Decreasing Ambulance Response Time Through an Optimal Base Location

Bernardo Villarreal, Edgar-Marco-Aurelio Granda-Gutierrez, Samantha Lankenau
Ana-Cristina Bastidas & Andrea Montalvo
Universidad de Monterrey
San Pedro Garza García, N.L., México 66238

Abstract

An important effort of academic research has been realized to improve ambulance response time (the time between the receipt of a call at the dispatch center and the arrival of the first emergency response vehicle at the scene). This study recommends the application of Lean Transportation concepts to achieve shorter ambulance response times. The approach considers the reduction of the time required to travel distance in excess waste by placing ambulances in optimal locations through mathematical modeling. In this project this is applied to the Monterrey metropolitan operations of the Mexican Red Cross. Results of the application are provided.

Keywords:

Emergency medical service, waste elimination; platinum ten; golden hour; transportation waste.

1. Introduction

As described by Glushak et al., (1997), the principal function of Emergency Medical Service (EMS) systems is to provide patients with urgently needed emergency medical care and to transport them to an emergency medical facility, if needed. Mexican EMS institutions started emerged formally at the beginning of the last century. According to Fraga-Sastrias et al., (2010) about a third of the institutions are non-governmental organizations financed through donations. One of the main private institutions that provide emergency medical services is the Mexican Red Cross which was founded in 1910. Also, as stated in Pinet-Peralta (2005), these institutions are not strictly supervised in terms of coordination, regulation and performance evaluation.

The efficiency of EMS systems is a major public concern worldwide. Its performance is highly related to health issues of life or death, in particular, with patients with strokes, severe trauma and cardiac arrest cases. This study recommends the application of the Lean Transportation approach (Villarreal 2012; Simmons et al., 2004; Sternberg et al., 2013) to achieve shorter ambulance response times. In this project this is applied for the improvement of ambulance response time for the Monterrey metropolitan operations of the Mexican Red Cross. The scheme suggested by Lankenau et al., (2016) to eliminate wastes related mainly to the ambulance performance efficiency is used to improve response time by placing ambulances in optimal locations through mathematical modeling. The organization´s operations in the Monterrey metropolitan area felt that the current average response time estimated in 19.9 minutes needed to be reduced significantly to satisfy international standards.

This paper consists of five sections. Sections one gives a brief review of the literature on lean healthcare and transportation, and a description of the scheme utilized to decrease waste. The application of this scheme is undertaken in section 3 and section 4 presents conclusions and recommendations.

2. Review of Lean Transportation and Healthcare Literature

Lean thinking has proven to be an effective way of improving healthcare organizations. The increasing number of implementations reported in the literature supports this view. An initial consideration about the use of lean concepts is offered by Heinbuch (1995), in the particular case of just-in-time. Some later publications that describe further applications are (Chalice 2007; Zidel 2006; Brandao 2009). Similarly to the case for the manufacturing and administrative processes, it is necessary to define the concept of waste or value for healthcare organizations. Two waste frameworks are provided by Bentley et al., (2008) and Graban (2016). The first suggest a framework of three types of waste; administrative; operational; and clinical. The second framework proposes an extension of the seven Toyota manufacturing wastes to the healthcare area. Both waste classifications are more suitable for process improvement inside the health institution.

2.1 Brief description of EMS operations

As previously stated, Emergency Medical Service (EMS) systems are to provide urgent medical care, such as pre-hospital care, and to transport the patient to the hospital if required. The activities involved to provide Emergency Medical Services are: Receive emergency call and ambulance assignation; ambulance preparation; transporting the ambulance to the customer site; serving the injured or sick patient until stabilized; transfer the patient to a health institution; delivering the patient to the health institution and; transportation back to ambulance base. The activities previously described are part of the ambulance cycle response. Transportation is an inherent element of the emergency medical service. Our main interest in this work is the time taken for executing the first three activities. This time must be reduced to comply with international standards. As a result, the common method to evaluate the EMS system is to measure coverage level that reflects the proportion of patients who experience a response time less than a given time standard. As stated by Pons et al., (2005), paramedic response time to the scene of a call for emergency medical assistance has become a benchmark measure of the quality of the service provided by emergency medical services (EMS) agencies. A suggested target response time of ≤ 10 minutes for at least 90% of emergent responses has evolved into a guideline that has been incorporated into operating standards for many EMS providers.

