, 2013.

[9] P. Sorensen and R. Church. Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences*, 44(1):8–18, 2010.

[10] H.H.R. Campbell, N. Bradshaw, and M. Porteous. Integrated care pathways. *British Medical Journal*, 316(133-144), 1998.

[11] J.A. Mesa, J. Puerto, and A. Tamir. Improved algorithms for several network location problems with equality measures. *Discrete Applied Mathematics*, 130(3):437–448, 2003.

[12] L.A. McLay and M.E. Mayorga. Evaluating emergency medical service performance measures. *Health Care Management Science*, 13(2):124–136, 2010.

[13] M.P. Larsen, M.S. Eisenberg, R.O. Cummins, and A.P. Hallstrom. Predicting survival from out-of-hospital cardiac arrest: A graphic model. *Annals of Emergency Medicine*, 22(11):1652–1658, 1993.

[14] E. Erkut, A. Ingolfsson, and G. Erdogan. Ambulance location for maximum survival. *Naval Research Logistics*, 55(1):42–58, 2008.

[15] H.K. Smith, P.R. Harper, and C.N. Potts. Bicriteria efficiency/equity hierarchical location models for public service application. *Journal of the Operational Research Society*, 64(4):500–512, 2013.

[16] S. Khodaparasti, H.R. Maleki, M.E. Bruni, S. Jahedi, P. Beraldi, and D. Conforti. Balancing efficiency and equity in location-allocation models with an application to strategic ems design. *Optimization Letters*, 2015. Published on line: 01 August 2015.

[17] S. Chanta, M.E. Mayorga, and L.A. McLay. Improving emergency service in rural areas: A bi-objective covering location model for EMS systems. *Annals of Operations Research*, 221(1):133–159, 2014.

[18] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, 1999.

[19] S. Chanta, M.E. Mayorga, M.E. Kurz, and L.A. McLay. The minimum p-envy location problem: A new model for equitable distribution of emergency resources. *IIE Transactions on Healthcare Systems Engineering*, 1(2):101–115, 2011.

[20] T. Drezner, Z. Drezner, and J. Guyse. Equitable service by a facility: Minimizing the Gini coefficient. *Computers & Operations Research*, 36(12):3240–3246, 2009.

[21] I. Espejo, A. Marin, J. Puerto, and A.M. Rodriguez-Chia. A comparison of formulations and solution methods for the minimum-envy location problem. *Computers & Operations Research*, 36(6):1966–1981, 2009.

[22] M.S. Daskin. *Network and Discrete Location: Models, Algorithms, and Applications*. Wiley, New York, 1995.

[23] S. Chanta, M.E. Mayorga, and L.A. McLay. The minimum p-envy location problem with requirement on minimum survival rate. *Computers & Industrial Engineering*, 74:228–239, 2014.

[24] S.A. Broverman. *Actex study manual, Course 1, Examination of the Society of Actuaries, Exam 1 of the Casualty Actuarial Society*. Actex Publications, Winsted, 2001.

[25] H. Toro-Diaz, M.E. Mayorga, L.A. McLay, H.K. Rajagopalan, and C. Saydam. Reducing disparities in large-scale emergency medical service systems. *Journal of the Operational Research Society*, 66(7):1–13, 2014.

[26] S. Budge, A. Ingolfsson, and E. Erkut. Technical note—approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location. *Operations Research*, 57(1): 251–255, 2009.

[27] V. Fragnelli and S. Gagliardo. Open problems in cooperative location games. *International Game Theory Review*, 15(3):1–13, 2013.

[28] L.S. Shapley and M Shubik. On market games. *Journal of Economic Theory*, 1:9–25, 1969.

[29] R.C. Larson. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, 1(1):67–95, 1974.

[30] E. Erkut, A. Ingolfsson, T. Sim, and G. Erdogan. Computational comparison of five maximal covering models for locating ambulances. *Geographical Analysis*, 41(1):43–65, 2009.

[31] R.C. Larson. Approximating the performance of urban emergency service systems. *Operations Research*, 23(5): 845–868, 1975.

[32] T.H. Burwell, J.P. Jarvis, and M.A. McKnew. Modeling co-located servers and dispatch ties in the hypercube model. *Computers & Operations Research*, 20(2):113–119, 1993.

[33] J.P. Jarvis. Approximating the equilibrium behavior of multi-server loss systems. *Management Science*, 31(2): 235–239, 1985.

