Operations Research for Health Care 11 (2016) 1–12

Contents lists available at ScienceDirect

Operations Research for Health Care

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

Coverage, survivability or response time: A comparative study of performance statistics used in ambulance location models via simulation–optimization

Muhammad Adeel Zaffar [a,*], Hari K. Rajagopalan [b], Cem Saydam [c], Maria Mayorga [d], Elizabeth Sharer [b]

[a] Lahore University of Management Sciences, Sector U, DHA Lahore 54792, Pakistan
[b] Francis Marion University, 4822 E Palmetto St, Florence, SC 29506, United States
[c] University of North Carolina-Charlotte, 9201 University City Blvd, Charlotte, NC 28262, United States
[d] North Carolina State University, 400 Daniels Hall, Raleigh, NC 27695, United States

ARTICLE INFO

Article history:
Received 23 April 2015
Accepted 18 August 2016
Available online 29 August 2016

Keywords:
Ambulance location
Covering models
Survivability
Simulation
Metaheuristics

ABSTRACT

Rapid response to medical emergencies is one of the main goals of Emergency Medical Service (EMS) systems. Ability to provide timely response is affected by fleet size and the locations of the ambulances. Literature on ambulance location has been dominated by models which either maximize coverage, or guarantee coverage within some threshold. Recent work has shifted the objective from maximizing coverage to improving patient survivability. In this paper we compare the performance of three recent ambulance location model objectives by applying a simulation–optimization framework. Our findings show that the maximum survivability objective performs better in both survivability and coverage metrics. Further, the results also support using the survivability objective for resource constrained ambulance operators.

1. Introduction

Rapid response to medical emergencies is one of the main goals of Emergency Medical Service (EMS) systems. Although, there is no global standardized response time (RT), in the US most EMS providers adopt the National Fire Protection Association's 1710 standard. [1], which is 8 min 59 s for 90% of life threatening calls. EMS providers routinely report the number of calls they reached within the response time thresholds (RTT) as a key performance statistic. Consequently, research of EMS models in the past has predominantly focused on improving performance against pre-specified RTT and "coverage" criteria [2–4].

There are two major drawbacks of the earlier models. First, they necessitate simplifying assumptions on fundamental issues, i.e., call coverage, relocation of ambulances, and busy probabilities in order to make the models mathematically tractable [5]. Second, coverage models are not sensitive to patient survivability outcomes [5–7]. For example, it is vital for a patient suffering from a cardiac arrest to receive care in the first four minutes [8,9]. However, coverage models consider a call to be covered as long as there is an ambulance available within the RTT, such that there is no distinction between a four-minute or a five-minute response time. Furthermore, coverage models do not differentiate between different locations within the same RTT. Recognizing the need to link patient outcomes to response times, there have been attempts recently to specifically incorporate survival functions into existing coverage models. Erkut et al. [6] were the first to develop such a model. Their work was extended by Knight et al. [7], who proposed incorporating multiple survival functions and developed the Maximal Survival Location model for heterogeneous patients. Bandara et al. [10] studied optimal dispatching policies to maximize patient survivability via a Markov decision process. McLay and Mayorga [11] also used a Markov decision process to make dispatching decisions; they reformulated the problem into a linear program and added equitability constraints, including survival probability. Bandara et al. [12] proposed a heuristic for dispatching ambulances to increase survival probability in real-world sized problems. Mayorga et al. [13] extended Bandara et al.'s work by

* Corresponding author.
E-mail addresses: adeel.zaffar@lums.edu.pk (M.A. Zaffar), hrajagopalan@fmarion.edu (H.K. Rajagopalan), saydam@uncc.edu (C. Saydam), memayorg@ncsu.edu (M. Mayorga), esharer@fmarion.edu (E. Sharer).

http://dx.doi.org/10.1016/j.orhc.2016.08.001
2211-6923/

incorporating integrated districting and dispatching policies, which theoretically increase patient survivability rates.

An important contribution of this paper is to incorporate a simulation–optimization approach for locating ambulances under a given objective. We are able to remove the majority of the assumptions employed by analytical approaches and develop a more realistic model that includes real-life operational practices; such as dispatching ambulances as soon as they leave an incident site, or as they are in transit to their assigned waiting station or location. We conduct a series of experiments in which a number of performance measures (e.g., *coverage*, *response time*, *patient survivability* and *busy probabilities* of the individual ambulances across several time periods) are compared using three different objective-optimization functions: *maximizing coverage*, *minimizing average response time* and *maximizing survivability*. Over 60,000 actual emergency call data received in a metropolitan area are used to test the objectives.