2.2 A lean EMS process

The improvement of the time taken to execute the pre-hospital process described earlier has been approached mainly by utilizing a re-location of facilities and identifying shortest routes from the ambulance sites to the site of interest (Brotcorne et al., (2003). According to Simmons et al., (2004), eliminating unnecessary transportation can also be achieved increasing transport efficiency. Therefore, the reduction of the ambulance response time can also be achieved eliminating transportation waste. Transport efficiency was originally suggested by Simmons et al., (2004) and proposed the measure called the Overall Vehicle Effectiveness (OVE). The previous measure has also been modified by Villarreal (2012). In this case, the OVE measure is adapted to consider total calendar time. The proposed measure is called Total Operational Vehicle Effectiveness and it is represented by the term TOVE. In summary, four efficiency factors for TOVE are suggested; Administrative or strategic availability, operating availability, performance and quality. The new measure would be obtained from the product of administrative availability, operating availability, performance and quality efficiency factors.

The TOVE index and related wastes are adapted to the EMS operations by Lankenau et al., (2016). The resulting index is called the Emergency TOVE and will be represented by E-TOVE hereafter. For this work, the goal will be to identify the wastes that impact the ambulance response time. Consider only the first three activities of the ambulance cycle time; Receiving an emergency call and assigning an ambulance; preparing the ambulance and crew for the service; and transporting the ambulance to the scene of the incident. Focusing on the time required to execute the previous activities, the wastes associated with it are those that determine the operating availability and performance efficiencies. The wastes considered in the performance efficiency are; speed loss; fill loss; and distance traveled in excess. In particular, distance traveled in excess is a result of a deficient ambulance transport planning; wrong ambulance site definition; deficient route planning; and inadequate ambulance assignment and dispatching policies.

On the other hand, the wastes considered in the operating availability efficiency factor are those resulting from the execution of non-value activities. Waiting and execution times in excess are the most common. These could happen before the ambulance departs to the point where it is required according to the call.

2.3 Description of the Waste Reduction Scheme

This work considers the scheme provided by Villarreal (2012) to guide waste elimination to generate projects for improving ambulance response time. The scheme consists of four general stages: The first stage begins with the mapping of the EMS services. The Value Stream Map (VSM) utilized in this work is a modified version of the one provided by Villarreal (2012) and will be denoted as the Ambulance VSM hereafter. This VSM is defined following and mapping the Ambulance flow. The waste identification phase is the second stage. This phase should be exhaustive to set a strong foundation for defining an effective strategy for waste elimination. The third phase consists on the determination of waste elimination strategies. Efficiency waste reduction strategies as well as the strategies for different ambulance dispatching rules, facility relocation and transportation mode change could be used to eliminate waste. Finally, the implementation stage is suggested to assess and select the improvement initiatives.

3.   Implementation and Results

This work applies the previous scheme to improve the level response time of the organization´s operations in the Monterrey metropolitan area. The operations count with ten fixed locations and seven mobile locations from which ambulances are sent

to service pre-hospital events. The organization has 34 ambulances but the financial resources allow only to operate 50% of them during any day. The number of services carried out in year 2015 was 35,400.

