

**MATHEMATICAL MODELING AND NUMERICAL ANALYSIS FOR
OPTIMIZING AMBULANCE ALLOCATION IN EMERGENCY SERVICES**

Sajedeh Norozpour

Faculty of Architecture and Engineering, Istanbul Gelisim University, Istanbul, Turkey

snorozpour@gelisim.edu.tr

Abstract

This study explores how to optimize the locations of ambulances to enhance overall coverage across a system. It employs a mathematical model that focuses on response time probabilities, which are crucial for effective emergency services. The response time is composed of two main components: pre-travel delays and the actual travel time. These elements are modeled using both normal and exponential distributions to capture the variability in response times accurately. To analyze the expected fraction of emergency calls that can be reached within a specified time frame, a differential equation is utilized. This equation is solved through the finite difference method, which allows for a detailed examination of response times. Moreover, an iterative numerical approach is implemented to maximize the expected coverage, denoted as $E(C)$. This approach takes into account how ambulances are allocated across different districts, ensuring that resources are used efficiently. The results of the study reveal significant trends in coverage performance as ambulance allocations change. These findings underscore the model's effectiveness in informing decisions about resource distribution. Overall, this study offers a thorough framework for assessing and optimizing emergency response systems, providing valuable insights for improving public health and safety.

Keywords: Ambulance location optimization, Emergency medical services (EMS), Numerical optimization, System-wide coverage.

Literature Review

Ambulance location and allocation optimization has been a widely researched topic in the fields of operations research and public health. A key foundational work in this area is the monograph on urban operations research, which offers a thorough framework for modeling and addressing urban service systems, including emergency medical services (EMS) [1]. This work introduced essential concepts for optimizing service facility locations in urban environments, forming the basis for many subsequent studies.

The formalization of emergency service facility locations as an optimization problem began in the early 1970s. Toregas et al. proposed a model aimed at determining the optimal placement of facilities to minimize the maximum distance or response time to service demand points [2]. This groundbreaking study established the foundation for covering location models, which seek to ensure that a significant portion of the population is within a

specified response time or distance of emergency services. Daskin later expanded on these ideas by introducing the Maximum Expected Covering Location Problem (MEXCLP), which incorporates stochastic demand and service availability, offering a more realistic representation of EMS systems [3].

In the years that followed, researchers delved into additional complexities within ambulance location models. Goldberg and Paz investigated how response times are influenced by call locations and proposed models that explicitly consider spatial variability in service times [4]. This research underscored the necessity of integrating geographic and operational factors into ambulance deployment strategies. Brotcorne, Laporte, and Semet conducted a comprehensive review of ambulance location and relocation models, categorizing them into static and dynamic approaches while identifying key challenges and opportunities for future research [5]. Their work continues to serve as a vital reference for understanding the evolution of EMS optimization models.

Recent developments in this field have concentrated on real-time ambulance redeployment and dynamic allocation. Jagtenberg et al. created an efficient heuristic for real-time ambulance redeployment, demonstrating its effectiveness in enhancing response times under dynamic demand conditions [6]. These studies highlight the significance of adaptability in EMS systems, as static allocation models may not adequately address the unpredictable nature of emergency incidents.

Beyond optimizing ambulance locations, considerable attention has been directed toward understanding the factors that influence EMS utilization and response. Honigman et al. conducted a national study on non-urgent emergency department visits and associated resource utilization, uncovering inefficiencies in EMS resource allocation and identifying opportunities for improvement [7]. Weaver et al. explored medical necessity in EMS transports, stressing the need for improved triaging systems to ensure that ambulances are dispatched to genuinely urgent cases [8]. These studies offer critical insights into the demand side of EMS operations, complementing the supply-side optimization models.

Other research has examined patterns and reasons behind emergency department visits, providing context for EMS utilization. Schappert and Rechtsteiner estimated ambulatory medical care utilization, while Weinick et al. investigated how many emergency department visits could potentially be managed at urgent care centers or retail clinics [9, 10]. These findings suggest that a substantial portion of EMS demand could be redirected to alternative care settings, alleviating the burden on emergency services and enhancing overall system efficiency.

Policy and operational guidelines have also significantly influenced the development of EMS systems. The American College of Emergency Physicians and the National Association of EMS Physicians jointly released a position paper on alternate ambulance transportation and destination, advocating for increased flexibility in EMS protocols [14]. Millin et al. provided a combined resource document for assessing the necessity of transport and reimbursement,

offering practical guidance for EMS providers [15]. These policy-oriented studies emphasize the connection between operational research and real-world implementation.

