Effects of ambulance dispatching and relocation decisions on EMS quality

Jenny Díaz-Ramírez
Engineering Department
Universidad de Monterrey
San Pedro Garza García, 66238, N.L. Mexico
jenny.diaz@udem.edu

María Gulnara Baldoquín De la Peña
Mathematical Sciences Department
Universidad EAFIT
Medellín, Colombia
mbaldoqu@eafit.edu.co

Abstract

When a call for a service is received by an Emergency Medical Services (EMS) provider, a decision must be made on which vehicle to send to provide assistance. The dispatched vehicle becomes busy and the system's preparedness to attend a new call decreases. Therefore, to increase this value, idle EMS vehicles could be relocated among the standby sites. This study aims to evaluate the effect on the critical performance measures of the system, the sequential application of a dispatching mathematical model and a relocation model, taking into account the operational requirements of an EMS provider in Colombia.

Four scenarios were simulated during a period of maximum activity to evaluate the dynamic state of the system given by the dispatch and relocation decisions. Dispatch decisions should follow the most common rule -dispatch the nearest vehicle- or the optimal solution of a dispatching model. Relocation options were either doing nothing and following the optimal solution of a mathematical relocation model. Both models consider the fleet and the services heterogeneous, and look for to improve the system's preparedness level. The simulated system showed improvements in response times and preparedness levels of up to 54% when the dispatch follows the solution of the mathematical model and no relocation occurs.

Keywords
Ambulances; dispatching models; relocation models; preparedness; ambulance response times.

1 Introduction

Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes, and are consistent with current professional knowledge (Sanders, 2002). Emergency Medical Services (EMS) are part of these services. The dispatch and relocation of the ambulance -when appropriate-, in a given area of coverage are two of the key elements to reduce the waiting time of a potential patient and the initiation a health care service. During the operation of an EMS provider, frequent decisions about which vehicle to send to respond to a service call are taken. The dispatched vehicle becomes busy and the system's preparedness for a new call decreases. An action to restore the system's service level might be to relocate some of the idle EMS vehicles waiting at their standby sites. The impact of these decisions is directly received by the patient, which may affect their chances of survival (McLay & Mayorga, 2010); (Pons, et al., 2005)). From the patient's point of view, response times are by far the most important performance measure for an EMS provider, and it is one of the measures most studied in the literature ((Nogueira, Pinto, & Silva, 2016), (Wei Lam, et al., 2014), (Pons, et al., 2005)). However, different performance measures have recently been studied, such as the system preparedness concept (Andersson & Värbrand, 2007).

When a call is received, a vehicle dispatching decision must be made; even while the call is in progress. The default approach is to identify the nearest vehicle to the position where the call originates, with the objective of responding to the specific customer (patient) in the minimum time. In addition, idle vehicles are usually waiting at their predetermined standby sites and do not move unless necessary to respond a call. However, this decision rule is not necessarily the best for overall performance ((Schmid, 2012); (Zhen, Whang, Hu, & Chang, 2014)). In this paper, the hypothesis is that the use of a decision system that includes the sequential application of a dispatching mathematical model and a relocation model significantly improves the measures of performance of the system, such as preparedness level and response times.

2 Methods

2.1 Study Design and Setting

A simulation of four scenarios was performed. The trigger event that changes the system state is a call requesting service. Historical call data was provided in a peak period of an EMS provider in Colombia. The two dispatching options considered were: 1) the default rule: dispatching the vehicle nearest to the customer's request for assistance and 2) following the dispatch decision indicated when solving a dispatching mathematical model. With regard to relocation decisions, the options were: 1) do nothing, that is, not relocate idle vehicles in their current standby sites and 2) to follow the optimal relocation decision obtained by solving a mathematical model of relocation, which includes the option of "not relocate idle vehicles". Table 1 summarizes these options.

The two mathematical models selected have been constructed taking into account the specific conditions of the EMS provider. Both models aim to improve the system preparedness level and both are deterministic, so only one replication is taken. Details of the mathematical models can be found in ... Historical data consist of the arrival times of the calls requesting service. The run with the rules "dispatch the nearest vehicle" and "do nothing" corresponds to the current baseline.

