12-2012

A Comprehensive Location-Allocation Method for Specialized Healthcare Services

Siddhartha S. Syam
Marquette University, siddhartha.syam@marquette.edu

Murray J. Côté
Texas A & M University System Health Science Center

Operations Research for Health Care, Vol. 1, No. 4 (December 2012): 73-83. DOI.

A Comprehensive Location-Allocation Method for Specialized Healthcare Services

Siddhartha S. Syam

Department of Management, Marquette University, Milwaukee, WI

Murray J. Côté

Department of Health Policy and Management, Texas A&M Health Science Center, College Station, TX

Abstract

This paper focuses on the development, solution, and application of a location-allocation model for specialized health care services such as the treatment and rehabilitation necessary for strokes or traumatic brain injuries. The model is based on our experience with the Department of Veterans Affairs' integrated service networks. The model minimizes the total cost borne by the health system and its patients and incorporates admission acuity levels, service proportion requirements, and admission retention rates. A common resource constraint is introduced at the facility level since treatment of multiple acuity levels involves the pooling of common resources. Realistic instances of the model with 20 potential service locations, 50 admission districts and up to five open treatment units for three levels of severity are solved in about 300 seconds. The applicability of the model is tested by an extensive managerial experiment using data derived from one of the Department of

Veterans Affairs specialized healthcare services. We investigate the effects of five critical factors: (1) the degree of service centralization, (2) service level mandates by acuity, (3) lost admission cost by acuity, (4) facility overload penalty cost by acuity and (5) target utilization level by acuity and treatment unit. We examine the countervailing forces present in making healthcare service location decisions and the resulting tradeoffs from the implicitly multiobjective nature of the system. The experiment and analysis demonstrate that the major factors of the experiment have a significant bearing on the optimal assignment of admission districts to treatment units.

Keywords

Design of experiments, Integer programming, Location-allocation, Simulated annealing

1. Introduction and motivation

In the United States, healthcare remains an area of crucial concern for millions as evidenced by the current ongoing debate over the federal healthcare reform bill. Healthcare providers are pressured by two conflicting dimensions: ever-increasing healthcare costs and the public demand for access to cutting-edge treatment. As a result, healthcare providers have virtually no choice but to constantly seek to become as efficient as possible in all aspects of their operations.

The original study that this paper extends is based on a perceived need to improve the delivery of specialized health services at the Department of Veterans Affairs (VA) in terms of effectiveness and efficiency [1]. The VA is the primary organization charged with providing healthcare to veterans and the original study was based upon a funded research project that aimed to determine the optimal location of traumatic brain injury (TBI) treatment units for VA medical centers. As a not-for-profit service organization, the VA has to define optimality using a multi-objective approach where the cost of providing care (i.e., efficiency) and the level and extent of healthcare provided in terms of access and availability (i.e., effectiveness) are generally viewed as equally critical objectives. The research described in this paper extends the previous study and model in meaningful and important ways that we expect have applicability to nonprofit healthcare providers beyond the VA.

The specific improvements to efficiency and effectiveness included in this paper are as follows: (a) Multiple levels of severity with capacity limits by facility and severity (sometimes called acuity). The justification for this is that treatment costs (including fixed costs of specialized equipment) are much higher for high levels of acuity. In turn, the high costs have to be offset by scale economies which make it impossible to replicate the same capacity to treat severe cases in every facility. (b) Common resource constraints at each facility. The need for this stems from the fact that many kinds of resources such as physicians, supplies, storage facilities, and operating theaters are common across severity levels but used at varying rates at different levels. Common resource availability constrains treatment at all severity levels in a facility. (c) Service level mandates, overload penalties, lost admission costs, and target utilization by severity level are factors introduced in the optimization model, whose impact is investigated in a managerial experiment described later. In this context, for each level of acuity (i) a service level mandate is a requirement to serve a particular proportion of eligible patients, (ii) a target utilization percentage is a proportion of capacity that management wishes to utilize in order to balance costly load on a facility versus the need to serve as many patients as possible (iii) an overload penalty is a cost associated with exceeding a facility's utilization target, and (iv) lost admission costs are federal funding not obtained if patients are not served (admitted).

In the model developed in this paper, the service level mandate is an explicit constraint while the other factors are included in the objective function so that violations of the corresponding managerial targets are penalized in accordance with the levels of the factors. The factors other than service level mandate are associated with targets that often trade off against each other rather than explicit requirements and the managerial experiment described later tracks the costs of not achieving the targets as the factor levels are varied. Instances of tradeoffs are service level versus overload and lost admissions versus target capacity utilization at a facility. Lost admissions and service level are related to effectiveness and overload and capacity utilization are related to efficiency (cost).

This paper is organized as follows. The following section provides the literature review germane to our research. Section 3 presents the optimization model and defines the relevant decision variables and model coefficients. Following the model, Section 4 describes an extensive managerial experiment meant to evaluate the impact of important managerial parameters and discusses the results that we obtained from that experiment. Last, concluding remarks and future research directions are provided in Section 5.