The rise of telemedicine has further broadened the scope of EMS optimization. Ellis et al. proposed a telemedicine model for emergency care in correctional facilities, showcasing its potential to enhance access to care while reducing resource utilization [16]. Langabeer et al. assessed the cost-benefit of telehealth-enabled EMS programs, demonstrating that such initiatives can significantly decrease ambulance transports and emergency department visits [17, 18]. Amadi-Obi et al. reviewed telemedicine applications in pre-hospital care, emphasizing their potential to transform EMS systems through improved triaging and decision-making [19, 20].

In summary, the literature on ambulance optimization and EMS utilization illustrates a rich interplay between theoretical models, empirical studies, and technological advancements. Foundational works have established the principles of location and allocation optimization, while recent progress has focused on dynamic redeployment, telemedicine integration, and demand management. Collectively, these studies provide a comprehensive framework for understanding and enhancing EMS systems, ensuring timely and efficient care for patients in need.

Mathematical Modeling

The optimization of ambulance locations involves maximizing the expected system-wide coverage by incorporating random delays and travel times. The key components of the mathematical model are as follows:

Let $Q_i(t)$ represent the expected fraction of calls reached within a given time standard by ambulance i at time t , and let $\lambda(t)$ denote the arrival rate of calls at time t . We can then write the following differential equation for $Q_i(t)$:

$$\frac{dQ_i(t)}{dt} = \lambda(t)[1 - F(T - t)]f_i(T - t) - Q_i(t)\lambda(t)f_i(T - t) \quad (1)$$

where F is the cumulative distribution function (CDF) of the travel time T , f_i is the probability density function (PDF) of pre-travel delay for the ambulance i , and T_t represents the remaining time after the delay before the time standard is reached. It is a dynamic system that evolves as time passes and call are dispatched.

The first term on the right-hand side of the equation 1 represents the expected number of calls that ambulance i is expected to reach at time t , according to the arrival rate $\lambda(t)$, the remaining time T_t , and the delay distribution function f_i . The second term represents the expected number of calls that an ambulance i could have reached if it were available at the start of the period (i.e., $Q_i(0) = 0$) and if there were no other ambulances in the system. The difference between these terms represents the expected fraction of calls that ambulance i will reach at time t .

The optimization problem involves finding the values of $Q_i(t)$ and the ambulance locations that maximize the expected system-wide coverage, and the objective is to constraints on the

total number of ambulances available and the maximum allowable travel time. The differential equation 1 can be solved numerically, using finite differences or finite elements methods, to compute the values of $Q_i(t)$ for different ambulance locations and arrival rates.

We first discretize the equation over a time grid to solve the following differential equations.

Using the finite difference method, $\frac{dQ_i(t)}{dt}$ is approximated as;

$$\frac{dQ_i(t)}{dt} = \frac{Q_i^{n+1} - Q_i^n}{\Delta t}$$

Where Q_i^n is the value of $Q_i(t)$ at the n-th time step and Δx is the time step size. Substitute this approximation in the differential equation

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} = \lambda(t^n)[1 - F(T - t^n)]f_i(T - t^n) - Q_i^n \lambda(t^n)f_i(T - t^n)$$

So,

$$Q_i^{n+1} = Q_i^n + \Delta t[\lambda(t^n)[1 - F(T - t^n)]f_i(T - t^n) - Q_i^n \lambda(t^n)f_i(T - t^n)]$$

To solve this equation numerically, we first choose a time grid $t_0, t_1, \dots, t_N$ with $\Delta t = t_{n+1} - t_n$ and let $Q_i^0 = 0$ as an initial condition, then we compute Q_i^{n+1} iteratively using the finite difference method. To implement this, we need a function or data for the call arrival rate ($\lambda(t)$), CDF of travel time $F(T - t)$, PDF of pre-travel delay $f_i(T - t)$, and remaining time to the standard $T - t$.

