8-2011

IMPROVING QUALITY OF SERVICE IN EMS SYSTEMS BY REDUCING DISPARITIES BETWEEN SERVICE ZONES

Sunarin Chanta
Clemson University, schanta@clemson.edu

IMPROVING QUALITY OF SERVICE IN EMS SYSTEMS BY REDUCING DISPARITIES BETWEEN SERVICE ZONES

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Engineering

---

by
Sunarin Chanta
August 2011

Accepted by:
Dr. Maria E. Mayorga, Committee Chair
Dr. Margaret M. Wiecek
Dr. Mary E. Kurz
Dr. William G. Ferrell

ABSTRACT

Emergency medical service (EMS) systems respond to emergency or urgent calls so as to provide immediate care, such as pre-hospital care and/or transportation, to hospitals. Care must be provided in a timely manner; in fact quality of service is usually directly associated with response time. To reduce the response time, the number and location of vehicles within the service area are important variables. However with limited capacity, increasing the number of vehicles is often an infeasible alternative. Therefore, a critical design goal is to decide at which facilities stations should be located in order to serve as much demand as possible in a reasonable time, and at the same time maintain equitable service between customers. This study aims to focus on locating ambulances which respond to 911 calls in EMS systems. The goals are to find the optimal base station location for vehicles so that the number of calls or customers served is maximized while disparity between those customers is minimized, to consider the survival rate of patients directly in the model, and develop appropriate meta-heuristics for solving problems which cannot be solved optimally.

DEDICATION

To my mother and father.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Mayorga for her support and guidance throughout this long process. She has been very patient and flexible, without her this dissertation would not have been completed.

I would also like to thank my committee members, Dr. Wiecek, Dr. Kurz, and Dr. Ferrell for their valuable suggestions and helpful feedback.

In addition, my thank goes to all Thai people, since my financial support came from their taxations. As a Royal Thai scholar student, I'm going to return this honored debt to them real soon.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PREFACE

EMS systems are specially organized systems that provide emergency medical service within a service area. The emergency medical services are varied depending on a call such as providing an emergency medical technician, a paramedic, or transportation. The EMS system is activated by an incident that generally causes serious illness or injury. Therefore in this situation, a very important factor is not only emergency medical care but also response time. Rapid response time by EMS can mean the difference between life and death. In urban areas, the most widely used ambulance response-time standard is 8 minutes and 59 seconds (Fitch, 2005). However in reality, not all incidents can receive service by this standard time depending on the area and the state. Especially in rural or remote areas the response times tend to be longer than that. One way to reduce the response time is by locating vehicles at the appropriate station locations so that they can serve the requested calls in time. To address this problem, we would like to develop a mathematical model for locating EMS vehicles. This problem is known as a covering problem where the service to customers depends on the distance between the customer and the facility to which the customer is assigned (Daskin, 1995). In the covering problem, we assume that demand location and potential facility sites are restricted to the nodes in the network with arcs specifying path between them. Moreover we also assume that demands are grouped at demand nodes and a demand node is covered when there exists at least one vehicle located within the coverage distance (Daskin, 1995). The problem can be formulated as an integer programming model by using binary decision variables which take a value of 1 if a demand node is covered or 0 otherwise. Because of 0-1 variables, covering location models are not easy to solve. Moreover our goal is not only to maximize the number of customers, but also minimize the disparity between customers. The resulting complex objective causes the model to be hard to solve to optimality. In this situation, a heuristic is preferred for solving the problem, especially in practical application in which the size of problem is large.

This dissertation is composed as a compilation of three journal papers that focused on reducing inequity in facility location problem for EMS systems. Any redundancies between chapters have been removed to make it easier to read. Furthermore, the dissertation chapters contain more material that did not necessarily get included in the submissions. An overview of each paper is presented as follows.

1. A bi-objective covering location model for EMS systems: Addressing the issue of fairness in rural areas.

This paper aims to balance the level of EMS service provided to patients in urban and rural areas by locating ambulances at appropriate locations. Traditional covering location models; whose objective is to maximize demand that can be covered, favor locating ambulances in urban areas; since urban areas have higher population densities. To address the issue of fairness in rural areas, we propose three bi-objective covering location models that directly consider fairness via a secondary objective; results are discussed and compared to provide alternatives to decision makers. (see publications related to this research in Chanta et al., 2009, 2011a).

2. The minimum *p*-envy location problem: A new model for equitable distribution of emergency resources

This paper aims to find optimal locations for Emergency Medical Service (EMS) vehicles in order to balance disparity in service between zones and at the same time maintain service coverage. Instead of carrying two conflicted objectives, we propose a minimum-envy covering location model which takes into account both issues by minimizing the sum of "envy" among all zones weighted by proportion of demand in each zone. Because of complexity in the objective function, a tabu search is developed for solving this problem. A case study using real-world data collected from Hanover County, VA is presented. The performance of the proposed model is compared to other location models. (see publications related to this research in Chanta et al., 2010a, 2011b).

3. The minimum *p*-envy location problem: Focusing on survivability of patients

This paper considers an extension to the minimum *p*-envy location model by evaluating the objective of the model based on the survival function instead of the distance function since survival probability is directly related to patient outcomes. The model was tested on a real world data set from the EMS system at Hanover County, VA, and also compared to the original minimum *p*-envy location problem including other location models. The results indicate that more lives can be saved by using the survival function objective and that the enhanced *p*-envy model outperforms other commonly used location models in terms of number of lives saved. (see publication related to this research in Chanta et al., 2010b).

CHAPTER 2

A BI-OBJECTIVE COVERING LOCATION MODEL FOR EMS SYSTEMS: ADDRESSING THE ISSUE OF FAIRNESS IN RURAL AREAS

2.1 Introduction

EMS systems are especially organized systems that provide emergency medical service to patients within a service area. This service area could be urban, rural, or some combination of the two depending on how the population is distributed in the geographical region. Unfortunately, rural communities often suffer from inadequate medical services, including emergency care. Such a problem is compounded with low expectations that emergency care in rural areas will be as fast and effective as in urban areas (NCSL, 2000). EMS systems are typically evaluated according to how they respond to and treat out-of-hospital cardiac arrest patients. In urban areas, EMS systems tend to have the highest known cardiac arrest patient survival rates. In contrast, EMS systems in rural and semi-rural areas have notably lower cardiac arrest patient survival rates (English, 2008).

People in rural areas have difficult time to face health disparities in health care. Because of less demands, the reimbursement in rural is low which this causes lack of service providers, hospitals, and technicians (Willams et al., 2001). Moreover geographic barriers and limitation of resources lead the patients take long time to access the service. All these reasons lead rural states had lower access in all types of emergency departments (Brendan et al., 2009). The quality of service is still concerned because of low call volumes, as a result in, less utilization, difficult to maintain medical operating skills, and lack of training. In addition, people in rural are aging, and many injuries are greater in severity than urban (OTA, 1989). Air transportation is another way to go when patients need immediately care. However, from previous report air transportation is faster than ground transportation if the patients are in the air zone, while ground transportation is also faster than air transportation if the patients are in the ground zone (Lerner, 1990). Therefore providing the air transportation is not enough; we still need the effective ground transportation. Although balancing equity between rural and urban almost impossible, we try to reduce cause of death happen at the scene which is the major cause of death of people in rural (Trevillyan et al., 1998).

A very important factor in determining EMS performance is not only the quality of emergency medical care provided but also the timeliness in which care is provided, or *response time*. A rapid response time by EMS can mean the difference between life and death. In urban areas, the most widely used ambulance response time standard is to respond to 90% of calls within 8 minutes and 59 seconds as compared to responding to 90% of calls within 14 minutes and 59 seconds in rural areas (Fitch, 2005). In practicality, it may not be possible to meet this standard depending on the geographical area, the EMS resources available, and the location of EMS resources at the time of the call. Response times may be much longer than the standard, especially in rural or remote areas. One way to reduce the response time is to locate ambulances at the appropriate station locations. This problem is known as a *covering problem* where the service to customers depends on the distance between the customer and the facility to which the customer is assigned (Daskin, 1995). A call is said to be covered if the response time is within the standard; for example, a call responded to within 8 minutes and 59 seconds or less is considered covered, while if the response time is 9 minutes or more the call is considered uncovered.

Most EMS systems locate and use their resources in a way that maximizes the number of persons (or calls) that can be served within a specified time or distance. Often this is translated to number of demand zones that can be covered, where a demand zone is a geographic region with associated call volume. However, when resources are limited, some demand zones may go uncovered. With a single objective that maximizes the number of covered demand zones, these uncovered demand zones tend to be located at the edges of the region. This results in an inequitable use of resources that impacts patient outcomes. Thus, patient survival rates in rural areas are observed to be significantly lower than in urban areas (English, 2008). As we will show in Section 2.6, applying a covering location model with the single objective of maximizing the number of covered demand zones to a semi-rural county results in optimal solutions that locate emergency ambulances at stations that leave the majority of rural demand zones uncovered.

We propose three bi-objective models for locating EMS ambulances in order to reduce the disparities in service among rural and urban areas. The first objective is the traditional covering problem objective of maximizing the number of covered calls while the second objective is aimed at improving service in rural areas. Since there is no universally accepted way to measure fairness in EMS systems, we propose three alternatives for the second objective as a means to identify how to best evaluate fairness in EMS systems. The three proposed alternatives to be the second objective are: (1) minimize the maximum distance between each uncovered demand zone and its closest opened station, (2) minimize the number of uncovered rural demand zones, and (3) minimize the number of uncovered demand zones (notice that this last objective is not the same as maximizing the number of covered calls, as each zone has a different demand). These three models are formulated as integer programs. Non-dominated solutions to each bi-objective model are generated using the ε-constraint approach.

The key contribution of this paper is a model that can be used to reduce the disparities in service between different demographics; in particular we focus on urban versus rural areas. The solution to the model provides decision makers with a set of solutions that can be chosen based upon performance measures of interest. Results indicate that the model that minimizes the maximum distance between each uncovered demand zone and its closest opened station as a secondary objective results in solutions that dominate those from the other models, when evaluating the average distance (or weighted average distance) between an uncovered zone to its closest station. On the other hand, a model that uses a secondary objective of minimizing the number of uncovered rural demand zones yields a larger solution set, which may be desirable to the decision maker as it provides more options. Moreover considering on the distribution of the distance from individuals to their closest stations, the solutions of the model that minimizes the maximum distance between uncovered demand zone and its closest open station as a secondary objective are equitably efficient as the solutions of the model that minimizes the number of uncovered demand zones as a secondary objective, since both models provide equally service among all individual customers.

2.2 Literature review and scope

Models for locating EMS resources typically use variations of the covering problem, where facilities are located at existing stations on the network to cover all the demand zones while minimizing the number of facilities. The basic covering model, the set covering location problem (SCLP), was developed by Toregas et al. (1971) with the objective of minimizing the number of ambulances needed to cover all demand nodes. Church and ReVelle (1974) extended the SCLP to address the situation in which the number of ambulances available is less than the number needed to cover all demand zones, this is called a maximal covering location problem (MCLP). These first two covering models are deterministic which assuming that vehicles are always available to serve calls. Daskin (1982) developed a stochastic model called maximum expected covering location problem (MEXCLP) model which is an extension of MCLP model by considering the probability of vehicle being busy, assumed that each server has the same probability. Batta et al. (1989) embedded the hypercube model (Larson, 1974, 1975) in to the MEXCLP, and differentiated probabilities of vehicle being busy of different location. Later, Hogan and ReVelle (1986) considered the issue of backup coverage, or secondary coverage of a demand zone. Backup coverage is required in high-demand areas to maintain a uniform level of service when EMS ambulances can respond to only one call at a time. All of these models are single-objective models.

Several papers do consider multiple objectives for ambulance location problems. Daskin et al. (1988) integrated different covering models such as multiple, excess, backup and expected covering models. For example, they reformulate the hierarchical objective set covering problem (Daskin and Stern, 1981) into a multi-objective set covering problem, which allows them to derive the trade-off between the number of facilities and the extra coverage. Pirkul and Schilling (1988) modeled the objective of maximizing covered calls while simultaneously considering workload capacities and backup service. Pirkul and Schilling (1991) extended this model with the addition of workload limits on the facilities and the quality of service delivered to the uncovered demand zones. The workload limit refers to the specific amount of demand that can be served by one facility. The workload limit condition is formulated as a constraint to make the model more realistic. The quality of service is modeled as the total distance from uncovered demand zones to the nearest facility, and the resulting model is solved using a solution procedure based on Lagrangian relaxation. Narasimhan et al. (1992) extended the model to consider multiple levels of backup.

ReVelle et al. (1996) considered extensions of the maximal conditional covering problem. In their models, the facility locations are supposed to be covered by other facilities and may not be used to cover their own zones. Berman and Krass (2002) presented the generalized maximal cover location problem which allows for partial coverage. The degree of coverage is defined as a non-increasing step function of the distance to the nearest facility. A greedy heuristic and an LP-relaxation are applied to solve the problem and provide bounds on the relative errors of the approximate solutions.

Karasakal and Karasakal (2004) introduced intermediate coverage or partial coverage; the model allows the coverage to change within a distance; that is, the demand points can be fully covered within the minimum critical distance, partially covered to a maximum critical distance, and not covered outside of the maximum critical distance. Araz et al. (2007) developed a multi-objective covering location model based on previously developed models (Hogan and ReVelle, 1986; Pirkul and Schilling, 1988). Their model has three objectives: (1) maximizing the population covered by one vehicle, (2) maximizing the population with backup coverage, and (3) minimizing the total distance from locations at a distance bigger than a specified distance standard for all zones. The problem is solved using a fuzzy goal programming approach. There are several multi-objective covering models can have been studied but most of them have the assumption that vehicles are always available to server calls.

Although there are many extensions to covering location models, there is no model that explicitly addresses fairness of service to patients in rural areas. Under a single covering objective (e.g., maximizing the expected covered demand), patients in urban areas are generally covered at the expense of rural patients, leading to adverse patient outcomes in rural areas. Even as more EMS ambulances become available, covering models tend to continue to concentrate EMS resources in urban areas.

In this paper, we propose three objective functions to locate EMS ambulances so as to reduce the disparities in service between rural and urban areas. Each of these objective functions is used in a bi-objective discrete optimization model that evaluates the tradeoffs between coverage and fairness. The first objective function minimizes the maximum distance between uncovered demand zones and opened stations. This objective function assigns the uncovered demand zones to the closest opened stations, in order to minimize the distance from an uncovered zone to an opened station. This is important because even if a zone cannot be responded to within the response time standard, patient survivability rates are directly related to response time (or equivalently, distance) (Larsen et al., 1993). The second objective function minimizes the total number of rural demand zones that cannot be covered. This objective function considers the trade-offs between the number of rural demand zones that can be covered and the amount of demands in these demand zones that can be covered. The third objective function minimizes the total number of uncovered demand zones, either urban or rural. The idea of the third objective function is similar to the second objective function, but it does not consider the type of uncovered zones (i.e., urban or rural). The proposed objective functions provide guidelines for locating ambulances, while allowing decision makers to simultaneously improve the quality of service in both rural and urban areas.

2.3 Covering location model formulation

This section introduces a bi-objective covering location model for locating EMS ambulances at preexisting rescue stations that balances the overall quality of service (i.e., coverage) with fairness. In this covering location problem, the goal is to cover as much demand as possible while reducing the disparity in service between urban and rural areas. To directly consider the issue of fairness, three bi-objective models are proposed. The first objective is to maximize the expected number of requested calls that can be covered, namely, Z_1 (Equation (2.1)). The second objective is to improve fairness. We propose three alternative objective functions for improving fairness in rural areas which are to

- minimize the maximum distance between uncovered demand zones and their closest opened stations (Z_{2a}, Equation (2.2a)),
- minimize the number of uncovered rural demand zones (Z_{2b}, Equation (2.2b)), and
- minimize the number of uncovered demand zones (Z_{2c}, Equation (2.2c)).

These three alternative objective functions are selected one at a time to be used as a second objective, resulting in three distinct models. We have three decision variables which are y_{ik} (a 0-1 variable that indicates if demand zone i is covered by at least k ambulances), and x_j (the number of ambulances at station j). There are three constraints, shown in Equations (2.3) to (2.5). The first constraint (3) limits the total number of ambulances available to be located to T. The second constraint (4) limits the maximum number of ambulances that can be located at a single station to S. In the third constraint (2.5), a demand zone can receive service from a station as long as that station is open (i.e., there is at least one vehicle located there). Equations (2.6) and (2.7) represent non-negativity and integrality constraints.

Maximize $$Z_1 = \sum_{i=1}^{n} \sum_{k=1}^{l_i} h_i w_k y_{ik} \tag{2.1}$$

Minimize $$Z_{2a} = \max_{i \in U} \{ \min_{j \in O} (d_{ij}) \} \tag{2.2a}$$

Minimize $$Z_{2b} = \sum_{i=1}^{n} (1 - y_{i1}) r_i \tag{2.2b}$$

Minimize $$Z_{2c} = \sum_{i=1}^{n} (1 - y_{i1}) \tag{2.2c}$$

Subject to:

$$\sum_{j=1}^{m} x_j \le p \qquad (2.3)$$

$$x_j \le s, \qquad j=1, \ldots, m \qquad (2.4)$$

$$\sum_{k=1}^{l_i} y_{ik} \le \sum_{j \in J_i} x_j, \qquad i=1,\ldots,n \qquad (2.5)$$

$$x_j \in \{0,1,\ldots,s\} \qquad j=1, \ldots, m \qquad (2.6)$$

$$y_{ik} \in \{0,1\} \qquad i=1,\ldots,n;\ k=1,\ldots,l_i \qquad (2.7)$$

Where the decision variables are:

$$y_{ik} = \begin{cases} 1 & \text{if demand zone } i \text{ is covered by at least } k \text{ ambulances} \\ 0 & \text{otherwise} \end{cases}$$

x_j = the number of ambulances located at station j

The following list summarizes the parameters used:

w_k = the probability that the k^{th} vehicle is available (see below)

h_i = the call volume in demand zone i

d_{ij} = the distance from station j to demand zone i

$$r_i = \begin{cases} 1 & \text{if demand zone } i \text{ is in rural area} \\ 0 & \text{otherwise} \end{cases}$$

p = the total number of ambulances to be located

s = the maximum number of ambulances allowed to be located at each station

J_i = $\{j \mid d_{ij} \le D\}$: set of stations that can cover demand zone i

D = the maximum distance that can be reached within 9 minutes (4 miles)

U = $\{i \mid y_{i1} = 0\}$: set of uncovered demand zones

O = $\{j \mid x_j \geq 1\}$: set of opened stations

l_i = $\min\{s \times |J_i|, p\}$ (upper bound on the number of vehicles that can cover a demand zone)

n = the number of demand zones

m = the number of stations

Note that a demand zone i can be covered by 1 up to l_i vehicles such that $y_{i1} \geq y_{i2} \geq ... \geq y_{il_i}$. For example, If demand zone i is covered by 1 vehicle, y_{i1} =1 and $y_{i2} = ... = y_{il_i} = 0$. If demand zone i is covered by 2 vehicles, y_{i1}=y_{i2}=1, and $y_{i3} = ... = y_{il_i} = 0$. However, it is not necessary to enforce this using a constraint of the form $y_{i,k} \geq y_{i,k+1}$ because of the definition of w_k given by Equation (2.11). By definition, w_k is the probability that the k^{th} vehicle is available (while k-1 vehicles are busy) such that w_k is greater than w_{k+1}, and since the objective is to maximize Z_1 which weighs each y_{ik} by w_k, then for each demand zone i it will be optimal to let $y_{i,k} \geq y_{i,k+1}$.

In our model, when calculating the expected number of calls that can be covered we account for the fact that, even if ambulances are stationed within the coverage distance, they may be busy and therefore unable to respond to a call. The probability that a randomly selected vehicle will be busy, p_b, depends on the number of ambulances that are deployed. To estimate p_b, we use actual data of the system, which is captured by Equation (2.8), where, λ is the average number of calls per hour, $1/\mu$ is the average service time per call (hours), and p is number of ambulances that are deployed. This definition of p_b assumes that all ambulances operate independently. This assumption can be relaxed using the correction factor given by Batta et al. (1989) in an embedded hypercube model. The hypercube model Larson (1974, 1975) has several underlying assumptions: 1) calls for service arrive according to a Poisson process, 2) if a call arrives while all servers are busy, it enters at the end of a queue and will be served in a FIFO manner. If there are k ambulances that may respond to a call, the probability that the k^{th} vehicle will be dispatched or is available is calculated from the probability that k-1 ambulances are busy and the k^{th} vehicle is available. The probability that the k^{th} vehicle is available (w_k) is shown in Equation (2.11) where Q (p, p_b, k-1) is the correction factor and Q (p, p_b, 0) =1.

$$p_b = \frac{\lambda}{p\mu} \tag{2.8}$$

$$p_0 = \left(\frac{p^p p_b{}^p}{p!(1-p_b)} + \sum_{j=0}^{p-1} \frac{p^j p_b{}^j}{j!} \right)^{-1} \tag{2.9}$$

$$Q(p, p_b, j) = \sum_{k=j}^{p-1} \frac{(p-j-1)!(p-k)p^k p_b^{k-j} p_0}{(k-j)!\,p!(1-p_b)} \qquad , j=0, \ldots, p-1 \tag{2.10}$$

$$w_k = Q(p, p_b, k-1)(1-p_b)(p_b^{k-1}) \qquad , k=1, \ldots, p \tag{2.11}$$

2.4 The ε-constraint method

Several approaches exist for solving multi-objective problems such as weighted-sum, ε-constraint, and weighted-norm; see a review on this in Ehrgott and Wiecek (2005). The weighted-sum method, while popular, is not suitable for our problem because our solution space is integer and it is known that when the solution space is not convex, the weighted-sum method cannot find all solutions. However, both ε-constraint and weighted-norm approaches can find all solutions of integer problems (Berube et al., 2009; Ehrgott and Wiecek, 2005). In this paper, we selected the ε-constraint method which was introduced by Haimes et al. (1971) and an extensive discussion can be found in Chankong and Haimes (1983). The idea of this technique is to minimize or maximize one objective while the other objectives are bounded at acceptable fixed values. If we have a bi-objective problem, the formulation of the ε-constraint method is given as follows, refer to Ehrgott (2005).

The Bi-Objective Problem:		The ε-Constraint Problem:	
Minimize	$[f_1(x), f_2(x)]$	Minimize	$f_j(x)$
Subject to	$x \in X$.	Subject to	$f_k(x) \le \varepsilon_k, k = 1, 2; k \ne j$
			$x \in X$.

We briefly discuss the concept of optimality as it relates to multi-objective problems. A feasible solution $x \in X$, where X is the set of feasible solutions, is called "efficient" or "Pareto optimal", if there is no other $x \in X$ such that $f(x) \le f(x)$. If x is efficient, the point $y = f(x)$ is called non-dominated. A feasible solution $x \in X$ is called weakly efficient or weakly Pareto optimal, if there is no other $x \in X$ such that $f(x) < f(x)$. If x is weakly efficient, the point $y = f(x)$ is called weakly non-dominated (Ehrgott, 2005). By varying the value of ε_k , the non-dominated front can be generated. Solving a multi-objective problem results in a set of solutions, and the decision-maker should be interested in the Pareto set because it represents a solution that is better than any other with respect to at-least one of the criteria of interest.

To apply the ε-constraint approach to solve this problem, we have to reformulate the problem in the ε-constraint form. Since we have two kinds of objectives which are objective 1 and objective 2, we have two choices. The first choice is maximizing the objective 1 while the objective 2 is bounded at the acceptable level, and the second choice is minimizing the objective 2 while the objective 1 is bounded at the acceptable level. In this case, we chose the first option because of the following reasons. If we consider the value or the range of the objectives (Z_1 and Z_2), the objective that has lower value or smaller range should be bounded at the acceptable epsilon value. Since we have to run the optimization model by vary the epsilon value, the smaller range of epsilon would be a computational efficient choice. For example, if the value of Z_1 is between 0 and 1000 and we selected Z_1 to be bounded at the epsilon value, then we have to run the model 1000 times to get all solutions. If we consider the integer programming in the previous section, we see that constraint (2.5), $\sum_{k=1}^{l_i} y_{ik} \leq \sum_{j \in J_i} x_j$, works with the Maximizing objective, $Z_1 = \sum_{i=1}^{n} \sum_{k=1}^{l_i} h_i w_k y_{ik}$. For example, if all x_j is 0, then all y_{ik} is 0 and if one of x_j =1, which means station j is opened or one vehicle located at station j, the y_{ik} can be either 0 or 1. In the case that we maximize term y, y_{ik} is set to 1. So, if one vehicle is located, at least one demand zone should be covered by that vehicle. But if Z_1 is bounded, we cannot guarantee that located vehicles will cover all the demand in their radius. Moreover, to make sure that a demand zone *i*, which is covered by *k* vehicles, also covered by *k-1* vehicles, we have to add the constraint $y_{i,k} \geq y_{i,k+1}$ in the model. Because of these reasons, in this case, we select to maximize the first objective while the second objective is bounded at acceptable value, ε_2. Since we have three choices to be the second objective as shown in Equations (2.2a), (2.2b), (2.2c), there are three models, denoted as (a), (b), (c), depending on the objective chosen to be bounded (Z_{2a}, Z_{2b}, Z_{2c}, respectively), which incorporate fairness. Model (a) is represented as Equation (2.1) subject to Equations (2.3)-(2.7) and (2.12), model (b) is represented as Equation (2.1) subject to Equations (2.3)-(2.7) and (2.13), and model (c) is represented as Equation (2.1) subject to Equations (2.3)-(2.7) and (2.14). The entire ε-constraint problem is represented as below.

Maximize $$Z_1 = \sum_{i=1}^{n} \sum_{k=1}^{l_i} h_i w_k y_{ik} \tag{2.1}$$

Subject to: (2.3) - (2.7)

$$\max_{i \in U} \{ \min_{j \in O} (d_{ij}) \} \leq \varepsilon_{2a} \tag{2.12}$$

$$\sum_{i=1}^{n} (1 - y_{i1}) r_i \leq \varepsilon_{2b} \tag{2.13}$$

$$\sum_{i=1}^{n} (1 - y_{i1}) \leq \varepsilon_{2c} \tag{2.14}$$

Where: ε_{2a}= the acceptable bound of objective Z_{2a}

ε_{2b}= the acceptable bound of objective Z_{2b}

ε_{2c}= the acceptable bound of objective Z_{2c}

2.5 Case study

Our case study uses data from the Hanover Fire/EMS department, which is located in Hanover County, VA. The Hanover EMS department responds to 911 calls 24 hours a day and serves a county of 474 square miles, with a population nearing 100,000 individuals. Based on zoning, all locations within Hanover County are classified as either rural or urban. The data are collected from the Fire/EMS department during 2007, which capture the life-threatening calls received during 2007. Instead of assuming each demand point is located in the middle of an area, we divided the coverage area into 175 distinct zones. In this way, we ensure that coverage is more accurate and that originating demand is represented realistically. Currently, there are m = 16 existing station facility locations for locating EMS ambulances. All demand zones and station locations are shown in Figure 2.1, as we can see from the figure there are 6 stations in urban areas and 10 in rural areas. Moreover, each demand zone is classified as either rural or urban. The number of requested calls is collected separately for each demand zone. Based on the current data, requested calls did not originate from all 175 zones. Therefore, we ignore the zones that have no demand and only considered the n = 122 zones in which demand existed in 2007.