Test results reveal that under real life like conditions the Maximum Survivability objective is statistically better than the Minimum Average Response Time and Maximum Coverage objectives in terms of survivability, as well as coverage. This exciting result further highlights the importance of developing emergency response systems that incorporate patient survivability functions instead of using proxy measures such as expected number of calls covered within an RTT that indirectly estimate patient survivability. An in-depth analysis of our test results reveals several additional interesting insights. First, and somewhat surprisingly, the Maximum Survivability objective proved to be superior to the Minimum Average Response Time objective in terms of coverage. The difference is statistically significant, in spite of the fact that survivability is essentially a function of response time. Second, an interaction effect was found between performance indicators of the system and fleet size. For example, if the fleet size increases the difference between the Maximizing Patient Survivability and the Minimizing Average Response Time objectives in terms of coverage reduces. Intuitively, this implies that emergency response managers with smaller fleet sizes (i.e., fewer ambulances) should adopt patient survivability objective instead of average response times. Third, the Maximum Survivability objective outperforms other objectives with respect to the percentage of calls covered within 3 min, as well as 3–6 min-margins with no reduction in the total coverage. These numbers are encouraging in light of the criticality of time sensitive response requirements for certain emergencies. Finally, we also shed light on the issue of workload balance within the context of public resource management by analyzing individual busy probabilities of ambulances across the different optimization objectives.

The remainder of this paper is organized as follows. In the next section we provide a brief review of the relevant literature on ambulance location and coverage models, patient survivability and simulation-based models in the area of emergency deployment. In Section 3 we present our research methodology. Section 4 contains an in-depth discussion of our results. Finally, Section 5 concludes with a summary of our findings and potential directions for future research.

2. Literature review

The literature on ambulance location problems began with covering problems in the 1960 sand has received significant attention over time. Interested readers are referred to ReVelle et al. [14] and Farahani et al. [3] for comprehensive reviews of location models. In addition, Brotcorne et al. [2], Goldberg [4], and Li et al. [15] provide in-depth reviews of recent developments regarding ambulance location problems and optimization techniques applied in this area.

Although coverage models can provide valuable information regarding location decisions, the necessarily simplified and restrictive assumptions regarding various operational aspects of the EMS system can limit the usefulness of these types of models, particularly with respect to our objective of increasing patient survivability. For example, coverage models do not differentiate between ambulances as long as the ambulance is within some given threshold, either with respect to time or distance. Hence, these models fail to consider the proximity of an available ambulance to the demand point, which can easily result in the suboptimal deployment of ambulances in some cases.

Rajagopalan and Saydam [16] proposed the Minimum Expected Response Location Problem (MERLP) to address this particular concern. They used expected time, or distance weighted coverage, measures to ensure that the search algorithm did not treat all ambulances located within the coverage distance homogeneously. Similarly, Erkut et al. [6] demonstrated the drawbacks of using binary coverage metrics in coverage models. The authors developed a survivability function based on the incidence of cardiac arrest events, and incorporated this function into existing coverage models. The Maximal Survival Location Problem (MSLP), developed by Erkut et al. maximizes the expected number of patients who survive. The authors conducted extensive experiments with data from Edmonton, Canada. Their findings showed that maximizing the expected number of survivors can in fact result in ambulance location decisions that can potentially save more lives. McLay and Mayorga [5] simplified the survival function developed by Larsen et al. [17] to make the probability of survival only a function of response time. The authors compared a discrete optimization model based on RTT with another model based on maximizing the survival function. Knight et al. [7] developed the Maximal Expected Survival Location Model for Heterogeneous Patients (MESLMHP), which was a notable extension of Erkut et al.'s seminal work. The authors used a novel approach and made two important contributions: (1) MESLMHP incorporates survival functions for capturing multiple-classes of heterogeneous patients thus enabling a more realistic analysis for various outcome measures, and (2) by employing queuing theory, the authors extended the MESLMHP to model traffic congestion, thus eliminating the need to compute each ambulances utilization a priori. Further, the authors demonstrated the efficacy of their proposed models using data from Wales.

In the EMS location literature, simulation has been generally utilized to verify the quality of solutions [18]. Savas [19] used simulation in New York City to show that a substantial improvement in mean response time could be achieved by the dispersal of ambulance depots away from hospitals and closer to high demand areas. Swoveland et al. [20] utilize the output from a simulation to construct an analytical approximation, or proxy, for mean response times. The resulting combinatorial optimization problem is then solved using a probabilistic branch and bound procedure to determine ambulance locations in Vancouver, Canada. Fitzsimmons [21] developed a model to predict response times and to find the deployment of ambulances that minimize average response times. Their model uses a Monte Carlo simulation to estimate conditional mean response times when two or more ambulances are busy. Berlin and Liebman [22] combined ambulance stations with fire stations by using the Set Covering Location Problem [23] and then allocated ambulances based on the result of a simulation model whose focus was response times. Fujiwara et al. [24] used the maximum expected coverage location MEXCLP [25] model to screen a large number of possible alternatives to derive a collection of solutions. Each of these solutions was then evaluated using a simulation. Liu and Lee [26], extending Uyeno and Seeberg's [27] work, employed a simulation to analyze an emergency call system for a hospital in Taipei. Repede and Bernardo [28] utilized simulation to evaluate their TIMEXLCP model which was applied in Louisville,