[34] R. Batta, J.M. Dolan, and N.N. Krishnamurthy. The maximal expected covering location problem: Revisited. *Transportation Science*, 23(4):277–287, 1989.

[35] R.A. Takeda, J.A. Widmer, and R. Morabito. Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. *Computers & Operations Research*, 34(3):727–741, 2007.

[36] A.P. Iannoni, R. Morabito, and C. Saydam. An optimization approach for ambulance location and the districting of the response segments on highways. *European Journal of Operational Research*, 195(2):528–542, 2009.

[37] S.H. Cho, H. Jang, T. Lee, and J. Turner. Simultaneous location of trauma centers and helicopters for emergency medical service planning. *Operations Research*, 62(4):751–771, 2014.

[38] A. Shariat-Mohaymany, M. Babaei, S. Moadi, and S.M. Amiripour. Linear upper-bound unavailability set covering models for locating ambulances: Application to Tehran rural roads. *European Journal of Operational Research*, 221(1):263–272, 2012.

[39] C. ReVelle and K. Hogan. The maximum availability location problem. *Transportation Science*, 23(3):192–200, 1989.

[40] M.S. Daskin. A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science*, 17(1):48–70, 1983.

[41] V. Marianov and C. ReVelle. The queueing maximal availability location problem: A model for the siting of emergency vehicles. *European Journal of Operational Research*, 93(1):110–120, 1996.

[42] A. Ingolfsson, S. Budge, and E. Erkut. Optimal ambulance location with random delays and travel times. *Health Care Management Science*, 11(3):262–274, 2008.

[43] R. Church and C. ReVelle. The maximal covering location problem. *Papers of the Regional Science Association*, 32(1):101–118, 1974.

[44] Michael O. Ball and Feng L. Lin. Reliability model applied to emergency service vehicle location. *Operations Research*, 41(1):18–36, 1993.

[45] P. Beraldi, M.E. Bruni, and D. Conforti. Designing robust emergency medical service via stochastic programming. *European Journal of Operational Research*, 158(1):183–193, 2004.

[46] Z.-H. Zhang and K. Li. A novel probabilistic formulation for locating and sizing emergency medical service stations. *Annals of Operations Research*, 229(1):813–825, 2014.

[47] P. Beraldi and M.E. Bruni. A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research*, 196(1):323–331, 2009.

[48] N. Noyan. Alternate risk measures for emergency medical service system design. *Annals of Operations Research*, 181(1):559–589, 2010.

[49] Z.-H. Zhang and H. Jiang. A robust counterpart approach to the bi-objective emergency medical service design problem. *Applied Mathematical Modelling*, 38(3):1033–1040, 2014.

[50] C. Araz, H. Selim, and I. Ozkarahan. A fuzzy multi-objective covering-based vehicle location model for emergency services. *Computers & Operations Research*, 34(3):705–726, 2007.

[51] T. Uno, K. Kato, and H. Katagiri. An application of interactive fuzzy satisficing approach with particle swarm optimization for multiobjective emergency facility location problem with a-distance. In *Computational Intelligence in Multicriteria Decision Making, IEEE Symposium on*, pages 368–373, 2007.

[52] M.L. Koc and M. Bostancioglu. A reliability based solution to an ambulance location problem using fuzzy set theory. *International Journal of Natural and Engineering Sciences*, 5(1):13–17, 2011.

[53] M. Gendreau, G. Laporte, and F. Semet. Solving an ambulance location model by tabu search. *Location Science*, 5(2):75–88, 1997.

[54] M.S. Maxwell, M. Restrepo, S.G. Henderson, and H. Topaloglu. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22(2):266–281, 2010.

[55] D. Bandara, M.E. Mayorga, and L.A. McLay. Priority dispatching strategies for EMS systems. *Journal of the Operational Research Society*, 65(4):572–587, 2014.

[56] R. Alanis, A. Ingolfsson, and B. Kolfal. A Markov chain model for an EMS system with repositioning. *Production and Operations Management*, 22(1):216–231, 2013.

[57] K. Sudtachata, M.E. Mayorga, and L.A. Mclay. A nested-compliance table policy for emergency medical service systems under relocation. *Omega*, 58:154–168, 2016.

[58] M. Gendreau, G. Laporte, and F. Semet. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, 27(12):1641–1653, 2001.