2. Literature review

The purpose of location-allocation models is to concurrently determine optimal facility locations and the assignment of customers to open facilities. Since the research literature in facility location is vast, no attempt is made here to provide a comprehensive review. Rather, we direct the reader to a full review of general facility location models and the methods used to solve them found in Love et al. [2] and Cornuéjols et al. [3]. Instances of the application of location-allocation models to healthcare issues include hospital location in rural regions [4], geographical considerations in healthcare planning [5], locating blood banks [6], service mix and location in managed healthcare [7], trauma care involving hospitals and ambulances [8], the reorganization of liver transplant regions [9], and optimizing the location of specialized treatment facilities [10]. An extensive survey of the research regarding the location of healthcare facilities may be found in Daskin and Dean [11].

The location-allocation model developed in this paper involves mixed integer programming and is rooted in the classic uncapacitated facility location (UFL) model [12]. In practice, the UFL model is often modified to provide feasible templates of frequently occurring service or business scenarios which correspond to more complicated models. For example, the introduction of facility capacity limits leads to the capacitated facility location problem, and limiting the permitted number of open facilities leads to the p-median problem. The different versions of the UFL model such as one in which every location can be a facility and demand node and another in which facilities are limited to a subset of nodes and variants of the UFL model are known to be difficult to solve (particularly for large instances) since they belong to the NP-complete class of problems [13]. However, it may be noted that the variants may be significantly harder to solve than the UFL model since they contain constraints not found in the UFL model. Consequently, optimal solutions to our type of problem are difficult to obtain and very large instances may necessitate specialized solution approaches or heuristics such as Lagrangian relaxation [14], simulated annealing [15], and dual ascent [16]. While these sophisticated heuristics offer the advantage of reduced computational times, they generally provide near-optimal rather than optimal solutions to complex problems. After some experimentation, we found that the commercial-grade general-purpose optimization software CPLEX-OPL [17] solved most instances of problems in our research environment in reasonable computing time. Consequently, we adopted CPLEX-OPL as our solver engine for the model developed here. This model is described in the next section.

3. The optimization model

The primary goal of the optimization model developed in this paper is to provide a mathematical framework that incorporates the primary criteria of the VA as it seeks to serve veterans: (1) the cost of providing service and (2) the service level provided to the VA's patients. The costs included in the model include fixed costs, treatment costs, travel costs, lodging costs, lost service costs, and overloading penalty costs. We note that patients or their families bear the cost of travel and hotel lodging of family members who accompany the patient to a given facility and other costs are borne by the VA. The service level, for each level of acuity, is defined as the proportion of eligible admissions served by the VA for a given geographical area. The model incorporates retention rates by distance traveled and these are incorporated according to multiple levels of acuity (i.e., reflecting the observation that patients are willing to travel relatively longer distances for higher levels of acuity). For the purposes of our model and its generalized application beyond the VA, we refer to acuity as representing the general medical condition of a potential patient where higher acuity patients will require longer lengths of stay and more resources than lower acuity patients. Our purpose behind this is simply to demarcate differing patient classes. A similar definition is applied to service levels in that we recognize the healthcare organization may need to maintain a certain volume to justify (either economically or by the decision-maker's prerogative) why a service is provided at a given facility. Last, we treat the eligibility of veterans as potential patients as detailed in [1].

An important application of our model is to analyze the tradeoff between a centralized capacity policy with a relatively small number of treatment units generally located in large metropolitan areas versus a decentralized capacity policy with a relatively large number of geographically disperse treatment units, some of which may be located in rural or low-population density areas. While the primary analysis is in terms of cost, our model also assists decision-makers in fulfilling the service mission of the VA, including examining secondary objectives such as patient travel and lodging costs. A by-product of the degree of centralization adopted by the organization that is captured by our model is the level of employment corresponding to various policies. In government and/or unionized work environments, the staffing level may be important enough to be a criterion in its own right. While we do not view staffing in that manner in our research, our model allows the decision-maker to analyze the consequences of alternative staffing level restrictions.

In addition to retention rates by acuity, the model includes constraints that ensure that the capacity of each potential treatment unit by acuity level is not exceeded and that mandatory service levels by acuity level are met. Parameter definitions ensure that the fixed cost of a treatment unit is a piecewise linear function of that unit's capacity. It also includes a common resource limit for each medical center where the treatment unit may be located such that the common resources are those that are used by all acuity levels. There are also restrictions on the number of open treatment units by acuity level. It should be noted that the service mission of the VA is maintained through constraints that enforce service level mandates for each level of acuity (e.g., at least 70% of a target patient population should be served at VA-based treatment units). In this paper, a cluster of patient demand locations that the VA feels should be served by a single facility (i.e., open treatment unit in an open medical center) is labeled as a "district" or "patient district". Typically, a district corresponds to an area code, but the definition allows some flexibility so that the units comprising the district may be aggregated upwards to represent larger geographic areas such as counties, states or regions, depending on need.