3.1. Mapping the ambulance cycle process and identifying relevant wastes

The first step of the methodology is the mapping of the operations. In this case, an Ambulance VSM for the cycle process of interest is elaborated. Figure 1 presents a shorter version of the Ambulance VSM of the Monterrey metro area operations. According to the Ambulance VSM, the average ambulance response time is estimated in 19.9 minutes, about 90% greater than the international standard. Under this service level, 16.4% of the emergency services do not satisfy the international standards. Therefore, given these results, the institution considered that it has a great challenge for improvement. From Figure 1, the average time observed before the ambulance departs to service an emergency call is 6.4 minutes. Therefore, the required international standard of 10 minutes is practically reached without the ambulance being used at all. The level of the average time taken to go from the ambulance base to the point where the service is requested is estimated in 13.6 minutes, which together with the time spent to dispatch the ambulance of 6.4 minutes, make the ambulance response time very undesirable.
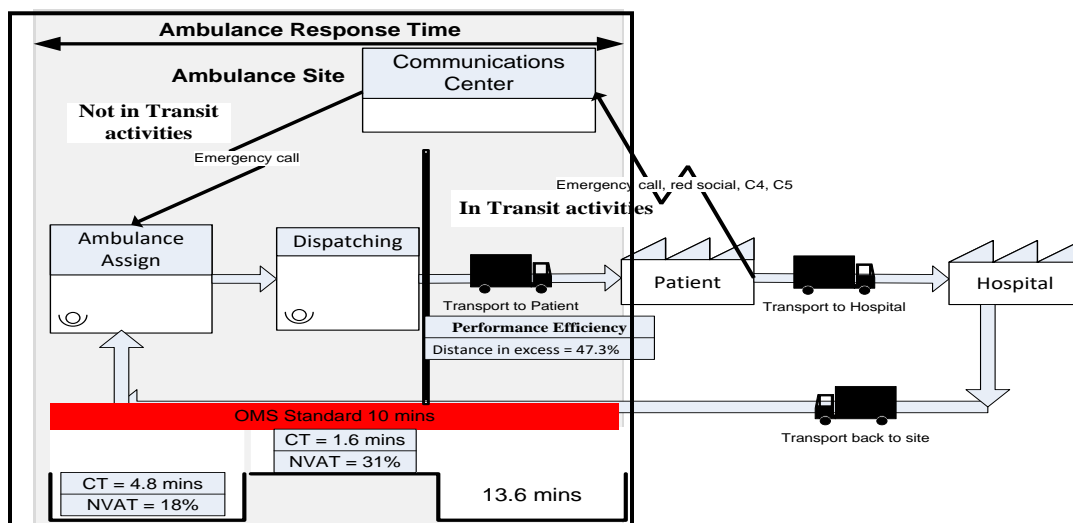

Figure 1 Description of the Ambulance VSM for the Monterrey Metro operations (Adapted from Lankenau et al., 2016)

3.2 Improving ambulance response time

After reviewing and analyzing a significant sample of emergency calls it was found that the excessive response time is due to the following causes; the deficient ambulance preparation procedures executed prior to dispatching them to service the call; the definition of the transport route from depot to the destination based on the experience of the drivers; and the location of the ambulance depots has not been updated in the last five years at least. The resulting distance travelled estimated exceeds 47.3% the optimal level. This work is focused on the improvement of response time by relocating ambulance bases.

The location of ambulances depends upon the behavior of service density and its dynamics throughout the day and the ambulance desired response time. Both conditions can be identified using hotspot analysis (Choudhary et al., 2015). Hotspot analysis uses past incident or event data to predict future patterns of events such as crimes, accidents or emergency calls. Hotspot identification can be done with several techniques as suggested by (Choudhary et al., 2015; Anderson 2013). In our case the analysis is realized using Getis-Ord Gi* statistics (Ord et al., 1995). Daily service demand requirements behavior is illustrated in Figure 2 (Lankenau et al., 2016). Two different patterns are identified; a low demand level occurring during the night shift from the 23:01 hrs of a day to 7:59 A.M. of the following day and; a high demand level occurring the rest of the following day. Therefore, from the demand behavior of emergency calls, two daily ambulance deployment strategies were developed for each day; one high-demand and one low-demand strategies.