To set up the location of the station and demand zone, we drew grid lines over the area of interest, with one block representing 2 miles. The coordinates (a, b) of the stations and demand zones are used to calculate the distance between each demand zone and each station. Distance between two points can be measured in many ways (Drezner and Hamacher, 2004). The most familiar two are rectilinear distance and Euclidean distance. In this case we use the Euclidean metric because approximately 70% of the Hanover County area is rural, and can thus be reached via highways or county roads. Given a demand zone i at (a_i, b_i) and a station location j at (a_j, b_j), the distance (d_{ij}) between demand zone i and station j is calculated by using the Euclidian metric. In this case, there are 1711 calls spread over 122 demand zones; given the set of possible station locations, there are 4 zones that cannot be covered, since they are more than 4 miles from the closest possible station. Therefore, the maximum percentage of coverage is 98.8%.

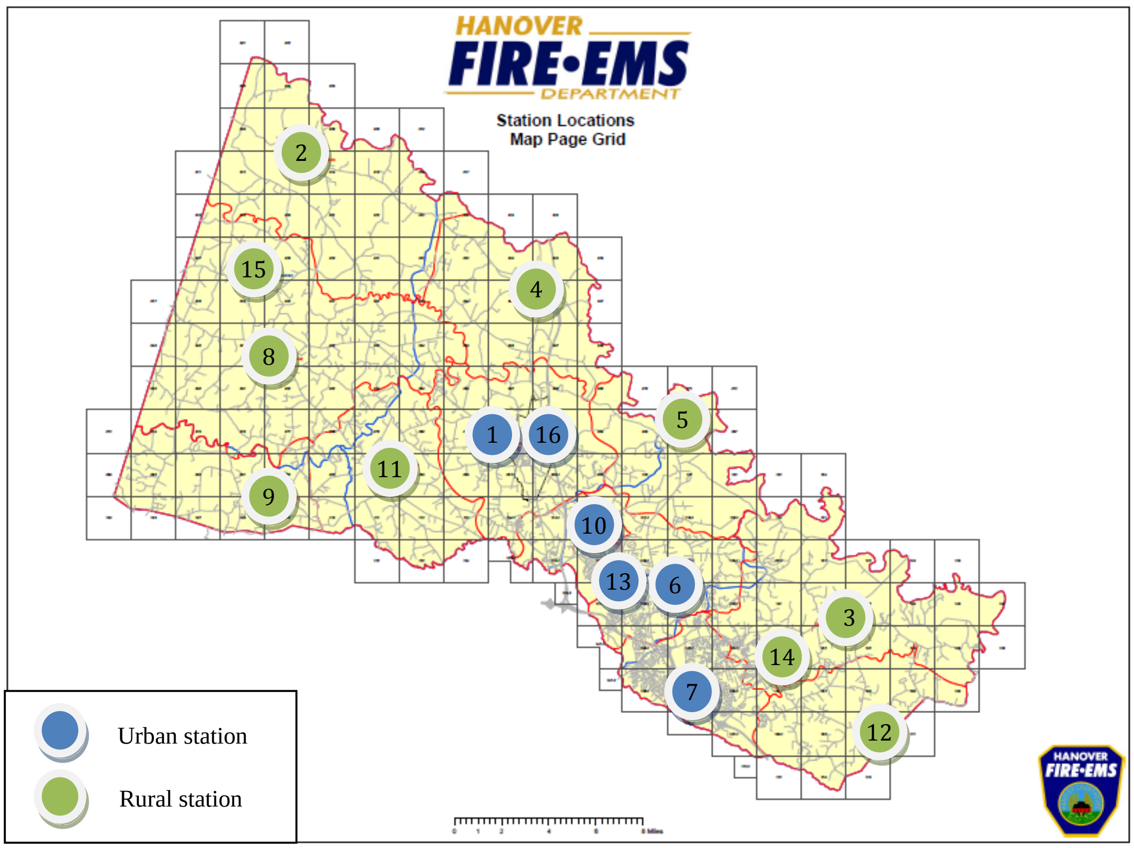


Figure 2.1: Map of existing station locations and demand zones

Based on the data during 2007, the average number of calls in Hanover is 1.2 calls/hour during the peak hours of operation when the call volume is essentially constant. The average service time per call is 74 minutes or 1.2 hours (note that this is not response time but also includes time in service). These data are used to compute the input parameters for our model.

2.6. Computational results

We use the data from the Hanover Fire/EMS department as described in Section 2.5, which is comprised of 122 demand zones and 16 possible station locations. We allow the total number of ambulances to be located in all stations to vary between 5 and 20, while the maximum number of ambulances that are allowed to be located at each station is 2. As a benchmark, we first consider the results using a single objective of maximizing the expected number of calls that are covered (Z_1). For clarity, the objective function value is rescaled by the total number of calls to reflect the proportion of calls that are covered, shown in Table 2.1. As we increase the number of ambulances, the probability of a randomly selected vehicle being busy decreases, and the number of calls that are covered increases. Using a single objective, at least two ambulances are always located at station 1 and 6, since it is located in an urban area and can serve a number of high call volume demand zones nearby. Stations 3, 4, 11, 14 and 15 are the next likely to be selected because they located near urban areas. Conversely, stations 2, 5 and 12, located in remote areas, are only selected when the number of available ambulances is high, or when coverage is already high.

Table2.1: Single-objective results that maximize covered demand Z_1

Number of ambulances	Prob. of vehicle being busy	Expected demands that covered (calls)	Opened stations (stations in bold face are located in rural areas)	Number of ambulances at each station	Coverage percentage
5	0.296	1260.3	{1 6 **14**}	{2 2 1}	73.66
6	0.247	1365.0	{1 6 **14 15**}	{2 2 1 1}	79.78
7	0.211	1448.4	{1 **4** 6 **14 15**}	{2 1 2 1 1}	84.65
8	0.185	1508.1	{1 **4** 6 7 13 **14 15**}	{2 1 1 1 1 1 1 }	88.14
9	0.164	1558.5	{1 **4** 6 7 **11** 13 **14 15**}	{2 1 1 1 1 1 1 1}	91.09
10	0.148	1589.3	{1 **3 4** 6 7 **9 11** 13 **15**}	{2 1 1 1 1 1 1 1 1}	92.89
11	0.135	1614.6	{1 **3 4 5** 6 7 **9 11 15**}	{2 1 1 1 2 1 1 1 1}	94.37
12	0.123	1636.5	{1 **2 3 4 5** 6 8 **9 11 14**}	{2 1 1 1 1 2 1 1 1 1}	95.65
13	0.114	1648.9	{1 **2 3 4 5** 6 7 8 **9 11** 13 **14**}	{2 1 1 1 1 1 1 1 1 1 1 1}	96.37
14	0.106	1659.4	{1 **2 3 4 5** 6 7 8 **9 11 12 15**}	{2 1 1 1 1 2 1 1 1 1 1 1}	96.98
15	0.099	1668.0	{1 **2 3 4 5** 6 7 8 **9 11 12 15**}	{2 1 1 2 1 2 1 1 1 1 1 1}	97.49
16	0.093	1675.3	{1 **2 3 4 5** 6 7 8 **9 11 12** 13 **14 15**}	{2 1 1 2 1 1 1 1 1 1 1 1 1 1}	97.91
17	0.087	1679.6	{1 **2 3 4 5** 6 7 8 **9** 10 **11 12** 13 **14 15**}	{2 1 1 2 1 1 1 1 1 1 1 1 1 1 1}	98.16
18	0.082	1682.7	{1 **2 3 4 5** 6 7 8 **9** 10 **11 12** 13 **14 15**}	{2 1 1 2 1 1 1 1 1 1 2 1 1 1 1}	98.35
19	0.078	1684.9	{1 **2 3 4 5** 6 7 8 **9** 10 **11 12 14 15**}	{2 1 1 2 2 2 1 1 1 1 2 1 1 1}	98.47
20	0.074	1686.8	{1 **2 3 4 5** 6 7 8 **9** 10 **11 12 14 15**}	{2 1 1 2 2 2 1 1 2 1 2 1 1 1}	98.59

When there are five ambulances, this single objective covering location model locates four ambulances in urban areas and one vehicle in a rural area, thus the majority of the uncovered demand zones are rural. To reduce the disparity between service in urban and rural areas, we would like to provide decision makers with more alternatives. Thus, we construct a bi-objective model which not only considers maximizing the expected number of calls that can be covered but simultaneously improves fairness to rural patients. As discussed in Section 4, we solve this problem using the ε-constraint method by first formulating the problem in the ε-constraint form. Then, we find bound of ε_{2a}, ε_{2b}, and ε_{2c}. If we solve the problem (Equations 2.3-2.7) with the objective Z_1 (Equations 2.1) to the optimality we get the upper bound of ε_{2a}, ε_{2b}, and ε_{2c} by calculating the values of Z_{2a}, Z_{2b}, and Z_{2c} after reaching optimal. Alternatively, we can solve the problem (Equations 2.3-2.7) with objective Z_{2a}, Z_{2b}, and Z_{2c} (Equations 2.2a, 2.2b, and 2.2c) one at a time, which yields a lower bound for ε_{2a}, ε_{2b}, ε_{2c}, respectively.

Detailed results are provided for the case with 5 ambulances though similar conclusions hold for the case of more ambulances; furthermore, a discussion is later provided regarding the effects of increasing the number of ambulances. With 5 ambulances to be located, we find the expected demand that can be covered is 1260.3 and the upper bounds of the maximum distance between uncovered demand zones and their closest opened stations (ε_{2a}), the number of uncovered rural demand zones(ε_{2b}) and the number of uncovered demand zones(ε_{2c}) are 18, 63 and 69, respectively. To get the solution points, we solve the problem by maximizing the first objective while varying the value of ε_2. Since we have three alternatives of the second objective, we have three models to solve. Figures 2.2-2.4 show all the solution points that are found by maximizing the first objective while decreasing the values of ε_{2a}, ε_{2b}, and ε_{2c} from their upper bounds down to the smallest values that still give the feasible solution incremented by 1. Note that if we choose to minimize the second objective and bound the first objective at the acceptable value ε_1, we have to solve the problem about 1000 times because the value of the first objective is in the range [0, 1260.3] while if we choose to maximize the first objective and bound the second objective at the acceptable value ε_2, we only solve the problem less than 100 times because value of the three choices to be the second objective are in the range [8, 18], [26, 63], and [36, 69]. These problems were solved using the optimization software ILOG OPL 5.5 on a Dell Latitude D410 machine with Intel Pentium processor 1.73 GHz, 1 GB of RAM, the run time was between 1 and 2 seconds per sub-problem. In Figures 2.2-2.4, an open circle represents a solution point of the ε-constraint method; a solid circle represents a non-dominated solution.

When Z_{2a} is selected, we build model (a) by selecting Z_{2a} to be the second objective. From Figure 2.2, we see that for Z_{2a} values between 8 and 18 blocks (16 and 36 miles) resulting the first objective values range between 1150.3 and 1260.3 calls. Similarly, we build Models (b) and (c) by selecting Z_{2b} and Z_{2c} as the second objective, respectively. Figure 2.3 shows all solution points for Model (b) which minimizes the number of uncovered rural demand zones as a secondary objective. From Figure 2.3, we see that for second objective values in the range of 26 and 63 zones the resulting second objective values are between 515.0 and 1260.3 calls. Figure 2.4 shows all solution points for Model (c) which minimizes the number of uncovered demand zones; when the second objective values are between 36 and 69 zones the resulting first objective values are between 1091.1 and 1260.3 calls.

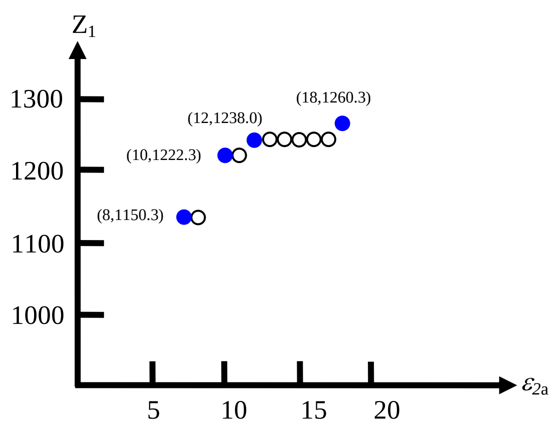


Figure 2.2: Solution points of Model (a) -- second objective is to minimize the distance between uncovered demand and opened stations, with solid circles representing non-dominated solutions

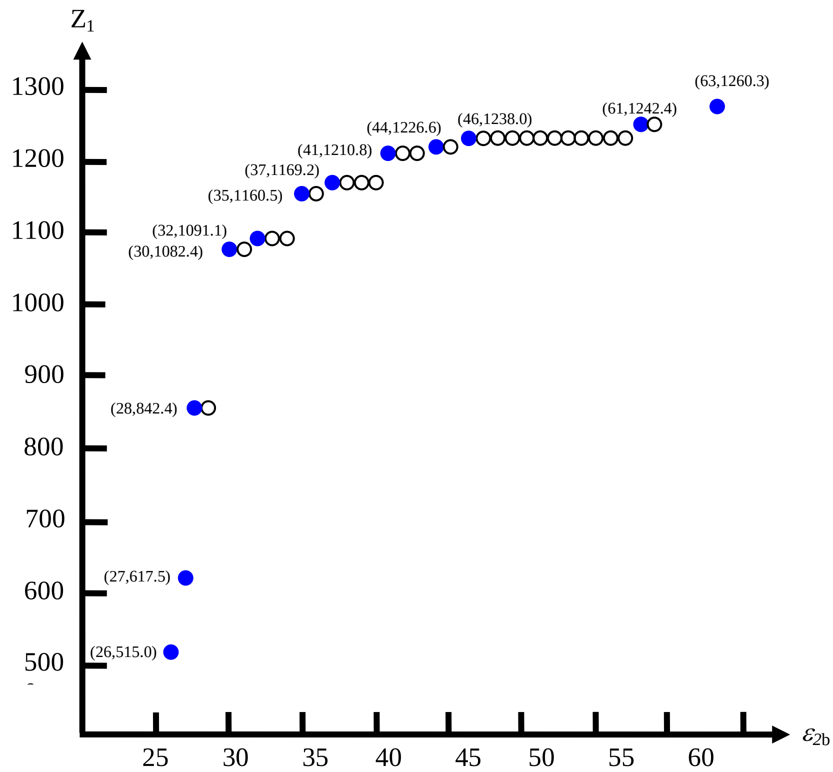


Figure 2.3: Solution points of Model (b) -- second objective is to minimize the number of uncovered rural demand zones, with solid circles representing non-dominated solutions

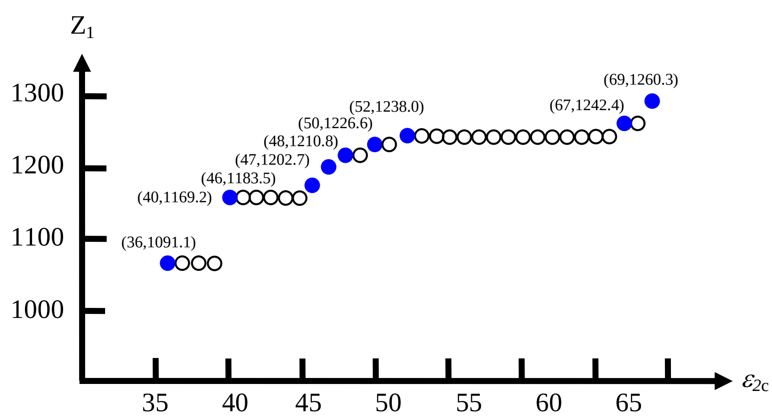


Figure 2.4: Solution points of Model (c) -- second objective is to minimize the number of uncovered demand zones, with solid circles representing non-dominated solutions

For all three models we see that the best first objective value Z_1 is reached at the maximum ε_2, and if we decrease the second objective function value by decreasing ε_2, the Z_1 objective function value deteriorates. The solution points shown in solid circles are the non-dominated solutions. All three models contain the single objective solution as part of the dominated set with $Z_1 = 1260.3$. The other solutions "improve" the issue of fairness by trading off for lower number of covered calls. The values of the objective functions and the corresponding non-dominated solution sets of Models (a)-(c) are shown in Tables 2.2-2.4, respectively. In general, if we decrease the number of calls that must be covered, more stations are opened in rural areas, decreasing disparity of service at the expense of losing patients in urban areas. Note that the use of Model (a) tends to locate ambulances at the most remote stations (i.e., those near the edges of the county) to reduce the maximum distance between uncovered demand zones and closest opened stations. Station 2 is the most isolated station, and the results of Model (a) indicate that station 2 is open when $Z_1 \leq 1150.3$ (this is not obvious from Table 2.2 since it shows only non-dominated solutions), however as Z_1 is increased more urban stations need to be included and the remote stations drop out of the solution set, since these have low call volumes. Also, Model (a) yields a small solution (efficient solution set). This implies that the distance between uncovered zones and stations can be minimized without sacrificing expected coverage when coverage is not required to be too high.

Model (b) also opens stations in rural areas, in contrast to the stations chosen in Model (a), stations are opened in order to cover as many as rural demand zones as possible, independent of how far away uncovered stations may be from opened stations.

Thus, Model (b) opens rural stations having relatively high demands. Model (c) minimizes the total number of uncovered demand zones, in either rural or urban areas. Thus Model (c) opens more urban stations than Models (a) or (b). In contrast to the single objective model, the decision maker may choose to sacrifice call volume to cover a larger geographical region. For example, instead of the single objective solution which places one ambulance at one rural station (and leaves 63 rural zones uncovered) for an expected coverage of 1260.3, the decision maker may choose to reduce the expected coverage to 1210.8 calls but leave only 41 rural stations uncovered; in other words, coverage may be decreased by only 5% while zones covered are increased by 35%. We also note that, Models (b) and (c) tend to "split" ambulances, rather than pairing ambulances at the same station. This is desirable when wanting to increase the number of covered zones, though it may be undesirable in the sense that it reduces backup coverage in high demand areas. For many decision makers a large solution set might be desirable, as it represents more options to choose from. In this case, Model (b) provides the largest set of efficient solutions.

Clearly, using any of the bi-objective models proposed in this paper provides more alternatives to the decision makers. However, comparing these solutions is difficult. In practice, if we cannot introduce specific metrics directly within the objective function, providing performance metrics for post-analysis evaluation is useful for comparing different portfolios of ambulance locations. Thus, for each model, we include the result of all three secondary objectives, as well as the average distance from uncovered zones to the closest open station (which captures the average distance between demand zones and the closest open station), and the weighted average distance from uncovered zones to the closest open station (which captures the average distance an ambulance can expect to travel per call). These latter two measures are reported in Tables 2.2, 2.3 and 2.4 for Models (a), (b) and (c), respectively. Note that they could not be included as objective functions in our models because they can only be formulated as nonlinear functions.

If we use the average distance or the weighted average distance from an uncovered zone to its closest station as supplementary criteria, the results from using Model (a) may be preferred to the results from using Models (b) and (c); we achieve the same objective function value Z_1, but improve these supplementary criteria simply by opening a different set of stations. That is, even though more rural zones are uncovered, these uncovered zones are located closer to open stations, improving the chance that they will receive service in a reasonable time. This is important since response time is directly linked to survivability (McLay and Mayorga, 2009).

The results shown here are for the case of five available ambulances. With a single objective the best, we can achieve is 73.6% coverage with 69 uncovered zones and an average distance of 10 blocks (20 miles) from uncovered demand zones to open stations. Using a bi-objective model, we can reduce the average distance from uncovered zones to open stations by sacrificing coverage. The only way to improve both metrics (decrease the average distance between uncovered demand zones and open stations and increase the number of calls that can be covered), is to increase the number of ambulances to be located. Increasing the number of ambulances improves all metrics.

Furthermore the reduction in disparities that can be achieved between a single objective and a bi-objective model are reduced. This is not necessarily because the model is less efficient but rather because there are enough ambulances to improve service for all regions. Therefore, the proposed bi-objective model may be most useful when there are fewer available ambulances, when identifying fair ambulance location portfolios is most difficult.

Table 2.2: The non-dominated solutions in the objective space of Model (a)

Expected demands covered (Z_1)	Maximum distance from uncovered zone to closest station ($Z_{2.a}$)	Opened stations {rural; urban}	Number of uncovered rural zones	Number of uncovered zones	Number of uncovered demands	Average distance from uncovered zone to closest station	Weighted average distance from uncovered zone to closest station	SDEV.
1150.3	8	{2 8 14; 6 1}	36	42	196	6.5	6.3	1.00
1222.3	10	{8 14; 6 13 1}	44	50	206	7.0	6.6	1.40
1238.0	12	{14 15; 1 6}	46	52	219	7.2	6.9	1.77
1260.3	18	{14; 1 6}	63	69	296	10.0	9.0	3.66

Table 2.3: The non-dominated solutions in the objective space of Model (b)

Expected demands covered (Z_1)	Number of uncovered rural zones ($Z_{2.b}$)	Opened stations {rural; urban}	Maximum distance from uncovered zone to closest station	Number of uncovered zones	Number of uncovered demands	Average distance from uncovered zone to closest station	Weighted average distance from uncovered zone to closest station	SDEV.
515.0	26	{2 3 8 9; 1}	10	45	982	6.6	6.7	1.14
617.5	27	{2 3 8 11; 13}	12	40	838	6.9	6.4	1.65
842.4	28	{3 9 15; 1 13}	10	37	529	6.4	6.0	1.02
1082.4	30	{3 9 15; 1 6}	10	37	186	6.4	6.1	1.05
1091.1	32	{3 11 15; 1 6}	10	36	180	6.6	6.3	1.32
1160.5	35	{9 14 15; 1 6}	10	41	180	6.5	6.2	1.08
1169.2	37	{11 14 15; 1 6 }	10	40	174	6.7	6.4	1.32
1210.8	41	{3 15; 1 6}	12	48	225	7.3	6.8	1.81
1226.6	45	{14 15; 1 6 13 }	12	50	200	7.3	7.0	1.80
1238.0	46	{14 15; 1 6}	12	52	219	7.2	6.9	1.77
1242.4	61	{14; 1 6 13}	18	67	277	10.1	9.2	3.67
1260.3	63	{14; 1 6}	18	69	296	10.0	9.0	3.66

Table 2.4: The non-dominated solutions in the objective space of Model (c)

Expected demands covered (Z_1)	Number of uncovered zones ($Z_{2.c}$)	Opened stations {rural; urban}	Maximum distance from uncovered zone to closest station	Number of uncovered rural zones	Number of uncovered demands	Average distance from uncovered zone to closest station	Weighted average distance from uncovered zone to closest station	SDEV
1090.1	36	{3 11 15; 1 6}	10	32	180	6.6	6.3	1.32
1169.2	40	{11 14 15; 1 6}	10	30	174	6.7	6.4	1.32
1183.5	46	{4 14 15; 1 6}	12	42	152	7.1	7.2	1.84
1202.7	47	{3 7 15; 1 6}	12	41	208	7.3	7.0	1.80
1210.8	48	{3 15; 1 6}	12	41	225	7.3	6.8	1.80
1226.6	51	{14 15; 1 6 13}	12	44	200	7.3	7.0	1.80
1238.0	52	{14 15; 1 6}	12	46	219	7.2	6.9	1.77
1242.4	67	{13 14; 1 6}	18	61	277	10.1	9.2	3.67
1260.34	69	{14; 1 6}	18	63	296	10.0	9.0	3.66

2.7 Equitably efficient solution

In this section, we proposed a way to analyze the solutions for the multi-objective problem by using the concept of equitable efficient solution and criteria aggregation which have been discussed by Ogryczak (2000). In a facility location problem in which we try to place facilities in order to serve customers, instead of looking at a problem as a whole system and trying to achieve the overall outcome of the system, we can view the problem individually in terms of need of each customer. Then, the facility location problem can be considered as a multi-criteria or multi-objective problem where an individual objective is defined for each customer. The effect of a location pattern on each customer can be defined as a traveled distance from each customer's location to the closest facility. The objective is to minimize the individual effect with respect to the distribution of facility location. By minimizing all individual objectives results in minimizing the effect of the system. This multiple criteria location model allows us to apply the concept of efficient solution, which is able to link to the equitably efficient solution. Ogryczak (2000) introduced the term of equitably efficient solution and presented some aggregations of criteria that can be applied to select equitably efficient solutions in multiple criteria analysis. The focus of equitable solution is on the distribution of outcomes. For example, consider a facility location problem that seeks to locate a vehicle among 3 zones so as to minimize the traveled distance from each zone to the facility. A solution can be evaluated as a distance vector **d**=(d_1, d_2, d_3), and we can formulate this problem as a multi-objective problem; Min [$f_1(x)$, $f_2(x)$, $f_3(x)$] subject to $x \in X$ where X is a set of candidate locations for locating a facility, and $f_i(x) = d_i$ = traveled distance from zone i to facility. Note that this formulation is not practical since we have to carry many objectives, so in most facility problems an aggregation of the objectives is more likely to be used. For example instead of minimizing each of three objectives we might minimize the summation of these three objectives. As we described in Section 2.3, a feasible solution $x \in X$, where X is the set of feasible solutions, is called "efficient" or "Pareto optimal", if there is no other $x \in X$ such that $f(x) \le f(x)$. Suppose there are three solutions; a:(0,5,3), b:(2,0,2), and c:(2,1,0). So we have $\boldsymbol{x}^{\mathrm{a}}$={1,0,0}, $\boldsymbol{x}^{\mathrm{b}}$={0,1,0}, $\boldsymbol{x}^{\mathrm{c}}$={0,0,1} and $\mathbf{d}^{\mathbf{a}} = (d_1^a, d_2^a, d_3^a)$ =(0,5,3), $\mathbf{d}^{\mathbf{b}} = (d_1^b, d_2^b, d_3^b)$ =(2,0,2), and $\mathbf{d}^{\mathbf{c}} = (d_1^c, d_2^c, d_3^c)$ =(2,1,0), respectively. In this case, solution (a) is the most preferable for zone 1 while solutions (b) and (c) are the most preferable for zone 2 and 3, respectively. Since we treat everyone equally, these three solutions are considered equally good. In fact, these solutions are efficient according to the definition of "efficient" in the multi-objective problem as we mentioned earlier. However, suppose we have another solution $\mathbf{d}^{\mathbf{e}}$:(2,2,2), it should be considered better than the previous three solutions in terms of providing equal distribution of traveled distance from each zone to the facility, but it is not efficient. Ideally, we want an efficient solution that also provides equal distribution of the traveled distance to all zones, or in other words, an equitably efficient solution.