Economies of scale and learning effects that result from increasing capacity may be modeled as nonlinear functions of volume. However, nonlinear elements tend to make already difficult integer combinatorial models quite intractable. On the other hand, linear approximations are tractable, and we represent fixed costs by piecewise linear functions of capacity. Capacity is divided into ranges from zero to some very high theoretical upper bound (conceivably, infinity). Each capacity range c is demarcated by a lower bound and an upper bound which depend on the acuity level. The piecewise linear approach is sufficiently flexible in that it can capture both economies and diseconomies of scale. We develop the model in two stages—in the first stage, the model has binary variables that model the piecewise linear functions of capacity and in the second stage we further eliminate these variables through a pre-processing stage. The reduction in the number of variables through this artifice greatly helps to achieve shortened computing times. Both the models are presented next, and the final version will be discussed in detail. The model developed in this study belongs to the category of binary integer programming models, which are widely prevalent in the facility location literature. The primary benefit of binary modeling in facility location analysis is that the binary (i.e., 0–1) variables act as on–off switches indicating whether or not treatment units (by acuity level) are available at a particular medical center.

The basic assumptions of the model are as follows: (1) a deterministic model is sufficient and the degree of variability observed in the empirical data is not high enough to justify the complexity of a stochastic model, (2) the linear (or linearized) structure of the model is sufficient to capture the primary complexities and tradeoffs of the operating environment for delivery of specialized healthcare services for the VA (or other service-oriented entity), and (3) the data required by the model are available to the VA (or other entity). The organization of the model is as follows. First, the objective function that is minimized contains the sum of fixed overhead and labor costs, variable admission treatment costs, travel, lodging, and labor costs, and the cost of lost service. The constraints of the model ensure that each district is served by only one medical center with a single retention rate. Variable definitions (implemented as constraints) ensure that the retention rate applied corresponds to the distance between the district and the medical center. Other constraints ensure that the capacity of each medical center is not exceeded and that mandatory service rates and employment levels are met. Parameter definitions in the model ensure that the fixed cost per unit of a medical center is a piecewise linear function of the medical center's capacity. The initial model, which has the additional difficulty of being a nonlinear mixed integer model (the nonlinearity is due to the first term of the objective function) is provided next.

Model variables.

$Y_c^s = 1$ if a treatment unit with acuity capability s is open in center c, 0 otherwise;
$\alpha_{cq}^s = 1$ if the qth capacity category applies to center c, acuity s, 0 otherwise;
$X_{cd}^{ks} = 1$ if district d, acuity s, is served by center c, with retention rate k, 0 otherwise.

Model parameters.

d_{cd} = Distance in miles between district d and center c;
Mxd_s = Maximum distance between a district and assigned center for acuity s;
t_{cd} = Transportation cost per mile between district d and center c;
a_{qs} = Fixed cost per admission associated with capacity category q for acuity s;
B_{qs}^u = Upper capacity bound for capacity category q for acuity s;
B_{qs}^l = Lower capacity bound for capacity category q for acuity s;
v_{cs} = Treatment variable cost per admission at center c with acuity s;
C_{cs} = Specialized healthcare service capacity (in number of admissions) at center c for acuity s;
R_{cd}^{ks} = Proportion of admissions from district d retained at center c for retention rate k and acuity s;
U^{ks} = Upper bound in miles for the kth retention rate;
L^{ks} = Lower bound in miles for the kth retention rate;
P_{ds} = Admission volume (total number of potential admissions) in district d, acuity s;
m_{cs} = Average length of stay in days per admission at center c for acuity s;
h_c = Average hotel charge per day for admission family at (near) center c;
ϕ_s = Mandated minimum proportion of admissions served by the VA system for acuity s;
Ω_s = Penalty cost to the VA of a potential admission with acuity s not treated by the system;
S_{cs} = Fixed minimum staffing level for acuity s at center c;
G_{cs} = Additional staff per admission for acuity s at center c;
e_{cs} = Average staff payment rate for acuity s at center c;
ω_s = Mandated number of open centers for acuity s;
σ_s = Overloading penalty per admission for acuity s;
μ_s = Target utilization percentage for acuity s;
δ_s = Common resource utilization per admission for acuity s;
Q_c = Common resource capacity at center c;
ρ = Common resource balance factor.

The location-allocation model (NIVA) is formally defined below, using the variables and parameters defined above.