Additionally, the location of areas with the highest incidence of emergency calls can also be identified. As shown in Figure 2, Emergency calls are concentrated on specific places of the Monterrey metropolitan suburban area. The most concentrated places include Monterrey downtown and a corridor that extends towards the northwest of the city close to the Garcia city. The other areas with concentrations of emergency services are Santa Catarina city, the frontier between Escobedo and San

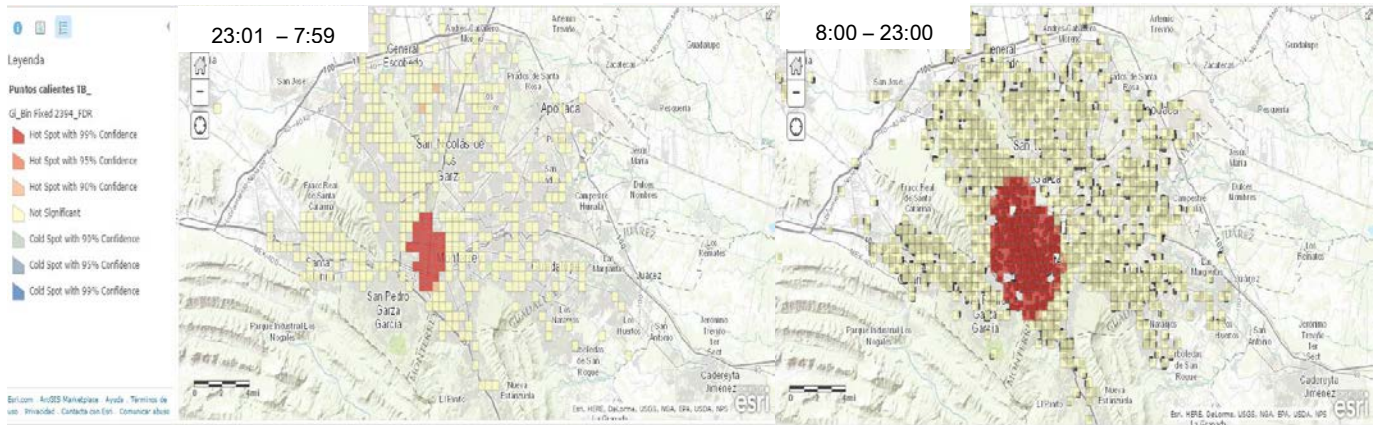Nicolas cities, downtown Guadalupe city and Apodaca city.

Figure 2 Hot spot analysis of emergency calls (adapted from Lankenau et al., 2016)

An excellent review of Operations Research contributions for optimizing patient flow in emergency operations is provided by Saghafiana et al., (2015). The authors identify ambulance location as an important issue for the improvement of flow into part of the operations.

The ambulance location problem has been studied for several years now, with a variety of solution methods and approaches proposed over the years (Brotcorne et al., 2003). Solution strategies to these problems have varied from mathematical programming techniques, such as Integer Linear Programming, to meta-heuristic approaches like genetic algorithms, recently reviewed in Li et al., (2011). In particular, one of the most successful approaches has been the Double Standard Model (DSM) of Gendreau et al., (1997) and illustrated in Figure 3. This model has been used to optimize the location of ambulance services in Belgium (Thirion, 2006), Canada (Gendreau et al., 1997) and Austria (Doerner et al., 2005), proving to be one of the most widely accepted and used models of the ambulance location problem. The same model was also used by Dibene et al., (2017) to optimize the location of ambulances in Tijuana, Mexico.

