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 prespecified 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, Kentucky. Zaki et al. [29] developed a simulation model to study, evaluate, and optimize the allocation of police patrol vehicles in the City of Richmond, Virginia. Goldberg et al. [30] did a similar study for ambulance location models. Borras and Pastor [18] also utilized a simulation to verify the precision of their ex-post evaluation method for the minimum local reliability levels of ambulance locations. Restrepo et al. [31], while proposing analytical models, recommend that simulation be employed to aid any final decision making. Maxwell et al. [32] and Alanis et al. [33] used simulation to find ambulance-to-base assignments. Mason [34] created a simulation–optimization algorithm to determine ambulance base locations in Copenhagen, Denmark. The author made a compelling argument for using simulation–optimization models for ambulance location and relocation problems. Recently, Kergosien et al. [35] developed a generic and flexible simulation model which explicitly considers EMS response to emergencies and patient transport requests [35]. A detailed review of simulation models applied to emergency medical service operations can be found in Aboueljinane et al. [36].

Our approach is based on prior research but extends it at different levels. First, unlike Mason [34], ambulances in our model are not tied to base stations. This allows for more realistic placement and deployment of ambulances. Second, similar to Kergosien et al.'s [35] simulation model we relax the assumption of ambulance dispatch made in Mayorga et al. [13] by allowing them to be dispatched enroute to their designated locations. Third, we perform in-depth analysis of recently developed models by comparing them on different measures: *coverage, response time, patient survivability* and *equitable workload distribution*. Using the simulation–optimization framework we compare and contrast the results of each of the three objectives (models). We demonstrate that focusing on patient survivability is superior to the other two objectives (i.e., Maximizing Coverage and Minimize Average Response Time) using various evaluation criteria.

3. Research methodology

The simulation approach allowed us to realistically capture the ambulance operations from the dispatch time to service completion time. We adapted the dispatch process presented by Mason [34] by relaxing the restriction of tying ambulances to their stations. Fig. 1 provides a flow diagram of the dispatch process employed in our study. When an emergency call is received by the dispatch center, the call is allocated to the closest available ambulance using the Manhattan distance. The RT is initialized with the average call taking plus chute times computed from the data. Next, the ambulance is dispatched to the scene. We assume that the ambulance travels approximately 36 mph, which is the average travel speed for ambulances from the empirical data. This assumption can be easily altered for future simulations, or even be replaced with actual travel times using the method developed by Kergosien et al. [35]. We compute the travel time via Manhattan distance formula between the call and ambulance location divided by the average speed and add it to determine the RT for the call. The real data shows that over 80% of the calls require transport to a hospital and the other require some on-scene time. Since our main goal is to compare the effectiveness of the three different objectives we opted for a simplified service completion time estimation. We use the weighted average of on-scene and travel and drop-off times, estimated from the data for each problem instance. The ambulance then becomes idle and available for next dispatch while traveling back to its original dispatch location (post).

As in Mason [34] we used trace driven simulation based on the actual data. We organized the data into twelve two-hour time blocks for each day, seven days a week. This gave us 84 different problem sets. Each replication is a two-hour block and the number of replications is equal to the number of weeks. This approach was adopted to accurately reflect the spatial and temporal distribution of calls.

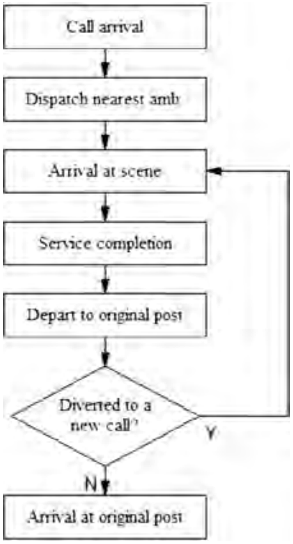


Fig. 1. Ambulance dispatch process.

The first objective is to maximize coverage for which we adopted the National Fire Protection Agency (NFPA) 1710 RT standard which is 8 min 59 s for 90% of the calls [1]. Interestingly, in 2009 about 64% of the largest 200 cities in the United States were using this standard and only about 28% of them reported being in compliance [37]. We compute the coverage by counting the number of calls reached under 9 min divided by the total number of calls. The second objective to minimize average response time which we implement by capturing actual RTs. The third objective is to maximize patient survivability where we capture the actual RT and evaluate survival probability, with $\max(0.594 - 0.055\text{RT}, 0)$ which is adapted from the function developed by McLay and Mayorga [5]. We also track individual ambulance busy probabilities in order to monitor workload balance. For each of these objectives the decision variables are ambulance locations and constraint is the fleet size.

Fig. 2 shows the experimental setup with 84 problem sets (7 days by 12 two-hour time blocks per day), which are run for a fleet size of 20, 21, 22 and 23 servers for each of the three different models. We consider the impact of the three different types of independent variables (1) Objective (Maximize Survivability, Minimize Average Response Time, and Maximize Coverage), (2) Number of Servers and (3) Days, and times, of the week on four different dependent variables (a) Coverage, (b) Survivability, (c) Response Time, and (d) Workload balance. Workload balance can also be viewed as equity among the crew members. Ideally there should be minimal differences among their workload. In this study we measure equity using the coefficient of variation, which is the ratio of the standard deviation of how busy ambulances are, and the average busy probability [38].

3.1. Optimization algorithm

There have been many heuristics successfully utilized in covering models. Galvao et al. [39], for example, used simulated annealing with the maximum expected coverage and the available coverage model. Rajagopalan et al. [40], in their survey of the performance of various meta-heuristics, concluded that, of those meta-heuristics examined, tabu search yielded both fast, and near-optimum solutions for problems with the objective of maximizing expected coverage. In general, tabu search based algorithms have

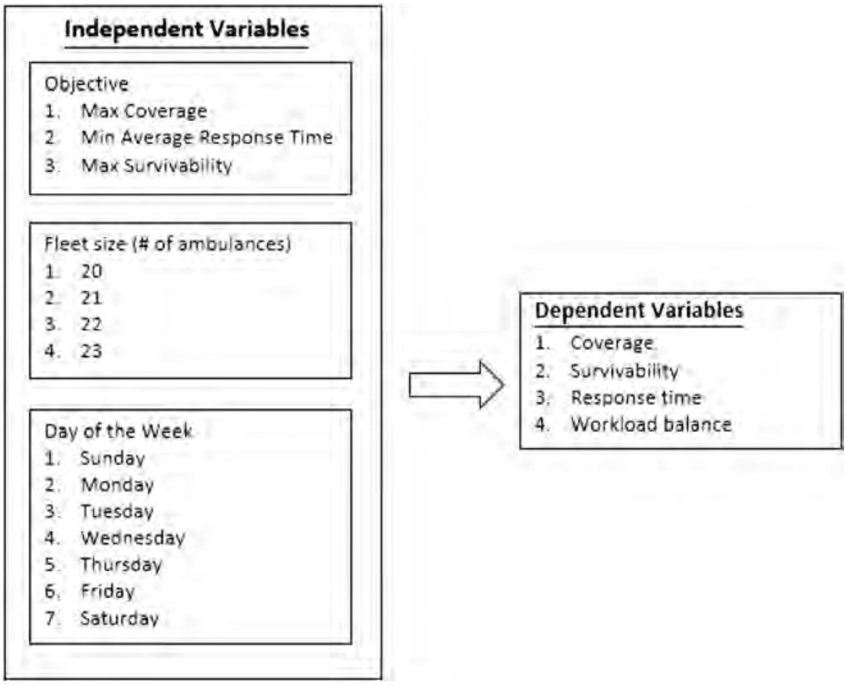


Fig. 2. Experimental setup.

been shown to be very fast in finding good solutions in this domain [41–43]. For an extensive review of covering model and corresponding optimization techniques, we refer the reader to Li et al.'s [15] recent comprehensive review of said problems. We chose to implement a variation of a tabu search meta-heuristic called Reactive Tabu Search (RTS) [44]. Objective 1 is solved using the RTS method. Then objectives 2 and 3 are initialized with the final solution from objective 1 and optimized via local search algorithm.