Table 1. Design table

Decision rule	Value	Meaning
Dispatching rule	-1	Default: assign the nearest vehicle
	1	Mathematical model's decision
Relocation rule	-1	Default: do not relocate
	1	Mathematical model's decision

2.2 Study Protocol

Before describing the simulation process, Figure 1 shows the map of the processes that an EMS provider follows to service a call requesting.

Process Map

In order to contextualize operative decisions related to the fleet of an EMS provider, a process map is presented in Figure 1. When a call enters to the system, the operator triages it, determines the need to send a vehicle, and if necessary, gives basic indications to the patient (Schmidt, et al., 2000). The decision to choose and send a vehicle to assist the patient is called dispatch. When the vehicle reaches its destination, the service begins and can end either at the patient's site or after the patient is transferred to an emergency center. In the latter case, another decision must be made: the selection of the medical center where to take the patient. Other studies have discussed this problem (Pham, Patel, Millin, Kirsch, & Chanmugan, 2006). The context of our experiment does not require this aspect, since the facilities are predetermined. When a vehicle is dispatched, it becomes busy and this event changes the system preparedness (e.g. the system has one vehicle less available to service a new call). Then, a new decision has to be made on idle and available vehicles: that is, if they should be relocated, and where to. Finally, when the service is finished, a newly idle vehicle is available again and the system status changes again.