From Ogryczak (2000), we briefly detail the concept of equitably efficient solution as follows. A feasible solution $x \in X$ is "equitably efficient" for the multiple criteria problem: Min { $f_i(x)$, $i \in N=\{1,2,\ldots,n\}$: $x \in X$}, if and only if there does not exist any $x' \in X$ such that $f(x') \prec_e f(x)$. Note that each equitably efficient solution is also a Pareto-optimal solution, but not vice versa. The relation of equitable dominance $\prec_e$ can be expressed as a vector of inequalities on the cumulative ordered outcomes. Let $\boldsymbol{v}=f(x)$, and $\boldsymbol{\Theta}(v) = (\theta_1(v), \theta_2(v), \ldots, \theta_n(v))$ where $\theta_1(v) \geq \theta_2(v) \geq \cdots \geq \theta_n(v)$, and there exists a permutation τ of set n such that $\theta_i(v) = v_{\tau(i)}$ for i=1, 2, …, n. The cumulative ordering map is defined as $\bar{\boldsymbol{\Theta}}(v) = (\bar{\theta}_1(v), \bar{\theta}_2(v), \ldots, \bar{\theta}_n(v))$ where $\bar{\theta}_i(v) = \sum_{j=1}^{i} \bar{\theta}_j(v)$ for i=1, 2, …, n. Achievement vector v' equitably dominates v'', if and only if $\bar{\theta}_i(v') \leq \bar{\theta}_i(v'')$ for all $i \in N$ where at least one strict inequality holds. In other word, a location pattern $x \in X$ is an equitably efficient solution of problem Min $\{f_i(x)$, $i \in N$: $x \in X\}$, if and only if it is an efficient solution of problem Min $\{\bar{\theta}_i(f(x))$, $i \in N$: $x \in X\}$. If we apply the concept of cumulative ordered outcome to our previous example, we get three ordered traveled distance vector of three solutions as $\boldsymbol{\Theta}^a(v)$=(5,3,0), $\boldsymbol{\Theta}^b(v)$=(2,2,0), $\boldsymbol{\Theta}^c(v)$=(2,1,0), and three cumulative ordered traveled distance vectors of three solutions as $\bar{\boldsymbol{\Theta}}^a(v)$=(5,8,8), $\bar{\boldsymbol{\Theta}}^b(v)$=(2,4,4), $\bar{\boldsymbol{\Theta}}^c(v)$=(2,3,3). In order to see which solution provides an equitably efficient solution, we plotted the cumulative ordered traveled distance values in Figure 2.5. We see that solution (a) is dominated by solutions (b) and (c) while solution (b) is dominated by solution (c). So, solution (c) is equitably efficient. Aggregation criteria helps provide us with further analysis that can be applied to multi-criteria problems to help decision makers decide between several alternatives. Instead of looking at the actual criteria (v_i), we can look at the aggregation criteria ($\bar{\theta}_i(v)$) for finding the equitably efficient solutions. The cumulative ordered outcome is one of several aggregations that have been mentioned in Kostreva et al. (2004).

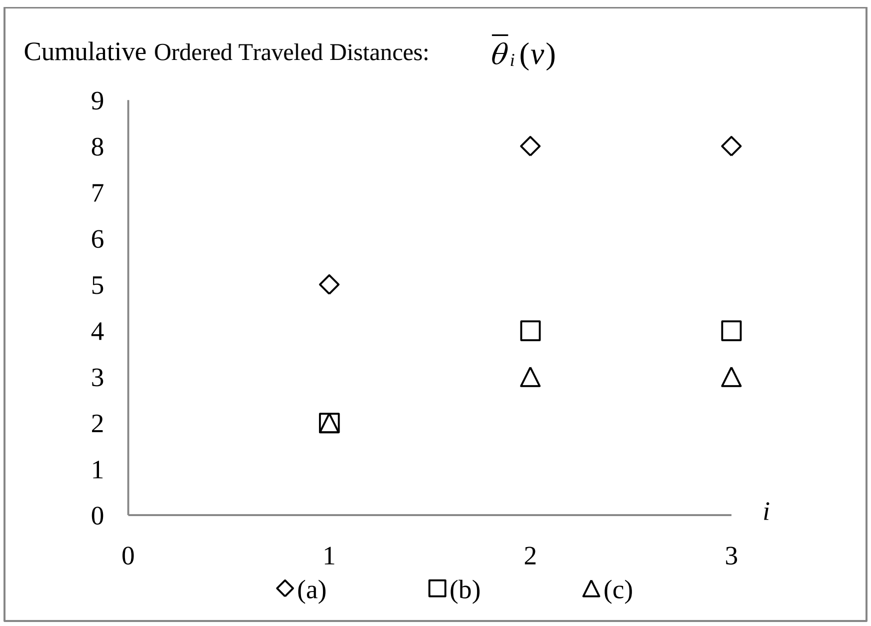


Figure 2.5: Cumulative ordered outcomes of a three-zone location problem

Previously, we have proposed three bi-objective models. In this section, we would like to apply the concept of cumulative ordered outcome to show which, if any, model yields an equitably efficient solution. Let v_i represents an individual outcome of our problem which is defined as the distance from location of a customer at zone i to its closest station, where number of zones=122 and number of stations=16. Therefore, we have a multi-objective problem as follow: Min $\{v_i=f_i(\boldsymbol{x}),\ i\in N:\ \boldsymbol{x}\in X\}$, where $v_i=f_i(\boldsymbol{x})=\min_{j\in O}\{d_{ij}\}$=traveled distance from zone i to its closest opened station according to a location solution $\boldsymbol{x}=(x_1, x_2, \ldots, x_{16})$=number of ambulances located at each station, N is a set of zones; $N=\{1, 2, \ldots, 122\}$, and O is a set of opened stations; $O=\{j: x_j\geq 1\}$. Note that a location solution $\boldsymbol{x}$ is obtained from each bi-objective model. Since each bi-objective model produced multiple optimal solutions, we selected a solution that yielded the best value of Z_2 objective (minimum disparity in service between rural and urban zones) of each model. For each model, the solution can be seen in row 1 of Tables 2.2-2.4, respectively. Note that one could also use this methodology to compare all efficient solutions in one model, or to compare solutions between models. Here, we choose to compare the solutions between models that yield the minimum disparity, as defined by that model, but the analysis described below can be applied directly to other comparisons. Thus, by picking one solution from each model, we have $\boldsymbol{x}^{\mathrm{a}}$, $\boldsymbol{x}^{\mathrm{b}}$, $\boldsymbol{x}^{\mathrm{c}}$, and next we calculate the outcome vector of each model, $\boldsymbol{v}^{\mathrm{a}}$, $\boldsymbol{v}^{\mathrm{b}}$, $\boldsymbol{v}^{\mathrm{c}}$. Then, we applied the cumulative ordered outcome by sorting the outcomes v_i of vector $\boldsymbol{v}=(v_1, v_2, \ldots, v_{122})$ in descending order to obtain vectors $\boldsymbol{\Theta}^a(v)$, $\boldsymbol{\Theta}^b(v)$, $\boldsymbol{\Theta}^c(v)$ and aggregating the sorted outcomes to obtain vectors $\overline{\boldsymbol{\Theta}}^a(v)$, $\overline{\boldsymbol{\Theta}}^b(v)$, $\overline{\boldsymbol{\Theta}}^c(v)$. Note that the first value in vector $\overline{\boldsymbol{\Theta}}(v)$ is the worst outcome, and then the second value is the sum of the worst and the second worst outcomes, and so on. The results of the cumulative ordered outcomes, $\overline{\boldsymbol{\Theta}}(v)$, generated by the three proposed bi-objective models are shown in Figure 2.6. We found that by considering the outcomes v_i, the solutions chosen from all three bi-objective models produce efficient solutions, but by considering the cumulative ordered outcomes $\bar{\theta}_i(v)$, the solution of Model (b) is dominated by the solution of Models (a) and (c), which means according to these three solutions chosen from each model, only the solution of Models (a) and (c) provide equitably efficient solutions. Particularly, Model (a) produced the lowest cumulative ordered outcomes among these three models in the first 44 worst outcomes while Model (c) also produced lowest cumulative ordered outcomes for the remaining outcomes 44 to 1711. Therefore, both Models (a) and (c) yielded efficient solutions of problem Min $\{\bar{\theta}_i(f(\boldsymbol{x}))$, $i \in N$: $\boldsymbol{x} \in X\}$ which result in equitably efficient solutions to the original location problem Min $\{f_i(\boldsymbol{x})$, $i \in N$: $\boldsymbol{x} \in X\}$.

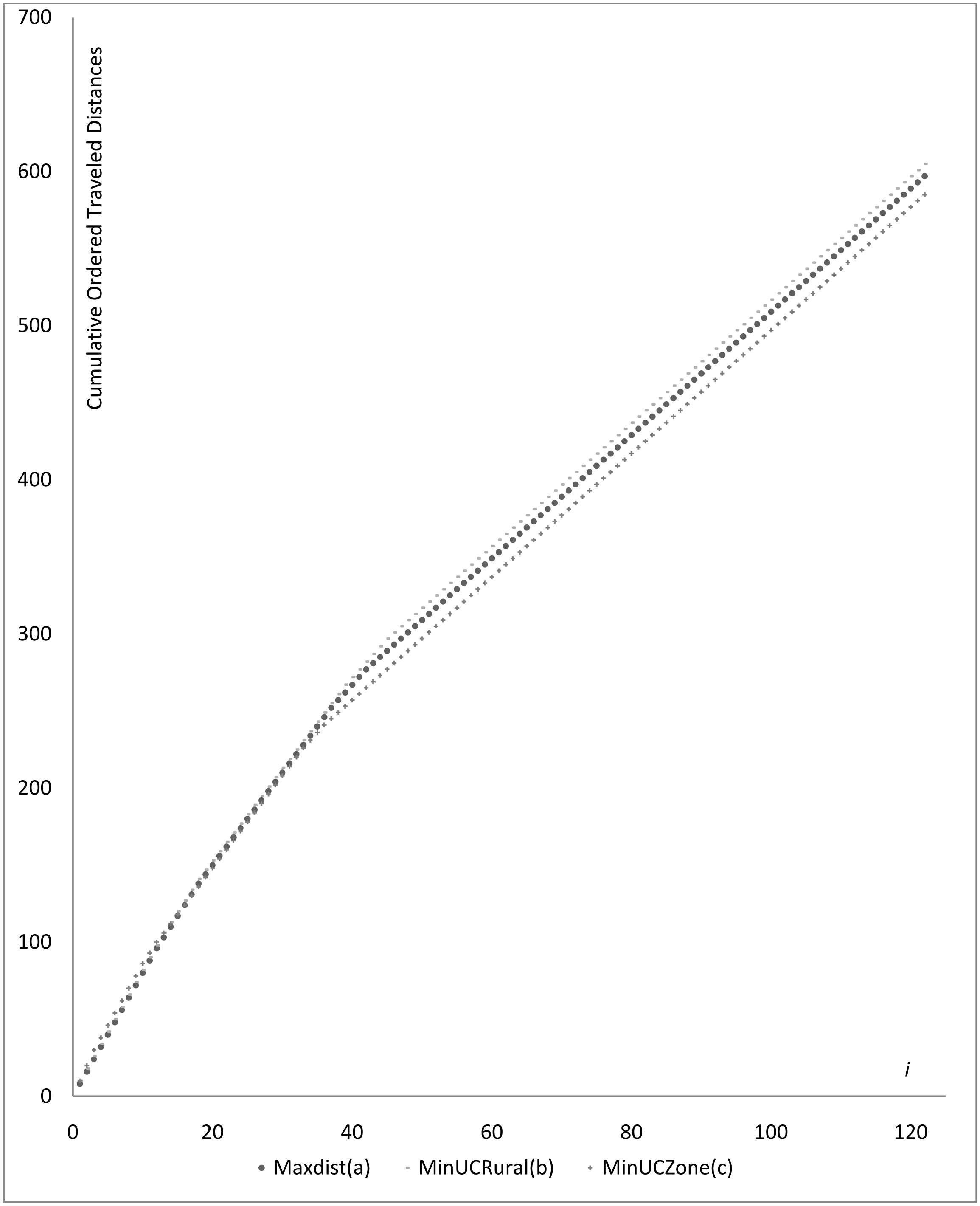


Figure 2.6: Comparison of the cumulative ordered outcomes generated by three proposed bi-objective models

2.8 Sensitivity Analysis

In this section, we study how the demand volume and probability of vehicle being busy impact the system. We varied the number of calls per hour (λ) from 1.0 to 1.5 (current λ=1.2), then recalculated the probability that a randomly selected vehicle will be busy, *p*, which also results in changes to the probability of the vehicle k^{th} being available (w_k). The values of probabilities p_b, w_1, and w_2 for all 16 cases; number of vehicles varied from 5 to 20, are shown in Table 2.5. Figure 2.7 shows probability of the first vehicle being available (w_1) at each case when λ is varied.

Table 2.5: Changes in probabilities of busy/available when λ is varied from 1.0 to 1.5

p	Demand decreases				Current Demand		Demand increases					
	λ=1.0		λ=1.1		λ=1.2		λ=1.3		λ=1.4		λ=1.5	
	p_b	w_1 w_2	p_b	w_1 w_2	p_b	w_1 w_2	p_b	w_1 w_2	p_b	w_1 w_2	p_b	w_1 w_2
5	0.2466	0.753 0.171	0.2713	0.729 0.180	0.296	0.704 0.188	0.3206	0.679 0.194	0.3453	0.655 0.199	0.3699	0.630 0.203
6	0.2055	0.794 0.155	0.2261	0.774 0.165	0.247	0.753 0.174	0.2672	0.733 0.182	0.2877	0.712 0.189	0.3083	0.692 0.195
7	0.1761	0.824 0.140	0.1938	0.806 0.150	0.211	0.789 0.159	0.2290	0.771 0.168	0.2466	0.753 0.176	0.2642	0.736 0.183
8	0.1541	0.846 0.127	0.1695	0.830 0.137	0.185	0.815 0.146	0.2004	0.800 0.155	0.2158	0.784 0.163	0.2312	0.769 0.170
9	0.1370	0.863 0.116	0.1507	0.849 0.125	0.164	0.836 0.134	0.1781	0.822 0.142	0.1918	0.808 0.150	0.2055	0.794 0.158
10	0.1233	0.877 0.106	0.1356	0.864 0.115	0.148	0.852 0.124	0.1603	0.840 0.132	0.1726	0.827 0.140	0.1849	0.815 0.147
11	0.1121	0.888 0.098	0.1233	0.877 0.107	0.135	0.866 0.115	0.1457	0.854 0.122	0.1569	0.843 0.130	0.1681	0.832 0.137
12	0.1027	0.897 0.091	0.1130	0.887 0.099	0.123	0.877 0.107	0.1336	0.866 0.114	0.1438	0.856 0.121	0.1541	0.846 0.128
13	0.0948	0.905 0.085	0.1043	0.896 0.093	0.114	0.886 0.100	0.1233	0.877 0.107	0.1328	0.867 0.114	0.1423	0.858 0.120
14	0.0880	0.912 0.080	0.0969	0.903 0.087	0.106	0.894 0.094	0.1145	0.885 0.100	0.1233	0.877 0.107	0.1321	0.868 0.113
15	0.0822	0.918 0.075	0.0904	0.910 0.082	0.099	0.901 0.088	0.1068	0.893 0.095	0.1151	0.885 0.101	0.1233	0.877 0.107
16	0.0770	0.923 0.071	0.0847	0.915 0.077	0.093	0.908 0.083	0.100	0.900 0.089	0.1079	0.892 0.095	0.1156	0.884 0.101
17	0.0725	0.927 0.067	0.0798	0.920 0.073	0.087	0.913 0.079	0.0943	0.906 0.085	0.1015	0.898 0.091	0.1088	0.891 0.096
18	0.0685	0.931 0.064	0.0753	0.925 0.069	0.082	0.918 0.075	0.0890	0.911 0.081	0.0959	0.904 0.086	0.1027	0.897 0.092
19	0.0649	0.935 0.060	0.0714	0.929 0.066	0.078	0.922 0.072	0.0843	0.916 0.077	0.0908	0.909 0.082	0.0973	0.903 0.087
20	0.0616	0.938 0.058	0.0678	0.932 0.063	0.074	0.926 0.068	0.0801	0.920 0.073	0.0863	0.914 0.078	0.0924	0.908 0.083

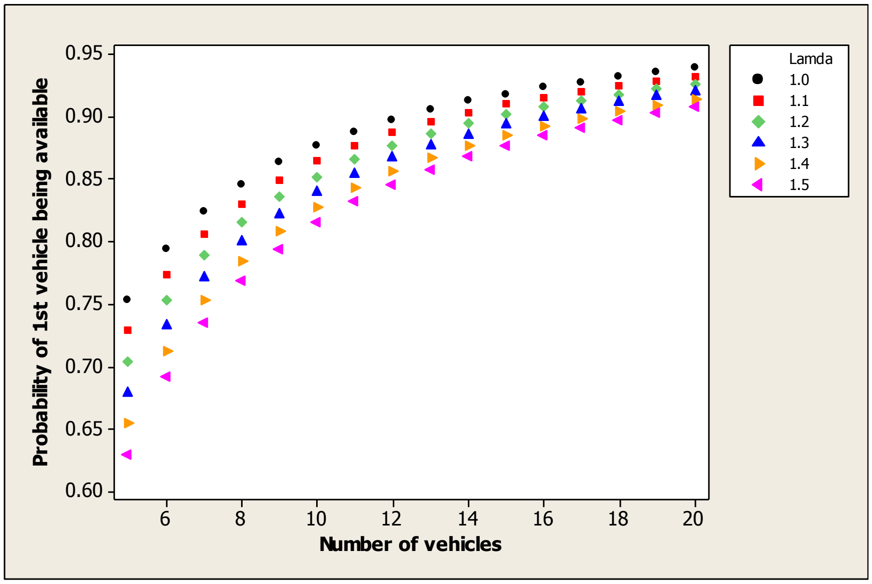


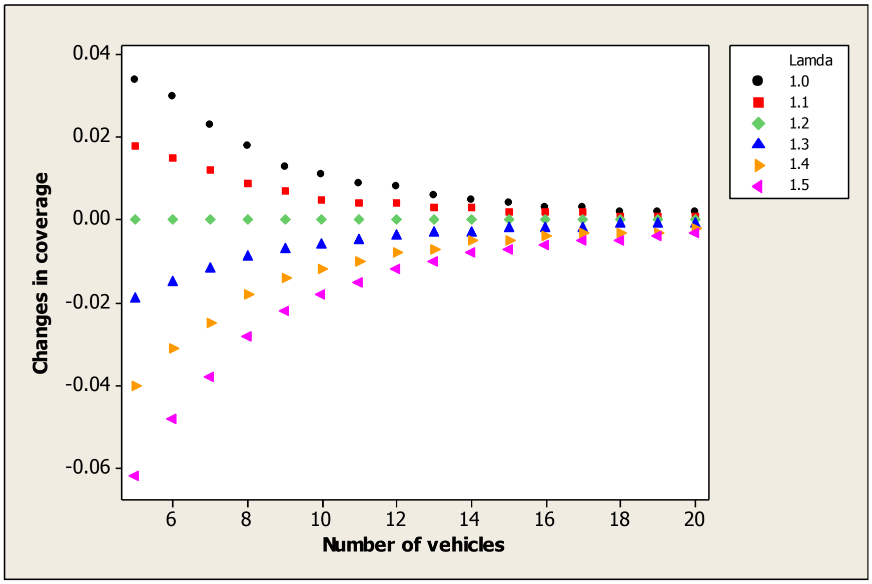
Figure 2.7: Probability of the first vehicle being available when λ is varied from 1.0 to 1.5

We see in Table 2.5 that changes in density of demand not only affect the value of p_b, it also affects the value of w at stations. As we see from Figure 2.7, the probability of the first vehicle being available increases when the arrival rate decreases, and this affect is decreased as number of vehicles increases. Then, we rerun the model with the objective Z_1 which is maximize the number of demand that can be covered to see how the changes affect the coverage and the locations of the current system. The details of results are reported in Table 2.6. Figure 2.8 and Figure 2.9 show the changes in coverage and the proportion of vehicles that need to be relocated in each case (p), respectively. By changing the number of calls or the probability of a particular vehicle being busy, the location of the facilities changed from 3% to 4.5% while the coverage changed from -1.4% to 0.9%. If the number of calls decrease (λ<1.2), the ambulances at primary stations tend to have more chance of being available, with results in higher coverage. On the other hand, if the number of calls increase (λ>1.2), the ambulance at primary stations tend to have less chance of being available, with also results in lower coverage. Moreover, over all cases the system is not much affected by decreases in demand.

Table 2.6 Changes in coverage and facility locations when λ is varied from 1.0 to 1.5

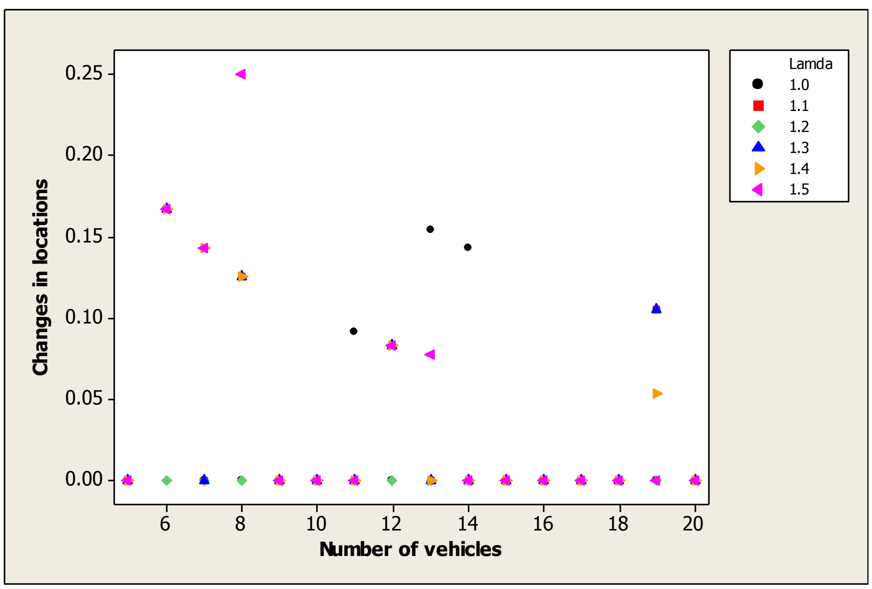
p	Changes in λ=1.0		Changes in λ=1.1		Current system λ=1.2		Changes in λ=1.3		Changes in λ=1.4		Changes in λ=1.5	
	Coverage	Locations	Coverage	Locations	Coverage	Locations	Coverage	Locations	Coverage	Locations	Coverage	Locations
5	0.000	0.000	0.000	0.000	0	0	0.000	0.000	0.000	0.000	0.000	0.000
6	0.030	0.167	0.015	0.167	0	0	-0.015	0.167	-0.031	0.167	-0.048	0.167
7	0.023	0.000	0.012	0.000	0	0	-0.012	0.000	-0.025	0.143	-0.038	0.143
8	0.018	0.000	0.009	0.125	0	0	-0.009	0.125	-0.018	0.125	-0.028	0.250
9	0.013	0.000	0.007	0.000	0	0	-0.007	0.000	-0.014	0.000	-0.022	0.000
10	0.011	0.000	0.005	0.000	0	0	-0.006	0.000	-0.012	0.000	-0.018	0.000
11	0.009	0.091	0.004	0.000	0	0	-0.005	0.000	-0.010	0.000	-0.015	0.000
12	0.008	0.000	0.004	0.083	0	0	-0.004	0.083	-0.008	0.083	-0.012	0.083
13	0.006	0.154	0.003	0.000	0	0	-0.003	0.000	-0.007	0.000	-0.010	0.077
14	0.005	0.143	0.003	0.000	0	0	-0.003	0.000	-0.005	0.000	-0.008	0.000
15	0.004	0.000	0.002	0.000	0	0	-0.002	0.000	-0.005	0.000	-0.007	0.000
16	0.003	0.000	0.002	0.000	0	0	-0.002	0.000	-0.004	0.000	-0.006	0.000
17	0.003	0.000	0.002	0.000	0	0	-0.002	0.000	-0.003	0.000	-0.005	0.000
18	0.002	0.000	0.001	0.000	0	0	-0.001	0.000	-0.003	0.000	-0.005	0.000
19	0.002	0.000	0.001	0.105	0	0	-0.001	0.105	-0.003	0.053	-0.004	0.000
20	0.002	0.000	0.001	0.000	0	0	-0.001	0.000	-0.002	0.000	-0.003	0.000
Avg.	0.009	0.035	0.004	0.030	0	0	-0.005	0.030	-0.009	0.036	-0.014	0.045

Change in coverage = (Current coverage-New coverage)/Current coverage
Change in locations = Number of vehicles need to be relocated/Number of total vehicles



Change in coverage = (Current coverage-New coverage)/Current coverage

Figure 2.8: Changes in coverage when λ is varied from 1.0 to 1.5



Change in locations = Number of vehicles need to be relocated/Number of total vehicles

Figure 2.9: Changes in locations when λ is varied from 1.0 to 1.5

2.9 Conclusion and discussion

Traditional covering location models can lead to solutions which result in disparity in service between different demographics. Optimally locating ambulances to improve fairness is an important issue, this paper proposes a bi-objective model to address this problem. In particular, we applied the ε-constraint method to solve a bi-objective covering location problem. The first objective is to maximize the number of requested calls that can be covered by the ambulances within a response time standard, the second objective is aimed at reducing disparity in service between rural and urban citizens. The second objective is modeled in three ways: to (a) minimize the maximum distance between uncovered demand zones and opened stations or to (b) minimize the number of uncovered rural demand zones or to (c) minimize the number of uncovered demand zones

The results are obtained using data from Hanover County, a rural/suburban county in Virginia. The results, therefore, should not be interpreted to provide a general policy for all types of EMS systems, since the results depend on travel distances and call locations that may not be characteristic of urban and other suburban areas. However, the proposed model can be used to reduce disparities in service for other types of EMS systems.

With one objective, we can only get the solution which maximizes number of requested calls that can be covered or we can get the solution which minimizes one of the secondary objectives. By using a bi-objective model, we can find all the solution points in between the best value of the first objective and the best value of the second objective.

The solution points we find provide a set of efficient (non-dominated) solutions, or alternatives, that are very useful for decision makers wishing to take into account issues of fairness when locating EMS ambulances. While each model yields a set of non-dominated solutions that are not directly comparable, we propose two performance metrics to use as selection criteria: the average distance or the weighted average distance from an uncovered zone to its closest station. Under these criteria, Model (a) which minimizes the maximum distance between uncovered zones and its closest open station always provides a better solution, without sacrificing the first objective, though Model (b) offers a larger Pareto set and therefore more options to the decision maker. The equitable preference analysis suggested that Model (a) and Model (c) which minimizes the number of uncovered zones is more preferable than Model (b) in terms of providing equal effects to individuals. The largest reduction in disparities is achieved when service is poor to average. This is important because it has been observed that levels of care are typically not as good in rural areas as compared to urban areas. Thus mediocre service in urban locations could translate to very poor service in rural areas; our model helps to provide solutions that mitigate these issues of fairness.

Analyzing models which consider issues of fairness in the delivery of EMS service is important. This paper proposes three bi-objective models to reduce disparities in service received by rural and urban citizens. Extending this approach to take into account more than two criteria, or including criteria which may be easier to interpret (but may result in non-linear formulations) are important future areas of research.