The main objective of standard DSM is to maximize the demand covered at least twice within a given time standard $r_1 > 0$ with the constraints that: All demand be covered at least once within a second time standard $r_2 > r_1$; A fraction $\alpha \in [0; 1]$ of the demand must be covered within r_1; All this must be accomplished using a total of $p > 0$ ambulances, and; At most $0 \leq p_j \leq p$ ambulances at site $j \in W$. The set of sites that cover point i within r_1 is $W_i^1 = \{j \in W : t_{ij} \leq r_1\}$ and within r_2 is $W_i^2 = \{j \in W : t_{ij} \leq r_2\}$. Figure 3 shows the structure of the DSM.

$$\text{Maximize} \quad \sum_{i \in V} d_i y_i^2 \tag{1}$$

Subject to

$$\sum_{j \in W_i^2} x_j \geq 1 \ (i \in V) \tag{2}$$

$$\sum_{i \in V} d_i y_i^1 \geq \alpha \sum_{i \in V} d_i \tag{3}$$

$$\sum_{j \in W_i^1} x_j \geq y_i^1 + y_i^2 \ (i \in V) \tag{4}$$

$$y_i^2 \leq y_i^1 \ (i \in V) \tag{5}$$

$$\sum_{j \in W} x_j = p \tag{6}$$

$$x_j \leq p_j \ (j \in W) \tag{7}$$

$$y_i^1, y_i^2 \in \{0,1\} \ (i \in V)$$

$$x_j \ integer \ (j \in W)$$

Figure 3 DSM model structure

where d_i is the weight of demand point i, x_j is the number of ambulances at site j and the binary variable y_{ki} is equal to 1 if and only if demand point i is covered at least k times within r_1. The objective function (1) computes the sum of the weights of the demand points covered at least twice within r_1. Constraint (2) enforces that all demand is covered within r_2 and (3) ensures that a fraction α of all demand is covered within r_1. Constraint (4) states that at least two ambulances are required for double coverage while (5) asserts that a demand point cannot be covered twice if it is not covered at least once. Constraints (6) and (7) ensure that p ambulances are used and at most p_j ambulances are located at site j respectively.

The demand points considered for the study are those obtained during the period of January to September of 2016. We selected a total of 1125 potential location sites in the metropolitan area of Monterrey that were located according to the geographical limits delineated by the hotspots identified and illustrated in Figure 2. Shopping malls, schools, government offices and parking areas are selected as potential sites including the locations corresponding to those currently used as bases by the company. The geographic coordinates of these points are obtained from the Google Places API (Google 2014). The solutions obtained from the model are computed using MOSEK as the integer program solver called from MATLAB MOSEK ApS (Mosek 2016). In order to optimize the location of ambulances for our case, we modified slightly the standards set by the United States EMS Act (Ball et al., 1993). The value prescribed for r_1 of 10 minutes is unchanged. The value for α of 0.95 is reduced to 0.90. The value for r_2 was set at 25 minutes. Furthermore, we set the number of ambulances, p, according to the scenario required to analyze.