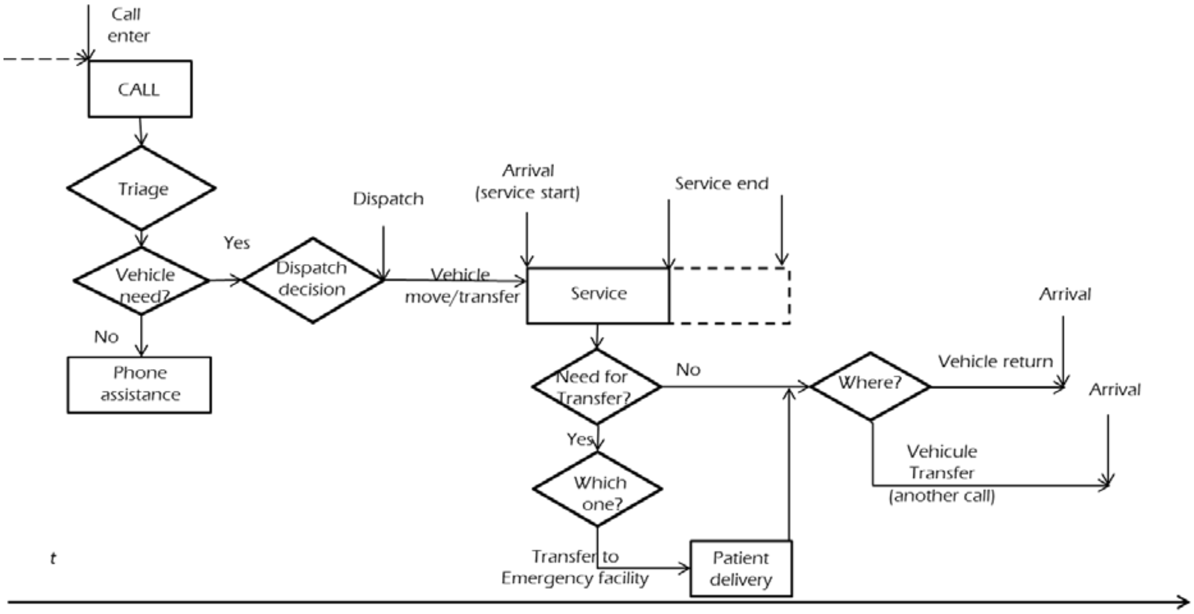

Figure 1. Process map of an EMS provider's operation

Simulation Process

Figure 2 depicts the protocol followed. The system status is initialized with parameters that define travel times, availability and location of the vehicle, allowed standby sites where vehicles wait for a service call, system preparedness, among others. Running time is set to zero. The scheduling and location of incoming calls are given, taken from an EMS provider's true historical data. So when the first call arrives, the system's status is updated as well as the runtime. This information is entered into a dispatch decision module. Outcomes of this process are: vehicle assigned (i.e. a newly busy one) and new preparedness values. Immediately after, the relocation decision module is used to decide if idle waiting vehicles are relocated among the possible standby sites, to restore preparedness values. Again, system parameters updating is carried out. If a new call enters the system before the last updating, it enters the queue of the dispatching module; otherwise it enters at the first system updating module.

The System Studied

The different types of services were classified according to their priority, in cases of emergency, urgency and consultation. A heterogeneous fleet (vehicles with different service assistance capabilities) was also identified. Statistical analysis was performed for three quality measures, for the whole system and specifically for the most critical calls (i.e. urgency), since no emergency calls were received during the period studied.

The system consists of a heterogeneous fleet of an EMS provider, with size | K |, a set of waiting sites I, in which a vehicle is allowed to wait for a service call, with | K | <| I |. The heterogeneous services are to be covered in a region divided into zones $j \in J$, where $Q_k \subseteq S$ is the set of services that are covered by vehicle k, and D_{js} is the historical weighted demand for service s of zone J. A vehicle is available when idling at a waiting site or moving to a known waiting site.

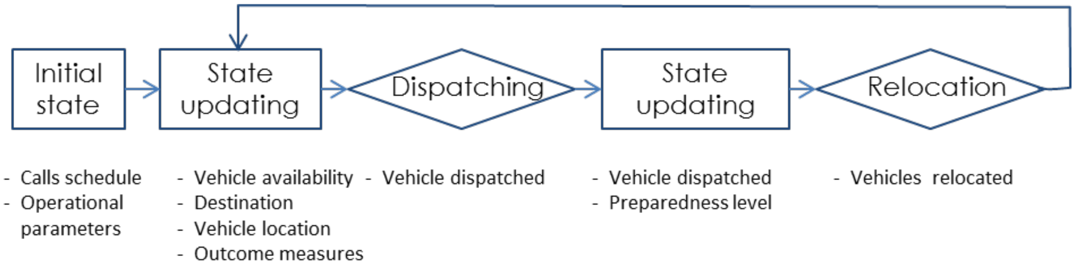

Figure 2. Simulation scheme

Optimization Models

Taking into account the limitations imposed by the operation of the EMS provider under study, vehicle dispatching is a particular case of the Vehicle Routing with Multiple Deposits and Time Windows problem. It was formulated using the Mixed Integer Linear Programming (MILP) model proposed in (Rengifo, Baldoquin, & Escobar, 2012). The relocation problem is also modeled as an MIP that considers moving idle vehicles between waiting sites to improve the minimum system preparedness per service type (or priority), subject to capacity, travel times and demand constraints as in (Herrera and Mosquera, 2016). Both optimization models 1) have a dynamic character, that is, they are deterministically solved, modifying the parameters and input data whenever there is a change in the state of the system, such as reception of a new call, failure of a vehicle, etc. 2) were developed in accordance with the EMS provider restrictions, and 3) optimize a customized system's preparedness measure.

Outcome Measures

At the end of each run the following indicators were calculated: 1) response time, defined as the length of the time interval between the start of a call and a vehicle arrival at the destination (i.e. the patient's location). 2) The occupation of the fleet, which measures the percentage of time the vehicles are occupied, either by answering a call or changing location; and 3) Preparedness level, general and per service. The preparedness level used in this work estimates the preparation of the system to respond to new calls. This extends the concept found in the literature (Andersson and Värbrand, 2007), (Zhen, Whang, Hu & Chang, 2014) to a preparation by type of service (or priority) as defined in equation (1), where β_{js} is the weighted demand for the service *s* of zone *j*, α_{ij} is a parameter that weighs travel times from a standby site *i* to a demand zone *j*, which considers the proximity of the site *i* to zone *j* with respect to the other standby sites. y_{is} can be defined as either the number of vehicles covering the service *s* at site *i* or as a binary variable indicating whether site *i* is covered with service *s* by some vehicle. In this work it is declared as binary.