[59] M. Moeini, Z. Jemai, and E. Sahin. Location and relocation problems in the context of the emergency medical service systems: A case study. *Central European Journal of Operations Research*, 23(3):641–658, 2014.

[60] A. Basar, B. Catay, and T. Unluyurt. A multi-period double coverage approach for locating the emergency medical service stations in Istanbul. *Journal of the Operational Research Society*, 62(4):627–637, 2011.

[61] V. Schmid and K.F. Doerner. Ambulance location and relocation problems with time-dependent travel times. *European Journal of Operational Research*, 207(3):1293–1303, 2010.

[62] M. Gendreau, G. Laporte, and F. Semet. The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society*, 57(1):22–28, 2006.

[63] C.J. Jagtenberg, S. Bhulai, and R.D. van der Mei. An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care*, 4:27–35, 2015.

[64] P.L. van den Berg and K. Aardal. Time-dependent MEXCLP with start-up and relocation cost. *European Journal of Operational Research*, 242(2):383–389, 2015.

[65] J. Naoum-Sawaya and S. Elhedhli. A stochastic optimization model for real-time ambulance redeployment. *Computers & Operations Research*, 40(8):1972–1978, 2013.

[66] H.K. Rajagopalan, C. Saydam, and J. Xiao. A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research*, 35(3):814–826, 2008.

[67] V. Marianov and C. ReVelle. The queueing probabilistic location set covering problem and some extensions. *Socio-Economic Planning Sciences*, 28(3):167–178, 1994.

[68] D. Degel, L. Wiesche, S. Rachuba, and B. Werners. Time-dependent ambulance allocation considering data-driven empirically required coverage. *Health Care Management Science*, 18(4):444–458, 2015.

[69] V. Schmid. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research*, 219(3):611–621, 2012.

[70] V. Bélanger, Y. Kergosien, A. Ruiz, and P. Soriano. An empirical comparison of relocation strategies in real-time ambulance fleet management. *Computers and Industrial Engineering*, 94:216–229, 2016.

[71] S. Lee. The role of centrality in ambulance dispatching. *Decision Support Systems*, 54(1):282–291, 2012.

[72] G.M. Carter, J.M. Chaiken, and E. Ignall. Response areas for two emergency units. *Operations Research*, 20(3): 571–594, 1972.

[73] T. Andersson and P. Varbrand. Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society*, 58(2):195–201, 2006.

[74] S. Lee. The role of preparedness in ambulance dispatching. *Journal of the Operational Research Society*, 62(10): 1888–1897, 2011.

[75] L.A. McLay and M.E. Mayorga. A model for optimally dispatching ambulances to emergency calls with classification errors in patient priorities. *IIE Transactions*, 45(1):1–24, 2012.

[76] L.A. McLay and M.E. Mayorga. A dispatching model for server-to-customer systems that balances efficiency and equity. *Manufacturing & Service Operations Management*, 15(2):205–220, 2013.

[77] S. Kanchala, M.E. Mayorga, and L.A. McLay. Recommendations for dispatching emergency vehicles under multitiered response via simulation. *International Transactions in Operational Research*, 21(4):581–617, 2014.

[78] A. Haghani, Q. Tian, and H. Hu. Simulation model for real-time emergency vehicle dispatching and routing. *Transportation Research Record: Journal of the Transportation Research Board*, 1882(1):176–183, 2004.

[79] S.G. Henderson and A.J. Mason. *Operations Research and Health Care*, volume 70 of *International Series in Operations Research & Management Science*, book section Ambulance Service Planning: Simulation and Data Visualisation, pages 77–102. Springer, New York, 2004.

[80] S. Lee. Centrality-based ambulance dispatching for demanding emergency situations. *Journal of the Operational Research Society*, 64(4):611–618, 2013.

[81] M. Zarkeshzadeh, H. Zare, Z. Heshmati, and M. Teimouri. A novel hybrid method for improving ambulance dispatching response time through a simulation study. *Simulation Modelling Practice and Theory*, 60:170–184, 2016.

[82] S. Lee. Role of parallelism in ambulance dispatching. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(8):1113–1122, 2014.

[83] C.S. Lim, R. Mamat, and T. Braunl. Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):624–632, 2011.

[84] S. Yang, M. Hamedi, and A. Haghani. Online dispatching and routing model for emergency vehicles with area coverage constraints. *Transportation Research Record: Journal of the Transportation Research Board*, 1923(1):1–8, 2005.