As identified previously with the demand analysis, two scenarios were evaluated. The optimal number of ambulances required for achieving an average transport time of less than 10 minutes is 23 for the low demand option and 29 for the high demand alternative. These new ambulance requirements represent increases from the initial operating status of 77% for the low demand shift and 71% for the rest of the day. Figure 4 illustrates the location and number of ambulances for both scenarios.
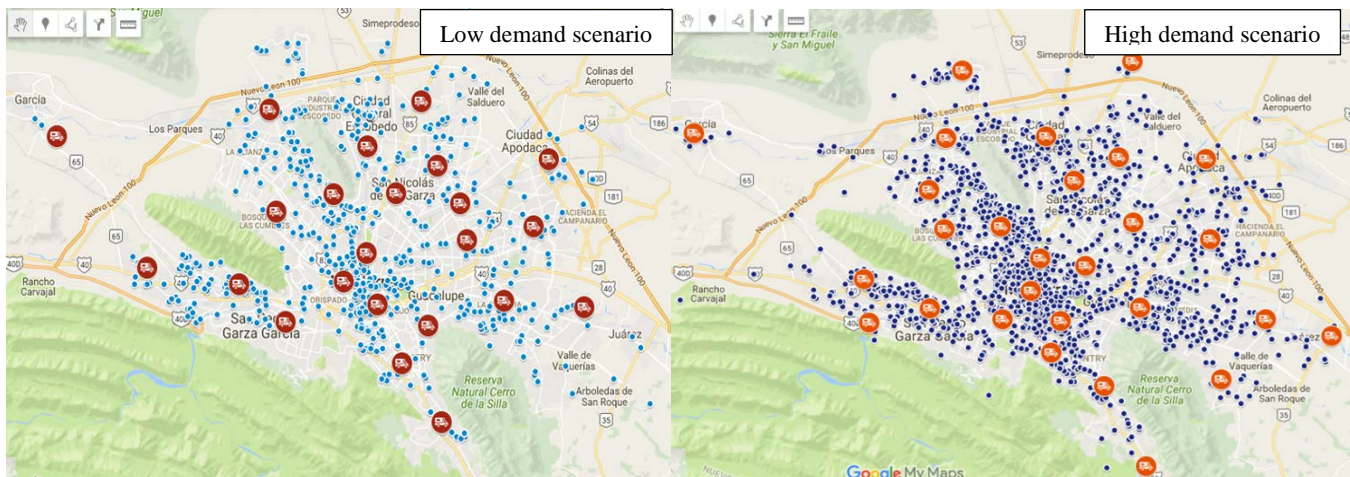

Figure 4 Summary of location and number of ambulances for low and high demand scenarios

3.3. Description of Implementation Results
Previously to carrying out full implementation of the selected strategy, the operations management of the institution decided to carry out a modified pilot program. The amount of ambulances in the pilot project was constrained by the shortage of well-trained crew members available at that moment and budgetary constraints. The management decided to increase the number of ambulances during the high demand shifts from 17 to 21. This implies an increase of 23.5% in the number of ambulances. In order to support this pilot program, it was necessary to determine new optimal locations generated by applying the mathematical model for $p = 21$ ambulances during the high demand shifts. This program was setup for running during November and December of year 2016.

After a month of operating the new base location structure, the average transport time from the ambulance bases to the patient decreased 17.6% from 13.6 minutes to 11.2 minutes. It is expected that, after implementing the optimal ambulance solution, the average transport time will be further reduced to a level lower than 10 minutes.

4.   Conclusions and recommendations

As previously mentioned, an important issue in current EMS operations concerns the improvement of ambulance response time. This study recommends the application of Lean Transportation concepts to achieve better ambulance response times. The approach considers the reduction of the time required to travel distance in excess waste by placing ambulances in optimal locations through mathematical modeling. In particular, the DSM model is applied to the Monterrey metropolitan operations of the Mexican Red Cross. The results of the model imply an increase in the number of required ambulances. In addition, it also suggests new locations for them.

After reviewing the recommendations, the management of the institution decided to implement a pilot program considering the limited resources available at the end of year 2016. The idea behind this move was to have further results to justify an increase in the operations budget for improving response time during year 2017. Further work related to this study includes the determination of the number and location of ambulances during holidays and summer. The same scheme is also considered to be used as an aid for strategic medium range planning.

Acknowledgements

References

Anderson, T., Review of Current Practices in Recording Road Traffic Incident Data: with Specific Reference to Spatial Analysis and Road Policing Policy, UCL Centre for Advanced Spatial Analysis, University College London 1 - 19 Torrington Place Gower St London WC1E 7HB.

Ball, M., Lin, F., (1993), A reliability model applied to emergency service vehicle location, Operations Research, Vol. 41, pp. 18–36.

Bentley, T.G.K., Effros, R.M., Palar, K. and Keeler, E.B., (2008), Waste in the U.S. Health Care System: A Conceptual Framework, The Milbank Quarterly, Vol. 86, No. 4, pp. 629 – 659.