CHAPTER 3

THE MINIMUM *p*-ENVY LOCATION PROBLEM: A NEW MODEL FOR EQUITABLE DISTRIBUTION OF EMERGENCY RESOURCES

3.1 Introduction

Emergency medical service (EMS) systems are public service systems that provide emergency medical service to patients within a service area. The services provided vary depending on the call such as providing emergency medical care via a technician or paramedic, or providing transportation. An important factor in determining EMS performance is not only the quality of emergency medical care provided but also the timeliness or *response time* in which care is provided (McGinnis, 2004). In urban areas, the most widely used ambulance response time standard is to respond to 90% of calls within 8 minutes and 59 seconds as compared to responding to 90% of calls within 14 minutes and 59 seconds in rural areas (Fitch, 2005). In practice, however, it may not be possible to meet this standard depending on the geographical area, the EMS resources available, and the location of EMS resources at the time of a call. In addition, response times may be much longer than the standard, especially in rural or remote areas. Even within a contained geographic area, guaranteeing the same (or similar) response times to all customers in the system may be infeasible.

Unlike private services, such as supermarkets or banks, which are free to locate their facilities in densely populated areas in order to maximize profits, public services such as EMS systems provided by governmental or non-profit agencies need to locate their facilities in a way that serves all residents (customers) fairly as they provide essential life-saving services (Savas, 1978; Stone, 2002). Locating ambulances in EMS systems is an important resource allocation problem that has many implications for equity.

We briefly provide a review of facility locations models that have been applied to public service problems and consider equity. Two well-known facility location models often used to locate ambulances are the *p*-median and *p*-center problems. We provide a short summary of the *p*-median and *p*-center problems here. In the facility location problem with *p* facilities, the *p*-median objective minimizes the total distance from demand points (customers) to their closest facility. Suppose a facility is to be located on a line between two demand points at the ends of the line, moving the facility from one end to another end does not change the total distance between the two demand points and the facility location. Thus, the *p*-median problem is reflective of aggregate level outcome rather than individual level outcomes; meaning that in the example given, it does not matter where the facility is located along the line. On the other hand, the *p*-center problem minimizes the maximum distance from demand points to their closest facilities. As with the previous example, if again the facility is moved along the line between two demand points, the distance to one demand point reduced while the distance to the other demand point is increased. Thus, the optimal solution of the *p*-center problem locates the facility equidistant to both demand points, which reflects one concept of equity (Leclerc et al., 2010). Although the *p*-center problem belongs to a family of equitable location design problems, its objective improves only the "worst" customer instead of explicitly reflecting the outcomes of all individuals. For a review of the *p*-center and *p*-median problems, see Daskin (1995).

Public services such as EMS systems have an expectation of fairness for their customers (Stone, 2002). The facility locations directly affect how customers access services. In order for all customers to have an equal chance to obtain services, inequity among all customers must be reduced. Several measures have been proposed to capture inequity of the system or the effect of distribution of the facilities to customers. The most common inequity measure is the maximum distance between customers and the closest facility, assuming that all customers are only serviced by their closest facility. Such a measure is reflected in the *p*-center problem. Other inequity measures suggested in the literature include range (see e.g. Brill et al., 1976; Erkut and Neuman, 1992), variance (see e.g. Maimon, 1986; Kincaid and Maimon, 1989; Berman, 1990), and mean absolute deviation (see e.g. Berman and Kaplan, 1990; Mulligan, 1991) in the distances between customers and their closest facility. Marsh and Schilling (1994) provide a comprehensive review of equity measures.

Range is a measure that considers the difference between the closest and the farthest customers, while variance and mean absolute deviation are two measures that consider minimizing the difference between individual outcomes and some system standard. However, even though one customer receives better access to care than a given standard, he feels dissatisfied if he is "worse off" than other customers. Another equity measure that considers the difference in the outcomes between individual customers is the sum of absolute differences in distance between customers and their closest facility (Keeney, 1980; Lopez-de-los-Mozos and Mesa, 2001; Lopez-de-los-Mozos, 2003). Similarly, the Gini coefficient and Lorenz curve are popular indexes that have been developed for evaluating inequity in economic and social welfare literature, and were applied for equalizing in facility location problem (Maimon, 1988; Erkut, 1993; Drezner, 2009). These measures are functions of the absolute value between individual differences, such that they penalize for any differences in individual outcomes (that is whether a customer is worse off or better off). Since people feel no dissatisfaction when they are better off than others, only negative effects are considered in the minimum envy location problem (MELP) introduced by Espejo et al. (2009) . They propose several ways to formulate the minimum envy problem; however, their formulations do not necessarily fit well with EMS models. In particular, those formulations provided by Espejo et al. (2009) assume that there is strict preference ordering information about customer's preferences or customer's dissatisfaction. This is not practical for application to EMS systems because, a customer is able to have two stations at the same preference ordering (equidistant). Furthermore, ordinal preferences lack information about distance which is an important metric when assessing quality of service. In our model we are able to relax the strict and ordinal preference order assumptions.

Furthermore, most inequity measure, including all equity location models mentioned above, consider customers' dissatisfaction based only on the closest facility. These inequity measures are appropriate for some public services, such as post or school locations where the customer travels to the facility, but not necessarily for EMS systems, where open facilities indicate the location where EMS ambulances are stationed. In an EMS system, the ambulance stationed at the closest facility is not always available to serve customers, and in that case the ambulance stationed at the next closest facility might instead be dispatched. To resolve this, many researchers account for the probability that a particular ambulance is available or busy at the time a call for service arrives. For probabilistic location models, see (Larson, 1974, 1975; Daskin, 1983; ReVelle and Hogan, 1989; Batta 1989; Galvao, 2005; Iannoni and Morabito, 2007). Other proposed location models explicitly consider backup or multiple coverage (Hogan and ReVelle, 1986; Daskin et al., 1988; Araz et al., 2005; Iannoni and Morabito, 2007)

Since EMS systems are an important public service that affects wellness of the service population, we are interested in developing a practical equitable location model that represents the inequity of all customers in the system, and more realistically represents the operations and performance criteria of EMS systems. To the best of our knowledge, this is the first equitable location model that integrates the concept of envy while taking into account the degree of importance of the different servers and incorporates the probability of servers being available to respond calls.

In particular, we propose the Minimum *p*-Envy Location Problem (M*p*ELP) for locating EMS ambulances at possible station locations in order to increase equity of receiving service among all demand zones. Envy is selected as a way to measure equity, where envy is defined as a function of the distance from a demand zone to its closest EMS station and the distance from a demand zone to its backup EMS stations weighted by priority of the serving stations and weighted by proportion of demand. The performance of our model is investigated by comparing it with two popular equity measures, *p*-center and Gini coefficient, and the well-known maximal covering location problem (MCLP). Because of its complexity, this problem cannot be solved efficiently to optimality, even for small test cases, using commercially available optimization software; thus a tabu search is developed which yields near-optimal solutions with little computational effort.

The rest of the paper is organized as follows. In Section 3.2, we describe the concept of envy, introduce notation, and formulate the model. An illustrative example is presented in Section 3.3. Section 3.4 details how to assign the station weights using the hypercube model. Section 3.5 presents the procedure of the solution method that we developed for solving the problem using a tabu search (TS). We conduct computational experiments for tuning the tabu search parameters in Section 3.6. In Section 3.7 a case study is selected to test the proposed approach using real-world data and computational results are reported in Section 3.8. Section 3.9 shows the performance of the minimum *p*-envy location model in comparison to other location models. Finally, conclusions and discussion are provided in Section 3.10.

3.2 Minimum *p*-envy location model

In this section, we modify the concept of envy to create an objective which is meaningful for the ambulance location problem. From Longman's English dictionary, envy is "the feeling of wanting something that someone else has." Therefore, customers in demand zone *i* feel envy when they receive worse service than others, but when they receive better service than others they have no feeling of envy. These concepts reflect definitional notions of equity in the social science domain (Stone, 2002) in that they clarify the recipients (the potential patients), what is being distributed (delivery of ambulances to patients according to the patients' relative dissatisfaction) and the process for equitably allocating resources (ambulance location). In our model, we define “envy” of demand zone i as a level of customers’ dissatisfaction in demand zone i as compared to other demand zones, where a demand zone is a demand point where customers are located. The dissatisfaction of customers in demand zone i is an ordered vector of the distance from demand zone i to its serving stations (facility locations) in decreasing order. That is, the distance to the station closest to demand zone i, which is the primary station, is the first element in the dissatisfaction vector, followed by the distance to the next closest station or the secondary station, and so on. The serving stations, except for the primary stations, are called *backup stations*, of which we can have one or more for each demand zone. Envy is defined as the difference in dissatisfaction between demand zones. Since different demand zones have different total number of customers (demand or call density), we weigh the total envy in each demand zone by the proportion of demand in that zone. An illustrative example of how envy is calculated is presented in the next section. We use the following notation:

n = the number of demand zones

m = the number of potential stations

p = the number of ambulances to be located (stations to be opened)

q = the number of serving stations which consists of one primary station and q-1 backup stations where $q \leq p$

w_l = weight assigned to the l^{th}-priority station, l=1, …, q

H_i = demand (call volume) in zone i

h_i = weight (proportion of demand) of zone i = $\frac{H_i}{\sum_{i=1}^{n} H_i}$

d_{ij} = the distance between zone i to station j

The objective of our equitable location model is to minimize the sum of weighted envy among all demand zones, as shown in Equation (1). Note that the proportion of demand at node i is the weight (h_i) that we assign to differentiate between call volume at different demand zones. As mentioned earlier, customers in each demand zone may have dissatisfaction with respect to all serving stations; first priority station, second priority station, and so on. Thus, we can differentiate the envy with respect to different serving stations by adding the different weights (w_l) to each level of priority; l=1, …, q, where q is the number of serving stations that are restricted to respond to a particular zone. Note that $q \leq p$ where p is the number of stations that will be opened. We introduce $q \leq p$ here because it may be that only certain number of back-up stations are allowed or that the decision maker only wants to consider envy with respect to some subset of stations; however, all stations need to be located and thus p cannot simply be replaced by q. A station is said to be opened when there is at least one ambulance stationed for serving customers. We note from (1) that since there is no contribution to the objective of locating more than one ambulance at the same station, the number of open stations and the number of ambulances are the same[1]. To avoid a trivial solution we assume that $m \geq p$ where m=the number of potential station locations, otherwise there are excess

[1] This assumption may be relaxed by incorporating constraints on the number of ambulances per station and modifying the envy calculation. For example, if up to two ambulances are allowed at each station, the first backup station is considered the same as the primary station when there are two ambulances at the primary station.

ambulances to be located. This assumption allows us to specify the effect that each station has on a demand zone through the vector **w**=(w_1, …, w_q), $w_l \geq 0 \quad \forall l$. Without loss of generality, we assume that $\sum_{l=1}^{q} w_l = 1$ and $w_1 \geq w_2 \geq \ldots \geq w_q$. Station priority weights can be assigned in various ways, depending on how the system administrator values backup service. For example, if a system only utilizes one backup station, we can set $q=2$, so **w**=(w_1, w_2) where $w_1 \geq w_2$, and $w_1+w_2=1$. How the weights **w** may be assigned is further discussed in Section 3.4.

The minimum *p*-envy location problem is introduced as an integer programming model. The objective function captures the total weighted envy among all demand zones as shown in Equation (3.1). The decision variable e_{ik}^{l} represents the envy of demand zone *i* compared with demand zone *k* based on their serving stations at the l^{th} priority level. Note that the l^{th} priority station serving demand zone *i* is not necessarily the same as the l^{th} priority station serving demand zone *k*. The index *l* of e_{ik}^{l} goes from 1 to *q*; if we consider the envy based on all of facilities in a system $q = p$ or if we consider the envy based on some facilities in a system $q \leq p$. Equations (3.2) - (3.3) work together to calculate the envy between all possible pairs of customers. The variable e_{ik}^{l} takes on value 0 when zone *i* is served by a closer facility than zone *k* compared with the same priority station, otherwise it is equal to the difference between the distance from zone *i* to its serving station and the distance from zone *k* to its serving station, that is $e_{ik}^{l} = \max\left\{0, \left(\sum_{j=1}^{m} d_{ij} y_{ij}^{l} - \sum_{j=1}^{m} d_{kj} y_{kj}^{l}\right)\right\}$. Equation (3.4) limits the number of ambulances that are available to be located, or equivalently, number of stations to be opened. Equation (3.5) ensures that a demand zone must be served by exactly one facility at each l^{th} priority station. Equation (3.6) ensures that a station can either serve as a 1st or 2nd or l^{th} priority of zone i. Equation (3.7) requires that a demand zone i can be served by facility j if station j is open. Equation (3.8) assigns a station to serve zone i by considering the distance from an open station to the zone; the closer station receives the higher priority to serve zone i.

The Minimum p-Envy Location Problem (MpELP):

Minimize $$Z = \sum_{l=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{n} w_l h_i e_{ik}^{l} \tag{3.1}$$

Subject to:

$$e_{ik}^{l} \geq \sum_{j=1}^{m} d_{ij} y_{ij}^{l} - \sum_{j=1}^{m} d_{kj} y_{kj}^{l} \quad \text{for } i,k=1,\ldots,n\text{: } k\neq i;\ l=1,\ldots,q \tag{3.2}$$

$$e_{ik}^{l} \geq 0 \qquad \text{for } i,k=1,\ldots,n;\ l=1,\ldots,q \tag{3.3}$$

$$\sum_{j=1}^{m} x_j = p \tag{3.4}$$

$$\sum_{j=1}^{m} y_{ij}^{l} = 1 \qquad \text{for } i=1,\ldots,n;\ l=1,\ldots,p \tag{3.5}$$

$$\sum_{l=1}^{p} y_{ij}^{l} \leq 1 \qquad \text{for } i=1,\ldots,n;\ j=1,\ldots,m \tag{3.6}$$

$$y_{ij}^{l} \leq x_j \qquad \text{for } i=1,\ldots,n;\ j=1,\ldots,m;\ l=1,\ldots,p \tag{3.7}$$

$$d_{ij} y_{ij}^{l} \leq d_{ij} y_{ij}^{l+1} \qquad \text{for } i=1,\ldots,n;\ j=1,\ldots,m;\ l=1,\ldots,p\text{-}1 \tag{3.8}$$

Where:

$$x_j = \begin{cases} 1 & \text{if a facility is located at station } j \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij}^l = \begin{cases} 1 & \text{if a facility at station } j \text{ assigned to serve zone } i \text{ as the } l^{th} \text{ priority station} \\ 0 & \text{otherwise} \end{cases}.$$

3.3 Illustrative example

In this section, a small example is provided to illustrate the concept of *p*-envy and how the objective function is calculated. Suppose there are three demand zones, three potential stations for locating EMS ambulances, and two ambulances. In this case, n=3, m=3, and p=2. Assume that one backup station is considered, q=2. The number of rows in the distance matrix (d_{ij}) represents the number of demand zones (n) while the number of columns represents the number of potential stations (m), where d_{ij} represents the distance from demand zone i to station j. Other inputs include the proportion of demand in each demand zone i (h_i), and the weights assigned each priority open station (w_l) . The inputs to this small example are given below in matrix form.

$$\mathbf{d} = \begin{bmatrix} 2 & 2 & 10 \\ 8 & 4 & 6 \\ 10 & 5 & 2 \end{bmatrix} \qquad \mathbf{h} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

The vector **h** denotes that 20%, 30% and 50% of customer calls originate in demand zones 1, 2, and 3, respectively. The vector $\boldsymbol{\lambda}$ indicates that a consumer's envy will be comprised of 60% resulting from envy regarding their primary serving station and 40% envy regarding their secondary serving station. Suppose ambulances are located at station 1 and station 2, demand zone 1 is 2 units away far from its 1st priority or primary station, and also 2 units away far from its 2nd priority or secondary station. Demand zone 2 is located closer to station 2, so station 2 serve as a 1st priority station of zone 2, and station 1 serves as a 2nd priority station of zone 2. The same with demand zone 3, it is served by station 2, and station 1 as 1st and 2nd priority stations respectively. Then, the envy of demand zone *i* with respect to demand zone *j*, in regards to their 1st priority station is calculated from the difference of the distance from demand zone *i* to its 1st priority station and the distance from demand zone *j* to its 1st priority station whereas if demand zone *i* is closer to its 1st priority station demand zone *j*, the envy of demand zone *i* with respect to *j* is equal to 0, because demand zone *j* does not have better access than demand zone *i*. If demand zone *i* is farther from its 1st priority station than demand zone *j* is to theirs, demand zone *i* envies demand zone *j* which we quantify as the difference in dissatisfaction between demand zone *i* and demand zone *j*. The envy matrix corresponding to locating ambulances at station 1 and 2 (e_{ik}^{l}**)** is calculated from the summation of max{0, $\sum_{j=1}^{m} d_{ij} y_{ij}^{l} - \sum_{j=1}^{m} d_{kj} y_{kj}^{l}$ } where l=1,2; i,k=1,2,3; $k \neq i$. For example, e_{12}^{1} =max{0, 2-4}=0, $e_{13}^{\mathbf{1}}$ **=**max{0, (2-5)}=0, and e_{23}^{1} =max{0, 4-5)}=0, e_{21}^{1} =max{0, 4-2)}=2. The total envy of all demand zones with respect to all serving stations is equal to the summation of all elements in the envy matrix multiply by the demand zone weight (h_i) and the station weight (λ_l). If we locate ambulances at station 1 and 2, the total envy of all demand zones is equal to 4.28. Our goal is to find the station locations that give the minimum total of envy. With this small-size example, one can easily enumerate all possible solutions, and the optimal solution is opening stations at locations {2, 3} with a total envy value of 1.56. Using the integer programming formulation of the minimum *p*-envy model, developed in the previous section, a solver found an optimal solution at x= {0,1,1}, y={[(0,0),(1,0),(0,1)],[(0,0),(1,0),(0,1)],[(0,0),(0,1),(1,0)]}, e={[(0,0),(0,4),(0,5)], [(2,0),(0,0),(2,1)],[(0,0),(0,0),(0,0)]}.

3.4 Determining appropriate station priority weights

The station weights should be assigned according to how a system administrator views the importance of the resources, or according to how they believe customers feel envy. The minimum *p*-envy problem is specifically designed to consider backup stations; thus, the number of backup stations should affect the values of the weights that are assigned. Suppose the system has no backup station, in other words only one station has 100% responsibility to serve a particular zone, the station weight should be set to 1 and $\mathbf{w}=(w_1, 0, 0, \ldots, 0)$ where $w_1=1$; in that case the minimum *p*-envy location problem becomes original minimum envy problem except that envy is measured nominally rather than with strict preference ordering. The only restriction on the weights assigned is that $\sum_{l=1}^{q} w_l = 1$ and $w_1 \geq w_2 \geq \ldots \geq w_q$. For example, these values could be are assigned to be linearly decreasing; that is, $w_l = \dfrac{q+1-l}{K}$ where $K = 1+2+...+q$, l=station priority, $l \in \{1, \ldots, q\}$. For example, if q=5, w_1=5/15, w_2=4/15, …, w_5=1/15, respectively. Next we provide a recommendation for how these weights might be assigned to reflect the actual performance of EMS systems.

In real EMS systems the closest vehicle may not be available to answer a call. Thus we suggest that the probability of vehicle being available be assigned as a station weight. Daskin (1983) developed the earliest probabilistic location model, the maximum expected coverage location problem (MEXCLP), which assumed that servers operate independently and have the same busy probability which is independent of their locations. Later, Batta et al. (1989) developed an adjusted MEXCLP (AMEXCLP) which relaxes some assumptions of the MEXCLP by embedding the hypercube queuing model into MEXCLP. The hypercube model developed by Larson (1974, 1975) considers a correction factor that accounts for busy probabilities depending on server locations. The model has several underlying assumptions: 1) calls for service arrive according to a Poisson process, 2) if a call arrives while all servers are busy; it enters at the end of a queue and will be served in a FIFO manner.

In this paper, the busy probability of vehicles is estimated by the hypercube queuing model. Let p_b denote the probability that a randomly selected vehicle will be busy which depends on the number of ambulances that are deployed (assuming q-1 backup stations). Using actual system data, one can we estimate probability p_b by $p_b=\lambda/p\mu$ where, λ is the average number of calls per hour, $1/\mu$ is the average service time per call (hours), and p is number of ambulances that are deployed. Constructing an M/M/p queuing model operating at steady state, we get the probability that all servers are available, p_0, as given in Equation (3.9). The correction factors Q are calculated as in Equation (3.10). If there are l ambulances that may respond to a call, the probability that the l^{th} vehicle will be dispatched or is available is calculated from the probability that l-1 ambulances are busy and the l^{th} vehicle is available. The probability that the l^{th} vehicle is available (w_l) is shown in Equation (3.11) where $Q(p, p_b, l\text{-}1)$ is the correction factor and $Q(p, p_b, 0) = 1$.

$$p_0 = \left(\frac{(pp_b)^p}{p!(1-p_b)} + \sum_{j=0}^{p-1} \frac{p^j p_b{}^j}{j!} \right)^{-1} \tag{3.9}$$

$$Q(p, p_b, j) = \sum_{k=j}^{p-1} \frac{(p-j-1)!(p-k)p^k p_b{}^{k-j} p_0}{(k-j)!\,p!(1-p_b)} \ , \qquad j=0, \dots, p\text{-}1 \tag{3.10}$$

$$w_l = Q(p, p_b, l-1)(1-p_b)(p_b{}^{l-1}) \, , \qquad l=1, \dots, p \tag{3.11}$$

3.5 Tabu search

Because of the complexity of the minimum p-envy location problem, finding the optimal solution via a commercial optimization sotfware is impractical due to the computational effort required to solve these problems, especially for real-world size problems. Although the model has been linearized to reduce computational effort, it requires a large number of additional variables and constraints to remove the nonlinear (maximization) terms involved in calcualting envy. The number of variables and constraints make the problem size grow exponentially as the number of demand zones and potential stations increase, which directly leads to increased computational costs. To illustrate the complexity, we generated 17 test problem sets with different combinations of parameters; $n \in \{5,10,20,30,50,100\}$, $m \in \{5,10,15,20,30\}$, $p \in \{2,3,5,10\}$, and assuming that q=p in all test cases. The integer programing formulations of minimum p-envy location problem were implemented in two commercial optimization solvers; ILOG

OPL 5.5. and AMPL 11.0, and both are running on a Dell Latitude D410 machine with Intel Pentium processor 1.73 GHz, 1 GB of RAM. With traditional branch and cut methods, the solver was able to find the optimal solutions in some cases as shown in Table 1. The running time limit was fixed to 1 hour. The results showed that AMPL performed better than OPL in terms of time and solution gap. However, based on this experiment, for problem sizes equal to or larger than 30 demand nodes, it is not practical to obtain the optimal solution via both optimization solvers. We also observed that in the case of n=30, m=15, p=q=5, it took about 6 hours to get the optimal solution. The notation >1H states that running time exceeded 1 hour and NA states that no feasible solutions have been found after running the solver for 1 hour.

Table 3.1: Results of solving *p*-envy location problem via optimization solvers

n	*m*	*p*	OPL		AMPL	
			Time (sec.)	Gap(%)	Time (sec.)	Gap(%)
5	5	3	1.06	0	0.25	0
		2	0.81	0	0.18	0
10	10	5	25.53	0	26.92	0
		3	5.57	0	4.45	0
		2	3.09	0	2.35	0
20	10	5	647.60	0	452.10	0
		3	121.26	0	177.26	0
		2	47.37	0	19.39	0
30	15	10	>1H	NA	>1H	NA
		5	>1H	NA	>1H	NA
		3	426.56	0	165.12	0
50	20	10	>1H	NA	>1H	NA
	15	5	>1H	NA	>1H	NA
		3	>1H	80.24	>1H	42.05
100	30	10	>1H	NA	>1H	NA
	20	5	>1H	NA	>1H	NA
	10	5	>1H	NA	>1H	NA

To overcome this problem, one might try to reduce the number of variables by improving the formulation. For a discussion of developing efficient minimum envy formulations see Espejo et al. (2009). However the integer programming formulations tend to have limitations depending on the problem structure, and they still suffer from dimensionality issues. Espejo et al. (2009) developed several formulations for the minimum envy location problem with the underlying assumption that a demand zone must has predefined strict preference for all potential stations, and the computational running time for the problem size n=40 was reported to be longer than 1 hour. In this paper, we are interested in providing a practical approach that will enable us to solve the real-world size problem, which tend to have a large number of demand zones ($n \geq 100$). Therefore, we developed a tabu search that enables us to find near-optimal solutions efficiently.

Tabu search (TS), a metaheuristic algorithm, was formalized in 1986 by Glover (1986). The characteristics of TS are based on the mechanism of human memory. During the search process, TS keeps memory of a predetermined number of solutions that have already been evaluated and records them on a tabu list. These solutions that have been evaluated are protected for a limited period of time using short-term memory in an attempt to escape local optima. If the new solution yields a better objective, a move is performed regardless of the tabu list, otherwise moving to the new solution will only occur when the new solution is not in the tabu list. The TS algorithm is composed of the following procedures 1) initializing a feasible solution 2) improving upon the current solution 3) managing the tabu list 4) checking the stopping criteria. The algorithm continues performing procedures 2 through 4 until the stopping criteria is satisfied.

3.5.1 Representation and Initialization

We choose a permutation representation for our solution. That is, suppose we have 2 ambulances to be located among 5 potential stations, and consider the solution of locating ambulance 1 at station 3 and ambulance 2 at station 5; the permutation representation string will be {3,5}. The initial solution is randomly generated using the concept of random keys as introduced by Bean (1994). We start with one feasible solution at the initial stage. Suppose we want to create a solution for a problem which has 2 ambulances and 3 stations, we first create 3 random numbers; 0.7, 0.4, 0.5. Next assign an order for each random number; 0.7(1), 0.4(2), 0.5(3). Then sort the random numbers as ascending order; 0.4(2), 0.5(3), 0.7(1). The initial solution is the first two ordered stations which are {2,3}.

3.5.2 Improving process

To improve a current solution, we consider all the solutions in the neighborhood of the current solution and replace it with its best neighbor. The swap neighborhood used in Ghosh (2003) is applied in which each neighbor is found by replacing one located ambulance with one non-located ambulance. In other words, an open station is replaced by a closed station. Suppose we have 2 ambulances to be located among 5 potential stations and the current solution is {3,5}. If we chose station 3 to be replaced, the possible neighbors are {1,5}, {2,5}, and {4,5}. The total number of possible neighbors to each solution are $(m\text{-}p)p$. Becasuse we only replace one station at each iteration, the number of neighbors at each iteration are $(m\text{-}p)$, and the best neighbor is the solution that yields the lowest total weighted envy.

3.5.3 Tabu list

To avoid selecting an old solution that has been recently evaluated, we create a tabu list to record the old moves or old solutions. We propose two types of tabu lists and apply each one to the TS algorithm we developed. These are swap record and solution record.

3.5.3.1 Swap record

As described in section 3.5.2, new solution is obtained by swapping an open station with a closed station. This tabu list consists of pairs of recent stations that have been replaced and the stations that replaced them. For example, if we have a current solution, {3,5}, and we want to move ambulance 1 from station 3 to station 2, our new solution is {2,5}. In this case, we record the move {3,2}. Thus the swap record tabu list is an mxm matrix where m is the number of the potential stations. We record the swap {3,2} by updating the value of element (3,2) and (2,3) in the swap record tabu list. This tabu list structure has the advantage of being convenient to manage; however, the size of the list grows as the number of the candidate stations increases.