[85] A. Oran, K.C. Tan, B.H. Ooi, M. Sim, and P. Jaillet. Location and routing models for emergency response plans with priorities. *Communications in Computer and Information Science*, 318:129–140, 2012.

[86] S. Panahi and M.R. Delavar. Dynamic shortest path in ambulance routing based on GIS. *International Journal of Geoinformatics*, 5(1):13–19, 2009.

[87] A. Jotshi, Q. Gong, and R. Batta. Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences*, 43(1):1–24, 2009.

[88] A. Beaudry, G. Laporte, T. Melo, and S. Nickel. Dynamic transportation of patients in hospitals. *OR Spectrum*, 32(1):77–107, 2010.

[89] L.H. Ardekani, D. Haight, A. Ingolfsson, M. Salama, and M. Stanton. Scheduling and routing ambulances that provide inter-facility patient transfers. Submitted, 2014.

[90] L. Talarico, F. Meisel, and K. Sorensen. Ambulance routing for disaster response with patient groups. *Computers & Operations Research*, 56:120–133, 2015.

[91] N.R. Hoot and D. Aronsky. Systematic review of emergency department crowding: Causes, effects, and solutions. *Annals of Emergency Medicine*, 52(2):126–136, 2008.

[92] J.C. Pham, R. Patel, M.G. Millin, T.D. Kirsch, and A. Chanmugam. The effects of ambulance diversion: A comprehensive review. *Academic Emergency Medicine*, 13(11):1220–1227, 2006.

[93] P. Sprivulis and B. Gerrard. Internet-accessible emergency department workload information reduces ambulance diversion. *Prehospital Emergency Care*, 9(3):285–291, 2005.

[94] M.K. Delgado, L.J. Meng, M.P. Mercer, J.M. Pines, D.K. Owens, and G.S. Zaric. Reducing ambulance diversion at hospital and regional levels: Systemic review of insights from simulation models. *Western Journal of Emergency Medicine*, 14(5):489–498, 2013.

[95] A.R. Nafarrate, J.W. Fowler, and T. Wu. Bi-criteria analysis of ambulance diversion policies. In *Proceedings of the 2010 Winter Simulation Conference*, pages 2315–2326, Baltimore, 2010.

[96] A. Ramirez, J.W. Fowler, and T. Wu. Analysis of ambulance diversion policies for a large-size hospital. In *Proceedings of the 2009 Winter Simulation Conference*, pages 1875–1886, Austin, 2009.

[97] A.B. Storrow, C. Zhou, G. Gaddis, J.H. Han, K. Miller, D. Klubert, A. Laidig, and D. Aronsky. Decreasing lab turnaround time improves emergency department throughput and decreases emergency medical services diversion: A simulation model. *Academic Emergency Medicine*, 15(11):1130–1135, 2008.

[98] A. Kolker. Process modeling of emergency department patient flow: Effect of patient length of stay on ED diversion. *Journal of Medical Systems*, 32(5):389–401, 2008.

[99] R. Hagtvedt, M. Ferguson, P. Griffin, G.T. Jones, and P. Keskinocak. Cooperative strategies to reduce ambulance diversion. In *Proceedings of the 2009 Winter Simulation Conference*, pages 1861–1874, Austin, 2009.

[100] A. Ramirez-Nafarrate, J.W. Fowler, and T. Wu. Design of centralized ambulance diversion policies using simulation-optimization. In *Proceedings of the 2011 Winter Simulation Conference*, pages 1251–1262, Phoenix, 2011.

[101] S. Deo and I. Gurvich. Centralized vs. decentralized ambulance diversion: A network perspective. *Management Science*, 57(7):1300–1319, 2011.

[102] A. Ramirez-Nafarrate, A.B. Hafizoglu, E.S. Gel, and J.W. Fowler. Comparison of ambulance diversion policies via simulation. In *Proceedings of the 2012 Winter Simulation Conference*, pages 1–12, Berlin, 2012.

[103] E. Almehdawe, B. Jewkes, and Q.-M. He. A Markovian queueing model for ambulance offload delays. *European Journal of Operational Research*, 226(3):602–614, 2013.

[104] A. Ramirez-Nafarrate, A.B. Hafizoglu, E.S. Gel, and J.W. Fowler. Optimal control policies for ambulance diversion. *European Journal of Operational Research*, 236(1):298–312, 2014.