Nonlinear initial model

$$\text{Minimize: } \sum_c \sum_q \sum_s \left\{ \alpha^{s}_{cq} a_{qs} C_{cs} + S_{cs} e_{cs} \right\} Y^{s}_{c} + \sum_c \sum_d \sum_s \{ v_{cs} + t_{cd} d_{cd} + m_{cs} h_c + G_{cs} e_{cs} \} P_{ds} R^{ks}_{cd} X^{ks}_{cd} + \sum_s \Omega_s \left\{ \sum_d P_{ds} - \sum_c \sum_d \sum_s P_{ds} R^{ks}_{cd} X^{ks}_{cd} \right\} + \sum_c \sum_s \sigma_s \left\{ \sum_d \sum_s P_{ds} R^{ks}_{cd} X^{ks}_{cd} - \mu_s C_{cs} Y^{s}_{c} \right\}$$

subject to:

(1) A district–acuity combination can be assigned only to an open center within the distance limit for the acuity

$$X^{ks}_{cd} * d_{cd} \leq Y^{s}_{c} * Mxd_s \forall c, d, k, s.$$

(2) A district–acuity combination is assigned to one and only one center

$$\sum_c \sum_k X^{s}_{cd} = 1 \forall d, s.$$

(3) Total number of admissions assigned to a center must not exceed capacity by acuity

$$\sum_d \sum_k P_{ds} R^{ks}_{cd} X^{ks}_{cd} \leq C_{cs} Y^{s}_{c} \forall c, s.$$

(4) Service level mandate for each acuity level

$$\sum_c \sum_d \sum_k P_{ds} R^{ks}_{cd} X^{ks}_{cd} \geq \phi_s \sum_d P_{ds} \forall s.$$

(5) Common resource usage limit by center

$$\sum_d \sum_s \delta_s \sum_k P_{ds} R^{ks}_{cd} X^{ks}_{cd} \leq \rho * Q_c \forall c.$$

(6) Number of open centers for each acuity level

$$\sum_c Y^{s}_{c} = \omega_s \forall s.$$

(7) Upper distance limit for retention rates

$$d_{cd} * X^{ks}_{cd} \leq U^{ks} X^{ks}_{cd} \forall k, s.$$

(8) Lower distance limit for retention rates

$$d_{cd} * X^{ks}_{cd} \geq L^{ks} X^{ks}_{cd} \forall k, s.$$

(9) Upper capacity limit for capacity rates

$$\alpha_{cq}^{s} C_{cs} \leq \alpha_{cq}^{s} B_{qs}^{u} \forall q, s.$$

(10) Lower capacity limit for capacity rates

$$\alpha_{cq}^{s} C_{cs} \geq \alpha_{cq}^{s} B_{qs}^{l} \forall q, s.$$

(11) Variable definitions

$$X_{cd}^{ks}, \alpha_{cq}^{s}, Y_{c}^{s} \in \{0,1\} \forall c, d, k, s.$$

Preprocessing steps.

This model is the logical representation of the decision problem faced by the medical system we investigated. However, as noted previously, it incorporates two major drawbacks: (a) it is nonlinear due to the fixed cost term in the objective function: $\sum_c \sum_q \sum_s \{\alpha_{cq}^{s} a_{qs} C_{cs} + S_{cs} e_{cs}\} Y_c^s$ and (b) it contains a very high number of binary variables of the types: α_{cq}^{s} and X_{cd}^{ks}. That these features are serious barriers to implementation was made evident to us when models coded in the fast CPLEX system took many hours to solve even for medium-sized problems. As a result of this, we redesigned the model by eliminating the α_{cq}^{s} variables altogether and replacing the X_{cd}^{ks} variables with X_{cd}^{s} variables. In effect, we precalculate the retention rates that (potentially) apply between medical centers and districts and also the capacity categories that apply to centers, both by acuity. As a result of these steps: (i) the nonlinearity in the model is eliminated (ii) the number of binary variables in the model is drastically reduced (iii) significant numbers of constraints in the model related to the eliminated variables are no longer needed and are therefore removed from the model. The result is a significantly more computationally tractable model.

The actual coding steps are: (a) For retention rates by acuity, the following applies: If $L^{ks} \leq d_{cd} \leq U^{ks}$, then parameter R_{cd}^{s} in the revised model (VA) below $= R_{cd}^{ks}$. (b) For capacity categories the following applies: If $B_{qs}^{l} \leq C_{cs} \leq B_{qs}^{u}$, then parameter a_{cs} in the revised model below $= a_{qs}$. The revised model is provided below.

General parameters.

N_s = Number of acuity (severity) levels
N_c = Number of centers
N_d = Number of districts.

Model variables.

$Y_c^s = 1$ if a treatment unit with acuity capability s is open in center c, 0 otherwise;
$X_{cd}^s = 1$ if district d, acuity s, is served by center c, 0 otherwise.