3.5.3.2 Solution record

Instead of recording the swap move, we can alternatively record the solutions that have been evaluated. Note that the swap record tabu list cannot protect some solutions that have recently been evalutated in the case that the order of the stations is different. For example, if the current solution is {1,2,3}, and the next solution is {1,2,5}; the swap record {3,5} would be added to the swap record list. However, this does not rule out the possibility that in two moves we would see solution {3,2,5} then {3,2,1}; this last solution is the same as the previous solution {1,2,3}. This problem is solved when using the solution record tabu list. However, this type of list structure requires more steps to create the list. To capture the different station order in each solution that yields the same objective, we convert the solution into a power of two form. In other words, each distinct set of ambulance locations yields the same value, despite the order in which the locations are listed in the solution. For example, the solution {1,2,3} will be recorded as a value of $2^1+2^2+2^3 = 14$. This way solution {3,2,1} or {2,3,1}, which also yield the value 14, are not selected as long as 14 is in the solution record tabu list. In this case, the length of the list is fixed at the number of the candidate stations at each iteration (*m*).

3.5.4 Short-term memory

Independent of which tabu list structure is used, the solutions in the tabu list will be protected for the next solution, which means we never have the same solution in the following iteration. This protection is set to be active for a limited time, called the tenure time. The tenure time works as a short-term memory of the TS algorithm which is one of the parameters that might effect the performance of the TS algorithm. There are three possible ways to manage the tenure time: fixed, dynamic, and random. In this study, we used fixed tenure time, and considered list lengths of 7, 10, 15, 20 as suggested by Glover (1990).

3.5.5 Aspiration Critera

An aspiration criteria is applied when the better move is tabu. In other words, a tabu move (solution that is in the tabu list) is allowed when this solution yields a better objective than the best found so far. This allow us to improve the performance of the TS algorithm and allows us to escape local optima.

3.5.6 Stoping Criteria

Several potential stopping criteria have been proposed such as maximum CPU time, a maximum number of solutions, a maximum number of iterations, or a a maximum number of iterations with no improvement. Based on preliminary experiments, we terminate the program after a fixed number of iterations which depends on the problem size and is dicussed further in later sections. For each scenario the TS is run for 30 replications. The steps of tabu search at each iteration are shown below:

```
Step 1:   Initialize solution
Step 2:   Best := Initial Solution
          Current := Initial Solution
Step 3:   While (Stopping criterion not met) do
                    Select a station to swap
                    Evaluate all possible neighbors
                    Best_nb := Best neighbor
                    If Best_nb is better than Best
                              Then Go to Step 5
                    Else      Go to Step 4
Step 4:                       If Best_nb is not  in the tabu list
                                        Then Go to Step 5
                              Else      Best_nb := Next best neighbor
                                        If Best_nb is the last neighbor
                                                  Then Go to Step 5
                                        Else Go to Step 4
Step5:              Current := Best_nb
                    Update tabu list
                    If Current is better than Best
                              Then Best := Current
          End while
```

While we realize that the proposed TS algorithm is quite simple, we will demonstrate below that it is both quite effective and efficient at finding solutions. Furthermore, the algorithm is robust in the sense that it works with any location model objective. Lastly, while we do perform parameter tuning for the TS, we do not test using alternate heuristic methods. The focus of the paper is the development and analysis of the M*p*ELP; the TS is developed here to allow us to analyze real-world size problems.

3.6 Parameter tuning experiments

In this section, we conducted experiments to find the best combination of two parameters: the type of tabu list structure and the choice of tenure time length. These parameters were identified as influential factors based on initial testing. Two data sets have been used. The first one is a real-world data set consisting of 122 demand zones and 16 potential stations (details regarding this data set are provided in Section 3.7). The second one is a publicly available data set with 30 nodes (or demand zones) and 30 stations, taken from Lorena's instances which accessible through the OR-Library (http://www.lac.inpe.br/~lorena/correa/Q_MCLP_30.txt). For each data set, we create 6 instances by varying the number of stations that can be opened, i.e. p varies from 5 to 10. Each case is tested under two types of tabu list structures and four tenure time lengths of 7, 10, 15 and 20, respectively. Our tabu search was coded in Visual Studio C. The resulting 96 test cases were run on a Dell Latitude D410 machine with Intel Pentium processor 1.73 GHz, 1 GB of RAM. We also obtain the optimal solution to each problem by enumerating all possible solutions[2]. The results are represented as the median and range of the solution gap (%gap= the relative difference between the best tabu search

[2] Full enumeration takes anywhere from 1 hour to 2 days depending on the problem size and is only used to evaluate the performance of our algorithm, not recommended as an approach to solving the p-envy problem.

solution value and the optimal solution value) over the 30 replications, which are reported in Table 3.2.

Table 3.2: Median solution gaps and solution gap range among the 30 replications for the parameter tuning experiments, expressed as Median (Min, Max).

Data set (*n*x*m*)	List	*p*	%Gap[a] [Median (Min, Max)] Tenure time 7	10	15	20
30QMCLP (30x30)	Swap	5	0.000 (0.000, 0.700)	0.000 (0.000, 3.397)	0.000 (0.000, 2.007)	0.000 (0.000, 1.188)
		6	0.716 (0.201, 1.833)	0.873 (0.000, 1.629)	0.5445 (0.000, 1.983)	1.045 (0.201, 1.886)
		7	2.170 (0.000, 3.022)	1.980 (0.720, 4.566)	2.101 (0.000, 3.199)	2.761 (0.747, 4.573)
		8	1.462 (0.000, 3.453)	1.596 (0.269, 2.877)	2.021 (0.827, 4.236)	1.596 (0.000, 3.926)
		9	1.712 (0.700, 4.764)	2.615 (0.778, 5.584)	2.159 (0.700, 3.940)	2.858 (0.489, 4.711)
		10	1.553 (0.000, 3.360)	1.263 (0.000, 3.300)	0.991 (0.061, 3.189)	1.869 (0.000, 3.302)
	Solution	5	0.000 (0.000, 1.003)	0.956 (0.000, 4.455)	0.694 (0.000, 1.310)	0.694 (0.000, 4.596)
		6	0.870 (0.201, 2.253)	0.876 (0.000, 2.240)	0.873 (0.000, 1.629)	1.056 (0.544, 2.936)
		7	2.514 (0.000, 5.238)	2.101 (0.000, 5.539)	2.723 (0.000, 4.116)	1.604 (0.720,3.877)
		8	1.955 (1.041, 3.995)	1.633 (0.973, 4.204)	1.966 (0.827, 4.546)	1.495 (0.269, 4.236)
		9	2.248 (0.700, 4.764)	2.194 (0.778, 5.584)	2.703 (0.700, 3.940)	2.896 (0.489, 4.711)
		10	2.135 (0.000, 3.300)	1.265 (0.000, 3.300)	0.757 (0.061, 3.1895)	1.265 (0.000, 3.3021)
Hanover County (122x16)	Swap	5	0.000 (0.000, 8.254)	0.000 (0.000, 10.088)	0.000 (0.000, 8.254)	0.000 (0.000, 5.507)
		6	0.000 (0.000, 3.801)	0.000 (0.000, 5.253)	0.000 (0.000, 3.752)	0.000 (0.000, 3.943)
		7	0.000 (0.000, 7.479)	0.000 (0.000, 3.556)	0.000 (0.000, 4.216)	0.569 (0.000, 2.772)
		8	1.022 (0.000, 1.617)	1.022 (0.000, 3.493)	0.000 (0.000, 1.767)	0.000 (0.000, 2.914)
		9	0.434 (0.000, 4.579)	0.433 (0.000, 5.039)	0.433 (0.000, 6.005)	0.433 (0.000, 4.878)
		10	1.492 (0.002, 6.944)	0.894 (0.002, 5.841)	0.894 (0.002, 5.039)	0.894 (0.002, 5.648)
	Solution	5	0.000 (0.000, 8.254)	0.000 (0.000, 5.004)	0.000 (0.000, 5.507)	0.000 (0.000, 8.254)
		6	0.000 (0.000, 3.943)	0.000 (0.000, 8.900)	0.000 (0.000, 3.943)	0.000 (0.000, 3.943)
		7	0.000 (0.000, 2.786)	0.284 (0.000, 3.673)	0.569 (0.000, 3.673)	0.000 (0.000, 6.491)
		8	1.022 (0.000, 1.617)	1.022 (0.000, 3.493)	0.000 (0.000, 1.767)	0.000 (0.000, 2.914)
		9	0.433 (0.000, 5.039)	0.433 (0.000, 3.374)	0.433 (0.000, 3.374)	0.000 (0.000, 3.374)
		10	0.894 (0.002, 4.218)	0.894 (0.002, 4.218)	0.894 (0.002, 5.039)	0.894 (0.002, 6.561)
Overall median	Swap		0.869 (0.000, 2.170)	0.883 (0.000, 2.615)	0.489 (0.000, 2.159)	0.731 (0.000, 2.858)
						0.800 (0.489, 0.883)
	Solution		0.882 (0.000, 2.514)	0.925 (0.000, 2.194)	0.725 (0.000, 2.723)	0.794 (0.000, 2.896)
						0.838 (0.725, 0.925)

a %Gap = [(Best known of TS - Optimal solution) *100] / Optimal solution

We report the median rather than mean because the solution gaps are not normally distributed, as will be later discussed. In this experiment we terminated each run after 50 iterations. We performed statistical analysis to identify if the tabu list structure and tenure time length significantly affect the performance of the TS. Because our results do not satisfy the assumptions required to use traditional ANOVA analysis (the solution gaps are not normally distributed and the variance in solution gaps is non-homogeneous), the Friedman test, a non-parametric statistical test, is selected to assess if differences in performance exist due to choice of list structure and tenure time length. At a significance level of 0.05, the Friedman test indicated that there is a statistically significant difference between using different types of tabu lists and among all levels of tenure time length. The swap record yielded the lowest overall median solution gap of 0.8%. We also observed that a tenure time equal to 15 yielded the best solutions with the smallest median solution gaps among all test cases regardless of the type of tabu list used. Therefore, the swap record tabu list structure with tenure time length of 15 is suggested as the best parameters for our TS.

3.7 Case study

Our case study uses real-world data from the Hanover Fire and EMS department, which is located in Hanover County, VA. The Hanover EMS department responds to 911 calls 24 hours a day and serves a county of 474 square miles, with a population of approximately 97,000 individuals. The data are collected from the Fire and EMS department during 2007, and captures the life-threatening calls received during 2007. We divided the coverage area into 175 distinct demand zones made up of approximately 2 by 2 mile areas. In this way, we ensure that originating demand is represented realistically. Currently, there are m = 16 existing potential stations for locating EMS ambulances. All station locations are shown in Figure 3.1. Based on the data, requested calls did not originate from all 175 zones. Therefore, we ignore the zones that have no demand and only considered the n = 122 zones in which demand existed in 2007.

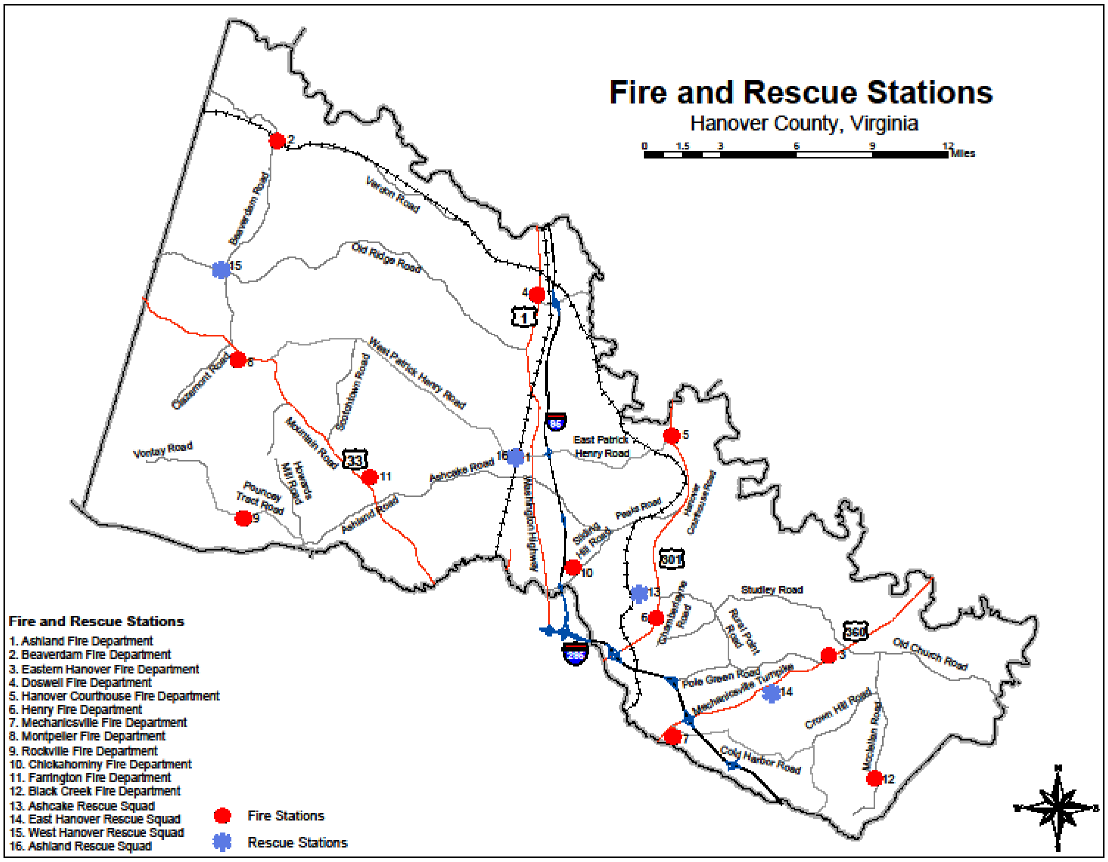


Figure 3.1: Map of fire and rescue stations in Hanover County, Virginia

The input data to the model are the number of the requested calls (or number of customers) in each demand zone, the geographical coordinates of the 122 demand zones and 16 potential stations, and the weights assigned to different priority stations. To set up the locations of the stations and demand zones, we drew grid lines over the area of interest, with one block representing 2 square miles. The coordinates (a, b) of the stations and center point of demand zone blocks are used to calculate the distance between each demand zone and each station. Distance between two points can be measured in many ways (see Drezner and Hamacher, 2004). The most familiar two are rectilinear distance and Euclidean distance. In this case we use the Euclidean metric because approximately 70% of the Hanover County area is rural, and can thus be reached via highways or county roads that do not conform to a grid. Given a demand zone i at (a_i, b_i) and a station location j at (a_j, b_j), the distance (d_{ij}) between demand zone i and station j is calculated using the Euclidian metric.

3.8. Computational results

In this section we test the performance of our tabu search heuristic using the same two data sets, after incorporating the parameter tuning results. Based on the parameter tuning experiments in Section 3.5, the swap record tabu list with a tenure time of 15 is used with both data sets. Since the numbers of neighborhoods in both cases are different we used different termination criteria for each data set. We terminated the program after 500 iterations for the 30QMCLP data set, and 100 iterations for the Hanover County data set. The solution gaps over 30 replications of both cases are shown in Table 3.3. We can see that the median and average solution gap is less than 1% for all cases and that, within a few seconds the TS obtained the optimal solution for all instances of the Hanover data set and for 2 out of 6 instances of the 30QMCLP data set (recall that a commercial solver was not able to obtain solutions to problems with n=30 in 1 hour).

Table 3.3: Experimental results of TS using tuned parameters

Data set (nxm)	p	%Gap[a]					CPU time (sec)				
		Median	Avg	SD	Min	Max	Median	Avg	SD	Min	Max
30QMCLP (30x30)	5	0	0	0	0	0	2.656	2.661	0.035	2.625	2.781
	6	0.544	0.351	0.274	0	0.873	3.742	3.778	0.081	3.703	4.079
	7	0	0.293	0.365	0	0.747	3.734	3.734	0.038	3.703	3.922
	8	0.269	0.408	0.382	0	1.254	3.703	3.706	0.020	3.672	3.750
	9	0.572	0.545	0.534	0	1.844	3.984	4.025	0.097	3.953	4.375
	10	0.061	0.098	0.148	0	0.757	5.078	5.085	0.031	5.062	5.234
Hanover County (122x16)	5	0	0	0	0	0	3.953	3.966	0.035	3.937	4.078
	6	0	0.080	0.304	0	1.201	3.531	3.569	0.132	3.406	3.922
	7	0	0.154	0.294	0	0.896	3.578	3.586	0.046	3.547	3.750
	8	0	0.102	0.312	0	1.022	4.172	4.188	0.039	4.156	4.344
	9	0	0.300	0.639	0	3.374	4.562	4.577	0.061	4.515	4.813
	10	0	0.446	0.681	0	2.087	3.453	3.477	0.085	3.406	3.828

[a] %Gap = [(Best known of TS - Optimal solution) *100] / Optimal solution

3.9 Performance of the minimum *p*-envy location problem model

While our model seeks to reduce inequity through the *p*-envy objective, we must be careful not to sacrifice efficiency of the current EMS system. Hanover County EMS measures efficiency in terms of coverage, where the coverage level is the total proportion of demand that can be reached within a response time threshold (RTT). Following current Hanover County standards, we use a response-time threshold of 9 minutes. Thus, a demand zone is said to be covered when there exists an EMS ambulance that is able to respond to a call in that demand zone within 9 minutes. In particular, we assume based on distance, average ambulance speed, and road conditions that for a call to be responded to within 9 minutes, at least one station should be open within 4 miles of the demand zone. In this case, there are 1711 calls spread over 122 demand zones; given the set of possible station locations, there are 4 zones that cannot be covered, since they are more than 4 miles from the closest possible station. Therefore, the maximum percentage of coverage for Hanover County is 98.8%.

We compare our model to a traditional covering location model, which maximizes efficiency, and to other equity models. In particular, we evaluate the performance of the minimum *p*-envy location model in terms of equity and coverage compared with other facility location models. Two standard measures of equity are selected for comparison, *p*-center and Gini coefficient. The *p*-center is a classic equity model that intends to improve the worst customer (minimizes the distance of the customer located the furthest away from their closest station). The Gini coefficient is an equity measure that considers the average dissatisfaction among all customers. The traditional maximal covering location (MCLP) model is selected as a baseline to measure coverage. The formulations of the models are provided below.

- Minimum *p*-envy location problem (M*p*ELP)

  Objective is to minimize sum of envy weighted by proportion of demand:

  $$\min Z = \sum_{l=1}^{q} \sum_{i=1}^{n} \sum_{k=1}^{n} w_l h_i e_{ik}^l$$

  Subject to (3.2) - (3.8).

- Maximal covering location problem (MCLP), see original version in Church and ReVelle (1974)

  Objective is to maximize proportion of demand that can be covered (reached within a given response time threshold):

$$\max Z = \sum_{i=1}^{n} y_i H_i$$

Subject to $\sum_{j=1}^{m} a_{ij} x_j \geq y_i$ for all i=1, 2, …, n (3.12)

and (3.4)

Where $y_i = \begin{cases} 1 & \text{if demand zone } i \text{ is covered by an open station} \\ 0 & \text{otherwise} \end{cases}$

$a_{ij} = \begin{cases} 1 & \text{if station } j \text{ can cover demand at zone } i \\ 0 & \text{otherwise} \end{cases}$

- *p*-center, see details in Daskin (1995)

  Objective is to minimize the maximum distance from customers to their closest station:

$$\min Z$$

Subject to $\sum_{j=1}^{m} d_{ij} y_{ij} \leq z$ for all i=1,2, …, n (3.13)

$\sum_{j=1}^{m} y_{ij} = 1$ for all i=1, 2, …, n (3.14)

$y_{ij} \leq x_j$ for all i=1,2, …, n, j=1, 2, …, m (3.15)

and (3.4)

Where $y_{ij} = \begin{cases} 1 & \text{if a demand zone } i \text{ is served by facility at station } j \\ 0 & \text{otherwise} \end{cases}$

- Gini coefficient measure (Gini), see details in Drezner (2009)

Objective is to minimize Gini coefficient (a weighted measure of absolute differences):

$$\min \frac{\sum_{i=1}^{n}\sum_{k=1}^{n}\left|\sum_{j=1}^{m} d_{ij}y_{ij} - \sum_{j=1}^{m} d_{kj}y_{kj}\right|}{2n\sum_{i=1}^{n}\sum_{j=1}^{m} d_{ij}y_{ij}}$$

Which is equivalent to minimizing the numerator:

$$\min \sum_{i=1}^{n}\sum_{k=1}^{n}\left|\sum_{j=1}^{m} d_{ij}y_{ij} - \sum_{j=1}^{m} d_{kj}y_{kj}\right|$$

Subject to (3.4), (3.14) - (3.15)

We use the Hanover County data, which contains 122 demand zones and 16 potential stations. We vary the total number of ambulances to be located from 5 to 10. Thus, in this case n=122, m=16, and p=q varies from 5 to 10. h_i is the proportion of demand at location i; i=1, …, 122 and all w_l values are assigned according to probability of vehicles being busy as described in Section 3.4. In this study we use a 9 minute response time threshold to evaluate coverage. The goal here is to gauge how much improving equity compromises typical EMS performance measures, such as coverage. We solved all four facility location models to optimality (optimal solution to the *p*-envy model was confirmed via full enumeration) and then compared the resulting equity measures and coverage. These results are shown in Tables 3.4 to 3.7. In these tables we present several metrics for evaluating the quality of a solution. We measure equity as the sum of weighted total envy, and we measure efficiency by the coverage of demand (this is the traditional measure of efficiency for EMS systems). We also report the maximum distance (Maxdist) between a demand zone and its closest open station (or the *p*-center objective) and the total covered demand. In these tables, larger values of covered demand are desirable and smaller values of inequity measures (Maxdist, the Gini coefficient, and total weighted envy) are desirable. The *p*-envy, Gini coefficient, and MCLP models produce unique optimal solutions while the *p*-center model often produces multiple solutions. In the case that the *p*-center object produces multiple optimal solutions, we report the average values of the covered demand, and equity measures from all optimal solutions.

Table 3.4: Results of *p*-envy

p	Opened stations	Maxdist	Gini coefficient	Total weighted envy	Covered demand
5	{1 4 6 7 8}	12	0.3139	63.7672	1524
6	{1 4 7 8 13 14}	10	0.3120	54.4391	1572
7	{1 3 4 7 9 13 15}	7	0.2810	47.5000	1628
8	{1 4 7 9 10 13 14 15}	8	0.2997	43.7513	1618
9	{1 2 4 7 8 9 10 13 14}	8	0.2995	38.9498	1637
10	{1 2 3 4 5 6 7 8 9 10}	6	0.2881	35.2600	1661

Table 3.5: Results of MCLP

p	Opened stations	Maxdist	Gini coefficient	Total weighted envy	Covered demand
5	{1 4 6 14 15}	12	0.3157	78.8367	1559
6	{1 4 6 11 14 15}	10	0.2960	75.9275	1604
7	{1 4 5 6 11 14 15}	10	0.3022	71.3908	1636
8	{1 4 5 6 9 11 14 15}	8	0.2925	71.8066	1657
9	{1 2 4 5 6 8 9 11 14}	8	0.2903	72.6818	1674
10	{1 2 4 5 6 8 9 11 12 14}	6	0.2735	72.5320	1688

Table 3.6: Results of *p*-center

p	Opened stations	Maxdist	Gini coefficient	Total weighted envy	Covered demand
5	{1 2 3 6 8}	8	0.2772	130.4130	1173
6	{1 3 4 9 13 15}	7	0.2623	125.9371	1153
7	{2 3 4 8 9 11 13}	6	0.2658	170.9044	978
8	{2 3 4 5 6 8 9 11}	6	0.2736	137.3588	1208
9	{1 2 3 4 5 6 8 9 10}	6	0.2790	109.6269	1397
10	{1 2 3 4 5 6 7 8 9 10}	6	0.2840	90.9555	1510

Table 3.7: Results of Gini

p	Opened stations	Maxdist	Gini coefficient	Total weighted envy	Covered demand
5	{3 4 9 10 15}	8	0.2533	146.4598	776
6	{3 4 9 11 13 15}	7	0.2588	154.1969	961
7	{1 3 4 9 13 15 16}	7	0.2640	119.1073	1249
8	{1 2 3 4 8 9 13 16}	7	0.2657	126.2046	1268
9	{2 4 5 8 9 10 11 12 14}	6	0.2677	137.6112	1293
10	{1 2 3 4 5 6 8 9 11 12}	6	0.2695	85.5668	1674

The four models are compared in Figures 3.2 to 3.4 in terms of the resulting equity and efficiency measures. Figure 3.2 shows the total weighted envy for each model for p=5 to 10. As expected, the minimum p-envy model has the lowest sum of total weighted envy among these four models. Interestingly, the p-center and Gini coefficient models, that also try to reduce inequity, do not dominate the MCLP model. A possible explanation for this is that neither the Gini or p-center models weight the demand zones by demand density, such that each zone is treated equally, which may be impractical in real systems, where demand density may vary widely by geographic location. Furthermore, the performance of the p-envy model is robust to the number of ambulances. For all models the resulting Gini coefficient is stable, ranging only from 0.2533 to 0.3157, while the maximum distance from a zone to its closest station (Maxdist) is quite variable, ranging from 6 to 12 miles. Figure 3.3 compares the four models in terms of coverage. In terms of coverage, we see that the Gini model performed much worse compared with the other models while the p-envy model performed very close to the MCLP model, whose objective is to maximize coverage. The performance of the p-center model largely depends on the number of ambulances. This is an undesired trait of the p-center model solutions because one would expect that coverage should increase as the number of ambulances increase. However, the *p*-center model does not weigh demand zones, and it sacrifices the coverage of densely populated areas in order to ensure better service to the demand zone that is "worse off".