3.2. The Reactive Tabu Search (RTS) procedure

The RTS algorithm we develop for this paper is similar to the one used in Rajagopalan et al. [45]. In our implementation, we use a one-dimensional array of varying size to store a given solution and the corresponding objective function value (Fig. 2). The initial tabu size is set to one. Similar to [42], the basic operation in RTS, or any other tabu search technique in this domain, involves moving an ambulance from node i to node j, where node j is the best location in the neighborhood. "Neighborhood" for this study is defined to include nodes immediately surrounding the selected node within a 9 mile radius. The first ambulance is selected for the basic operation, and the best node from its neighborhood is selected only if the (j, k) pair is not already on the tabu list. The best node is selected based on the best coverage generated by the simulation model. Once selected the (j, k) pair goes into the tabu list. All moves (pairs) are stored in long-term memory. We consider this as a single iteration. The second ambulance is then selected for the basic neighborhood search, and the process repeats until the last ambulance within the neighborhood is selected, after which the first ambulance to the last ambulance in the neighborhood is selected again. The process continues for 100 iterations.

Throughout this search, the size of the tabu list changes according to the exploration or exploitation pressure needed. For example, if a move (i.e., (j, k) pair) is selected but is already stored in long-term memory, the tabu size increases to include that move. However if the (j, k) pair in the tabu list is not repeated for $2m$ iterations, where m is the current number of servers, then that (j, k) pair is removed from the list. As stated earlier, the terminating rule for the current implementation of RTS is 100 iterations. This number was selected after running a set of preliminary tests for long periods of time (e.g., 1000 iterations or more). The results of these preliminary tests showed that the incremental gains after 100 iterations were negligible. The set of locations generated during the first 100 iterations, which resulted in the maximum coverage, is then stored.

3.3. A local search algorithm

As stated earlier, solutions derived for objective 1 are used as the initial, feasible, starting solution for objectives 2 and 3. A local search algorithm is used to either minimize average response times (objective 2) or maximize survivability (objective 3). The total response time, or survivability, respectively, is generated by running the simulation for a set of ambulance locations. The local search algorithm takes the location of the first server and searches through all possible locations within the neighborhood to identify the node where the server should be moved so that the objective function improves. Once the first server has been moved to a "better" location, we continue the same process on the second server. This is an iterative process that continues until the last server (i.e., server m) has been evaluated and perhaps moved to a new node. The search process continues again, through all m servers. The process terminates when we go through all m servers without changing the location of any server. There are two reasons why we systematically move from the first server to the last and then back again to the first: (1) to ensure that all servers are selected an equal number of times, and (2) to establish a search termination point where all servers have been checked without any changes in their locations.

4. Results

The simulation model was coded in java (jdk 1.7) and run on a laptop Intel Core i7-2630QM CPU @ 2 GHz with 16 GB of RAM. The average run time for each one of the 84 problem sets was 120 s, and the range was from 30 to 300 s. We tested the three objectives

Table 1
Demand distribution per two-hour time intervals per day.

Time period	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.
12 am–2 am	743	428	446	427	442	459	643
2 am–4 am	684	380	338	356	386	352	565
4 am–6 am	382	297	265	302	304	313	361
6 am–8 am	399	587	524	577	553	544	420
8 am–10 am	622	850	816	850	854	797	660
10 am–12 pm	780	1015	959	942	1033	951	822
12 pm–2 pm	863	994	941	1044	1041	1049	934
2 pm–4pm	870	1026	992	993	1014	1091	927
4 pm–6 pm	821	1029	1067	1033	1063	1108	949
6 pm–8 pm	866	883	916	884	911	888	970
8 pm–10 pm	847	728	757	760	764	876	875
10 pm–12 am	648	591	648	612	673	812	906
Total	8525	8808	8669	8780	9038	9240	9032

using actual data from a metropolitan county approximately 540 square miles with a population of 801,137 in 2004. In that year, the county received a total of 77,292 calls, of which 62,092 were classified as "medical emergency." As mentioned previously we superimposed a two mile by two mile grid over the county for call aggregation purposes. This generated a total of 168 demand nodes and we assumed that the ambulances could be posted at any of the nodes except those that constituted a boundary node, which resulted in 104 potential ambulance locations. Most boundary nodes are typically less than the four sq. miles and contain very few calls, if any. Data analyses (See Table 1) showed that the county's call demand distribution fluctuates significantly by day of the week and time of day [46]. Table 1 displays the yearly demand for each of the two-hour time intervals.

We can also see that the volume of calls begins to increase around 8 a.m. and the peak is usually between 4 pm and 6 pm, before slowly declining.

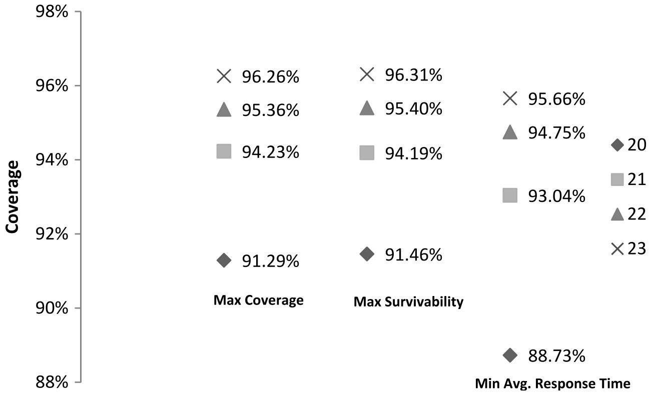


Fig. 3. Impact on coverage for 20, 21, 22 and 23 ambulances.

4.1. Impact on coverage

We consider the performance of the three different objectives (1) Maximizing Coverage, (2) Minimizing Average Response Time and (3) Maximizing Survivability with respect to coverage for different fleet sizes (20–23 ambulances). Coverage, computed via the simulation, is the percentage of calls covered within a given response time threshold. We also know from our data that different days in a week have different demand distributions. Therefore, we attempt to determine whether there is an interaction between the different days in a week and each of the three independent variables (Fig. 2). An analysis of variance is used to determine if the independent variables (model, fleet size, and day and time of the week) are statistically significant with respect to coverage. The results of the analysis of variance are given in Table A.1 in the Appendix.

Table A.1 in the Appendix shows the impact of the independent variables, on the dependent variable, coverage. The type of objective (Maximize Coverage, Minimize Average Response Time, and Maximize Survivability) has a statistically significant effect on percentage of calls covered ($p < 0.05$). We also see that the fleet size (20, 21, 22, or 23 ambulances) has a significant impact on percentage of calls covered, as well as the day and time of the week ($p < 0.05$). In addition, notice that there is a statistically significant interaction between the type of objective and the number of ambulances ($p < 0.05$). These effects are explained in further detail by analyzing Figs. 3–7.