Brandao, D.S.L. (2009) 'Trends and approaches in lean healthcare', *Leadership in Health Services*,Vol. 22, No. 2, pp.121–139.

Brotcorne, L., Laporte, G. & Semet, F., (2003), Ambulance Location and Relocation Models, European Journal of Operational Research, Vol. 147, pp. 451 – 463.

Chalice, R. (2007) *Improving Healthcare Quality using Toyota lean Production Methods: 46 Steps for Improvement*, 2nd ed., Quality Press, Milwaukee, WI.

Choudhary, J., Anurag Ohri, A. and, Brind Kumar, B., (2015) , Spatial and statistical analysis of road accidents hot spots using GIS, 3rd Conference of Transportation Research Group of India (3rd CTRG), Kolkata, India.

Dibene, J.C., Maldonado, Y., Vera, C. Oliveira, M. Trujillo, L. & Schutze, O., (2017), Optimizing the location of ambulances in Tijuana, Mexico, Computers in Biology and Medicine, Vol 80, pp. 107 – 113.

Doerner, K.F., Gutjahr, W.J., Hartl, R.F., Karall, M., Reimann, M., (2005), Heuristic solution of an extended double-coverage ambulance location problemfor austria.CEJOR, Vol. 13, pp. 325–340.

Fraga-Sastrías, J.M., Aguilera-Campos, A.A. and Lafuente, E.A., (2010), Motivos de Llamada a los Servicios Médicos de Emergencia en México: Definiendo Prioridades, Archivos de Medicina de Urgencia de México, Vol. 2, No. 2, pp. 60 – 67.

Gendreau, M., Laporte, G. and Semet, F., (1997), Solving an Ambulance Location Model by Tabu Search, Location Science, Vol. 5, No. 2, pp. 75 – 88.

Glushak, C., Delbridge, T.R., Garrison, H.G., (1997), Ambulance Diversion, Prehospital Emergency Care, Vol. 1 No. 2, pp. 100 – 103.

Google (2014). Google places API. https://developers.google.com/places/. (2014. 11. 30).

Graban, M., (2016), Lean Hospitals: Improving Quality, Patient Safety and Employee Engagement, 3rd Edition, CRC Press.

Heinbuch, S.E. (1995) 'A case of successful technology transfer to health care: total quality materials managements and just-in-time', *Journal of Management in Medicine*, Vol. 9 No. 2, pp.48–56.

Lankenau, S., Montalvo, A. & Bastidas A.C., (2016), Reducción del Tiempo de Respuesta de Ambulancias de Emergencias Médicas de Cruz Roja, Tesis de Profesional, Universidad de Monterrey, Departamento de Ingeniería.

Li, X., Zhao, Z., Zhu, X., Wyatt, T., (2011), Covering models and optimization techniques for emergency response facility location and planning: a review. Math. Meth. Oper. Res., Vol. 74, pp. 281–310.

MOSEK ApS (2015). The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).

Ord, J.K. and Getis, A., (1995), Local Spatial Autocorrelation Statistics: Distributional Issues and an Application, Geographical Analysis, Vol. 27, No. 4.

Pinet-Peralta, L.M., (2005), The Prehospital Emergency Care System in Mexico City: A Systems Performance Evaluation, Prehospital and Disaster Medicine, Vol. 21, No. 2, pp.104 – 111.

Pons, P.T., Jason S. Haukoos, J.S., MS, Whitney Bludworth, M.S., Thomas Cribley, T., A. Pons, A. & Markovchick, V, (2005), Paramedic Response Time: Does It Affect Patient Survival? ACAD EMERG MED d July 2005, Vol. 12, No. 7

Saghafiana, S., Austina, G., Stephen J. & Traubb, SJ., (2015), Operations research/management contributions to emergency department patient flow optimization: Review and research prospects, IIE Transactions on Healthcare Systems Engineering, Vol. 5, No.2, pp.101-123.