To apply a numerical solution, consider $\Delta t = 1$ to be step size, and the call arrival time rate to be $\lambda(t) = 0.05$ and discretize t into 16 points $t = 0, 1, 2, \dots, 15$. Let the travel time follow a normally distributed with a mean of 10 minutes and a standard deviation of 2 minutes, then the probability distribution function and cumulative probability distribution function $f_i(T - t)$ and $F(T - t)$ are given by;

$$F(T - t) = \Phi\left(\frac{t - T - \mu_T}{\sigma_T}\right)$$

And,

$$f_i(T - t) = \frac{1}{\sqrt{2\pi}\sigma_T} e^{-\frac{(t-T-\mu_T)^2}{2\sigma_T^2}}$$

Where $\Phi(x)$ is the standard normal cumulative distribution function below;

$$\Phi(x) = p(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Let us do iterations. At $t = 0$ we have, $T - t_0 = 10$

$$F(T - t_0) = \Phi(0) = 0.5$$

$$f_i(T - t_0) = \frac{1}{\sqrt{2\pi}2} e^0 = 0.19947$$

That implies;

$$\begin{aligned} Q_i^1 &= Q_i^0 + \Delta t[\lambda(0)[1 - F(T - t_0)]f_i(T - t_0) - Q_i^0 \lambda(0)f_i(T - t_0)] \\ &= 0 + 1[0.05(1 - 0.5).0.19947 - 0.005(0.19947)] = 0.004987 \end{aligned}$$

And similarly, at $t = 1$ we have $T - t_0 = 9$ so;

$$F(T - t_1) = \Phi\left(\frac{9 - 10}{2} = -0.5\right) = 0.3085$$

$$f_i(T - t_1) = \frac{1}{\sqrt{2\pi}2} e^{-\frac{(9-10)^2}{8}} = 0.17603$$

That gives;

$$\begin{aligned} Q_i^2 &= 1 + \Delta t[\lambda(1)[1 - F(T - t_1)]f_i(T - t_1) - Q_i^1 \lambda(1)f_i(T - t_1)] \\ &= 0.004987 + 1[0.05(1 - 0.3085).0.17603 - 0.004987.0.05(0.17603)] \\ &= 0.01203 \end{aligned}$$

We continue the iterations in this way and here is the table 1 of results in 6 different iterations,

t	$T - t$	$F(T - t)$	$f_i(T - t)$	$Q_i(t)$
0	10	0.5	0.19947	0.00499
1	9	0.3085	0.17603	0.01203
2	8	0.1587	0.12098	0.01755
3	7	0.0668	0.06599	0.02102
4	6	0.0228	0.02918	0.02273
5	5	0.0062	0.01043	0.02302

Table 1: the values of $Q_i(t)$ for different ambulance locations and arrival rates

The results from the table 1 demonstrate the evolution of $Q_i(t)$ over time. Initially, $Q_i(t)$ increases rapidly due to the higher probabilities of calls being reachable as indicated by large values of PDF $f_i(T - t)$ and moderate value of $F(T - t)$. However, as time progresses, the growth of $Q_i(t)$ slows down, this is because $T - t$ decreases, leading to lower values of $f_i(T - t)$ and higher cumulative probabilities $F(T - t)$, reducing the chance of successfully reaching new calls. Eventually, $Q_i(t)$ approaches a plateau as the fraction of reachable calls stabilizes. This trend highlights the importance of response time and travel time distribution in determining coverage performance over time.

Scenario of Optimizing Ambulance Locations

In this section, to calculate simplicity we first provide a scenario of optimizing ambulance locations and apply equation (1) to maximize the expected system-wide coverage and then generalize the result.

Consider the following hypothetical setup;

Consider a hypothetical city with a population of 500,000 people. The city is divided into 10 districts, and each district has a hospital. The city has a total of 20 ambulances, which are stationed at different hospitals based on their geographical location. The ambulance service aims to provide timely emergency medical assistance to the citizens of the city. However, due to the limited number of ambulances and the random nature of emergency calls, it is challenging to optimize the allocation of ambulances to minimize response time and maximize system-wide coverage (for example 10 minutes).

Note that ambulances are allocated to districts to maximize the expected system-wide coverage fraction $C(E(C))$, where C is the total response time including pre-travel delays and travel time.

$$E(C) = \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} P(C > d, t, a) f_D(d) f_A(a) f_T(t) dt da dd$$

Here, $P(C > d, t, a)$ is the probability that the time to reach a call exceeds the time threshold d , (this part combines the PDFs of d, a , and t to model the probability that the total response time exceeds d , t is the travel time, a is the activation delay, d is the pre-travel delay and the functions $f_D(d)$, $f_T(t)$ and $f_A(a)$ are as follows;

$f_D(d)$ = the probability density function of the pre-travel delay at the dispatch stage =

$$f_D(d) = \frac{1}{\sigma_D \sqrt{2\pi}} \exp \left(-\frac{(d - \mu_D)^2}{2\sigma_D^2} \right)$$

where D is the random variable representing the pre-travel delay at the dispatch stage, μ_D , and σ_D are the mean and standard deviation of the delay, respectively.