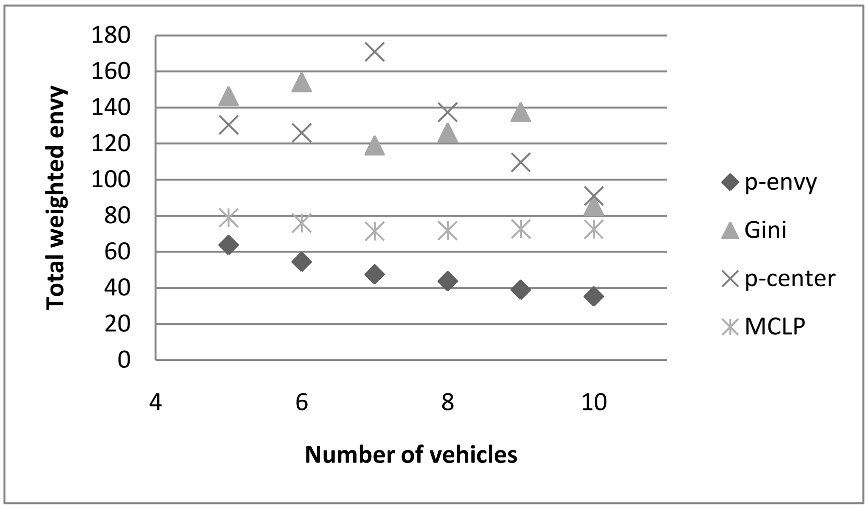


Figure 3.2: Equity comparison—*p*-envy measure for each model

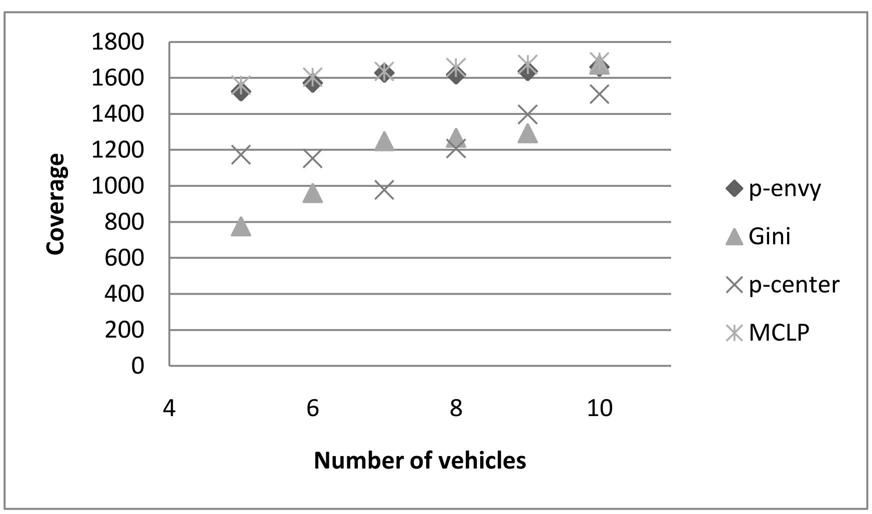


Figure 3.3: Efficiency comparison—resulting coverage for each model

To illustrate the tradeoff between equity and coverage, we plot the performance of all four models with respect to equity and coverage. Figure 3.4 shows the results of all four models where the *p*-envy model uses the station available probabilities (see Section 3.4) for station weights. Interestingly, we see that the minimum *p*-envy location model not only yields the lowest total envy, but attains almost the same coverage as MCLP. Therefore, the *p*-envy model allows us to reduce inequity without sacrificing coverage, for this data set. This is an unexpected outcome for the equity model presented, as equity and coverage tend to be conflicting objectives which necessitate a multi-objective approach, such as the one undertaken by Chanta et al. (2011a). The results depend on the weights assigned to the priority of the stations (vector **w**). For example, we note that Maxdist could be reduced in the *p*-envy model by giving more weight to the closest station (increasing w_1). Figure 3.5 shows the results when we have equal weight of station priorities ($w_l = 1/q$ for all l).

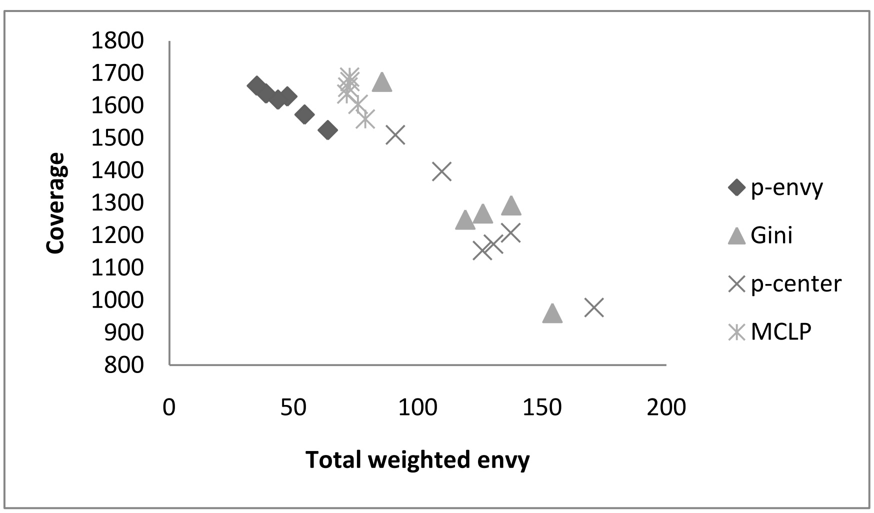


Figure 3.4: Coverage—equity trade off (with available probability station weights)

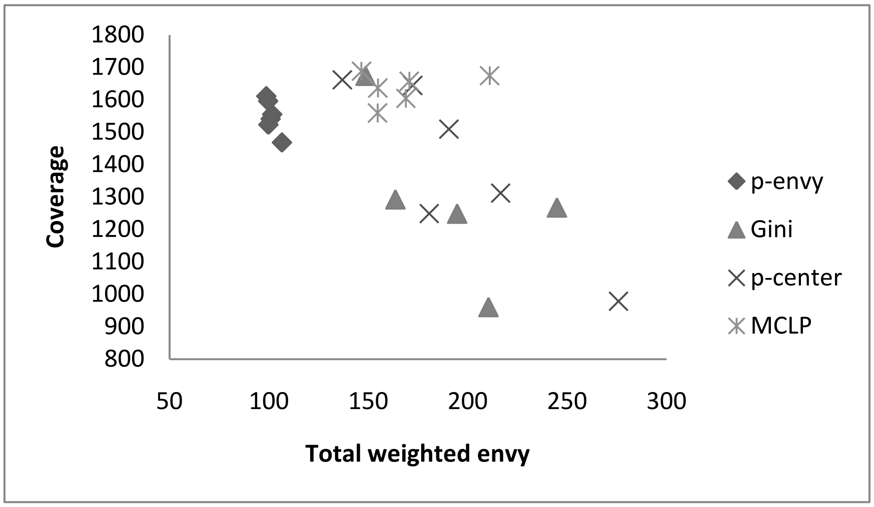


Figure 3.5: Coverage—equity trade off (with equal station weights)

We see that solutions of minimum *p*-envy model dominate solutions of other equity models, and this difference increases as the weights assigned to the backup stations become increasingly important.

3.10. Conclusion and Discussion

In this paper, we have proposed the minimum *p*-envy location problem (M*p*ELP) for EMS systems using the concept of envy which minimizes the inequity of access to service among all zones between all serving facilities (stations). Our model is different in that we consider the effect that all serving stations have on all customers, unlike most equity measures that only consider the effect of the closest facility. Because this objective is complex it results in a problem that cannot practically be solved with commercial optimizers. Thus, a tabu search is developed to solve the problem. Solutions are obtained in a few seconds and the performance of the heuristic is very effective with respect to both computational time and quality of solutions. We also compare the minimum *p*-envy location model with other equity models such as *p*-center and Gini coefficient to see how well the proposed model performs. The results show that the proposed model not only yields the lowest total weighted envy compared with other equity models, but also yields highly efficient solutions in terms of coverage. In fact the coverage of the minimum *p*-envy location model is very close to the coverage resulting from the standard maximal covering location model (MCLP). These results are unexpected, as equity and coverage are usually conflicting objectives (Chanta et al., 2011a). The proposed model is helpful for facility location planners, especially in the realm of public service where reducing inequity is of high importance, though not at the expense of efficiency.

CHAPTER 4

A PROBABILISTIC MINIMUM *p*-ENVY LOCATION PROBLEM: FOCUSING ON SURVIVABILITY OF PATIENTS

4.1 Introduction

Emergency medical service (EMS) is a public service that involves life-or-death situations which often require immediate medical assistance. The EMS system is designed to be able to respond to a 911 emergency call to provide either urgent medical treatment or transport. The system is activated by an emergency call, and then the EMS center dispatches the appropriate medical units to the call. Most EMS systems' performance is measured by the percentage of calls responded to (covered) within some fixed time standard, known as the response time threshold (RTT). Ideally, a system should be able to respond to a call with in the RTT. However, it may not be possible to deliver care within the RTT for all customers; people who live in remote areas usually have to wait longer. For example, Fitch (2005) notes that 90% of calls in urban areas are responded within a 9 minute RTT while 90% of calls in rural areas are responded within 15 minutes. Moreover, when considering coverage, there is no difference between a call responded to within one minute and 8.59 minutes. This is not reflective of patient outcomes; for example, patients who have cardiac arrest need help within 6 minutes otherwise; brain damage is likely to occur (Mayer, 1980).

Since EMS systems provide important basic services, they are expected to serve the public fairly. A patient's chance of receiving timely service is directly affected by the locations and availability of service facilities. Many performance measures in facility location models have been introduced to equalize the chance of access to service between customers. Typically, the objective of these models is to minimize inequity of the system in terms of distance, or to minimize the variation of the distances between demand locations and facilities that serve them. The standard statistical dispersion measures such as range (see e.g. Brill et al., 1976; Erkut and Neuman, 1992), variance (see e.g.; Maimon, 1986; Kincaid and Maimon, 1989; Berman, 1990), mean absolute deviation (see e.g. Berman and Kaplan, 1990; Mulligan, 1991), and sum of absolute differences (see e.g. Keeney, 1980; Lopez-de-los-Mozos and Mesa, 2001; Mesa, 2003) are used as an inequity measure for equitably locating facilities. Moreover, the Gini coefficient, which is commonly used to measure inequity of income, has been brought into the field of equitable facility location design (Maimon, 1988; Erkut, 1993; Drezer et al., 2009). For a review of measures for equity in facility location, see Marsh and Schilling (1994).

In this paper, we apply the concept of envy as one way to capture inequity of the system. The minimum envy model was first introduced in location problems by Espejo et al. (2009). Envy is a measure that considers the differences in service quality between all possible pairs of customers. Since people feel no dissatisfaction when they are better off than others, only negative effects are considered in the minimum envy model. Unlike other measures, the envy measure takes into account all individual effects compared with each other which results in overall satisfaction to the whole system. To say that one customer is better than another customer, we need to define a standard way to quantify the dissatisfaction of each individual which can be done in several ways. Most location models included in Espejo et al.(2009) 's work considers customers' dissatisfaction based on the distance from the customers' locations to their closest facilities, assuming that all customers are only serviced by their closest facilities. This representation is appropriate for some public services, such as post office locations or school locations where the customer travels to the facility, but not necessarily for EMS systems. In an EMS system, the ambulance stationed at the closest facility is not always available to serve the customers, and in that case the ambulance stationed at the next closest facility might instead be dispatched.

To take this into account Chanta et al. (2011b) developed the minimum *p*-envy model which defines dissatisfaction of customer in zone *i* as a function of distance from zone *i* to all *p* serving facility locations weighted by priority of the serving stations. In this paper, we propose an enhancement to the *p*-envy model presented in Chanta et al. (2011b) which focus more directly on patient outcomes. We redefine envy as differences of customers' satisfaction between zones (as opposed to dissatisfaction), and we consider satisfaction is measured by the survival probability of each demand zone (as opposed to distance from a station), which more accurately reflects patient outcomes. The differences of calculating envy based on dissatisfaction or satisfaction is presented along with a study of assignment of priority weights to the *p* serving stations. Moreover, the performance of the model is evaluated regarding of patients' outcomes.

The traditional way to measure performance an EMS system is by considering the coverage or the number of calls that can be responded to within a standard time. That is, a call is considered as "covered" if a vehicle located at a facility is able to reach the call location within the RTT, otherwise it is considered as "uncovered." This measure is called 0-1 coverage, which is commonly used in many facility location models. The 0-1 coverage is simple and easy to interpret, but it cannot distinguish systems with response times faster than the RTT; that is, for a 9 minute RTT, reaching a call in 4 or 9 minutes yields the same coverage. Moreover, the 0-1 coverage considers a call responded to within the RTT as a 100% covered call while it considers a call responded one second later as a 0% uncovered call which is not reflective of patient outcomes. Several ways have been proposed to improve how to calculate the coverage such as using a step function or a gradual function (see e.g. Church and Roberts, 1983; Pirkul and Schiling, 1991; Berman et al., 2003), for review see Eiselt and Marianov (2009). Another way to relax the 0-1 coverage objective is to integrate survival function into the model. Erkut etal. (2008) first introduced using survival function to evaluate the performance of the covering facility location models especially for the EMS systems. McLay and Mayorga (2010) also proposed a way to evaluate performance of the EMS system based on survival probability with respect to a piece-wise function of distance. Since response time directly affect the patients' survival rate; it makes more sense to evaluate the performance of the system based on the overall survival probability instead of standard response time. In our model, survival probability is incorporated into the objective as customers' satisfaction. The performance of our model is evaluated against other well know location models in terms of the expected number of lives saved.

The rest of the paper is organized as follows. In Sections 4.2 - 4.3 we discuss two important model inputs. In Section 4.2 we briefly describe how we estimate survival probability of a demand zone using existing models from the literature; followed by the details of calculating vehicle being busy using probabilities using the hypercube model in

Section 4.3. The notation and formulation of the minimum *p*-envy location model are presented in Section 4.4. Section 4.5 provides an illustrative example. Section 4.6 shows the performance of the *p*-envy location model in comparison to other location models. Section 4.7 discusses the sensitivity of the *p*-envy location model when using different quality measures and different choices of priority assigned to serving facilities. Finally, Section 4.8 provides a conclusion.

4.2 Estimating survival function

Typically, 911-emergency calls are classified by their degree of urgency into three types; priority 1, 2, 3. Priority 1 calls involve with life-threatening emergencies such as cardiac arrest, priority 2 calls may involve life-threatening emergencies, and priority 3 calls are believed to be non-life-threatening. This study focuses on the priority 1 calls for which patient's survival is highly correlated with EMS response time. In particular, the survival probability of a patient who has cardiac arrest depends on the response time. The survival probability at the time of collapse decays linearly to zero if there is no assistance. However, survival probability may remain stable or decreasingly decay when EMS staff arrives and provides pre-hospital administration such as cardiopulmonary resuscitation (CPR), defibrillation, or medications. Early EMS response time leads to early sequence of therapy which yields higher chance of survival. Other factors that might affect survival probability of patient are type of trauma, age, sex, etc. Several studies focus on how to estimate the survival probability of patients who have cardiac arrest based on influential variables including response time. For a review see Erkut (2008).

In this study, we selected the survival function estimated by Valenzuela et al. (1997). The authors found that age, initial of CPR by bystanders, interval time from collapse to CPR, interval time from collapse to defibrillation, bystanders CPR/collapse to CPR interval interaction, and collapse to CPR/collapse to defibrillation interval interactions were significantly associated with survival, they also provided a simplified version of the predictive model in which only collapse to CPR and collapse to defibrillation intervals were used as variables; this model performed comparably to the initially more complex model. The simplified model for estimating survival function is shown as follows.

$$s(t_{CPR}, t_{Defib}) = (1 + e^{-0.260+0.106t_{CRP}+0.139t_{Defib}})^{-1} \tag{4.1}$$

Where s denotes the patient survival probability, t_{CPR} is the interval time from collapse to CPR and t_{Defib} is the interval time from collapse to defibrillation.

For our purposes, let t_{Res} denotes the response time or the travel time of EMS vehicle from station to incident. Assume that it takes 1 minute after collapse to make a call for EMS dispatching, and CPR is performed immediately upon EMS arrival as well as defibrillation which is used by a paramedic or EMT resulting in t_{CPR} = t_{Defib} = 1+ t_{Res} (these assumptions are similar to those made in Mclay and Mayorga (2010)). Then, the model in Equation (4.1) can be rewritten as follows.

$$s(t_{Res}) = (1 + e^{-0.015+0.245t_{Res}})^{-1} \tag{4.2}$$

Figure 4.1 shows the relationship between response time and probability of survival from Equation (4.2).

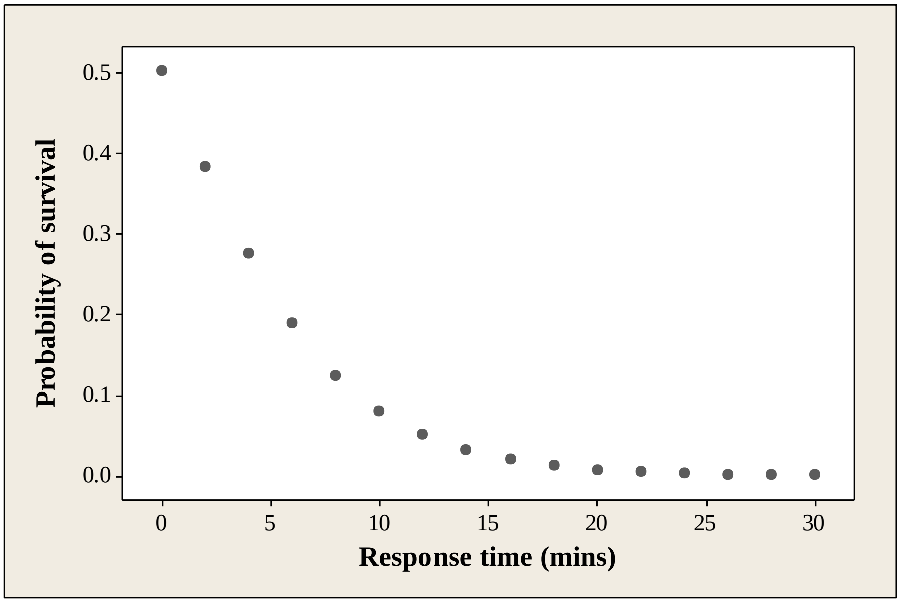


Figure 4.1: Scatter plot of probability of survival vs. response time based on Equation (4.2)

4.3 Estimating probability of vehicle being busy using the hypercube model

Even though an ambulance is stationed close to an incident, it is possible that the ambulance might be busy and unable to serve the call. In order to estimate the probability of ambulance being busy, we used the hypercube model. Let p_b denotes the probability that a randomly selected vehicle will be busy which depends on the number of vehicles that are deployed. Using the actual data of a system, we can estimate the probability p_b by $p_b = \lambda/\mu$ where, λ is the average number of calls per hour to the entire system, $1/\mu$ is the average service time per call (hours), and p is number of ambulances that are deployed. This definition of p_b assumes that all ambulances operate independently. This assumption can be relaxed using the correction factor given by Batta et al. (1989) in an embedded hypercube model. The hypercube model by Larson (1974, 1975) has several underlying assumptions: 1) calls for service arrive according to a Poisson process, 2) if a call arrives while all servers are busy, it enters at the end of a queue and will be served in a FIFO manner. Constructing an M/M/p queuing system operating at steady state, we get the probability that all servers are available, p_0, as given in Equation (4.3). If there are l ambulances that may respond to a call, the probability that the l^{th} vehicle will be dispatched or it is available is calculated from the probability that l-1 ambulances are busy and the l^{th} vehicle is available. The probability that the l^{th} vehicle is available (w_l) is shown in Equation (4.5) where Q (p, p_b, l-1) is the correction factor which is given in Equation (4.4) and Q (p, p_b, 0)=1.

$$p_0 = \left(\frac{(pp_b)^p}{p!(1-p_b)} + \sum_{j=0}^{p-1} \frac{p^j {p_b}^j}{j!} \right)^{-1} \tag{4.3}$$

$$Q(p, p_b, j) = \sum_{k=j}^{p-1} \frac{(p-j-1)!(p-k)p^k {p_b}^{k-j} p_0}{(k-j)!\,p!(1-p_b)} \,, \qquad j=0, \ldots, p\text{-}1 \tag{4.4}$$

$$w_l = Q(p, p_b, l-1)(1-p_b)({p_b}^{l-1}) \,, \qquad l=1, \ldots, p \tag{4.5}$$

4.4 The model

The p-envy model was first proposed by Chanta et al. (2011b), in which the concept of envy is modified to create an objective which is meaningful for the ambulance location problem. A demand zone is a demand point where customers are located. Customers in demand zone i are said to feel envy when they receive inferior service as compared to others, but when they receive superior service they have no feeling of envy. In other words, if customers in zone i have higher (lower) dissatisfaction (satisfaction) than customers in other zones, they feel envy. In the original p-envy model, envy was measured in terms of distance, where longer distances were associated with dissatisfaction. In our model, we define envy in terms of survival probabilities, such that higher survival probabilities are associated with satisfaction. Thus envy of demand zone i is the level of customers' satisfaction in demand zone i as compared to other demand zones. The satisfaction of customers in demand zone i is an ordered vector of the survival probability of demand zone i calculated based on its serving stations (facility locations) in decreasing order. That is, the survival probability of demand zone i when serviced by its closest station, which is the primary station, is the first element in the satisfaction vector, followed by the survival probability of demand zone i when serviced by its next closest station or the secondary station, and so on. The serving stations, except for the primary stations, are called backup stations, of which we can have one or more for each demand zone. Since different demand zones have different total number of customers (demand or call density), we weigh the total envy in each demand zone by the proportion of demand in that zone. An illustrative example of how envy is calculated is presented in the next section. We use the following notation:

n = the number of demand zones

m = the number of potential stations

p = the number of ambulances to be located (stations to be opened)

q = the number of serving stations which consists of one primary station and q-1 backup stations where $q \leq p$

w_l = weight of the k-priority station

H_i = demand (call volume) in zone i

h_i = weight (proportion of demand) of zone i = $\dfrac{H_i}{\sum_{i=1}^{n} H_i}$

s_{ij} = the survival rate of customers in zone i when serviced by station j

The p-Envy Location Model is introduced as an integer programming model. The objective of the equitable location model is to minimize the sum of weighted envy among all demand zones, as shown in Equation (4.6). Equations (4.7) - (4.8) work together to calculate the envy between all possible pairs of customers. The variable e_{ik}^{l} takes on value 0 when zone i is served by a closer facility than zone j compared with the same priority station, otherwise it is equal to the difference between the distance from zone i to its serving station and the distance from zone j to its serving station, that is $e_{ik}^{l} = \max\left\{0, \left(\sum_{j=1}^{m} s_{kj} y_{kj}^{l} - \sum_{j=1}^{m} s_{ij} y_{ij}^{l}\right)\right\}$. Equation (4.9) limits the number of ambulances that are available to be located, or equivalently, number of stations to be opened. Equation (4.10) ensures that a demand zone must be served by exactly one facility at each l^{th} priority station. Equation (4.11) ensures that a station can either serve as a 1st or 2nd or l^{th} priority of zone i. Equation (4.12) requires that a demand zone i can be served by facility j if station j is open. Equation (4.13) assigns a station to serve zone i by considering the survival chance of receiving service from an open station to the zone; the station that provides higher survival chance receives the higher priority to serve zone i.

The Minimum p-Envy Location Model (MpELP):

Minimize $$Z = \sum_{l=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{n} w_l h_i e_{ik}^l \qquad (4.6)$$

Subject to: $$e_{ik}^l \geq \sum_{j=1}^{m} s_{kj} y_{kj}^l - \sum_{j=1}^{m} s_{ij} y_{ij}^l \quad \text{for } i,k=1,\ldots, n\text{: } k\neq i;\ l=1,\ldots, q \qquad (4.7)$$

$$e_{ik}^l \geq 0 \quad \text{for } i,k=1,\ldots, n;\ l=1,\ldots, q \qquad (4.8)$$

$$\sum_{j=1}^{m} x_j = p \qquad (4.9)$$

$$\sum_{j=1}^{m} y_{ij}^l = 1 \quad \text{for } i=1,\ldots,n;\ l=1,\ldots,p \qquad (4.10)$$

$$\sum_{l=1}^{p} y_{ij}^l \leq 1 \quad \text{for } i=1,\ldots, n;\ j=1,\ldots, m \qquad (4.11)$$

$$y_{ij}^l \leq x_j \quad \text{for } i=1,\ldots, n;\ j=1,\ldots, m;\ l=1,\ldots, p \qquad (4.12)$$

$$s_{ij} y_{ij}^l \geq s_{ij} y_{ij}^{l+1} \quad \text{for } i=1,\ldots, n;\ j=1,\ldots, m;\ l=1,\ldots, p\text{-}1 \qquad (4.13)$$

Where:

$$x_j = \begin{cases} 1 & \text{if a facility is located at station } j \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij}^l = \begin{cases} 1 & \text{if a facility at station } j \text{ assigned to serve zone } i \text{ as the } l^{th} \text{ priority station} \\ 0 & \text{otherwise} \end{cases}$$

Note that the e_{ik}^l takes on positive value when customers in zone i have less satisfaction than customers in zone j, which means that customers in zone i envy customers in zone j. Otherwise the e^{l}_{ik} takes on value zero, which means customers in zone i have higher survival rate compared to zone j so they have no feeling of envy. The proportion of demand at node i is the weight (h_i) that we assign to differentiate between call volume at different demand zones. As mentioned earlier, customers in each demand zone may have satisfaction with respect to all serving stations; first priority station, second priority station, and so on. Thus, we can differentiate the envy with respect to different serving stations by adding the different weights (w_l) to each level of priority; l=1, ..., q, where q is the number of serving stations that are restricted to respond to a particular zone. Note that $q \leq p$ where p is the number of stations that will be opened. A station is said to be opened when there is at least one ambulance stationed for serving customers. We note from Equation (4.6) that since there is no contribution to the objective of locating more than one ambulance at the same station, the number of open stations and the number of ambulances are the same. To avoid a trivial solution we assume that $m \geq p$ where m=the number of potential station locations, otherwise there are excess ambulances to be located. This assumption allows us to specify the effect that each station has on a demand zone through the vector $\mathbf{w}=(w_1, \ldots, w_q)$, $w_l \geq 0, \forall_l$. Without loss of generality, we scale and order the w_l's such that $\sum_{l=1}^{q} w_l = 1$ and $w_1 \geq w_2 \geq \ldots \geq w_q$.

Station priority weights can be assigned in various ways, depending on how the system administrator values backup service. For example, if a system only utilizes one backup station, we can set w_1 and w_2 to be active, and the rest to be 0.