We can see in Fig. 3 that the Maximum Coverage objective is significantly better than the Minimum Average Response Time objective ($p < 0.05$) in terms of coverage. However there is no statistically significant difference between the Maximum Survivability objective and Maximum Coverage objective. Therefore, we

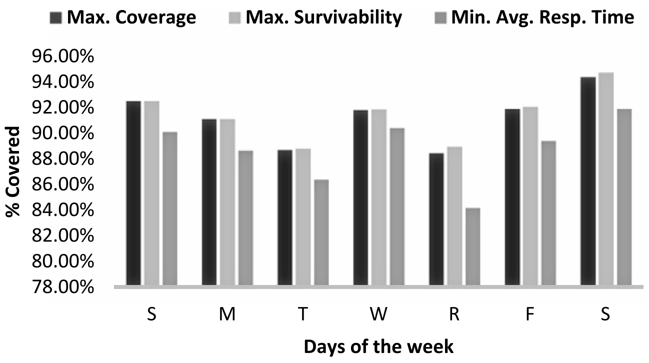


Fig. 4. Impact on coverage (20 ambulances).

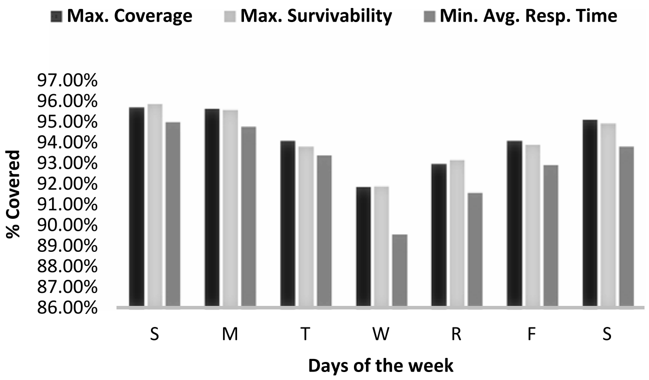


Fig. 5. Impact on coverage (21 ambulances).

Table 2
% coverage at different response time intervals.

Time (min)	0–3	0–6	0–9	0–12	0–15
Fleet size: 20 ambulances					
Max. Coverage	9.32%	35.88%	65.06%	82.52%	91.07%
Min. Avg. Response Time	9.62%	37.23%	65.37%	81.54%	89.14%
Max. Survivability	10.10%	38.74%	67.23%	83.34%	91.19%
Fleet size: 21 ambulances					
Max. Coverage	8.88%	36.71%	66.40%	85.04%	93.99%
Min. Avg. Response Time	9.53%	38.70%	67.95%	84.96%	92.93%
Max. Survivability	9.80%	39.37%	68.91%	85.88%	93.96%
Fleet size: 22 ambulances					
Max. Coverage	10.23%	41.73%	72.18%	88.20%	95.20%
Min. Avg. Response Time	11.16%	44.00%	74.07%	88.71%	94.68%
Max. Survivability	11.36%	44.57%	74.54%	89.05%	95.24%
Fleet size: 23 ambulances					
Max. Coverage	12.13%	45.74%	74.96%	90.22%	96.07%
Min. Avg. Response Time	13.09%	48.51%	76.92%	90.62%	95.62%
Max. Survivability	13.37%	48.86%	77.37%	91.00%	96.13%

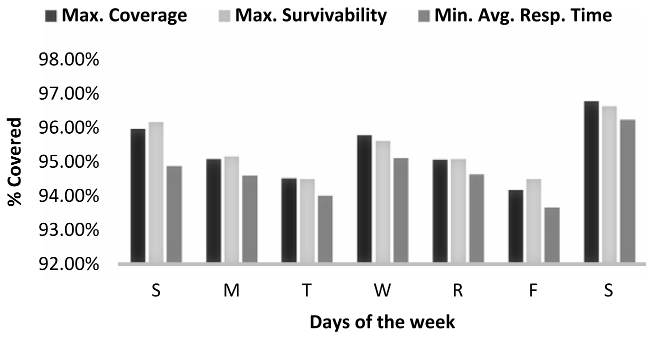


Fig. 6. Impact on coverage (22 ambulances).

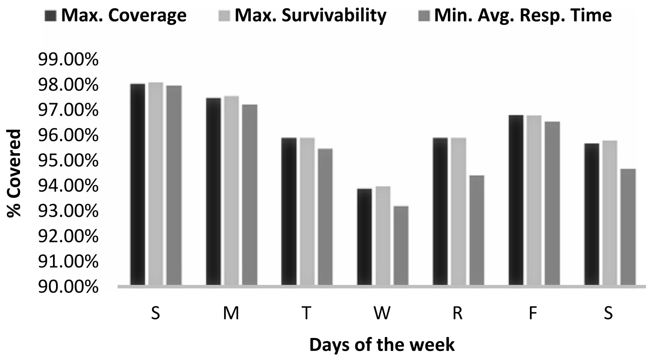


Fig. 7. Impact on coverage (23 ambulances).

can conclude that the Maximum Survivability objective is clearly superior to the Minimum Average Response Time objective, which sacrifices coverage in order to achieve its objective.

Fig. 3 also shows the interaction effect. As the fleet size increases from 20 to 23 ambulances the Minimum Average Response Time objective performs much closer to the other two objectives. As expected, coverage is worst for all three objectives when the fleet size is 20.

Figs. 4–7, show the impact of coverage for different days of the week with respect to fleet size (20, 21, 22, and 23 ambulances). The patterns in Figs. 4 through 7 are consistent with what is seen in Fig. 3. The Maximum Coverage and the Maximum Survival objectives give us the best coverage and the Minimum Average Response Time objective performs worse. However, the difference in performance between Minimum Average Response Time and the other two objectives is reduced as the number of ambulances is increased. We can also see in days where the system is stressed (i.e., heavy demand) such as on Tuesdays, Wednesdays and Thursdays, and number of servers is small the Minimum Average Response Time objective performs noticeably worse than the Maximum Coverage and the Maximum Survival objectives.

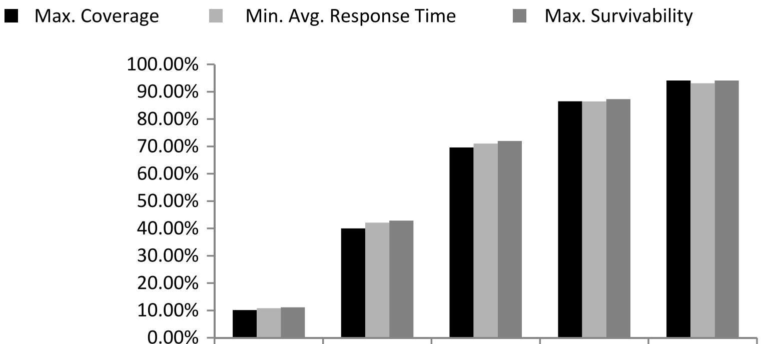
These figures also confirm that the performance of the Minimum Average Response Time objective improves as the number of servers increase. We also consider the impact of these objectives on the percentage of calls covered within 3, 6, 9, 12 and 15 min in Table 2 for a varying fleet size (20, 21, 22 and 23 ambulances).

As expected, the Minimum Average Response Time and the Maximum Survivability objectives cover a greater percentage of calls within 3 and 6 min interval than the Maximum Coverage objective. This is because these later objectives value shorter distances, or times, whereas the Maximum Coverage objective assumes calls are covered as long as the RTT (or distance to call) is within the specified threshold.