$$P_{js} = \frac{1}{\beta_{js}} \sum\nolimits_{i \in I} \alpha_{ij} \cdot y_{is} \quad \forall j \in J, s \in S \qquad (1)$$

2.3 Run setting

To test the effect of dispatch and relocation decisions, we use the busiest hour of a working day from a private EMS provider in Colombia. During this period 21 calls were received, for which 21 runs were used as input to the model. The models were programmed on the AMPL ide® interface and solved with Gurobi 3.0®, using an HP Envy 4 Notebook, RAM: 4GB, an Intel® Core ™ i3-3217U processor and a 1.80GHz CPU.

The procedure for calculating the travel times between centroids of demand zones *j* and standby site locations *i* takes into account (Herrera & Mosquera, 2016): 1) the Cartesian coordinates of the different standby site *i* locations and the 21 points where the calls originate; 2) an experimental design, using distances given by the Google Maps ® tool, to estimate an adjustment factor when comparing these distances with the Euclidian ones for each pair of points; and 3) weighting factors for the calculation of speeds, considering peak and off periods, as well as type of service (i.e. speed to attend an emergency call may be different from urgency or consultation).

Travel speeds and standard times were considered, according the priority, as shown in table 2. Table 2 also shows the fleet size per ambulance type; an "X" indicates whether the vehicle can serve a specific service. The demand was weighted taking into account the number of calls received in the last year per zone and service.

Table 2. Input parameters

Service		Speed in peak hours [km/min]	Standard response time [min]	EMS Vehicles		
				I	II	III
1	Consultation	0.77	120	X		
2	Urgency	0.77	30	X	X	
3	Emergency	1.33	15	X	X	X
Fleet size				7	3	2

3 Results

After taking a random sample of 30 distances between centroids, an adjustment factor of 0.48 km with SD 0.26 was found. This value was used to correct the Euclidian distance between each pair of points.

Five performance indicators were analyzed, for the four possible combinations of the dispatch and relocation rules, according to Table 2. The indicators are: 1) RT: average response time, in seconds; 2) FO: occupation of the fleet (%) and 3) PL: Preparedness level. Since no emergency call was received during the study period, attention was paid to the most critical service: urgency. Therefore, the other two variables are: 4) PLu: Preparedness level for urgency services and 5) RTu: response time for urgency calls, in seconds. Table 3 shows the results obtained. Columns 1 and 2 show the corresponding dispatching and relocation rules according to Table 2. Thus, for example, the first row refers to the nearest vehicle dispatch rule without subsequently applying a relocation of idle vehicles. For each combination, average and standard deviation values are given. The last row in Table 3 shows the relative change when using the third scenario with respect the first (default) scenario.

Table 3. Results of the performance indicators analyzed

D	R		y_1: $\overline{\text{RT}}$	y_2: $\overline{\text{FO}}$	y_3: $\overline{\text{PL}}$	y_4: $\overline{\text{PLu}}$	y_5: $\overline{\text{RTu}}$
			[s]	[]	[]	[]	[s]
-1	-1	Ave	1,154.7	59%	0.432	0.4197	423.7
		SD	1,465	19%	0.348	1.844	105.0
-1	1	Ave	1,174.8	59%	0.468	0.4162	962.3
		SD	1,426	19%	0.322	1.298	285.5
1	**-1**	Ave	**526.9**	**44%**	**0.665**	**0.6008**	**233.7**
		SD	**489.8**	**44%**	**0.297**	**0.279**	**66.9**
1	1	Ave	897.7	49%	0.592	0.4362	742.0
		SD	923.5	34%	0.265	0.319	332.3
Change*		Ave	-54.4%	-25.4%	53.9%	43.1%	-44.8%
		SD	-66.6%	131.6%	-14.7%	-84.9%	-36.3%

D: Dispatching rule, R: Relocation rule, Ave: Average, SD: Standard Deviation, RT: Response time, FO: fleet occupancy percentage, PL: preparedness level, PLu: preparedness for urgency calls. RTu: response times for urgency calls. * Relative changes of row 3 with respect to row 1.