Model parameters.

d_{cd} = Average distance in miles between district d and center c;
Mxd_s = Maximum distance between a district and assigned center for acuity s;
t_{cd} = Travel cost per mile between district d and center c;
a_{cs} = Fixed cost per admission associated with center c and acuity s;
v_{cs} = Treatment variable cost per admission at center c with acuity s;
C_{cs} = Specialized healthcare service capacity (in number of admissions) at center c for acuity s;
R_{cd}^s = Proportion of admissions from district d retained at center c rate for acuity s;
P_{ds} = Admission volume (total number of potential admissions) in district d, acuity s;
m_{cs} = Average length of stay in days per admission at center c for acuity s;
h_c = Average hotel charge per day for admission family at (near) center c;

ϕ_s = Mandated minimum proportion of admissions served by the VA system for acuity s;
Ω_s = Penalty cost to the VA of a potential admission with acuity s not treated by the system;
S_{cs} = Fixed minimum staffing level for acuity s at center c;
G_{cs} = Additional staff per admission for acuity s at center c;
e_{cs} = Average staff payment rate for acuity s at center c;
ω_s = Mandated number of open centers for acuity s;
σ_s = Overloading penalty per admission for acuity s;
μ_s = Target utilization percentage for acuity s;
δ_s = Common resource utilization per admission for acuity s;
Q_c = Common resource capacity at center c;
ρ = Common resource balance factor.

The location-allocation model (LRVA) is formally defined below, using the variables and parameters defined above.

Linear revised model (LRVA).

$$\text{Minimize:} \sum_{c} \sum_{s} \{a_{cs} C_{cs} \underset{c}{\overset{s}{+}} S_{cs} e_{cs}\} Y$$
$$+ \sum_{c} \sum_{d} \sum_{s} \{v_{cs} + t_{cd} d_{cd} + m_{cs} h_c + G_{cs} e_{cs}\} P_{ds} \underset{cd}{\overset{s}{R}} \underset{cd}{\overset{s}{X}}$$
$$+ \sum_{s} \Omega_s \left\{ \sum_{d} P_{ds} - \sum_{c} \sum_{d} \sum_{s} P_{ds} \underset{cd}{\overset{s}{R}} \underset{cd}{\overset{s}{X}} \right\}$$
$$+ \sum_{c} \sum_{s} \sigma_s \left\{ \sum_{d} \sum_{s} P_{ds} \underset{cd}{\overset{s}{R}} \underset{cd}{\overset{s}{X}} - \mu_s C_{cs} \underset{c}{\overset{s}{Y}} \right\}.$$

subject to:

(1) A district–acuity combination can be assigned only to an open center within the distance limit for the acuity

$$X_{cd}^s * d_{cd} \leq Y_c^s * Mxd_s \forall c, d, s.$$

(2) A district–acuity combination is assigned to one and only one center

$$\sum_{c} \underset{cd}{\overset{s}{X}} = 1 \forall d, s\,.$$

(3) Total number of admissions assigned to a center must not exceed capacity by acuity

$$\sum_{d} P_{ds} \underset{cd}{\overset{s}{R}} X_{cd}^s \leq C_{cs} Y_c^s \forall c, s\,.$$

(4) Service level mandate for each acuity level

$$\sum_{c} \sum_{d} P_{ds} \underset{cd}{\overset{s}{R}} \underset{cd}{\overset{s}{X}} \geq \phi_s \sum_{d} P_{ds} \forall s\,.$$

(5) Common resource usage limit by center

$$\sum_{d}\sum_{s}\delta_s P_{ds} \underset{cd}{\overset{s}{R}} X_{cd}^{s} \leq \rho * Q_c \forall c\,.$$

(6) Number of open centers for each acuity level

$$\sum_{c} \underset{c}{\overset{s}{Y}} = \omega_s \forall s$$

$$X_{cd}^{s}, Y_c^s \in \{0,1\} \forall c, d, s.$$

Objective function.

The first term in the objective function is the sum of fixed treatment cost and fixed staffing cost. The second term is the sum of variable treatment, travel, lodging, and staffing costs. The third term is lost admission cost and the fourth is the overloading penalty cost. It may be noted that lost admission cost decreases while the overloading penalty cost increases as the number of admissions increases. It can also be observed that the overloading penalty cost decreases when the target capacity utilization proportion (by acuity level) goes up. These tradeoffs imply that the model is, at least in some ways, multiobjective in nature. The tradeoffs are further explored in the managerial experiment described in Section 4.

Constraints.

The first set of constraints ensures that an admission district can be assigned to only open centers for each level of acuity and that, further, these candidate open centers must be within the system's distance limit for the level of acuity. The second set of constraints further restricts the assignment of each admission district to exactly one open center by acuity. Constraint set (3) ensures that the total number of admissions assigned to a center must not exceed capacity by acuity.

The fourth set of constraints imposes minimum service level mandates by acuity on the system while constraint set (5) ensures that no medical center violates its limits for common resources (such as physicians and operating theaters). Finally, the sixth set of constraints restricts the number of open centers for each level of acuity to those mandated by the system design. Binary restrictions on the variables are also part of the model.

The managerial experiment described in Section 4 entails twenty candidate centers ($N_c = 20$) and fifty medical districts ($N_d = 50$) for each of three levels of acuity ($N_s = 3$). As a result, the LRVA model has ($N_s \times N_c$) or 60 Y_c^s location variables and ($N_s \times N_c \times N_d$) or 3000 allocation variables for a total of 3060 binary variables. The corresponding number of constraints in the LRVA model is: $(N_s \times N_c \times N_d) + (N_d \times N_s) + (N_c \times N_s) + N_s + N_c = 3000 + 150 + 60 + 3 + 20 = 3233$.