$f_A(a)$ = the probability density function of the pre-travel delay at the activation stage =

$$f_A(a) = \frac{1}{\sigma_A \sqrt{2\pi}} \exp \left(-\frac{(a - \mu_A)^2}{2\sigma_A^2} \right)$$

where A is the random variable representing the pre-travel delay at the activation stage, μ_A , and σ_A are the mean and standard deviation of the delay, respectively.

$f_T(t)$ = the probability density function of the travel time:

$$f_T(t) = \frac{1}{\sigma_T \sqrt{2\pi}} \exp \left(-\frac{(t - \mu_T)^2}{2\sigma_T^2} \right)$$

where T is the random variable representing the travel time, μ_T and σ_T are the mean and standard deviation of the travel time, respectively.

To maximize $E(C)$ using differential equation $Q_i(t)$, let us leverage the result of $Q_i(t)$ obtained in table 1, and let $P(C > d, t, a)$ be $1 - Q_i(t)$. Then,

$$E(C) = \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} (1 - Q_i(t)) f_D(d) f_A(a) f_T(t) dt da dd$$

Discretize $E(C)$ for numerical maximization. To do so, we have to divide the time variables d , a and t into small intervals, in this case $E(C)$ is approximated as follows;

$$E(C) \approx \sum_{d=0}^{d_{max}} \sum_{a=0}^{a_{max}} \sum_{t=0}^{t_{max}} (1 - Q_i(t)) f_D(d) f_A(a) f_T(t) \Delta t \Delta a \Delta d$$

Where Δt , Δa , Δd are the step sizes for discretization. Let us consider A_{ij} as the fraction of ambulances allocated o station i for district j as below;

$$A_{ij} = \frac{C_{ij}}{\sum_{i=1}^n \sum_{k=1}^n C_{k,i}}$$

Where C_{ij} is the expected coverage fraction for station i and call zone j , n is the number of stations and m is the number of call zones or districts.

Then, the total coverage $E(C)$ becomes as below;

$$E(C) = \sum_{i=1}^{n_{ambulances}} \sum_{j=1}^{n_{districts}} A_{ij} \sum_{d,a,t} (1 - Q_i(t)) f_D(d) f_A(a) f_T(t) \Delta t \Delta a \Delta d \quad (2)$$

The aim is to maximize $E(C)$ by adjusting the ambulance allocation matrix A_{ij} subject to $\sum_{j=1}^{n_{districts}} A_{ij} \leq 1$ for each ambulance, since an ambulance cannot exceed 100% allocation and also $\sum_{i=1}^{n_{ambulance}} A_{ij} \geq demand\ j$, ensuring each district receives sufficient coverage.

We have used the method of Lagrange multipliers to incorporate the constraints into the objective function. Define the Lagrangian as;

$$L(A, \lambda) = E(c) + \sum_{i=1}^{n_{ambulances}} \lambda_i (1 - \sum_{j=1}^{n_{districts}} A_{ij})$$

Where λ_i is the Lagrange multiplies for the equality $\sum_{j=1}^{n_{districts}} A_{ij} = 1$.

Compute the partial derivatives of $L(A, \lambda)$ with respect to A_{ij} and λ_i and set them to zero, then we get the following;

$$\frac{\partial L}{\partial A_{ij}} = \frac{\partial E(c)}{\partial A_{ij}} - \lambda_i = 0$$

Corresponding

$$\frac{\partial E(c)}{\partial A_{ij}} = \lambda_i$$

And

$$\frac{\partial L}{\partial \lambda_i} = 1 - \sum_{j=1}^{n_{districts}} A_{ij} = 0$$

That gives us $\sum_{j=1}^{n_{districts}} A_{ij} = 1$. Since the partial derivative of $E(c)$ in equation 2 with respect to A_{ij} is;

$$\frac{\partial E(c)}{\partial A_{ij}} = \sum_{d,a,t} (1 - Q_i(t)) f_D(d) f_A(a) f_T(t) \Delta t \Delta a \Delta d$$