[105] A.S. Go, D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, W.B. Borden, D.M. Bravata, S. Dai, E.S. Ford, C.S. Fox, S. Franco, H.J. Fullerton, C. Gillespie, S.M. Hailpern, J.A. Heit, V.J. Howard, M.D. Huffman, B.M. Kissela, S.J. Kittner, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, D. Magid, G.M. Marcus, A. Marelli, D.B. Matchar, D.K. McGuire, E.R. Mohler, C.S. Moy, M.E. Mussolino, G. Nichol, N.P. Paynter, P.J. Schreiner, P.D. Sorlie, J. Stein, T.N. Turan, S.S. Virani, N.D. Wong, D. Woo, and M.B. Turner. Heart disease and stroke statistics-2013 update: A report from the American Heart Association. *Circulation*, 127(1):e6–e245, 2013.

[106] A. Di Carlo. Human and economic burden of stroke. *Age and Ageing*, 38(1):4–5, 2009.

[107] T. Krafft, L. García Castrillo-Riesgo, S. Edwards, M. Fischer, J. Overton, I. Robertson-Steel, and A. König. European Emergency Data Project (EED project): EMS data-based health surveillance system. *European journal of Public Health*, 13(3):85 – 90, 2003.

[108] M. Fischer, J. Kamp, L. García-Castrillo Riesgo, I. Robertson-Steel, J. Overton, A. Ziemann, and T. Krafft. Comparing emergency medical service systems – A project of the European Emergency Data (EED) Project. *Resuscitation*, 82(3):285 – 293, 2011.

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

[110] A.A. Siddiq, S.C Brooks, and T.C.Y. Chan. Modeling the impact of public access defibrillator range on public location cardiac arrest coverage. *Resuscitation*, 84(7):904–909, 2013.

[111] M.B. Mandell and L.R. Becker. A model for locating automatic external defibrillators. *Socio-Economic Planning Sciences*, 30(1):51–66, 1996.

[112] M.S. Rauner and N. Bajmoczy. How many AEDs in which region? An economic decision model for the Austrian red cross. *European Journal of Operational Research*, 150(1):3–18, 2003.

[113] D.C. Myers and M. Mohite. Locating automated external defibrillators in a university community. *Journal of the Operational Research Society*, 60(6):869–872, 2009.

[114] T.H.D. Dao, Y. Zhou, J.-C. Thill, and E. Delmelle. Spatio-temporal location modeling in a 3D indoor environment: The case of AEDs as emergency medical devices. *International Journal of Geographical Information Science*, 26 (3):469–494, 2012.

[115] T.C.Y. Chan, H. Li, G. Lebovic, S.K. Tang, J.Y.T. Chan, H.C.K. Cheng, L.J. Morrison, and S.C. Brooks. Identifying locations for public access defibrillators using mathematical optimization. *Circulation*, 127(17):1801–1809, 2013.

[116] T.C.Y. Chan, D. Demirtas, and R.H. Kwon. Optimizing the deployment of public access defibrillators. *Management Science*, 2016. Published Online: January 28, 2016.

[117] L.A. McLay and M.E. Mayorga. Evaluating the impact of performance goals on dispatching decisions in emergency medical service. *IIE Transactions on Healthcare Systems Engineering*, 1(3):185–196, 2011.

[118] V. A. Knight, P. R. Harper, and L. Smith. Ambulance allocation for maximal survival with heterogeneous outcome measures. *Omega*, 40(6):918–926, 2012.

[119] L. Aboueljinane, E. Sahin, and Z. Jemai. A review on simulation models applied to emergency medical service operations. *Computers & Industrial Engineering*, 66(4):734–750, 2013.

[120] J. Goldberg, R. Dietrich, J.M. Chen, M. Mitwasi, T. Valenzuela, and E. Criss. A simulation model for evaluating a set of emergency vehicle base locations. *Socio-economic Planning Sciences*, 24(2):125–141, 1990.

[121] J. Goldberg, R. Dietrich, J.M. Chen, M. Mitwasi, T. Valenzuela, and E. Criss. Validating and applying a model for locating emergency medical vehicles in Tucson, AZ. *European Journal of Operational Research*, 49(3):308–324, 1990.

[122] A. Zaki, H. Cheng, and B. Parker. A simulation model for the analysis and management of an emergency service system. *Socio-Economic Planning Sciences*, 31(3):173–189, 1997.