4. Managerial experiment

This section describes a managerial experiment that we conducted using the model presented in the previous section in order to investigate the nature of system behavior particularly in the context of the tradeoffs alluded to in Section 2. It was conducted on a Windows XP personal computer with 4 GB of RAM operating at 2.66 GHz. The experiment also evaluated the relative performance of a decentralized service system versus a centralized service system. Recapping the primary tradeoffs of the model, we have: (i) lost admission costs decrease while the overloading penalty cost increases as the number of admissions served increases and (ii) the overloading penalty cost decreases when the target capacity utilization proportion (by acuity level) increases. Since one of the main purposes of the managerial experiment is to evaluate tradeoffs between parameters such as the lost admission cost and service level mandate, we provide the values of those parameters in Table 1. Other details regarding data values are provided in the Appendix. The data are simulated based on a project funded by the VA but do not correspond to any specific values at the VA or other healthcare organization.

Table 1. Parameters for the managerial experiment.

Factor	**Level**	**Acuity level**		
		1	**2**	**3**
Target utilization (TU) (in %)	1	45	50	55
	2	50	60	70
	3	55	65	75
Service level mandate (SLM) (in %)	1	50	60	70
	2	55	65	75
	3	60	70	80
Lost admission cost (LAC) (in $)	1	100	200	300
	2	400	500	600
	3	700	800	900
Overload penalty cost (OPC) (in $)	1	100	200	300
	2	400	500	600
	3	700	800	900

Our initial analysis of the model was conducted by generating two instances for each combination of target utilization (TU), service level mandate (SLM), lost admission cost (LAC), and overload penalty cost (OPC). Each factor, as shown in Table 1, had three levels so the resulting experiment had $2 \times 3 \times 3 \times 3 \times 3 = 162$ observations. This was done in order to conduct a full fourway experimental design on the model to identify both the significant main effects and significant interactions so as to better guide decision-makers as they undertake location decisions.

As shown in Table 2, Table 3, we provide the p-values for the fourway analysis of variance (ANOVA) results for seven key outputs from our model: (1) total objective function cost, (2) variable treatment cost, (3) variable staffing cost, (4) lodging cost, (5) lost admission cost, (6) overload capacity cost, and (7) percentage of admissions served. Scanning across both Table 2, Table 3, we note that all four factors contribute to the total objective function and the overload capacity cost but were not all equally significant for the other outputs. This is not surprising given how the four factors affect each of the outputs. Equally important is the presence (or absence) of significant interaction effects between the factor levels for the outputs. Overload capacity cost is strongly influenced by all four factors and their interactions regardless of number of treatment units. At the other end of the spectrum, admissions treated tended to be only affected by changes to the service level mandate in the two-treatment unit instance and only the service level mandate and lost admission cost in the five-treatment unit instance.

Table 2. Analysis of variance results for two treatment units.

Source	**DF**	**Response variable**						
		Total cost	**Treatment cost**	**Staffing cost**	**Lodging cost**	**Lost admission cost**	**Overload cost**	**Admissions served**
		p-**value**	p-**value**	p-**value**	p-**value**	p-**value**	p-**value**	p-**value**
TU	2	0.0003	0.2264	0.3014	0.5990	0.3630	< 0.0001	0.5116

SLM	2	< 0.0001	< 0.0001	0.0084	0.3780	<0.0001	0.0010	< 0.0001
LPC	2	< 0.0001	0.5323	0.4728	0.0491	<0.0001	< 0.0001	0.1951
OPC	2	< 0.0001	0.0026	0.3500	0.8768	0.2458	< 0.0001	0.1857
TU×SLM	4	0.5981	0.5862	0.7949	0.8487	0.7360	0.1289	0.3264
TU×LPC	4	0.7171	0.9114	0.3541	0.5688	0.9242	0.0035	0.3975
TU×OPC	4	0.7450	0.5628	0.5828	0.3704	0.2871	< 0.0001	0.4990
SLM×LPC	4	< 0.0001	0.0750	0.3231	0.3369	<0.0001	0.0010	0.3996
SLM×OPC	4	0.6778	0.4479	0.1055	0.6763	0.6487	< 0.0001	0.2402
LPC×OPC	4	0.6718	0.2667	0.4758	0.5230	0.7190	0.0086	0.3328
TU×SLM×LPC	8	0.4741	0.6513	0.3362	0.5589	0.9426	0.0239	0.5208
TU×SLM×OPC	8	0.9550	0.6876	0.7476	0.6007	0.3990	0.0052	0.2584
TU×LPC×OPC	8	0.7849	0.4893	0.7397	0.6010	0.6606	0.0563	0.5343
SLM×LPC×OPC	8	0.2350	0.7473	0.3270	0.8706	0.7415	0.0022	0.5268
TU×SLM×LPC×OPC	16	0.6004	0.6359	0.7441	0.8526	0.7615	< 0.0001	0.5477
Model	80	< 0.0001	0.0577	0.4531	0.8739	<0.0001	< 0.0001	0.1039

Table 3. Analysis of variance results for five treatment units.