Then, from the Lagrange conditions $\lambda_i = \sum_{d,a,t} (1 - Q_i(t)) f_D(d) f_A(a) f_T(t) \Delta t \Delta a \Delta d$, for simplicity let the right side of λ_i be denoted by w_{ij} , demonstrated the expected contributions of ambulance i allocated to distinct j . To operate, we use the initial value as given in our scenario, so the order of the matrix w_{ij} will be 20*10 (20 ambulances and 10 districts). The hypothetical values for w_{ij} are,

$$w = \begin{pmatrix} 0.8 & 0.6 & 0.7 & 0.9 & 0.5 & 0.4 & 0.6 & 0.8 & 0.7 & 0.6 \\ 0.5 & 0.7 & 0.6 & 0.8 & 0.6 & 0.5 & 0.4 & 0.7 & 0.6 & 0.5 \\ 0.9 & 0.8 & 0.6 & 0.7 & 0.7 & 0.6 & 0.8 & 0.9 & 0.7 & 0.8 \\ 0.5 & 0.6 & 0.8 & 0.9 & 0.9 & 0.9 & 0.8 & 0.7 & 0.6 & 0.5 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

From $w_{ij} = \lambda_i$ all A_{ij} for a given ambulance i must be proportional to w_{ij} as follows;

$$A_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n_{districts}} w_{ij}}$$

Let us consider ambulance 1 and use the first row of the matrix w_{ij}

$$w_{1j} = [0.8 \ 0.6 \ 0.7 \ 0.9 \ 0.5 \ 0.4 \ 0.6 \ 0.8 \ 0.7 \ 0.6]$$

Then we have $\sum_{j=1}^{10} w_{1j} = 6.6$. Normalize A_{1j}

$$A_{1j} = \frac{w_{1j}}{6.6} = \left[\frac{0.8}{6.6} \ \frac{0.6}{6.6} \ \dots \dots \frac{0.6}{6.6} \right]$$

$$A_{1j} = [0.121 \ 0.091 \ 0.106 \ 0.1360 \ 0.076 \ 0.061 \ 0.091 \ 0.121 \ 0.106 \ 0.091]$$

By applying the same way for the second ambulance and using the second row of the matrix $w_{ij} = [0.5 \ 0.7 \ \dots \dots 0.5]$ we will get;

$$A_{2j} = \frac{w_{2j}}{5.9} = \left[\frac{0.5}{5.9} \ \frac{0.7}{5.9} \ \dots \dots \frac{0.5}{5.9} \right]$$

$$A_{1j} = [0.085 \ 0.119 \ 0.102 \ 0.1360 \ 0.102 \ 0.085 \ 0.068 \ 0.119 \ 0.102 \ 0.085]$$

Repeat this process for all 20 ambulances to calculate the full allocation matrix A_{ij} . It is found that the maximum expected coverage $E(C)$ is while we substitute A_{ij} into $E(C)$ to get

$$E(C) = \sum_{i=1}^{20} \sum_{j=1}^{10} A_{ij} w_{ij}$$

Generalize it to all zones and ambulances to get the maximum expected coverage $E(C)$ as below;

$$E(C) = \sum_{i=1}^{n_{ambulances}} \sum_{j=1}^{n_{districts}} A_{ij} w_{ij}$$

This numerical optimization approach for maximizing the expected coverage $E(C)$, particularly through the iterative adjustment of the allocation matrix A_{ij} , leads to significant insights and practical conclusion.

The optimization process ensures that the limited number of ambulances is distributed to maximize system-wide coverage. By focusing on the weight of the matrix w_{ij} , the model emphasizes districts with the greatest need or the highest probabilities of improving response time. This results in an efficient allocation strategy that prioritizes districts with higher demand and a higher likelihood of success. To get the best model the optimization showed that the ambulances must not be evenly distributed but be concentrated in areas with high-weight w_{ij} , representing districts with higher emergency call volume or shorter travel time. Accurate modeling of w_{ij} including realistic pre-travel delay distributions, travel time and demand rate is critical for the success of the optimization.

Conclusion

In conclusion, this study highlights how effective a mathematical optimization model can be in improving ambulance location strategies aimed at maximizing overall coverage. By applying finite difference methods to solve the differential equation for $Q_i(t)$, we were able to uncover temporal trends in emergency response coverage. Additionally, our iterative optimization process underscored the critical role of strategic resource allocation in maximizing $E(C)$.