[123] A. Ingolfsson, E. Erkut, and S. Budge. Simulation of single start station for Edmonton EMS. *Journal of the Operational Research Society*, 54(7):736–746, 2003.

[124] C.-H. Wu and K.P. Hwang. Using a discrete-event simulation to balance ambulance availability and demand in static deployment systems. *Academic Emergency Medicine*, 16(12):1359–1366, 2009.

[125] L. Aboueljinane, E. Sahin, Z. Jemai, and J. Marty. 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, 2014.

[126] A. Bettinelli, R. Cordone, F. Ficarelli, and G. Righini. Simulation and optimization models for emergency medical systems planning. *Journal of Emergency Management*, 12(4):287–302, 2014.

[127] M. van Buuren, R. van der Mei, K. Aardal, and H. Post. Evaluating dynamic dispatch strategies for emergency medical services: TIFAR simulation tool. In *Proceedings of the 2012 Winter Simulation Conference*, Berlin, 2012.

[128] N. Gilbert. *Agent-Based Models*, volume 153 of *Quantitative Applications in the Social Sciences*. SAGE Publications, London, 2008.

[129] N. Gilbert and P. Terna. How to build and use agent-based models in social science. *Mind & Society*, 1(1):57–72, 2000.

[130] M. C. Macal and J. M. North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3): 151–162, 2010.

[131] R. Aringhieri. An integrated DE and AB simulation model for EMS management. In *Proceedings of the 2010 IEEE Workshop on Health Care Management*, pages 1–6, 2010.

[132] R. Aringhieri, G. Carello, and D. Morale. Supporting decision making to improve the performance of an Italian emergency medical service. *Annals of Operations Research*, 236:131–148, 2016.

[133] A. Anagnostou, A. Nouman, and S.J.E. Taylor. Distributed hybrid agent-based discrete event emergency medical services simulation. In *Proceedings of the 2013 Winter Simulation Conference*, pages 1625–1636, Washington, 2013.

[134] E. Blanc and I. Reychav. *Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care*, chapter Potential of GIS and spatial knowledge in health care and public safety, pages 734–749. IGI Global, Hershey, 2013.

[135] K. Peleg and J.S. Pliskin. A geographic information system simulation model of EMS: Reducing ambulance response time. *American Journal of Emergency Medicine*, 22(3):164–170, 2004.

[136] D.-C. Huang, H.-H. Liu, A.Y. Chen, and W.-Z. Sun. GIS assessment of emergency medical response infrastructure. In *Proceedings of the 2013 ASCE International Workshop on Computing in Civil Engineering*, pages 508–515, Los Angeles, 2013.

[137] A. Pinciroli, G. Righini, and M. Trubian. An interactive simulator of emergency management systems. In *Proceedings of the 2010 IEEE Workshop on Health Care Management*, pages 1–5, Venice, 2010.

[138] M.H. Azizan, C.S. Lim, W.A.L.W.M. Hatta, T.L. Go, and S.S. Teoh. Simulation of emergency medical services delivery performance based on real map. *International Journal of Engineering and Technology*, 5(3):2620–2627, 2013.

[139] L.R. Pinto, P.M.S. Silva, and T.P. Young. A generic method to develop simulation models for ambulance systems. *Simulation Modelling Practice and Theory*, 51:170–183, 2015.

[140] Y. Kergosien, V. Bélanger, P. Soriano, M. Gendreau, and A. Ruiz. A generic and flexible simulation-based analysis tool for ems management. *International Journal of Production Research*, 2015. Article in Press.

[141] M. C. Fu. Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing*, 14(3):192–215, 2002.

[142] M.C. Fu, F.W. Glover, and J. April. Simulation optimization: A review, new developments, and applications. In *Proceedings of the 2005 Winter Simulation Conference*, pages 83–95, Orlando, 2005.

[143] R. Aringhieri, G. Carello, and D. Morale. A simulation based tool for ambulances management evaluation. In X. XIe, F. Lorca, and É. Marcon, editors, *Operations Research for Health Care Delivery Engineering*, Selected Proceeding of the 33rd International Conference on Operational Research Applied to Health Service (ORAHS 2007), pages 379–392. Publications de l'Université de Saint-Étienne, 2008. ISBN 978-2-86272-507-9.

[144] M. S. Maxwell, S. G. Henderson, and H. Topaloglu. Ambulance redeployment: An approximate dynamic programming approach. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 1850–1860, Dec 2009.