Fig. 8 shows the average cumulative percentage of calls covered within five time intervals. The Survivability objective covers an average of 1.02% more calls within the first 3 min than the Coverage objective and given that it does not sacrifice in total coverage to achieve this, it makes it the superior with respect to coverage. In comparison the Minimum Average Response Time objective covers only an average of 0.71% more calls within the first 3 min than the Coverage objective. Therefore, while it does cover more calls within 3 min than the Coverage objective's performance, it does not match that of the Survivability objective.

4.2. Impact on survivability

Next, we consider the impact of using the three different objectives with varying fleet size (20, 21, 22, and 23 ambulances) on survivability. Table A.2 in the Appendix shows the result of the analysis of variance on survivability. We find that all three independent variables (i.e., Objective, Days, and Ambulances) significantly affect the chance of survival ($p < 0.05$). There is no significant interaction between the objective and day and time of the week which indicates that objectives perform consistently across different days and times of a week. There is also no significant interaction between objectives and number of ambulances.



	3.00	6.00	9.00	12.00	15.00
Max. Coverage	10.14%	40.02%	69.65%	86.50%	94.09%
Min. Avg. Response Time	10.85%	42.11%	71.08%	86.45%	93.09%
Max. Survivability	11.16%	42.89%	72.02%	87.32%	94.13%

Fig. 8. Average cumulative coverage (%) for the three objectives.

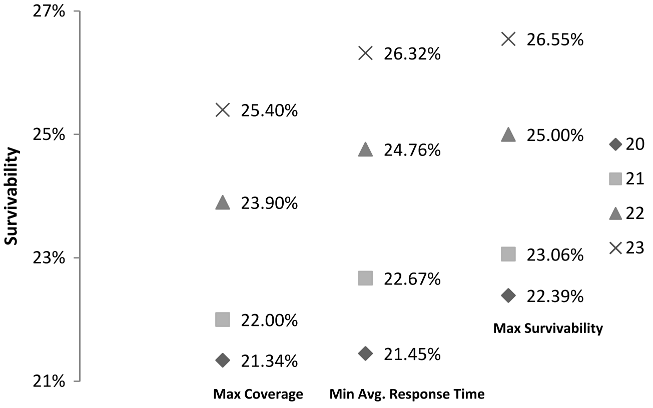


Fig. 9. Impact on survivability with 20, 21, 22 and 23 ambulances.

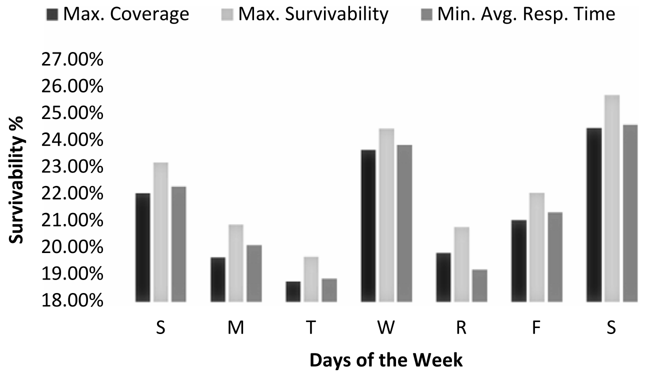


Fig. 10. Impact on survivability (20 ambulances).

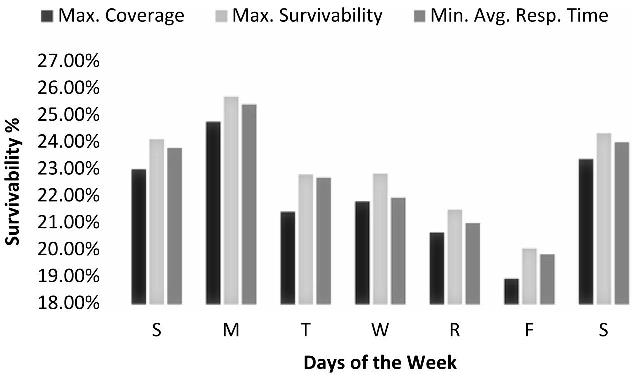


Fig. 11. Impact on survivability (21 ambulances).

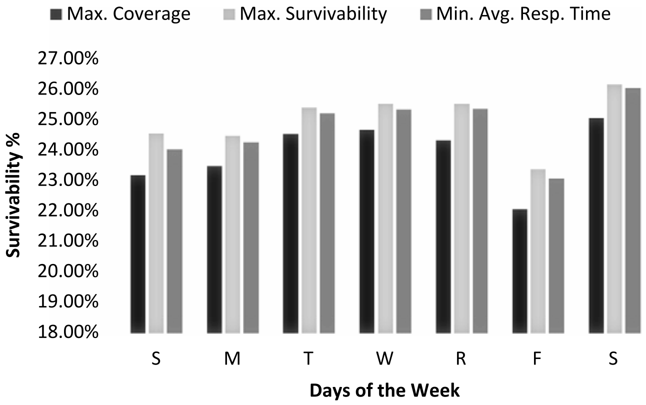


Fig. 12. Impact on survivability (22 ambulances).

Fig. 9 shows that the Maximum Survivability objective clearly gives the best results followed by Minimum Average Response Time objective and then Maximum Coverage objective. Also, as the fleet size (20, 21, 22, and 23 ambulances) increases, the probability of survival increases.

There is no statistically significant interaction between objective type and number of ambulances. However, we can see some trends when we investigate by day of week in detail. Figs. 10–13 show the impact of each objective on survivability for each day of the week and number of servers. We can see for small number of servers (20) in Fig. 10, the Minimum Average Response Time objective and the Maximum Coverage objective produce similar results. However, as the number of servers increases (21, 22, 23) in Figs. 11–13, the Minimum Average Response Time objective performs better than the Maximum Coverage objective. However, this is not statistically significant because the performance fluctuates on different days of week. This is clearly seen in Fig. 11 where the Minimum Average Response Time objective matches the performance of the Maximum Survivability objective on some days and on others days this objective performs as poorly as the Maximum Coverage objective. This is because the Minimum Average Response Time objective performs better when more servers are available or demand is not high.

For smaller fleet sizes and on days when demand is high the solutions from the Minimum Average Response Time objective do not match the quality of the solutions obtained using the Maximum

Table 3
Survivability for different response time intervals.

Time (min)	0–3	3–6	6–9	9–12	12–15
Fleet size: 20 ambulances					
Max. Coverage	4.52%	9.94%	6.11%	0.77%	0.00%
Min. Avg. Response Time	4.62%	10.30%	5.83%	0.71%	0.00%
Max. Survivability	4.94%	10.78%	5.97%	0.70%	0.00%
Fleet size: 21 ambulances					
Max. Coverage	4.42%	10.54%	6.24%	0.81%	0.00%
Min. Avg. Response Time	4.77%	11.05%	6.12%	0.73%	0.00%
Max. Survivability	4.90%	11.22%	6.20%	0.73%	0.00%
Fleet size: 22 ambulances					
Max. Coverage	5.00%	11.82%	6.37%	0.71%	0.00%
Min. Avg. Response Time	5.48%	12.37%	6.27%	0.64%	0.00%
Max. Survivability	5.57%	12.53%	6.26%	0.63%	0.00%
Fleet size: 23 ambulances					
Max. Coverage	5.99%	12.63%	6.12%	0.66%	0.00%
Min. Avg. Response Time	6.47%	13.34%	5.92%	0.59%	0.00%
Max. Survivability	6.63%	13.37%	5.97%	0.59%	0.00%

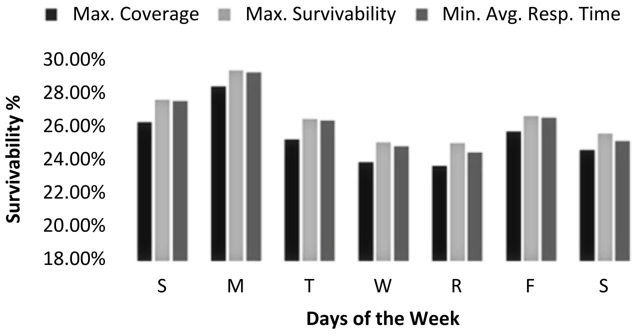


Fig. 13. Impact on survivability (23 ambulances).