After applying the ANOVA technique, Table 4 shows the p_values of the partial test of each factor when the solution of both optimization models are used. P_values below the significance level α=0.05 are considered to be statistically significant. Table 5 also shows the corresponding adjusted coefficient of determination R_a^2.

Table 4. ANOVA results

Test: Decision rule		Response variables				
		y_1: RT	y_2: FO	y_3: PL	y_4: PLu	y_5: RTu
Partial: Dispatching	P_value	0.236	0.138	0.189	0.059	0.010
Partial: Relocation	P_value	0.466	0.530	0.792	0.870	0.004
Total: Both	R_a^2	66.3%	86.7%	76.6%	97.5%	99.9%

4 Discussion

Recalling that the period analysed was the busiest one of the year, Table 4 shows that the best scenario for all performance measures is the third one, which consists of dispatching an ambulance according the optimal solution indicated by the mathematical model without relocating idle ambulances afterwards. This makes sense in a peak activity period where most of the ambulances are busy and the time between call arrivals may be small enough to not allow idle vehicles to change their positions. This makes sense in a period of maximum activity where most ambulances are busy and the time between call arrivals may be so small that they do not allow inactive vehicles to change their positions.

Improvements on response times and preparedness levels reached 54%, by comparing this scenario with the default scenario (i.e. the nearest dispatching rule without relocation). Reductions on the standard deviation of these variables were obtained, with a greater reduction (85%) on the standard deviation of the preparedness level for urgency calls. On other hand, fleet occupancy was also improved (25% reduction) but the increase in the standard deviation reflects low balance of workload.

The first thing to be observed from the ANOVA results is that the decisions of dispatch have greater influence in the performance measures than the ones of relocation. Both types of decisions can explain 66.3% of response time variation and 76.6% of preparedness levels, and have an even more impact in the fleet occupancy (86.7%). It should be noted that the two main quality performance measures (i.e. response times and preparedness level) for the more critical cases, -the urgency calls-, are much more sensible to both dispatching and relocation decisions. The variability of preparedness level for urgency calls, can be estimated with a R_a^2= 0.987, and the response times with a R_a^2= 0.999, which are very high values. In contrast, response times are only explained 66.3% with both factors. This value makes sense, since both decisions are taken at the beginning of the EMS vehicle’s travel, and the response time also greatly depends on the distance to the destination.

4.1 Limitations

Only four complete runs were simulated in the same horizon, which was the most congested one in observed the period. Taking more replicates makes no sense, since the models are deterministic. The results shown here are only valid to the EMS provider under study, and for the two mathematical models used, which were taken from the literature and they were adapted to the operating conditions of that EMS provider. However, general conclusions are a call for EMS provider companies to consider the use of operations research models to improve their quality performance. During the simulation, the probability of vehicle breaking or maintenance was not included. During the selected horizon, no emergency calls were received.

5 Conclusions

The performance measures of the most critical calls -the urgency ones-, were found to be significantly affected by both, the dispatching and relocation methods. In particular, the preparedness of the system and the response times for to handle urgency calls were the most sensitive performance measures. The scenario that reports a better performance for a period of peak activity of the EMS provider under study was to dispatch an ambulance according the optimal solution indicated by the mathematical model without relocating idle vehicles, because the times between call arrivals are very short. Then, an EMS provider should consider alternatives to the most common rule of assigning the vehicle closest to a call. It will be necessary to study other scenarios, with less demand, to verify if the use of the mathematical model to relocate can influence favorably the average times of response and the preparation of the system. It was found that the use of optimization models, such as those used in this work to dispatch and relocate EMS vehicles, can improve not only the average response time but also the system preparedness to handle new calls, especially when major services are requested priority.