4.5 Illustrative example

The objective function described in (4.6) is difficult to calculate. Here we present an illustrative example. Suppose there are three demand zones, three potential stations for locating EMS ambulances, and two ambulances. In this case, n=3, m=3, and p=q=2. To estimate the survival probability of each demand zone, we have to know the location of its serving facilities. Once we know which facility is open, we can calculate the probability of survival using the relationship between the response time and the survival probability provided in Equation (4.2), assuming that response time is a function of the distance. Matrix **d** is an input distance matrix in which each element d_{ij} represents the distance (in miles) from demand zone i to station j where the number of rows represents the number of demand zones (n) while the number of columns represents the number of potential stations (m). Assuming that exactly 2 minutes are required to travel 1 mile, we get the response time to be used to estimate the probability of survival with respect to all stations (**s**). In a previous paper, Chanta et al. (2011b) directly used the distance matrix as customer's dissatisfaction to calculate envy, but in this paper we attempt to more realistically reflect patient outcomes by using the survival probability matrix as customer's satisfaction to calculate envy. Other inputs include the proportion of demand in each demand zone i (**h**), and the weights assigned each priority open station (**w**). The inputs to this small example are given below in matrix form below.

$$\mathbf{d}=\begin{bmatrix} 2 & 2 & 10 \\ 8 & 4 & 6 \\ 10 & 5 & 2 \end{bmatrix} \quad \mathbf{s}=\begin{bmatrix} 0.277 & 0.277 & 0.007 \\ 0.019 & 0.125 & 0.050 \\ 0.007 & 0.080 & 0.277 \end{bmatrix} \quad \mathbf{h}=\begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix} \quad \mathbf{w}=\begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$$

The vector **h** denotes that 20%, 30% and 50% of customer calls originate in demand zones 1, 2, and 3, respectively. The vector **w** indicates that a patient's envy will be comprised of 60% resulting from envy regarding their primary serving station and 40% envy regarding their secondary serving station. Suppose ambulances are located at station 1 and station 2, the third column of the matrix **s** is neglected, and then customers in demand zone 1 have probability of survival of 0.277 if it is reached from its 1st priority or primary station, and also 0.277 chances of survival if it is reached from its 2nd priority or secondary station. Demand zone 2 is located closer to station 2, so station 2 serve as a 1st priority station of zone 2, and station 1 serves as a 2nd priority station of zone 2, with survival probability of 0.125, and 0.019, respectively. The same with demand zone 3, it is served by station 2, and station 1 as 1st and 2nd priority stations respectively. Next, the envy of demand zone *i* with respect to demand zone *j*, in regards to their 1st priority stations is calculated from the difference of survival probability of demand zone *i* regarding to the service provided by its 1st priority station and survival probability of demand zone *j* regarding to the service provided by its 1st priority station whereas if demand zone *i* has higher probability of survival than demand zone *j* regarding to their 1st priority stations, the envy of demand zone *i* with respect to *j* is equal to 0, because demand zone *j* does not have higher chance of survival than demand zone *i*. If demand zone *i* has survival probability regarding to the service from its 1st priority station lower than demand zone *j* has to theirs, demand zone *i* envies demand zone *j* which we quantify as the difference in satisfaction between demand zone *i* and demand zone *j*. The envy matrix (e_{ik}^{l}**)** corresponding to locating ambulances at station 1 and 2 is calculated from the summation of max{0, $\sum_{j=1}^{m} s_{kj} y_{kj}^{l} - \sum_{j=1}^{m} s_{ij} y_{ij}^{l}$ } where l=1,2; i,k=1,2,3; $k \neq i$. For example, e_{12}^{1} =max{0, 0.125-0.277}=0, e_{13}^{1} =max{0, (0.080-0.277)}=0, and e_{23}^{1} =max{0, 0.080-0.125)}=0, e_{21}^{1} =max{0, 0.277-0.125)}=0.152. The total envy of all demand zones with respect to all serving stations is equal to the summation of all elements in the envy matrix multiply by the demand zone weight (h_i) and the station weight (w_l). If we locate ambulances at station 1 and 2, the total envy of all demand zones is equal to 0.1872. Our goal is to find the station locations that give the minimum total of envy. With this small-size example, one can easily enumerate all possible solutions, and the optimal solution is opening stations at locations {2, 3} with a total envy value of 0.0676. So locating ambulances at stations 1 and 2, we balanced chances of survival among all customers in the system. Using the integer programming formulation of the minimum *p*-envy model, developed in previous section, a solver found an optimal solution at *x*={0,1,1}, *y*={[(0,0), (1,0), (0,1)], [(0,0), (1,0), (0,1)], [(0,0), (0,1), (1,0)]}, *e*={[(0,0), (0,0.043), (0,0.073)], [(0.152,0), (0,0), (0.152,0.030)], [(0,0), (0,0), (0,0)]}.

4.6 Performance of the minimum *p*-envy location model with survival function

In this section, we would like to compare our *p*-envy model to other location models. Three well-known location models selected are maximal covering location problem (MCLP), *p*-center, and Gini coefficient. Since the *p*-envy belongs to a class of equitable location models; thus, we want to compare it with other equity measures such as *p*-center, which improves the service quality of the customer who received the worst service, and Gini coefficient, which minimizes the differences in service quality between customers. Moreover, most equity location models tend to give low coverage, thus we also compare the *p*-envy with the MCLP model.

For fair comparison, instead of evaluating the objective function based on the distance matrix as the original version of these three models do, we use the survival probability. That is, in this paper the *p*-center model maximizes the customer who receives the lowest survival rate (we elaborate further on this later). Because we focus on the EMS system, using the survival rate is more meaningful to represent real patient outcomes than distance. Most location models use the distance traveled from the facility to the demand zone or the response time as an input metric. However, distance is a linear function while survival probability is a nonlinear function. Thus, based on the distance matrix, if an ambulance is located closer to a demand zone the contribution to the objective is increasing linearly as distance is decreased; while the survival probability of patient is increasing non-linearly. The survival function estimated by Valenzuela et al. (1997) gives us one way to convert the response time to survival probability. This can lead to different solutions to the facility location problem. While there are many other possible survival functions that could be used, we choose this one as a way to that illustrate the resulting solution can be very different when some direct metric for patient outcomes (such as survival probability) is used instead of distance in the objective of location models.

Below we review each location model used to compare with the *p*-envy model enhancement proposed here. As mentioned earlier, for fair comparison, each model has the same input metric to the objective function; that is survival probability. For some models, using distance and survival does not affect the solution. For example, the MCLP maximizes the number of calls that can be responded to within some time standard, which can be reinterpreted in terms of survival as maximizing the number of calls that can responded to in order to achieve at least the survival probability associated with that time standard. For example, if an ambulance is able to reach the call within 9 minutes, a patient has survival probability equal or greater than 0.125 according to Equation (4.2). In *p*-center, the goal is to improve the customer that receives the worst service in the system. Traditionally, the worst service refers to the customer farthest from a facility, in this case, it is the one with the lowest survival probability. So, the objective of the *p*-center model in this context is to maximize the minimum survival probability. For both the MCLP and *p*-center models, using distance as opposed to survival as a metric does not change the solution. This is not so for the Gini coefficient. This model minimizes the differences between individuals. Instead of distance, here each individual has different survival probabilities that depend on the station locations. The formulations of each model are provided below. Note that the Gini coefficient only considers differences in quality of service from the closest serving station, while the *penvy* envymodel considers all *p* serving stations.

- Minimum *p*-envy location problem (M*p*ELP)

  Objective is to minimize sum of envy weighted by proportion of demand:

$$\min Z = \sum_{l=1}^{q} \sum_{i=1}^{n} \sum_{k=1}^{n} w_l h_i e_{ik}^{l}$$

Subject to (4.7) - (4.13).

- Maximal covering location problem (MCLP), see original version in Church and ReVelle (1974)

  Objective is to maximize proportion of demand that can be covered (reached within a given response time threshold):

$$\max Z = \sum_{i=1}^{n} y_i H_i$$

Subject to $\sum_{j=1}^{m} a_{ij} x_j \geq y_i$ for all i=1, 2, …, n (4.14)

and (4.9)

Where $y_i = \begin{cases} 1 & \text{if demand zone } i \text{ is covered by an open station} \\ 0 & \text{otherwise} \end{cases}$

$$a_{ij} = \begin{cases} 1 & \text{if station } j \text{ can cover demand at zone } i \\ 0 & \text{otherwise} \end{cases}$$

- *p*-center, see details in Daskin (1995)

  Objective is to minimize the maximum distance from customers to their closest station:

$$\max Z$$

Subject to $\sum_{j=1}^{m} s_{ij} y_{ij} \geq z$ for all i=1,2, …, n (4.15)

$\sum_{j=1}^{m} y_{ij} = 1$ for all i=1, 2, …, n (4.16)

$y_{ij} \leq x_j$ for all i=1,2, …, n, j=1, 2, …, m (4.17) and (4.9)

Where $y_{ij} = \begin{cases} 1 & \text{if a demand zone } i \text{ is served by facility at station } j \\ 0 & \text{otherwise} \end{cases}$

- Gini coefficient measure (Gini), see details in Drezner (2009)

  Objective is to minimize Gini coefficient (a weighted measure of absolute differences):

$$\min \frac{\sum_{i=1}^{n}\sum_{k=1}^{n}\left|\sum_{j=1}^{m} s_{ij} y_{ij} - \sum_{j=1}^{m} s_{kj} y_{kj}\right|}{2n\sum_{i=1}^{n}\sum_{j=1}^{m} s_{ij} y_{ij}}$$

  Which is equivalent to minimizing the numerator:

$$\min \sum_{i=1}^{n}\sum_{k=1}^{n}\left|\sum_{j=1}^{m} s_{ij} y_{ij} - \sum_{j=1}^{m} s_{kj} y_{kj}\right|$$

  Subject to (4.9), (4.15) - (4.16)

We compare the performance of these models by considering the equity and efficiency. The total envy represents the equity of the system while the number of lives expected to be saved represents efficiency of the system, which is calculated by using the survival function. The number of lives saved in each zone is calculated based on the survival probability with respect to the distribution of their serving facility, and the summation of all zones represents the expected number of lives saved of the whole system.

We use real world data from the Hanover County, VA Fire and EMS department, which contains 122 demand zones and 16 potential stations, to serve a county of 474 square miles as shown in Figure 4.2. The area is divided into 175 demand zones of 2 by 2 mile squares, where each zone has requested calls aggregated at the center of the zone. The distance between existing facilities and each demand zone is estimated by using the Euclidian distance since about 70% of the county is rural. We excluded the zones that have no demand, so the total number of the zones that demand exists is 122. Based on data during the year 2007, the average number of requested calls in Hanover is 1.2 calls/hour during the peak hours when the call volume is constant. The call volume of interest is from the evening weekend data. This time period was selected for two reasons. First, the data analysis suggests that these times operate in steady state, with the customer arrival rate approximately constant per unit time. The call volume used in this example is 1711 calls (note that the total call volume during the year is >6000). The average service time per call is 74 minutes or 1.2 hours. This data is necessary for estimating the probability that vehicle will be busy. The total number of ambulances to be located is varied from 5 to 10. Thus, in this case n=122, m=16, and p varies from 5 to 10. We use a 9 minute RTT as it is what was in place in Hanover County. h_i is the proportion of demand at location i ; i=1, ..., 122 and all w_l values are estimated by using hypercube model M/M/p as describes in Section 4.3, where l=station priority; $l \in i, \ldots, q$, and q=number of serving stations. Note that $w_1 \geq w_2 \geq \ldots \geq w_q$, respectively. The probability of the l^{th} priority vehicle being available (while if all other higher priority vehicles are busy) is shown in Table 4.1.

Table 4.1: Probability that vehicle l^{th} available (w_l)

p	l									
	1	2	3	4	5	6	7	8	9	10
5	0.7040	0.1879	0.0585	0.0214	0.0091					
6	0.7533	0.1739	0.0464	0.0145	0.0053	0.0022				
7	0.7886	0.1593	0.0368	0.0098	0.0031	0.0011	0.0005			
8	0.8150	0.1459	0.0295	0.0068	0.0018	0.0006	0.0002	0.0001		
9	0.8356	0.1340	0.0240	0.0048	0.0011	0.0003	0.0001	0.0000	0.0000	
10	8.8520	0.1237	0.0198	0.0036	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000

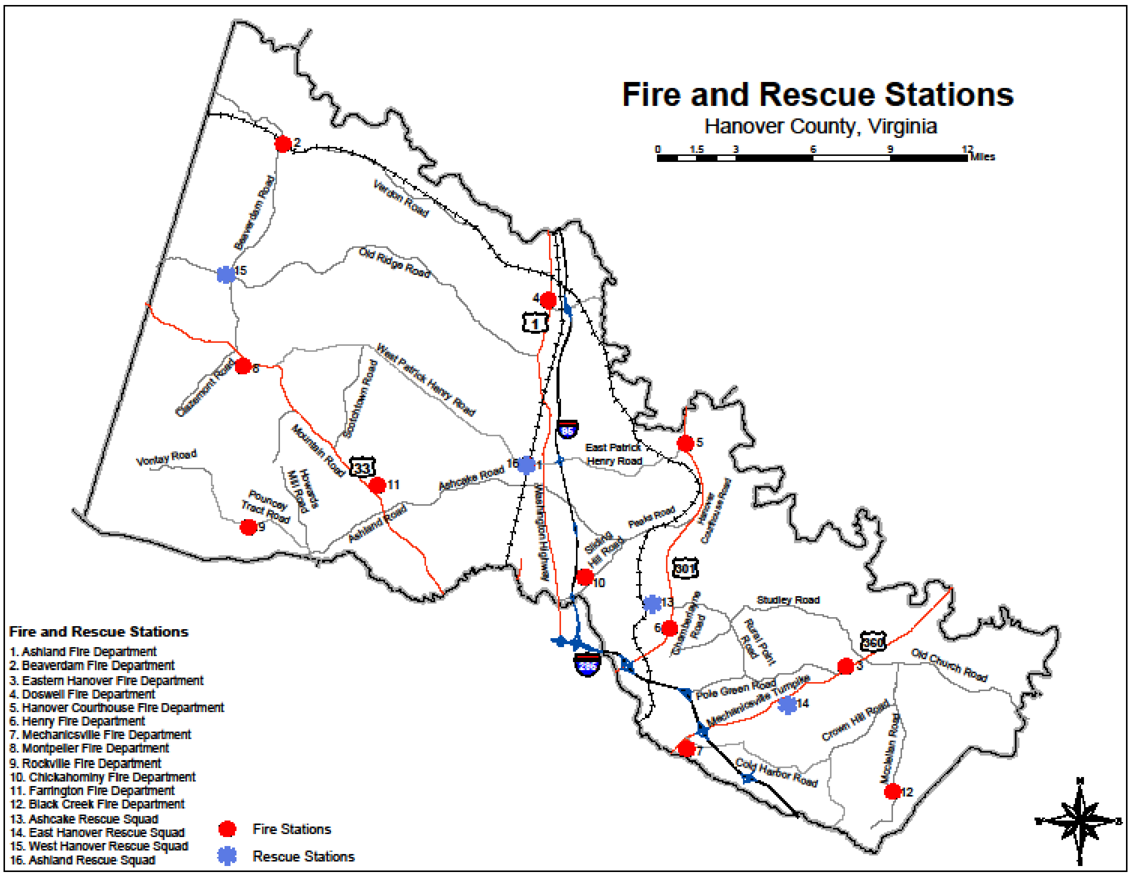


Figure 4.2: Map of fire and rescue stations in Hanover County, Virginia

We tested four facility location models with 5 cases each, for a total of 20 cases. These are solved on a Dell Latitude D410 machine with Intel Pentium processor 1.73

GHz, 1 GB of RAM. The Tabu search, which developed according to Chanta (2011b), obtained the optimal solution in each case in 3 to 5 seconds. We also verified the solution obtained by the Tabu search to the optimal solution obtained by enumerating all possible solutions to make sure that each solution is optimal. The results are reported in terms of each objective value: the minimum survival probability (Minrate) between demand zones and their closest open stations which is the *p*-center objective, the Gini coefficient (Gini) which is the Gini objective, the total covered demand (coverage) which is the MCLP objective, and the total weighted envy (total envy) which is the *p*-envy objective. We also do post-processing to report other relevant performance measures: the number of lives saved (livesaved), the average survival probability (avg), the weighted average survival probability (wavg). The results are shown in Tables 4.2-4.5.

Table 4.2: Min *p*-envy with survival rate

p	Opened stations	Minrate	Gini	**Total envy**	Coverage	Post-processing		
						livesaved	avg	wavg
5	{1 4 7 8 13}	0.0028	0.4645	2.2179	1543	508.4248	0.1428	0.2972
6	{1 4 6 7 8 10}	0.0028	0.4701	2.3563	1525	536.7084	0.1477	0.3137
7	{1 4 5 6 7 8 10}	0.0028	0.4566	2.4740	1557	545.2211	0.1574	0.3187
8	{1 3 4 6 7 8 9 10}	0.0075	0.3805	2.5527	1597	556.0802	0.1833	0.3250
9	{1 3 4 5 6 7 8 10 11}	0.0075	0.3346	2.5723	1637	566.2083	0.2030	0.3309
10	{1 2 3 4 5 6 7 8 10 11}	0.0197	0.2786	2.5591	1666	575.1165	0.2231	0.0197

Table 4.3: Max MCLP with survival rate

p	Opened stations	Minrate	Gini	Total envy	**Coverage**	Post-processing		
						livesaved	avg	wavg
5	{1 4 6 14 15}	0.0028	0.4242	3.3349	1559	398.1236	0.1525	0.2327
6	{1 4 6 11 14 15}	0.0075	0.361	3.9310	1604	410.1897	0.1763	0.2397
7	{ 1 4 5 6 11 14 15}	0.0075	0.3464	4.2469	1636	418.7129	0.186	0.2447
8	{ 1 4 5 6 9 11 14 15}	0.0197	0.3194	4.6383	1657	424.4131	0.1987	0.248
9	{ 1 2 4 5 6 8 9 11 14}	0.0197	0.2863	4.9833	1674	436.1053	0.2149	0.2549
10	{ 1 2 4 5 6 8 9 11 12 14}	0.0509	0.2557	5.4090	1688	439.8288	0.2285	0.2571

Table 4.4: Max *p*-center with survival rate

p	Opened stations	**Minrate**	Gini	Total envy	Coverage	Post-processing		
						livesaved	avg	wavg
5	{1 2 3 6 8}	0.0197	0.3815	3.8798	1509	364.9606	0.1621	0.2133
6	{1 3 4 9 13 15}	0.0318	0.3458	5.2742	1249	333.4969	0.1746	0.1949
7	{2 3 4 8 9 11 13}	0.0509	0.3156	7.7409	978	256.6419	0.1915	0.1500
8	{2 3 4 5 6 8 9 11}	0.0509	0.3019	6.5561	1312	315.8165	0.201	0.1846
9	{1 2 3 4 5 6 8 9 10}	0.0509	0.2922	4.4116	1644	434.5122	0.2103	0.254
10	{1 2 3 4 5 6 7 8 9 10}	0.0509	0.2849	2.5884	1661	572.0506	0.2187	0.3343

Table 5: Min Gini with survival rate

p	Opened stations	Minrate	**Gini**	Total envy	Coiverage	Post-processing		
						livesaved	avg	wavg
5	{3 4 11 13 15}	0.0075	0.3628	6.2300	940	239.2495	0.1626	0.1398
6	{2 3 4 8 11 13}	0.0197	0.3354	7.1103	958	251.1741	0.1798	0.1468
7	{1 2 3 4 8 11 13}	0.0197	0.3107	6.1343	1273	348.8485	0.1965	0.2039
8	{1 2 3 4 8 9 11 13}	0.0509	0.2879	6.5704	1293	354.3163	0.2083	0.2071
9	{1 2 4 8 9 11 12 13 14}	0.0509	0.2703	5.6100	1593	398.3211	0.2196	0.2328
10	{1 2 3 4 7 8 9 11 12 13}	0.0509	0.2530	2.9830	1678	546.8512	0.2305	0.3196

In these tables, larger values of covered demand and maxi-min survival rate are desirable while smaller values of Gini coefficient and total weighted envy are desirable. The four models are compared in Figures 4.3 to 4.5 in terms of the resulting equity and efficiency measures. From Figure 4.3, we see that *p*-envy yields the lowest total envy as expected. Interestingly, it also yields high coverage as shown in Figure 4.4. These results are interesting, as equity location models tend to trade off coverage in order to achieve higher equity. As expected, since the MCLP focuses only on efficiency; it yields the best coverage while the other three location models, which belong to the category of equitable location models, are expected to yield lower coverage when compared to the MCLP. The results show that the *p*-envy model yields highest coverage among the three equitable location models, and in fact, its coverage is almost as good as the optimal coverage provided by the MCLP.

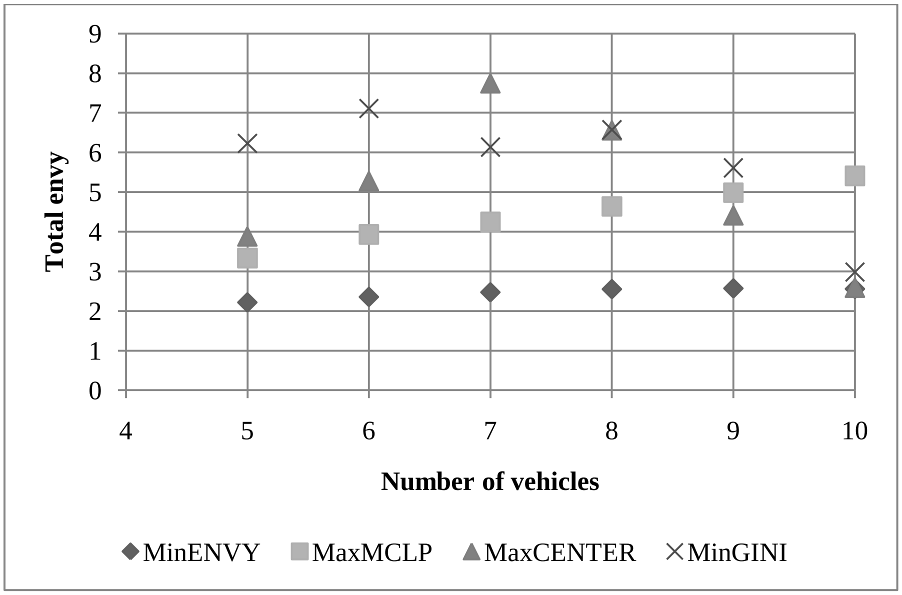


Figure 4.3: Equity comparison of location models

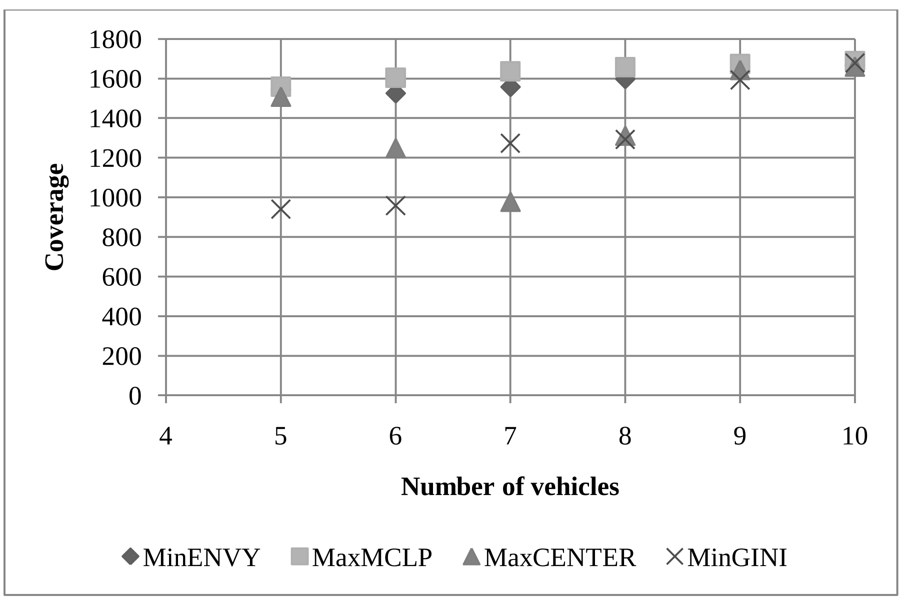


Figure 4.4: Coverage comparison of location models

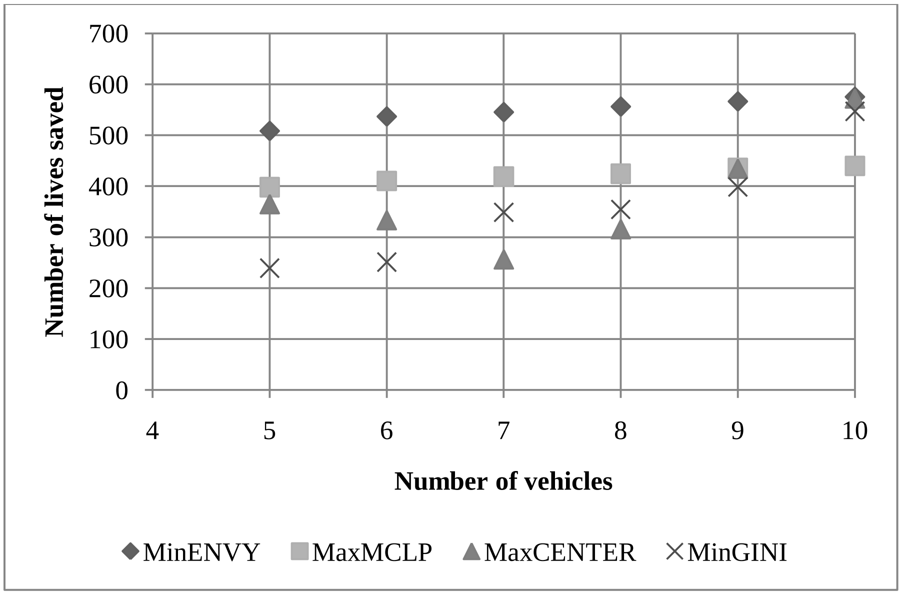


Figure 4.5: Efficiency comparison of location models

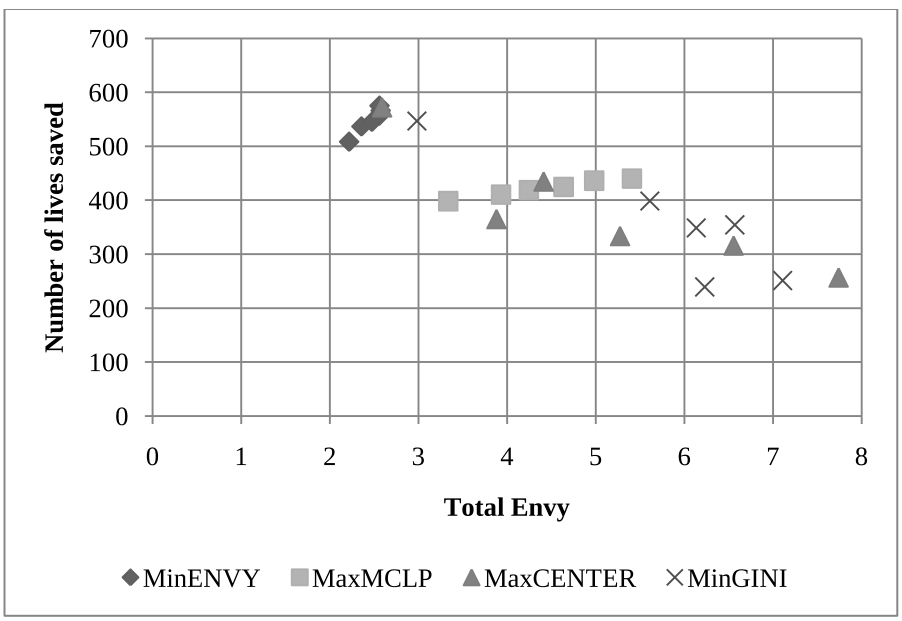


Figure 4.6: Equity-Efficiency trade off among location models

As mentioned earlier, coverage alone may not be a suitable criterion to measure the performance of an EMS system. We are interested in the number of lives saved as shown in Figure 4.5. The *p*-envy model yields the highest number of lives saved which means that by reducing the envy of the system with respect to survival probability we are able to save more lives than by focusing on other measures. The benefits (in terms of number of lives saved) of using the *p*-envy model increases as the number of vehicles decreases. If a system has high resource capacity (in this case ambulances), locating ambulances by any one of the three equity measures tends to yield the same number of lives saved. However, if the system has limited resources, in this example, less than 10 ambulances, locating ambulances by different equity measures could drastically reduce the number of lives saved (out of 1711 calls).

To illustrate the tradeoff between equity and efficiency, we plot the performance of all four models with respect to equity and number of lives saved. Figure 4.6 shows the results of all four models where the *p*-envy model uses the station available probabilities (see Section 4.3) for station weights. Interestingly, we see that the minimum *p*-envy location model with survival function not only yields the lowest total envy, but also yield the highest number of lives saved. Therefore, the *p*-envy model allows us to reduce inequity and at the same time maintain efficiency in term of saving lives. These results suggested that *p*-envy is more preferable to EMS systems than the other three. The results depend on the weights assigned to the priority of the stations (vector **w**).