Survivability objective. Table 3 shows the survivability data from time intervals from 0 to 3 min, 3 to 6 min, 6 to 9 min, 9 to 12 min and 12 to 15 min.

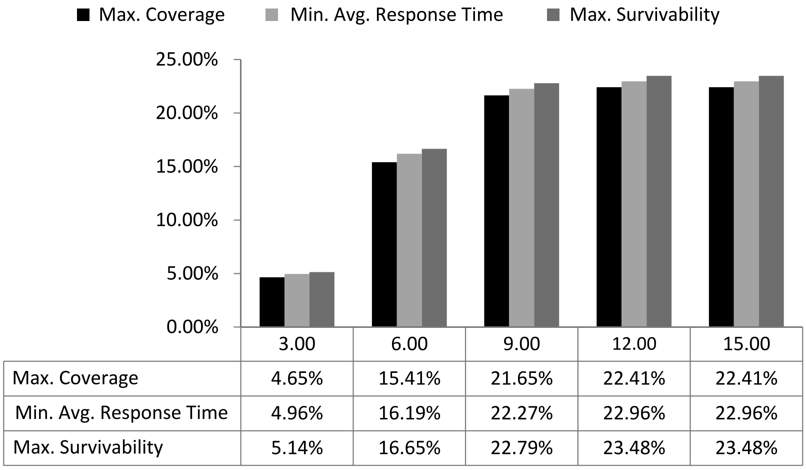
Given the survivability function, the chance of survival after 10 min drops to 0%. Since the survivability depends on the calls covered within a certain time period, we calculate the survivability measure by multiplying the product of the percentage of calls covered within a certain time interval by the probability of survival. Table 3 follows a similar pattern to Table 2. Ambulance locations provide the Maximum Survivability objective the best results within the first 3 min followed by the Minimum Average Response Time objective and then, lastly, by the Maximum Coverage objective. With 20 servers the Minimum Average Response Time objective's performance is closer to the Maximum Coverage objective and it improves as the number of servers increase.

Fig. 14 shows the average performance of all three objectives on survivability. Within the first three minutes, the Maximum Survivability objective outperforms the other two objectives. Within the critical 3 min time interval the chance of survival increases on average by 0.5%, as compared to the Maximum Coverage objective, and 0.31% over the Minimum Average Response Time objective, or an additional 50 and 31 lives saved per 10,000 cardiac arrest calls, respectively.

4.3. Impact on average response times

Fig. 15, shows the impact of the three different objectives on average response times per call. Table A.3 in the Appendix displays the results of the Analysis of Variance conducted on the data.

All three independent variables are statistically significant ($p < 0.05$). As expected the Maximum Coverage objective gives us the worst average response times and is significantly ($p < 0.05$) different from Minimum Average Response Time and Maximum Survivability. There is no statistically significant difference between



	3.00	6.00	9.00	12.00	15.00
Max. Coverage	4.65%	15.41%	21.65%	22.41%	22.41%
Min. Avg. Response Time	4.96%	16.19%	22.27%	22.96%	22.96%
Max. Survivability	5.14%	16.65%	22.79%	23.48%	23.48%

Fig. 14. Average cumulative survivability (%) for three objectives.

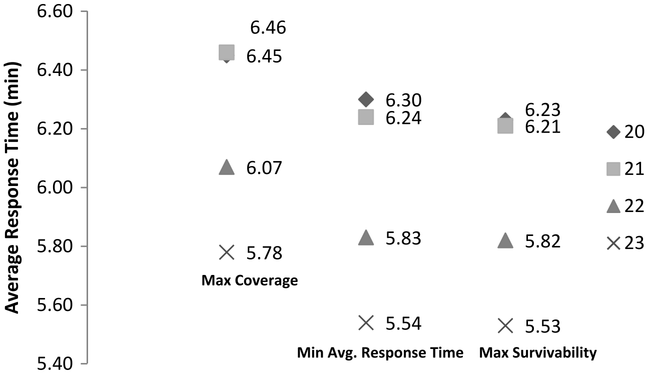


Fig. 15. Impact on average response time for 20, 21, 22, and 23 servers.

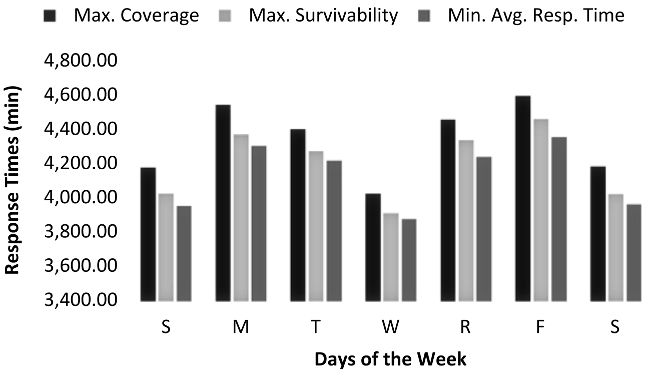


Fig. 16. Impact on cumulative response times (20 amb.)

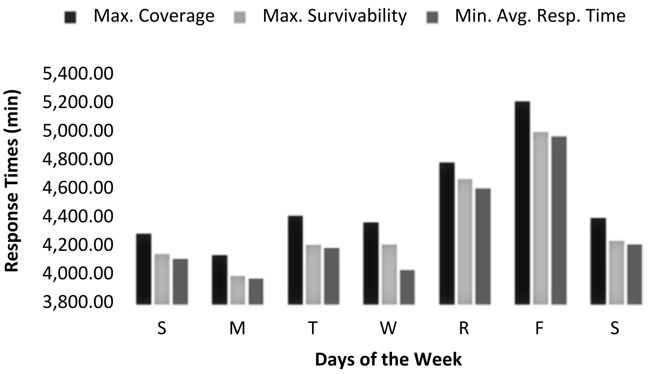


Fig. 17. Impact on cumulative response times (21 amb.)

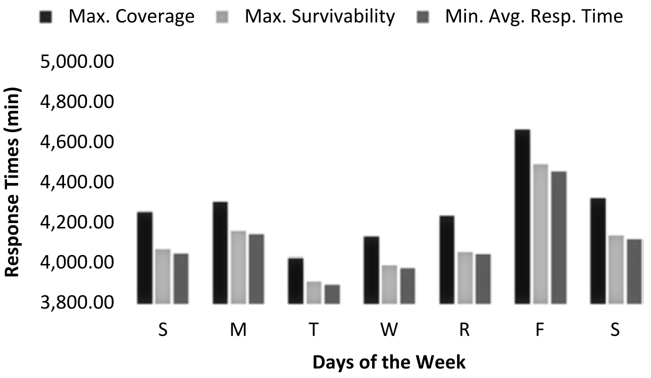


Fig. 18. Impact on cumulative response times (22 amb.)