The analysis demonstrated that optimized ambulance allocations lead to significant improvements in coverage; however, we also observed diminishing returns when resource constraints became limiting factors. Furthermore, district-specific elements, such as pre-travel delays and travel times, were found to greatly influence coverage, which points to the necessity for customized allocation strategies. The findings affirm the reliability of the numerical methods used and highlight the model's potential as a valuable decision-making tool in emergency response systems. Looking ahead, future research could integrate real-time dynamics and additional stochastic variables to further enhance system performance.

References

1. Larson, R. C., & Odoni, A. R. (1981). Urban operations research (No. Monograph).
2. Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The location of emergency service facilities. *Operations research*, 19(6), 1363-1373.

3. Daskin, M. S. (1983). A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation science*, 17(1), 48-70.
4. Goldberg, J., & Paz, L. (1991). Locating emergency vehicle bases when service time depends on call location. *Transportation science*, 25(4), 264-280.
5. Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European journal of operational research*, 147(3), 451-463.
6. Jagtenberg, C. J., Bhulai, S., & van der Mei, R. D. (2015). An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care*, 4, 27-35.
7. Honigman, L. S., Wiler, J. L., Rooks, S., & Ginde, A. A. (2013). National study of non-urgent emergency department visits and associated resource utilization. *Western Journal of Emergency Medicine*, 14(6), 609.
8. Weaver, M. D., Moore, C. G., Patterson, P. D., & Yealy, D. M. (2012). Medical necessity in emergency medical services transports. *American Journal of Medical Quality*, 27(3), 250-255.
9. Schappert, S. M., & Rechtsteiner, E. A. (2008). National health statistics reports. Hyattsville, MD: National center for health statistics.
10. Weinick, R. M., Burns, R. M., & Mehrotra, A. (2010). Many emergency department visits could be managed at urgent care centers and retail clinics. *Health affairs*, 29(9), 1630-1636.
11. Capp, R., Rooks, S. P., Wiler, J. L., Zane, R. D., & Ginde, A. A. (2014). National study of health insurance type and reasons for emergency department use. *Journal of general internal medicine*, 29, 621-627.
12. Young, G. P., Wagner, M. B., Kellermann, A. L., Ellis, J., & Bouley, D. (1996). Ambulatory visits to hospital emergency departments: patterns and reasons for use. *Jama*, 276(6), 460-465.
13. NEHI, N. (2010). A matter of urgency: reducing emergency department overuse. URL https://www.nehi.net/writable/publication_files/file/nehi_ed_overuse_issue_brief_032610final edits.pdf.
14. Policy Statement of American College of Emergency Physicians and the National Association of EMS Physicians. Alternate ambulance transportation and destination. National association of EMS Physicians/American College of Emergency Physicians joint position paper. *Ann Emer Med*. 2008;52(5):594.
15. Millin, M. G., Brown, L. H., & Schwartz, B. (2011). EMS provider determinations of necessity for transport and reimbursement for EMS response, medical care, and transport: combined resource document for the National Association of EMS Physicians position statements. *Prehospital Emergency Care*, 15(4), 562-569.
16. Ellis, D. G., Mayrose, J., Jehle, D. V., Moscati, R. M., & Pierluisi, G. J. (2001). A telemedicine model for emergency care in a short-term correctional facility. *Telemedicine journal and e-health*, 7(2), 87-92.
17. Langabeer, J. R., Gonzalez, M., Alqusairi, D., Champagne-Langabeer, T., Jackson, A., Mikhail, J., & Persse, D. (2016). Telehealth-enabled emergency medical services program reduces ambulance transport to urban emergency departments. *Western Journal of Emergency Medicine*, 17(6), 713.

18. Langabeer, J. R., Champagne-Langabeer, T., Alqusairi, D., Kim, J., Jackson, A., Persse, D., & Gonzalez, M. (2017). Cost–benefit analysis of telehealth in pre-hospital care. *Journal of telemedicine and telecare*, 23(8), 747-751.
19. Amadi-Obi, A., Gilligan, P., Owens, N., & O'Donnell, C. (2014). Telemedicine in pre-hospital care: a review of telemedicine applications in the pre-hospital environment. *International journal of emergency medicine*, 7, 1-11.
20. Winburn, A. S., Brixey, J. J., Langabeer, J., & Champagne-Langabeer, T. (2018). A systematic review of prehospital telehealth utilization. *Journal of telemedicine and telecare*, 24(7), 473-481.