Simmons, D., Mason, R. and Gardner, B., 2004, "Overall Vehicle Effectiveness", International journal of Logistics: Research and Applications, Vol. 7, No. 2, pp. 119-34.

Sternberg, H., Stefansson, G., Westernberg, E., Boije af Gennas, R., Allenstrom, E., Nauska, M.L. (2013), "Applying a Lean Approach to Identify Waste in Motor Carrier Operations", *International Journal of Productivity and Performance Management*, Vol. 62 No. 1, 2013, pp. 47-65.

Thirion, A., (2006), Modeles de localisation et de reallocation d'ambulances. Application aux communes en provinces de Namur et Brabant Wallon, PhD thesis, Facultes Universitaires Notre-Dame de la Paix, Namur, Belgium.

Villarreal, B. (2012), "The Transportation Value Stream Map (TVSM)", *European Journal of Industrial Engineering*, Vol. 6, No. 2, pp. 216-233.

Zidel, T. (2006) *A lean Guide to Transforming Healthcare: How to Implement Lean Principles in Hospitals, Medical Offices, Clinics, and other Healthcare Organizations*, Quality Press, Milwaukee, WI.

Biographies

Bernardo Villarreal is a full professor of the Department of Engineering of the Universidad de Monterrey. He holds a PhD and an MSc of Industrial Engineering from SUNY at Buffalo. He has 20 years of professional experience in strategic planning in several Mexican companies. He has taught for 20 years courses on industrial engineering and logistics in the Universidad de Monterrey, ITESM and Universidad Autónoma de Nuevo León. He has made several publications in journals such as Mathematical Programming, JOTA, JMMA, European Journal of Industrial Engineering, International Journal of Industrial Engineering, Production Planning and Control, International Journal of Logistics Research and Applications, Industrial Management and Data Systems and the Transportation Journal. He is currently a member of the IIE, INFORMS, POMS and the Council of Logistics Management.

Edgar Aurelio Marco Granda is a full professor of the Department of Engineering of the Universidad de Monterrey. He holds a PhD of Industrial Engineering from ITESM. He has 18 years of professional experience in logistics, operations and supply chain in several Mexican companies. He has taught for 5 years courses on industrial engineering and logistics in the Universidad de Monterrey, ITESM, UMIN and Universidad Autónoma de Nuevo León. As a consultant, he has carried out projects on logistics and supply chain for different company in México.

Samantha Lankenau Delgado is a CUM LAUDE Industrial Engineer graduated from Universidad de Monterrey (UDEM). She is currently a graduate student at the Master Degree program in Supply Chain Management. Her specialty is strategic planning and the operations and logistics improvement. She has participated on several projects such as The Redesign of the Supply Process of Drugs on a Medical Center and the Improvement of the routing operations of a soft drink bottling firm. Nowadays, she works at FEMSA S.A. de C.V., developing operations strategies for improving quality and productivity. Samantha is a member of the IISE, ASQ and APICS Societies.

Andrea Montalvo is a CUM LAUDE Industrial Engineer just graduated from Universidad de Monterrey (UDEM). She has participated on several projects such as the Improvement of the routing operations of a leading convenience store firm. She also applied Lean Thinking principles for Improving the Productivity of several metal assembly lines for a Mexican metal mechanic company. Currently, she has started to work at a Mexican firm leader in the manufacturing of frozen and refrigerated food as a transportation and traffic analyst. Andrea is a member of the IIE and ASQ Societies.

Ana Cristina Bastidas Lopez is a CUM LAUDE Industrial Engineer just graduated from Universidad de Monterrey (UDEM). She has participated on several projects such as the Improvement of the routing operations of a leading

convenience store firm. She also applied Lean Thinking principles for Improving the Productivity of several metal assembly lines for a Mexican metal mechanic company. Currently, she has started to work at a Mexican firm leader in the manufacturing of frozen and refrigerated food as a transportation and traffic analyst. Ana Cristina is a member of the IIE and ASQ Societies.