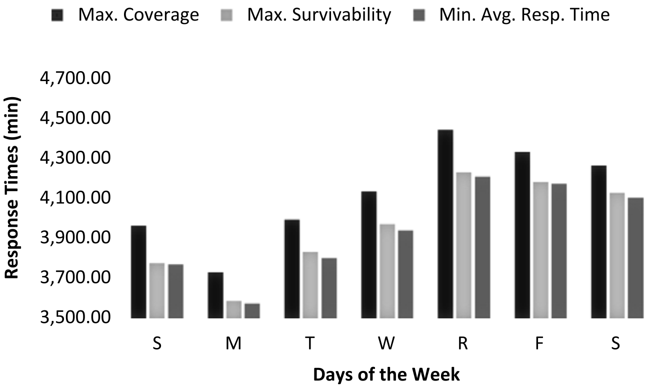


Fig. 19. Impact on cumulative response times (23 amb.)

Minimum Average Response Time and Maximum Survivability. We can also conclude that as the number of ambulances increase the average response time decreases. There is no statistical difference in average response time when the number of ambulances is 20 and 21. However, when the number of ambulances increase to 22 and 23 the decrease in average response time is statistically significant ($p < 0.05$).

Upon further investigation, when we break the data up by day as shown in Figs. 16–19 the response time for different days of the week is also statistically significant ($p < 0.05$) due to the different demand patterns each day. We can also see a significant interaction between number of ambulances and different days of the week ($p < 0.05$). The Maximum Coverage model consistently has the biggest response time for every day of the week for the different kinds of servers while the other two objectives have similar response times. We also can see that Thursdays and Fridays consistently have a higher average response time than other days. This is due to higher call volume. We can also see that the difference between Thursdays and Fridays is higher when we have 20 or 21 ambulances than when we have 22 or 23 ambulances showing the interaction between the number of ambulances and the days of the week.

4.4. Impact on workload balance

The coefficient of variation (CV), which is the ratio of the standard deviation of how busy ambulances are and the average busy probability, is used to calculate workload balance [38]. The larger the CV the more variation there is among the different ambulances deployed, hence less workload balance among the crews. Table A.4 in the Appendix shows the results of an analysis of variance on the CV for the ambulances. We can see that while the day and time of the week and number of ambulances are statistically significant ($p < 0.05$), the kind of objective used is not statistically significant.

Fig. 20 shows the average CV for the three different objectives for varying fleet size (20, 21, 22 and 23 ambulances). Here we observe that the Minimum Average Response objective has the highest CV for 20 and 21 servers, and then declines quite sharply for 22 and 23 servers. In Figs. 21–24 we show that there is a high degree of fluctuation between the three different objectives during different day and times of the week.

If we average the CV for each objective as shown in Fig. 20, the Maximum Coverage is the worst (70.33%), followed by Minimum Average Response objective (69.90%), and the Maximum Survivability objective (69.52%). However, since there is so much variation with respect to the day and time of the week with respect to the three objectives, and the fleet size, the interaction as shown in Table A.4 is not statistically significant.

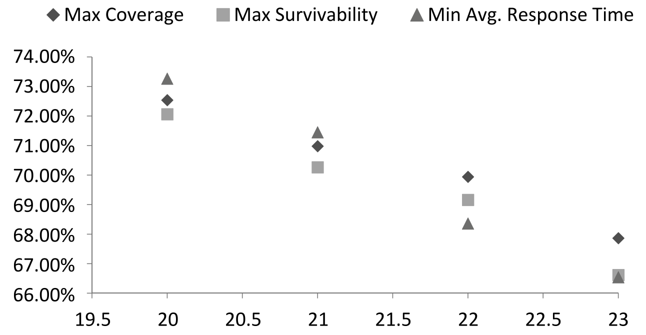


Fig. 20. Impact of fleet size on workload balance (CV).

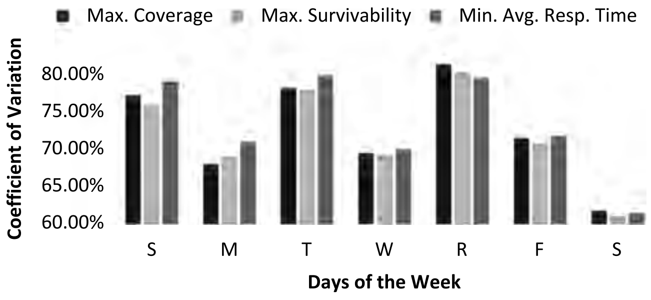


Fig. 21. Impact on equity (20 ambulances).

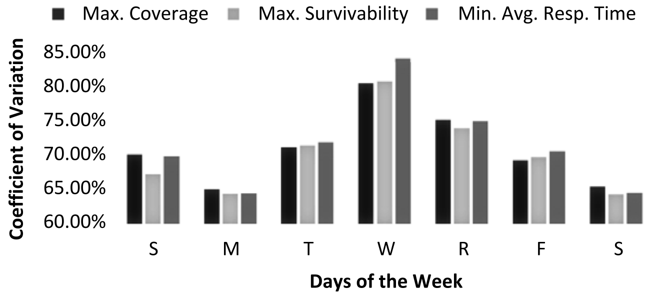


Fig. 22. Impact on equity (21 ambulances).

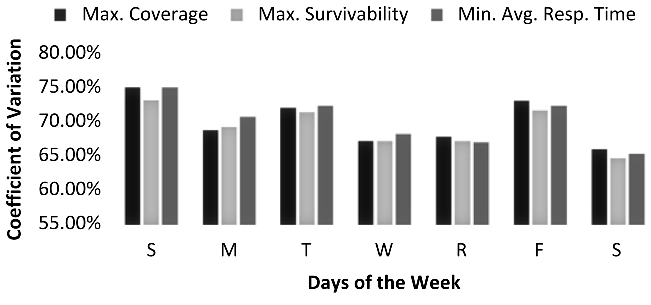


Fig. 23. Impact on equity (22 ambulances).

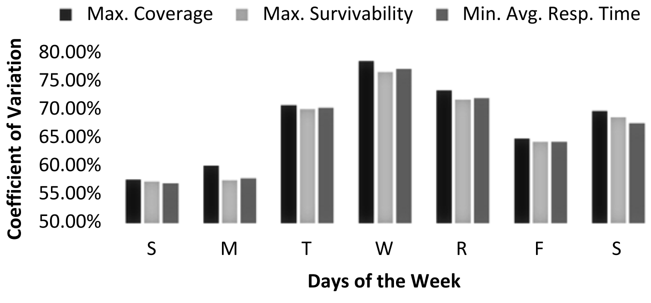


Fig. 24. Impact on equity (23 ambulances).

4.5. *Comparing simulated response times to actual response times*

Table 4 shows the percentage of actual calls covered by in year 2004 versus the percentage of calls covered by our simulation–optimization model with a fleet size of 23 ambulances.

Overall the coverage statistics from the simulation model with the three objectives follow closely the same coverage pattern obtained from the actual data. The net differences in coverage between the actual data and maximum survivability objective ranges from 2.75% to 7.96%. There is about a net 3% difference between calls covered by our model and the MEDIC for the first six minutes. Some of the differences can be attributed to flexible deployment practices of MEDIC such as ambulances originally intended (enroute) to low priority calls are at times diverted towards high priority calls. On some other occasions, idle ambulances returning to the headquarters to end their shift can be dispatched to a high priority emergency or posted at a location temporarily while another unit is rushed to another emergency. For high priority calls when the available ambulances are significantly outside the target RT or in an exceptionally rare moment when all ambulances are busy, the dispatcher can reach out to the private ambulance operators and Charlotte Fire Department to dispatch an ambulance and a fire truck (with an EMT) at the same time. Similar issue is faced in county border areas as well. Like most EMS operators MEDIC has 'mutual aid agreements' with surrounding county EMS operators for handling high priority calls.