Source	**DF**	**Response variable**						
		Total cost	**Treatment cost**	**Staffing cost**	**Lodging cost**	**Lost admission cost**	**Overload cost**	**Admissions served**
		p-**value**	p-**value**	p-**value**	p-**value**	p-**value**	p-**value**	p-**value**
TU	2	< 0.0001	0.3360	0.6114	0.5100	0.6692	< 0.0001	0.9719
SLM	2	< 0.0001	< 0.0001	< 0.0004	0.0087	<0.0001	< 0.0001	0.0010
LPC	2	< 0.0001	0.0001	0.0227	0.2431	<0.0001	< 0.0001	0.0578
OPC	2	< 0.0001	0.3982	0.1964	0.7307	<0.0001	< 0.0001	0.8216
TU×SLM	4	0.8120	0.8486	0.1477	0.1792	0.0903	0.0341	0.4389
TU×LPC	4	0.5322	0.5459	0.4378	0.9469	0.3316	0.3784	0.4144
TU×OPC	4	0.0620	0.2776	0.5797	0.3254	0.6900	< 0.0001	0.1851
SLM×LPC	4	0.0282	0.2613	0.5930	0.7177	<0.0001	0.7791	0.0784
SLM×OPC	4	0.8613	0.4832	0.5930	0.1630	0.6819	< 0.0001	0.5153
LPC×OPC	4	0.5090	0.3532	0.8049	0.0758	0.0016	< 0.0001	0.4947
TU×SLM×LPC	8	0.7968	0.6443	0.3257	0.3669	0.0003	0.4739	0.7177
TU×SLM×OPC	8	0.7652	0.5997	0.1680	0.3023	0.6197	0.0906	0.5127
TU×LPC×OPC	8	0.2579	0.7362	0.0188	0.1369	0.1263	0.4237	0.3180
SLM×LPC×OPC	8	0.9369	0.5307	0.6650	0.8828	0.0373	0.9177	0.7901
TU×SLM×LPC×OPC	16	0.8370	0.8889	0.3233	0.8814	0.4633	0.6204	0.2406
Model	80	< 0.0001	0.0008	0.0375	0.3173	<0.0001	< 0.0001	0.1813

Next, we discuss the primary managerial results by presenting them for a centralized two open medical center (for each level of acuity) system and also a more decentralized five open medical center (for each acuity level) system. This presentation facilitates a comparison between the two systems for each system tradeoff that is evaluated.

First, we consider the objective value (i.e., total cost) as a function of a primary performance driver, the service level mandates. Fig. 1, which compares the four factors for two-center and five-center systems. All panels indicate that the objective value rises slowly as the service level mandates rise from level 1 to level 2 but subsequently rises much faster from level 2 to level 3 although the jump is much larger for the two-center system compared to the five-center system—this is because lost admission costs actually fall when service levels rise for a five-center system partially compensating for the increase of other costs. In turn, the decline in lost admission costs is a result of the fact that retention rates are primarily a function of distance. Thus, the relatively shorter travel distances of a five-center system (compared to a two-center system) correspond to higher retention for every level of acuity.

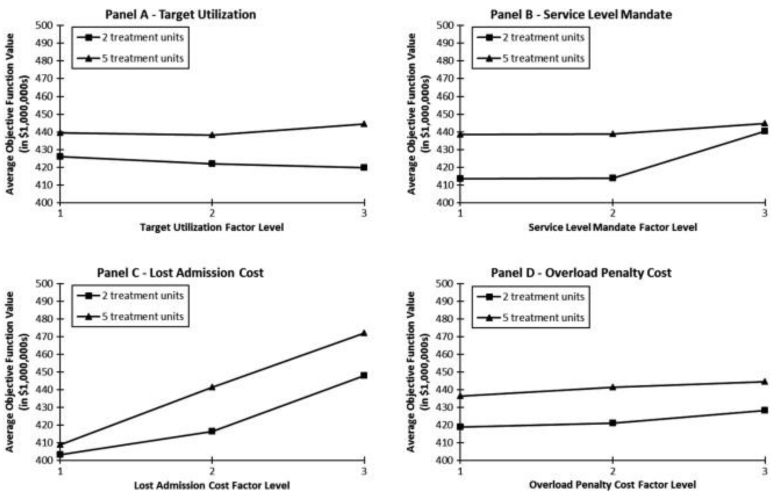

Fig. 1. Interaction plots for average objective function value.

We also provide other results that shed light on the impact of various factors on the functioning of the model. These include overloading cost by service level mandate, overloading cost by capacity utilization target, lost admission cost by utilization target, objective function value by lost admission penalty, and objective function value by overload penalty. These results are illustrated in Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6.