[145] Lei Zhang. *Simulation optimisation and Markov models for dynamic ambulance redeployment.* PhD thesis, Engineering Science, University of Auckland, 2012. Available at http://hdl.handle.net/2292/20319.

[146] Andrew James Mason. *Handbook of Healthcare Operations Management: Methods and Applications*, chapter Simulation and Real-Time Optimised Relocation for Improving Ambulance Operations, pages 289–317. Springer New York, New York, NY, 2013. ISBN 978-1-4614-5885-2.

[147] L. Zhen, K. Wang, H. Hu, and D. Chang. A simulation optimization framework for ambulance deployment and relocation problems. *Computers & Industrial Engineering*, 72(1):12–23, 2014.

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

[149] W.K. Hall. Management science approaches to the determination of urban ambulance requirements. *Socio-Economic Planning Sciences*, 5(5):491–499, 1971.

[150] C.A. Aldrich, J.C. Hisserich, and L.B. Lave. An analysis of the demand for emergency ambulance service in an urban area. *American Journal of Public Health*, 61(6):1156–1169, 1971.

[151] K.F. Siler. Predicting demand for publicly dispatched ambulances in a metropolitan area. *Health Services Research*, 10(3):254–263, 1975.

[152] T.O. Kvalseth and J.M. Deems. Statistical models of the demand for emergency medical services in an urban area. *American Journal of Public Health*, 69(3):250–255, 1979.

[153] C.E. McConnel and R.W. Wilson. The demand for prehospital emergency services in an aging society. *Social Science and Medicine*, 46(8):1027–1031, 1998.

[154] J.R. Baker and K.E. Fitzpatrick. Determination of an optimal forecast model for ambulance demand using goal programming. *Journal of the Operational Research Society*, 37(11):1047–1059, 1986.

[155] N. Channouf, P. L'Ecuyer, A. Ingolfsson, and A.N. Avramidis. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Management Science*, 10(1):25–45, 2007.

[156] H. Setzler, C. Saydam, and S. Park. EMS call volume predictions: A comparative study. *Computers & Operations Research*, 36(6):1843–1851, 2009.

[157] J.L. Vile, J.W. Gillard, P.R. Harper, and V.A. Knight. Predicting ambulance demand using singular spectrum analysis. *Journal of the Operational Research Society*, 63(11):1556–1565, 2012.

[158] A. Micheletti, D. Morale, D. Rapati, and P. Nolli. A stochastic model for simulation and forecasting of emergencies in the area of Milano. In *Proceedings of the 2010 IEEE Workshop on Health Care Management*, pages 1–6, Venice, 2010.

[159] H. Emir-Farinas and R.L. Francis. Demand point aggregation for planar covering location models. *Annals of Operations Research*, 136(1):175–192, 2005.

[160] R.L. Francis, T.J. Lowe, M.B. Rayco, and A. Tamir. Aggregation error for location models: Survey and analysis. *Annals of Operations Research*, 167(1):171–208, 2009.

[161] X. Zhou, S. Shekhar, P. Mohan, S. Liess, and P.K. Snyder. Discovering interesting sub-paths in spatiotemporal datasets: A summary of results. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 44–53, New York, 2011.

[162] P. Mohan, S. Shekhar, J.A. Shine, J.P. Rogers, Z. Jiang, and N. Wayant. A neighborhood graph based approach to regional co-location pattern discovery: A summary of results. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 122–131, New York, 2011.

[163] S. Shang, B. Yuan, K. Deng, K. Xie, and X. Zhou. Finding the most accessible locations: Reverse path nearest neighbor query in road networks. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 181–190, New York, 2011.

[164] P. Kolesar, W. Walker, and J. Hausner. Determining the relation between fire engine travel times and travel distances in New York City. *Operations Research*, 23(4):614–627, 1975.

[165] B.S. Westgate, D.B. Woodard, D.S. Matteson, and S.G. Henderson. Travel time estimation for ambulances using Bayesian data augmentation. *Annals of Applied Statistics*, 7(2):1139–1161, 2013.

[166] S. Budge, A. Ingolfsson, and D. Zerom. Empirical analysis of ambulance travel times: The case of Calgary emergency medical services. *Management Science*, 56(4):716–723, 2010.

[167] D.S. Kc and C. Terwiesch. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, 55(9):1486–1498, 2009.