5. Conclusions

In this paper, we developed a simulation model of the EMS operations to compare the performance of three well-known location model objectives in the recent literature: Maximum Coverage, Minimum Average Response Time, and Maximum Survivability. The simulation model removed several simplifying assumptions that are necessary in analytical models. The output generated from the simulation model enabled us to analyze coverage and survivability statistics for various response time intervals. This enabled us to better understand the differences between the objectives of the three models with respect to the given data which was highly variable depending on the day and time of the week. We evaluated the objectives on four criteria: (1) Percentage of calls covered, (2) Survivability, (3) Average Response Time, and (3) Workload balance among the fleet. Our findings show that the Maximum Survivability objective is superior to the other two objectives. The Maximum Survivability objective matched or performed significantly better than the other two objectives in all the four criteria. Table 5 summarizes our experimental results and rankings of the four objectives according to their performance with respect to the four criteria. It should be noted that the Minimum Average Response Time objective, while performing well in three of the four criteria, requires a larger sacrifice in coverage to achieve its goals. The Maximum Coverage objective performed the worst in three of the four criteria.

Most EMS agencies use coverage measures for deployment planning, and performance tracking and reporting purposes [1]. Our results suggest that using a survivability objective tends to produce better coverage results than the maximum coverage objective. EMS agencies can benefit from the use of such models for planning purposes such as determining the posts (staging locations) of their fleet. However, for reporting and performance tracking, survivability statistics could be difficult to collect and explain to the governing body (city or county boards) and even more so to the public. First, estimating patient survival is difficult since it is assessed at the hospital and a patient may be discharged several days after delivery to the emergency department. Second,

Table 4
Actual versus simulated coverage statistics.

Response Time (min)	0–3	0–6	0–9	0–12	0–15
Actual	16.12%	52.02%	85.33%	96.34%	98.91%
Max. Coverage	12.13%	45.74%	74.96%	90.22%	96.07%
Min. Avg. Response Time	13.09%	48.51%	76.92%	90.62%	95.62%
Max. Survivability	13.37%	48.86%	77.37%	91.00%	96.13%

Table 5
Summary of the results.

	Maximum survivability	Minimum Avg. Response Time	Maximum coverage
Coverage[a]	1	3	2
Survivability[b]	1	2	3
Response Time[c]	1	2	3
Equity[d]	1	2	3

[a] Maximum survivability and maximum coverage are not statistically different but minimum average response time is.
[b] All three objectives are statistically different from each other.
[c] Maximum survivability and minimum average response time are not statistically different but maximum coverage is.
[d] None of the three objectives are statistically different from each other.

patient survival information is not readily available due to medical privacy regulations. Third, the few published survivability functions are derived from data on out-of-hospital cardiac arrests. These numbers indicate that the probability of survival is about 55% if there is an immediate intervention (a response time of nearly zero minutes) with a rapid decline in probability of survival after the first few minutes and almost zero chances of survival past the standard 9–10 min response time. In case of heart attacks (myocardial infarction with ST elevation) the patient's one-year mortality increases by 7.5% for every 30 min of delay in receiving percutaneous coronary intervention (PCI) which can only be delivered at an appropriate receiving facility. Unlike survivability statistics, response times are easy to obtain and evaluate. Thus, based on our results and taking into account the difficulty of estimating survivability rates ambulance operators should consider using a "maximum survivability model" for deployment planning purposes while simultaneously reporting compliance statistics, such as 90% of calls covered within 9 min.

There are some limitations of this study that are important to acknowledge. First, due to the nature of the problem we are constrained to use meta-heuristics. Therefore, the solutions are not guaranteed to be optimal. Second, the problem domain has been discretized. All demand is assumed to be aggregated at the center of a 2 by 2 square mile zone and the servers are also assumed to be located at the center of these demand zones and the travel times are computed using Manhattan distance metric between zone centers. Although this is consistent with analytical approaches, ideally one would prefer to use the actual road network in the simulation model, which is likely to add a major computational burden. Third, when the optimization routine determines the new locations for ambulances at the beginning of each two-hour block, the actual movement of the ambulances to their new locations is not simulated. And, fourth, we assumed all calls from our available data are of equal priority, which is perhaps reflected in the low survivability values.

Future research may consider using the exact longitude and latitude of the calls, and allowing the ambulances to be located anywhere on the road network. Also, it would be of interest to look at the stratification of calls as classified by their priority and based on this consider alternative dispatch policies while maximizing survivability for priority 1 patients. For example, the coverage performance measure could be divided into hard and soft coverage for high to low priority calls. This will in effect make the coverage performance measure a proxy of a survival function. It will be worthwhile to explore the performance of such a coverage measure with the survival functions found in literature.

Table A.1
Analysis of variance for the dependent variable coverage.

Source	Sig. (p-value)	Observed power[a]
Objective*	0.000	1.000
Day*	0.000	1.000
Ambulances*	0.000	1.000
Objective × Day	0.998	0.135
Objective × Ambulances*	0.010	0.887
Day × Ambulances*	0.000	1.000
Objective × Day × Ambulances	1.000	0.167

[a] Computed using alpha = 0.05.
* Refers to variables which are statistically significant at ($p < 0.05$).

Table A.2
Analysis of variance for the dependent variable survivability.

Source	Sig. (p-value)	Observed power[a]
Objective*	0.000	1.000
Day*	0.000	1.000
Ambulances*	0.000	1.000
Objective × Day	1.000	0.081
Objective × Ambulances	0.563	0.324
Day × Ambulances*	0.000	1.000
Objective × Day × Ambulances	1.000	0.097

[a] Computed using alpha = 0.05.
* Refers to variables which are statistically significant at ($p < 0.05$).

Table A.3
Analysis of variance for the dependent variable response time.

Source	Sig. (p-value)	Observed power[a]
Objective*	0.00	1.000
Day*	0.00	1.000
Ambulances*	0.00	1.000
Objective × Day	1.000	1.000
Objective × Ambulances	0.970	1.000
Day × Ambulances*	0.00	0.081
Objective × Day × Ambulances	1.000	0.324

[a] Computed using alpha = 0.05.
* Refers to variables which are statistically significant at ($p < 0.05$).

Acknowledgments

The authors thank MEDIC, the EMS agency in Mecklenburg County (Charlotte), for providing data and operational insights. The authors also thank the anonymous reviewer for his detailed and thoughtful comments that have greatly improved this manuscript.

Table A.4
Analysis of variance dependent variable coefficient of variation.

Source	Sig. (p-value)	Observed power[a]
Objective	0.601	0.134
Day[*]	0.000	1.000
Ambulances[*]	0.000	0.997
Objective × Day	1.000	0.073
Objective × Ambulances	0.995	0.078
Day × Ambulances[*]	0.000	1.000
Objective × Day × Ambulances	1.000	0.070

[a] Computed using alpha = 0.05.
[*] Refers to variables which are statistically significant at ($p < 0.05$).

Appendix. ANOVA tables (significance and observed power columns only)

See Tables A.1–A.4.