4.7 Sensitivity analysis of the *p*-envy model inputs

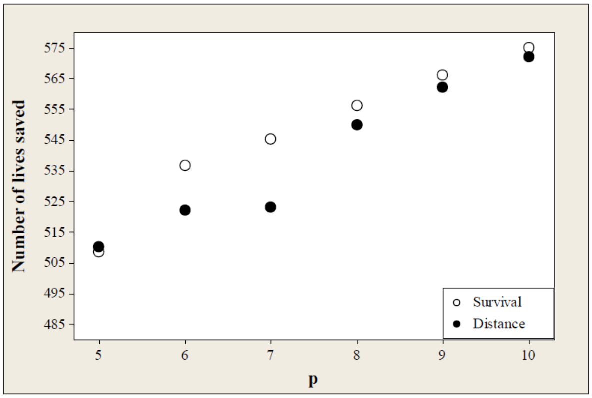
In this section, the results of the *p*-envy model when using different measures to calculate the total envy of the system are compared when different priority weights are given to stations. In Section 3.6, we used the survival function to calculate the envy of each demand zone and the total envy of the system. In this section, the distance matrix has been used to quantify the envy of each demand zone and then the summation of the envy at each demand zone is the total envy of the system. The distance from a demand location to its serving facility represents customer's dissatisfaction. This is opposite to the way we calculate envy using the survival probability; in this case, people feel envy when they are further away from a facility. Let **d** be the distance matrix from a demand zone to all existing stations, where d_{ij} is the distance from demand zone i to facility at station j. The objective function is changed when we are working with dissatisfaction data instead of satisfaction data. Let e_{ik}^{l} be the envy of zone i compared to zone k with respect to their l^{th} priority stations. Then $e_{ik}^{l} = \max\left\{0, d_{ij}y_{ij}^{l} - d_{kj}y_{kj}^{l}\right\}$, where the positive value of the max function represents the feeling envy. So the constraints (4.7) and (4.13) need to be changed as the following.

$$e_{ik}^{l} \geq \sum_{j=1}^{m} d_{ij}y_{ij}^{l} - \sum_{j=1}^{m} d_{kj}y_{kj}^{l} \qquad \text{for } i,k=1,\ldots,n\text{: } k\neq i;\ l=1,\ldots,q \qquad (4.18)$$

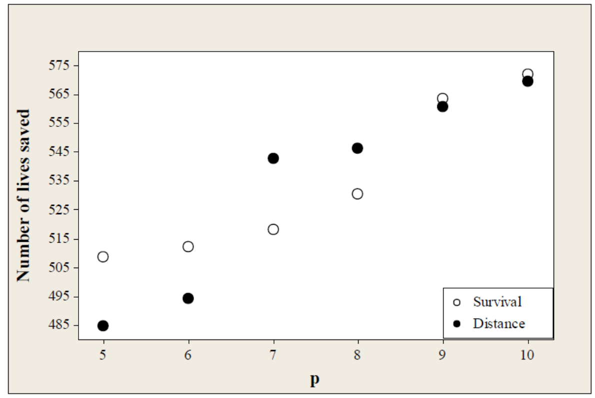
$$d_{ij}y_{ij}^{l} \leq d_{ij}y_{ij}^{l+1} \qquad \text{for } i=1,\ldots,n;\ j=1,\ldots,m;\ l=1,\ldots,p-1 \qquad (4.19)$$

Since we focus on EMS systems, the survival function is a reasonable measure to quantify envy. However, distance is most often used in location models. Here, we investigate how different the solution based on the distance matrix is from the solution based on the survival function in the *p*-envy model.

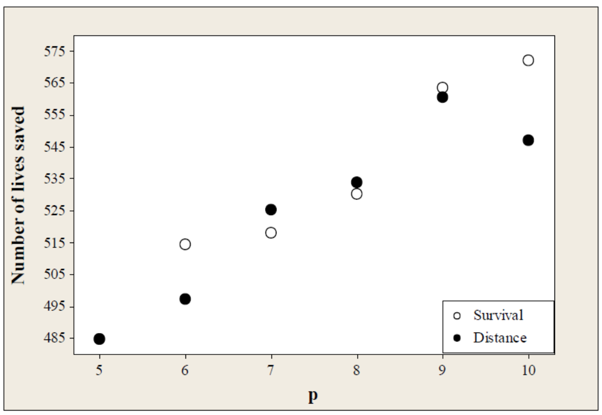
Another factor that would effect on the solution of the *p*-envy model is the station weight; thus, we tested 4 different ways to assign the weights (or priorities) to stations. The first way is assigning the probability that a vehicle at each station will be available as a station weight (**w**). The probability of vehicles being available is estimated by the hypercube model as described in Section 4.3. Note that the weights estimated by hypercube model are nonlinearly decreasing. An important property of the vector **w** is $w_i \geq w_j$ if $i \leq j$ since an ambulance at a higher priority station should be more likely to be dispatched. One might design a simple way to choose the weight vector by making it linearly decreasing by assigning $w_l = \frac{q+1-l}{K}$, where K=1+2+ ... +q, l=station priority ; $l \in \{i, \ldots, q\}$ and q=number of serving stations. If all serving stations are equally likely to be dispatched, we can set all w_k values to be the same; $w_k = \frac{1}{q}$ for all $k \in \{i, \ldots, q\}$. If a system doesn't have a backup station or in other words, only the closest station is always dispatched, we can set the **w**=(1,0, ...,0). Figure 4.7 shows the solutions of the *p*-envy model when using different input matrices and different station weights.



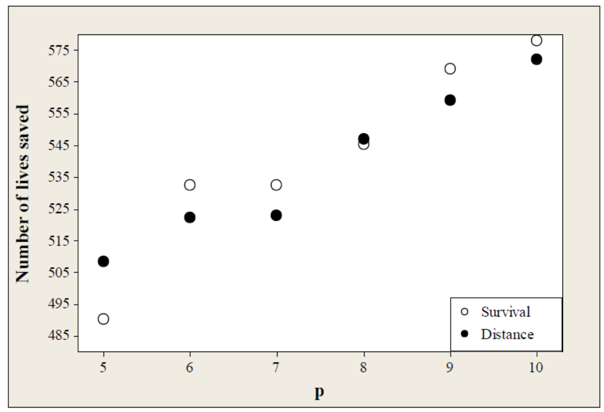
(a) Hypercube probability weight



(b) linear decreasing



(c) equal weight



(d) weighted on only the closest station

Figure 4.7: The *p*-envy model with different measures and different weight vectors

From Figure 4.7 we make several observations. First, note that using the probabilities of vehicles being available as priority weights yields the highest number of lives saved among all four cases regardless of the envy measure used. That is, assigning station weights according to either (b) or (c) or (d) either attaches to much or too little importance to the backup stations. The difference between the performance of distance versus survival as a measure of envy depends on the priority weights assigned to stations. Figure 4.7(a) shows that using distance as a measure of envy leads to degraded performance of the system. On the other hand, panels (a)-(d) in Figure 4.7 show that the results of using the distance matrix depend on the number of stations in use. Over all cases ($p \in [5, 6, \ldots, 10]$), the average benefit of using survival as a measure of envy instead of distance when priority weights are assigned based on busy probabilities is 8 more lives saved. The probability based weight assignment yields similar benefits, resulting in 12 (8) more lives saved over using other weight assignments when survival (distance) is used to measure envy. The details of other cases are summarized in Table 4.6.

Table 4.6: Average number of lives saved gained by using survival objective

weight (**w**)	avg number of lives saved		benefit gained
	survival	distance	(out of1711 calls)
a) hypercube probability	547.96	539.93	8.03
b) linearly decreasing	534.10	532.99	1.10
c) equal	530.49	524.75	5.75
d) on only the closest station	541.22	538.67	2.55
average benefit of (a)	12.69	7.80	

4.8 Conclusion and discussion

Minimum *p*-envy is one of many equity measures that have been used to minimize inequity of a system in facility location problems in which quality of service depends on the distribution of the facility locations. Most facility location models represent the quality of service as the distance traveled from a demand location to its closest facility location. In this paper, we discuss another way to represent quality of service by relating to patient outcomes. In particular we consider envy with respect to survival probability, which is attained as a function of response time. Furthermore, since the analysis is focused on an EMS system, the expected number of lives saved is calculated ex-post to assess the performance of the system. Four different location models, which are maximal covering location problem (MCLP), *p*-center, Gini coefficient, and *p*-envy, are selected to be studied and their performance is compared using the survival probability as a quality of service measure instead of distance traveled. The optimal solution for each problem is then further analyzed in order to gauge the performance of each model side by side. Measures of interest included number of lives saved: the average survival probability and the weighted average survival probability of the system.

The *p*-envy model yielded the lowest total weighted envy of the system while maintaining high coverage; the coverage is almost as high as the MCLP. Moreover, the *p*-envy yielded the highest number of lives saved among these four location models. From sensitivity analysis, we found that the solution of the *p*-envy model depends on the quality of service measures and the station weights. Using distance instead of survival probability may result in overestimation or underestimation of performance of the system. The solution gap depends on how the station weights are assigned. A station weight assigned to a given station should be associated with the proportion of time that a vehicle at the station is likely to be dispatched in the real situation. Thus including survival probability as well as busy probabilities in the *p*-envy model can results in many additional lives saved at no additional costs. The benefits of using the *p*-envy model over other facility location models, in terms of number of lives saved, are similar. These benefits increase as resources become more limited.

CHAPTER 5

CONCLUSION

5.1 Summary

We have presented three different location models that deal with the equity issue in EMS systems. All three location models are formulated as integer programs. The objective is to minimize inequity of service among customers.

In Chapter 2, we proposed three bi-objective location models that focused on balancing equity of service between rural and urban areas. Each model is formulated as a bi-objective programming where the first objective is to maximize the number of covered calls, while the second objective is to reduce disparity between urban and rural areas. We proposed three ways to reduce inequity of the system: a) minimize the maximum distance between the uncovered zone to its closest stations, b) minimize the number of uncovered rural zones, c) minimize the number of uncovered zones either it is a rural or urban, which result in three bi-objective location models. We accounted for the probability of a vehicle being busy and considered partial coverage by using the hypercube queuing model. We solved the problem with the ε-constraint approach via an optimization software, and the optimal solutions were found in several seconds. The results showed that all three bi-objective location models have ability to balance disparities between rural and urban areas. In particularly, Model (a) yielded the lowest average weighted distance from all call locations to their closest stations, Model (b) produced the largest number of non-dominated solution points, and Models (a) and (c) yielded equitably efficient solutions in terms of providing equally individual effects of system (Chanta et al., 2009, 2011a).

In Chapter 3, we proposed a new equitable location model, namely, the minimum *p*-envy location problem, which focused on minimizing inequity of a system by reducing the differences of dissatisfactions among all customers in the system. The concept of envy was applied and incorporated into the objective which allowed us to consider the inequity based on the effect of everyone in the system based on the distribution of facility locations. The model considers the probability of a vehicle being busy using the hypercube model and specifies the priority of the serving stations; primary station and backup stations. The problem was solved via the developed heuristic, tabu search, since optimization software cannot handle large size problems. The tabu search obtained near-optimal solutions in a few seconds. The result of the *p*-envy model was compared to other location models, and it showed that *p*-envy yielded lowest total envy in the system while maintaining as high as coverage as the maximal covering location model (Chanta et al., 2010a, 2011b).

In Chapter 4, we extended the performance of the minimum *p*-envy location problem by using an input metric for evaluating the objective which is more directly related patients' outcome. The probability of survival was incorporated into the objective, and the inequity of the system is still minimized by reducing the total envy of the system as original minimum *p*-envy location problem. But instead of minimized the differences of dissatisfactions, we minimized the differences of satisfactions among all customers in the system. The hypercube model is used to estimate the probability of a vehicle being busy at each station, as well as the priority of serving stations that was taken into account. The results of the *p*-envy model with survival probability compared to other location models showed that higher number of lives are saved when locating facilities based on the proposed model at the same capacity of resources (Chanta et al., 2010b).

5.2 Concluding Remarks

Minimizing inequity in a facility location problem can be done in several ways with different objective functions. Designing the objective function is the first important step that we have to consider. The objective function should be able to represent the overall inequity of a whole system. An effective objective function leads to improve both equity of overall system and individual effect. Minimizing the number of uncovered rural zones reduces overall inequity of the system but does not provides small individual effect compared to minimizing the maximum distance from an uncovered zone to its closest station, which reduces overall inequity of the system and also reduces the effects of individuals.

Most of facility location models evaluate their objective functions based on the traveled distance from customers to facilities. This measure is not appropriate for EMS system which is related to life/death situation. Survival chance of patients is a key thing that should be considered and incorporated in to the model. We have shown that using survival probability to evaluate the objective instead of traditional distance can greatly improve the performance of the minimum *p*-envy location model. However, not all facility location models can improve their performances by simply replacing the survival probability with the traveled distance.

A good facility location model should be able to capture realistic situations in the system, so we ensure that it represents the real system. The proposed bi-objective models are able to account for the chance of vehicle being available according to system busyness or queuing, and partial coverage with facilities at the same or different stations. The minimum *p*-envy location model is able to translate real customers' feelings in to an equitable location model. It is able to account for the chance of vehicle being available according to system busyness, the priority weights of serving stations; primary and backup stations, including the chance of patients' survival. Fail to capture the realities of the system may lead to an underestimate or overestimate the performance of the system.

5.3 Future Work

Incorporating how a system operates its facilities into the facility location model could be an interesting area for facility location model for EMS systems. EMS systems operate their facilities differently, depending on available resources, capacity of staffs, geographical area, etc. Customizing the model to reflect the system operations leads to more accuracy of the model. For example, combining a dispatching rule and a districting zone in to the model makes the model more realistic. Different zones might have different dispatching rules.

In real life, emergency calls request for different kinds of helps; from a basic life rescue to a serious injury. Moreover, one might require immediately help while another one might be able to wait for a period of time. Considering priority of calls, type of resources, including response regarding with patients' need or patients' priority should be beneficial in increasing performance of the system.

A pattern of demand in several zones tends to change during day and week, which affects the optimal facility locations. In order to serve the calls more efficiently, a future facility location model should be able to adjust its solution according to the change of demand pattern. Relocating facilities to match demand or recruiting temporally staffs or volunteers in some zones could be an alternative.

REFERENCES

Araz, C., Selim, H., and Ozkarahan, I. (2007) A fuzzy multi-objective covering-based vehicle location model for emergency services. *Computers & Operations Research*, 34(3), 705-726.

Ball, M.O., Lin F.L. (1993). A reliability model applied to emergency service vehicle location. *Operations Research* 41(1), 18-36.

Batta, R., Dolan, J. M., and Krishnamurthy, N. N. (1989) The maximal expected covering location problem: Revisited. *Transportation Science*, 23(4), 277-287.

Bean, J.C. (1994) Genetic algorithms and random keys for sequencing and optimization. *INFORMS Journal on Computing*, 6(2), 154-160.

Berman, O. (1990) Mean-Variance location problems. *Transportation Science*, 24(4), 287-293.

Berman, O. and Kaplan, E.H. (1990) Equity maximizing facility location schemes. *Transportation Science*, 24(2), 137-144.

Berman, O., Krass, D. (2002) The generalized maximal covering location problem. *Computers & Operations Research*, 29, 563-581.

Berman, O., Krass, D., and Drezner, Z. (2003) The gradual covering decay location problem on a network. *European Journal of Operational Research* 151(3), 474-80.

Berube, J.F., Gendreau, M. Potvin, J.Y. (2009) An exact ε-constraint method for bi-objective combinatorial optimization problems: Application to the Traveling Salesman Problem with Profits. *European Journal of Operational Research* 194(1), 39-50.

Brendan G. Carr, Charles C. Branas, Joshua P. Metlay, Ashley F. Sullivan, and Carlos A. Camargo Jr. (2009) Access to emergency care in United States. *Annals of Emergency Medicine*, 54(2), 261-269.

Brill, E. D., Jr., Liebman, J. C., and ReVelle, C.S. (1976) Equity measures of exploring water quality management alternatives. *Water Resources Research*, 12(3), 845-851.

Brotcorne, L., G. Laporte and F. Semet. (2003) Ambulance location and relocation models. *European Journal of Operational Research*. 147, 451-463.

Chankong, V. and Haimes, Y. (1983) *Multiobjective decision making: Theory and methodology*, Elsevier Science Publishing Co., New York, NY.

Chanta, S., Mayorga, M.E., McLay, L.M., Wiecek, M.M. (2009) A bi-objective covering location model for EMS systems. *Industrial Engineering Research Conference Proceedings*, Miami, FL.

Chanta, S., Mayorga, M. E., Kurz, M. E., and McLay, L. A. (2010a) The minimum *p*-envy location problem. *Industrial Engineering Research Conference Proceedings*, Cancun, Mexico.

Chanta, S., Mayorga, M. E., and McLay, L. A. (2010b) The minimum *p*-envy location problem: Focusing on survivability of patients. Submitted to *Socio-Economic Planning Sciences*.

Chanta, S., Mayorga, M.E., McLay, L.M. (2011a) Improving rural emergency services: A bi-objective covering location model for EMS systems. Accepted to appear in *Annals of Operations Research*.

Chanta, S., Mayorga, M. E., Kurz, M. E., and McLay, L. A. (2011b) The minimum *p*-envy location problem: A new model for equitable distribution of emergency resources. Accepted to appear in *IIE Transactions on Healthcare Systems Engineering*.

Church, R. L. and ReVelle, C. (1974) The maximal covering location problem. *Papers of Regional Science Association*, 32, 101-118.

Church, R. L., and Roberts, K. L. (1983) Generalized coverage models and public facility location. *Papers in Regional Science* 53(1), 117-35.

Daskin, M. S. (1982) Application of an expected covering model to emergency medical service system design. *Decision Sciences*, 13(3) 416-439.

Daskin, M. S. (1983) A maximal expected covering location model: formulation, properties, and heuristic solution. *Transportation Science*, 17, 48-70.

Daskin, M. S. (1995) Network and discrete location models, algorithms, and applications, John Wiley & Sons, Inc., NY.

Daskin, M. S., Hogan, K., and ReVelle, C. (1988). Integration of multiple, excess, and expected covering models. *Environment and Planning*, 15, 15-35.

Daskin, M.S. and Stern, E.H. (1981) A hierarchical objective set covering model for emergency medical service vehicle deployment. *Transportation Science*, 15, 137-152.

De-los Mozos, M. C. L., and Mesa, J. (2001) A. The maximum absolute deviation measure in location problems on networks. *European Journal of Operational Research* 135, 184-94.

Drezner, T., Drezner, Z., and Guyse, J. (2009) Equitable service by a facility: Minimizing the Gini coefficient. *Computers & Operations Research*, 36(12), 3240-3246.

Drezner, Z. and Hamacher, H. W. (2004) *Facility location: Application and theory*, Springer, New York.

Ehrgott, M. (2005) *Multicriteria optimization*, Edition 2nd, Springer Verlag, Berlin.

Ehrgott, M. and Wiecek, M.M. (2005) Multiobjective programming, in: Figueira, J., Greco, S., Ehrgott, M. (Eds.), *Multiple criteria decision analysis: State of the art surveys*, Kluwer Academic Publishers, pp. 667–722.

Eiselt, H. A., and Marianov, V. (2009) Gradual location set covering with service quality. *Socio-Economic Planning Sciences* 43(2), 121-30.

English, F. (2008) Improving cardiac survival rates in rural communities by ownership of health: what the evidence shows. (2008) *Journal of Emergency Primary Health Care*. 6(3), 21.

Erkut, E. (1993) Inequality measures for location problems. *Location Science*, 1, 199-217.

Erkut, E., Ingolfsson, A., and Erdogan, G. (2008) Ambulance location for maximum survival. *Naval Research Logistics* 55, 42-58.

Erkut, E., Ingolfsson, A., Sim, T., and Erdoğan, G. (2009). Computational comparison of five maximal covering models for locating ambulances. *Geographical Analysis* 41, 43-65.

Erkut, E. and Neuman, S. (1992) A multiobjective model for location of undesirable facilities. *Annals of Operations Research*, 40, 209-227.

ERS/USDA. *Data- Rural Definitions*. September 4, 2007. Accesses Feb, 17, 2010. <http://www.ers.usda.gov/data/ruraldefinitions/>

Espejo I., Marin, A., Puerto, J., and Rodriguez-Chia, A. M. (2009) A Comparison of formulations and solution methods for the minimum-envy location problem. *Computers & Operations Research*, 36, 1966-1981.

Fitch, J. (2005) Response times: Myths, measurement and management. *Journal of Emergency Medical Services*. 30(9), 46-56.

Galvao, R. D., Chiyoshi, F. Y., and Morabito, R. (2005) Towards unified formulations and extensions of two classical probabilistic location models. *Computers & Operations Research*, 32(1), 15-33.

Ghosh D. (2003) Neighborhood search heuristics for the uncapacitated facility location problem. *European Journal of Operational Research*, 150, 150-162.

Glover, F. (1986) Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13, 533-549.

Glover, F. (1990) Tabu search: A tutorial. *Interfaces*, 20(4), 74-94.

Goldberg, J.B. (2004). Operations research models for the deployment of emergency service vehicles. *EMS Management Journal* 1(1), 20 – 39.

Haimes, Y. Y., Lasdon, L. S., and Wismer, D. A. (1971) On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man and Cybernetics*, 1, 296–297.

Hogan, K. and ReVelle, C. (1986) Concepts and applications of backup coverage. *Management Science*, 32, 1434-1444.

Iannoni, A. P. and Morabito, R. (2007) A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways. *Transportation Research, Part E*, 43, 755–771.

Karasakal, O. Karasakal, E. K. (2004) A maximal covering location model in the presence of partial coverage. *Computers & Operations Research*, 3, 1515-1526.

Kariv, O., Hakimi, S. L. (1979) An algorithmic approach to network location problems. Part I: The p-centers. *SIAM Journal on Applied Mathematics*, 37(3), 513-538.

Keeney, R. L. (1980) Equity and public risk. *Operations Research*, 28(3), 527-534.

Kincaid, R.K. and Maimon O. (1989) Locating a point of minimum variance on triangular graphs. *Transportation Science*, 23, 216–219.

Kostreve, M.M., Ogryczak, W., and Wierzbicki, A. (2004) Equitable aggregations and multiple criteria analysis. *European Journal of Operational Research*, 158(2), 362-377.

Larsen, M.P., Eisenberg, M.S., Cummins, R.O., and Hallstrom, A. (1993) Predicting survival from out-of-hospital cardiac arrest – a graphic model. *Annals of Emergency Medicine*, 22, 1652-1658.

Larson, R. C. (1974) A hypercube queuing model for facility location and redistricting in Urban Emergency Services. *Computers & Operations Research*, 1, 67-95.

Larson, R. C. (1975) Approximating the performance of urban emergency service systems. *Computers & Operations Research*, 23, 845-868.

Leclerc, P. D., McLay, L. A., and Mayorga, M. E. (2010) Modeling equity for allocation in public resources. Forthcoming, *Community-Based Operations Research: Decision Modeling for Local Impact and Diverse Populations*, Springer.

Lerner, E. B., Billittier, A. J., Sikora, J. et al. (1999) Use of geographic information system to determine appropriate means of trauma patient transport. Academic Emergency Medicine 6(11), 1127-1133.

Lopez-de-los-Mozos, M. (2003) The sum of absolute differences on a network: Algorithm and comparison with other equity measures. *Information Systems & Operational Research*, 41(2), 195-210.

Lopez-de-los-Mozos, M. and Mesa, J. A. (2001) The maximum absolute deviation measure in location problems on networks. *European Journal of Operational Research*, 135, 184-194.

Lorena, L. A. N. Problem Instances: Queuing Maximal Covering Location-Allocation Problem, available at http://www.lac.inpe.br/~lorena/correa/Q_MCLP_30.txt, last accessed February 5, 2011.

Maimon, O. (1986) The variance equity measure in locational decision theory. *Annals of Operations Research*, 6, 147-160.

Maimon, O. (1988) An algorithm for the Lorenz measure in locational decisions on trees. *Journal of Algorithms*, 9(4), 583–596.

Marsh, M.T., Schilling, D.A. (1994). Equity measurement in facility location analysis: A review and framework. *European Journal of Operations Research* 74(1), 1 – 17.

Mayer, J. D. (1980) Response time and its significance in medical emergencies. *Geographical Review* 70, 79-87.

McGinnis, K.K. (2004) State EMS Rural Needs Survey 2004. National Association of State EMS Directors, Falls Church, VA.

McLay, L.A. and Mayorga, M.E. (2010) Evaluating emergency medical service performance measures. *Health Care Management Science*, 13, 124-36.

Mesa, J. A., Puerto, J., and Tamir, A. (2003) Improved algorithms for several network location problems with equality measures. *Discrete Applied Mathematics* 130, 437-48.

Mulligan, G. F. (1991) Equity measures and facility location. *Papers in Regional Science*, 7(4), 345-365.

Narasimhan, S., Pikul, H., Schilling D. (1992) Capacitated emergency facility sitting with multiple levels of backup. *Annals of Operations Research*, 40(1), 323-37.

National Conference of State Legislatures. (2000) Emergency Medical service in rural areas: How can states ensure their effectiveness? Available at http://www.ncsl.org/programs/health/forum/ruralems.htm, August 2000 (Accessed on 25 May 2009).

Office of Technology Assessment. (1989) Rural emergency medical services—Special report. OTA Publ. No. OTA-H-445. Washington, D.C., U.S. Government Printing Office.

Ogryczak, W. (2000) Inequality measures and equitable approaches to location problems. *European Journal of Operational Research*, 122(2), 374-391.

Pirkul, H., Schilling D. A. (1988) the sitting of emergency service facilities with workload capacities and backup service. *Management Science*, 34, 896-908.

Pirkul, H., Schilling D. A. (1991) The maximal covering location problem with capacities on total workload. *Management Science*, 37, 233-248.

Rawlinson, C., and Crews, P. (2003). Access to Quality Health Services in Rural Areas—Emergency Medical Services: A Literature review. Rural Healthy People 2010: A comparison document to Healthy People 2010. Volume 2. College Station, TX: The Texas A&M University System health Science Center, School of Rural Public Health, Southwest Rural Health Research Center." Downloadable from:http://www.srph.tamhsc.edu/centers/rhp2010/03Volume2accessems.pdf

ReVelle, C., Hogan, K. (1989). The maximum availability location problem. *Transportation Science* 23(3), 192 – 200.

ReVelle, C., Schweitzer, J., and Snyder, S. (1996) The maximal condition covering problem. *INFOR*, 34(2), 77-91.

Savas, E. S. (1978) On equity in providing public services. *Management Science*, 24(8), 800-808.

Stone, D. (2002) *Policy paradox: The art of political decision making*, W. W. Norton & Company, New York, NY.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971) The location of emergency service facilities. Operations Research, 19, 1363-1373.

Trevillyan, J., Abbott, J., Dixon, M., et al. (1998) Trauma deaths in a rural area. *Journal of the Arkansaa Medical Society* 95(7), 290-293.

U.S. Department of Health and Human Services. Leading Health Indicators. <http://healthypeople.gov/LHI/> 2010.

Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W., and Larsen, M. P. (1997) Estimating effectiveness of cardiac arrest intervention-a logistic regression survival model. *Circulation* 96(10), 3308-13.

Willams, J. M., Ehrlich, P. F., and Prescott, J. E. (2001) Emergency medical care in rural America. *Annals of Emergency Medicine*, 38(3), 323-327.