Acknowledges

Authors thank to Sebastián Herrera and Cristian Mosquera for their participation in the modeling and numerical experimentation.

6 References

Andersson, T., & Värbrand, P. 'Decision support tools for ambulance dispatch and relocation'. *Journal of the Operational Research Society, 58*, 195-201, 2007

Herrera, J., & Mosquera, C. *'Modelo matemático para la solución de un problema de relocalización de vehículos de emergencia médica en una empresa de SEM en la ciudad de Cali'.* Thesis, Pontificia Universidad Javeriana, Engineering School, Cali, Colombia, 2016.

McLay, L., & Mayorga, M. 'Evaluating emergency medical service performance measures'. *Health Care Management Science, 13*(2), 124-136, 2010

Nogueira, L., Pinto, L., & Silva, P. Reducing Emergency Medical Service response time via the reallocation of ambulance bases. *Health Care Management Science, 19*(1), 31-42, 2016

Pham, J., Patel, R., Millin, M., Kirsch, T., & Chanmugan, A. 'The Effects of Ambulance Diversion: A Comprehensive Review'. *Academic Emergency Medicine, 13*(11), 1220–1227, 2006

Pons, P., Haukoos, J., Bludworth, W., Cribley, T., Ponss, K., & Marovchick, V. 'Paramedic Response Time: Does it affect patient survival?'. *Academic Emergency Medicine, 12*(7), 594-600, 2005

Rengifo, A., Baldoquin, M., & Escobar, J. 'Diseño de un modelo matemático para el despacho de vehículos de emergencias médicas en Colombia, ST - Logistics and Transport'. *Annals XVI CLAIO September 24-28,*, (pp. 1880-1891). Rio de Janeiro, Brazil, 2012

Sanders, A. 'Quality in Emergency Medicine: An Introduction'. *Academic Emergency Medicine, 9*(11), 1064–1066, 2002

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

Schmidt, T., Atcheson, R., Federiuk, C., Clay Mann, N., Pinney, T., Fuller, D., & Colbry, K. 'Evaluation of Protocols Allowing Emergency Medical Technicians to Determine Need for Treatment and Transport'. *Academic Emergency Medicine*, *7*(6), 663-669, 2000

Wei Lam, S. S., Zhang , Z. C., Oh , H. C., Ng, Y. Y., Wah, W., Hock Ong, M. E., & Cardiac Arrest Resuscitation Epidemiology (CARE) S. Reducing Ambulance Response Times Using Discrete Event Simulation. *Prehospital Emergency Care, 18*, 207-216, 2014

Zhen, L., Whang, K., Hu, H., & Chang, D. (2014). 'A simulation optimization framework for ambulance deployment and relocation problems'. *Computers and Industrial Engineering, 72*, 12-23, 2014

7 Biography

Jenny Díaz-Ramírez is professor at the University of Monterrey. She has worked previously as a professor at Tecnologico de Monterrey, Mexico and Pontificia Universidad Javeriana Cali, Colombia. She got a MSc in operations research from Georgia Tech and the PhD in Industrial Engineering from Tecnologico de Monterrey, Campus Toluca in 2007. Her research areas are optimization and statistics applied in topics such as health systems, air quality and logistics.

María Gulnara Baldoquin de la Peña is retired Professor of the Department of Mathematical Sciences, School of Sciences, EAFIT University. Prior to this, she worked as Professor in the Department of Civil and Industrial Engineering, Faculty of Engineering, Pontificia Universidad Javeriana, Cali (2011-2015) and as Professor in the Department of Mathematics, Industrial Engineering Faculty, Technical University of Havana (ISPJAE), Cuba (1978-2010). She obtained the Bachelor degree in Mathematics, Speciality Operations Research, Havana University, Cuba (1978) and the PhD in Technical Sciences in Technical University of Havana, Cuba (1995). Her research works include Development of mathematical models, particularly in Combinatorial Optimization and Mixed Integer Linear Programming Models; Heuristic and metaheuristics algorithms, to solve complex models arising in Operations Research, as in Hospital Logistics.