References

[1] N.F.P. Association, Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations, and Special Operations to the Public by Career Fire Departments, NFPA 1710, NFPA, Quincy, MA, 2010.

[2] L. Brotcorne, G. Laporte, F. Semet, Ambulance location and relocation models, European J. Oper. Res. 147 (3) (2003) 451–463.

[3] R.Z. Farahani, et al., Covering problems in facility location: A review, Comput. Ind. Eng. 62 (1) (2012) 368–407.

[4] J.B. Goldberg, Operations research models for the deployment of emergency services vehicles, EMS Manage. J. 1 (1) (2004) 20–39.

[5] L. McLay, M. Mayorga, Evaluating emergency medical service performance measures, Health Care Manage. Sci. 13 (2) (2010) 124–136.

[6] E. Erkut, A. Ingolfsson, G. Erdogan, Ambulance location for maximum survival, Nav. Res. Logist. 55 (1) (2008) 42–58.

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

[8] T.H. Blackwell, J.S. Kaufman, Response Time Effectiveness: Comparison of Response Time and Survival in an Urban Emergency Medical Services System, Acad. Emerg. Med. 9 (4) (2002) 288–295.

[9] T.D. Valenzuela, et al., Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos, New Engl. J. Med. 343 (17) (2000) 1206–1209.

[10] D. Bandara, M.E. Mayorga, L.A. McLay, Optimal dispatching strategies for emergency vehicles to increase patient survivability, Int. J. Oper. Res. 15 (2) (2012) 195–214.

[11] L.A. McLay, M.E. Mayorga, A dispatching model for server-to-customer systems that balances efficiency and equity, Manuf. Serv. Oper. Manage. 15 (2) (2013) 205–220.

[12] D. Bandara, M.E. Mayorga, L.A. McLay, Priority dispatching strategies for EMS systems, J. Oper. Res. Soc. 65 (4) (2014) 572–587.

[13] M.E. Mayorga, D. Bandara, L.A. McLay, Districting and dispatching policies for emergency medical service systems to improve patient survival, IIE Trans. Healthcare Syst. Eng. 3 (1) (2013) 39–56.

[14] C. ReVelle, H.A. Eiselt, M.S. Daskin, A bibliography for some fundamental problem categories in discrete location science, European J. Oper. Res. 184 (3) (2008) 817–848.

[15] X. Li, et al., Covering models and optimization techniques for emergency response facility location and planning: A review, Math. Methods Oper. Res. 74 (3) (2011) 281–310.

[16] H.K. Rajagopalan, C. Saydam, A minimum expected response model: formulation, heuristic solution, and application, Socio-Econ. Plan. Sci. 43 (4) (2009) 253–262.

[17] M.P. Larsen, et al., Predicting survival from out-of-hospital cardiac arrest: A graphic model, Ann. Emerg. Med. 22 (11) (1993) 1652–1658.

[18] F. Borras, J.T. Pastor, The ex-post evaluation of the minimum local reliability level: an enhanced probabilistic location set covering model, Ann. Oper. Res. 111 (2002) 51–74.

[19] E. Savas, Simulation and cost-effectiveness analysis of New York's emergency ambulance service, Manage. Sci. 15 (1969) 608–627.

[20] C. Swoveland, et al., Ambulance location: A probabilistic enumeration approach, Manage. Sci. 20 (4) (1973) 686–698.

[21] J.A. Fitzsimmons, A methodology for emergency ambulance deployment, Manage. Sci. 19 (6) (1973) 627–636.

[22] G. Berlin, J. Liebman, Mathematical analysis of emergency ambulance locations, Socio-Econ. Plan. Sci. 8 (1974) 323–328.

[23] C. Toregas, et al., The location of emergency service facilities, Oper. Res. 19 (6) (1971) 1363–1373.

[24] O. Fujiwara, T. Makjamroen, K.K. Gupta, Ambulance deployment analysis: A case study of Bangkok, European J. Oper. Res. 31 (9–18) (1987).

[25] M.S. Daskin, A maximal expected covering location model: Formulation, properties, and heuristic solution, Transp. Sci. 17 (1) (1983) 48–69.

[26] M.S. Liu, J.T. Lee, A simulation of a hospital emergency call system using SLAM, Simulation 51 (6) (1988) 216–221.

[27] D. Uyeno, C. Seeberg, A practical methodology for ambulance location, Simulation 43 (2) (1984) 79–87.

[28] J. Repede, J. Bernardo, Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky, European J. Oper. Res. 75 (3) (1994) 567–581.

[29] A.S. Zaki, H.K. Cheng, B.R. Parker, A simulation model for the analysis and management of an emergency service system, Socio-Econ. Plan. Sci. 31 (3) (1997) 173–189.

[30] J.B. Goldberg, et al., A simulation model for evaluating a set of emergency vehicle base locations: Development, validation, and usage, Socio-Econ. Plan. Sci. 24 (2) (1990) 125–141.

[31] M. Restrepo, S.G. Henderson, H. Topaloglu, Erlang loss models for the static deployment of ambulances, Health Care Manage. Sci. 12 (2009) 67–79.

[32] M.S. Maxwell, et al., Approximate dynamic programming for ambulance redeployment, INFORMS J. Comput. 22 (2) (2010) 266–281.

[33] R. Alanis, A. Ingolfsson, B. Kolfal, A Markov chain model for an EMS system with repositioning, Prod. Oper. Manage. 22 (1) (2013) 216–231.

[34] A.J. Mason, Simulation and real-time optimised relocation for improving ambulance operations, in: B. Denton (Ed.), Handbook of Healthcare Operations Management: Methods and Applications, Springer, New York, 2013, pp. 289–317.

[35] Y. Kergosien, et al., A generic and flexible simulation-based analysis tool for EMS management, Int. J. Prod. Res. 53 (24) (2015) 7299–7316.

[36] L. Aboueljinane, E. Sahin, Z. Jemai, A review on simulation models applied to emergency medical service operations, Comput. Ind. Eng. 66 (4) (2013) 734–750.

[37] D.M. Williams, M. Ragone, 2009 JEMS 200-City Survey: Zeroing in on what matters, J. Emerg. Med. Serv. 34 (2) (2010) 38–42.

[38] H. Toro-Díaz, et al., Joint location and dispatching decisions for emergency medical services, Comput. Ind. Eng. 64 (4) (2013) 917–928.

[39] R.D. Galvao, F.Y. Chiyoshi, R. Morabito, Towards unified formulations and extensions of two classical probabilistic location models, Comput. Oper. Res. 32 (1) (2005) 15–33.

[40] H.K. Rajagopalan, et al., Developing effective meta-heuristics for a probabilistic location model via experimental design, European J. Oper. Res. 177 (2) (2007) 365–377.

[41] L. Brotcorne, G. Laporte, F. Semet, Fast heuristics for large scale covering location problems, Comput. Oper. Res. 29 (6) (2002) 651–665.

[42] M. Gendreau, G. Laporte, F. Semet, Solving an ambulance location model by tabu search, Locat. Sci. 5 (2) (1997) 75–88.

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

[44] R. Battiti, G. Tecchiolli, The reactive tabu search, J. Comput. 6 (2) (1994) 126–140.

[45] H.K. Rajagopalan, C. Saydam, J. Xiao, A multiperiod set covering location model for dynamic redeployment of ambulances, Comput. Oper. Res. 35 (3) (2008) 814–826.

[46] H. Setzler, C. Saydam, S. Park, EMS call volume predictions: a comparative study, Comput. Oper. Res. 36 (6) (2009) 1843–1851.