[168] A.T. Ernst, H. Jiang, M. Krishnamoorthy, B. Owens, and D. Sier. An annotated bibliography of personnel scheduling and rostering. *Annals of Operations Research*, 127:21–144, 2004.

[169] A.T. Ernst, H. Jiang, M. Krishnamoorthy, and D. Sier. Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, 153:3–27, 2004.

[170] B. Cheang, H. Li, A. Lim, and B. Rodrigues. Nurse rostering problems - a bibliographic survey. *European Journal of Operational Research*, 151:447–460, 2003.

[171] D. L. Kellogg and S. Walczak. Nurse scheduling: From academia to implementation or not? *Interfaces*, 37(4): 355–369, 2007.

[172] P.V.G. Bradbeer, C. Findlay, and T.C. Fogarty. An ambulance crew rostering system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1803: 267–279, 2000.

[173] G. Erdogan, E. Erkut, A. Ingolfsson, and G. Laporte. Scheduling ambulance crews for maximum coverage. *Journal of the Operational Research Society*, 61(4):543–550, 2010.

[174] J.L. Vile, J.W. Gillard, P.R. Harper, and V.A. Knight. Time-dependent stochastic methods for managing and scheduling emergency medical services. *Operations Research for Health Care*, 8:42–52, 2016.

[175] H. Rajagopalan, C. Saydam, H. Setzler, and E. Sharer. Ambulance deployment and shift scheduling: An integrated approach. *Journal of Service Science and Management*, 4(1):66–78, 2011.

[176] Yuan Li and Erhan Kozan. Rostering ambulance services. In *Industrial Engineering and Management Society*, pages 795–801, 2009.

[177] C. Laine and F. Davidoff. Patient-Centered Medicine. AProfessional Evolution. *Journal of the American Medical Association*, 275(2):152–156, 1996.

[178] International Alliance of Patients' Organizations. What is patient-centred healthcare? a review of definitions and principles, June 2004.

[179] R. Aringhieri. Composing medical crews with Equity and Efficiency. *Central European Journal of Operations Research*, 17(3):343–357, 2009.

[180] A. D. Auerbach. Localizing care to high-volume centers. In *Making Health Care Safer*, chapter 18, pages 205–212. Agency for Healthcare Research and Quality, 2001.

[181] Teodor P Grantcharov, Linda Bardram, Peter Funch-Jensen, and Jacob Rosenberg. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *The American Journal of Surgery*, 185(2):146 – 149, 2003. ISSN 0002-9610.

[182] B. McIntosh and B. Sheppy. Skill maximisation: The future of healthcare. *British Journal of Health Care Management*, 19(3):118–122, 2013.

[183] T. De Feyter. Modeling heterogeneity in manpower planning: dividing the personnel system into more homogeneous subgroups. *Applied stochastic models in business and industry*, 22:321–334, 2006.

[184] T. De Feyter. Modeling mixed push and pull promotion flows in Manpower Planning. *Annals of Operations Research*, 155:25–39, 2007.

[185] B. Addis, R. Aringhieri, G. Carello, A. Grosso, and F. Maffioli. Workforce Management based on forecasted demand. In E. Tanfani and A. Testi, editors, *Advanced Decision Making Methods applied to Health Care*, volume 173 of *International Series in Operations Research & Management Science*, pages 1–11. Springer, 2012. ISBN 978-88-470-2320-8.

[186] Alain Billionnet. Integer programming to schedule a hierarchical workforce with variable demands. *European Journal of Operational Research*, 114(1):105–114, 1999.

[187] Júlíus Atlason, Marina A. Epelman, and Shane G. Henderson. Call center staffing with simulation and cutting plane methods. *Annals of Operations Research*, 127:333–358, 2004. ISSN 0254-5330.

[188] Julius Atlason, Marina A. Epelman, and Shane G. Henderson. Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science*, 54(2):295–309, 2008.

[189] Itai Gurvich, James Luedtke, and Tolga Tezcan. Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science*, 56(7):1093–1115, 2010.

[190] R. Aringhieri, B. Addis, G. Carello, M. Gribaudo, and A. Grosso. Advanced workforce management in healthcare. In P. Harper, V. Knight, I. Vieira, and J. Williams, editors, *Operational Research Informing National Health Policy*, Selected Proceeding of the 37rd International Conference on Operational Research Applied to Health Service (ORAHS 2011), pages 308–319. School of Mathematics, Cardiff University, 2011. ISBN 978-0-9569158-0-1.