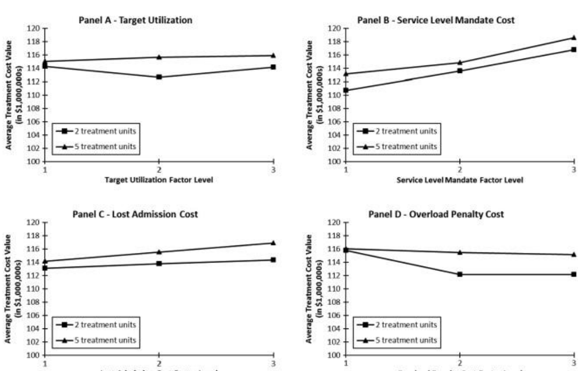

Fig. 2. Interaction plots for average treatment cost.

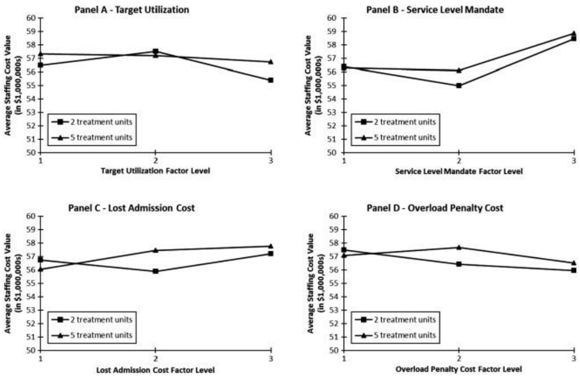

Fig. 3. Interaction plots for average staffing cost.

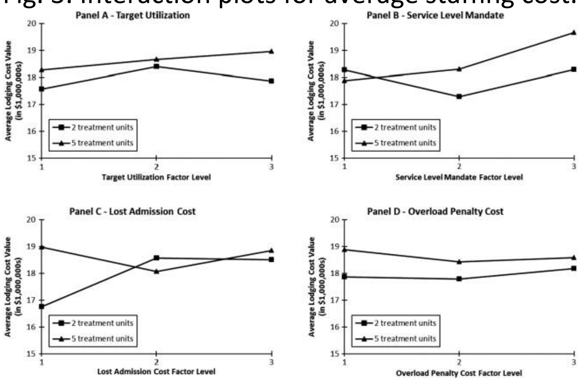

Fig. 4. Interaction plots for average lodging cost.

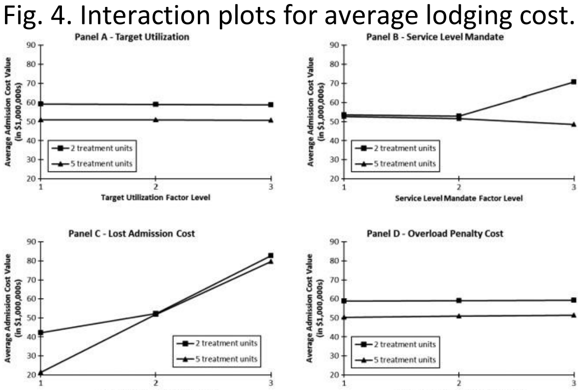

Fig. 5. Interaction plots for average lost admission cost.

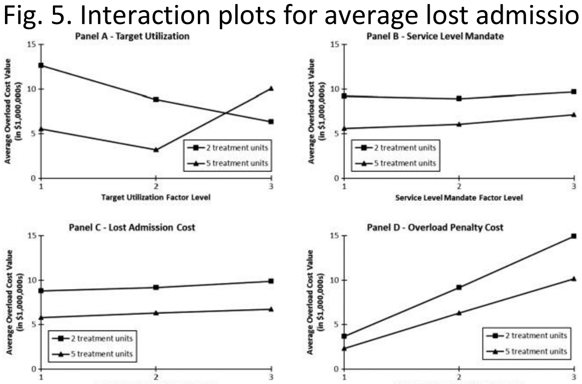

Fig. 6. Interaction plots for average overload capacity cost.

Another result of interest is the impact of the facility utilization target on the objective function (cost) as well as the proportion served. Capacity utilization targets are generally increased or decreased depending on the tradeoff between 'stress' levels on the system (which lead to system breakdowns) and the need for high service levels (or output in a manufacturing context). For instance, the desire to keep a relatively older facility functioning without too many breakdowns will lead decision-makers to set the capacity utilization targets (by severity) for that facility relatively low even though this may lead to lower current admissions. A relatively new facility, however, can withstand greater capacity utilization and this will lead decision-makers to set higher targets for this facility which leads to higher service levels and, implicitly, higher return on capital investment. The general expectation is that cost should go down with higher targets due to lower utilization penalties and lower lost admission costs. This expectation is met in both the two-center and five-center systems as seen in Fig. 7. The explanation is that the initial relaxation of the utilization targets and therefore the utilization penalties is not enough to compensate for the increase in treatment and other costs associated with serving a larger number of patients.

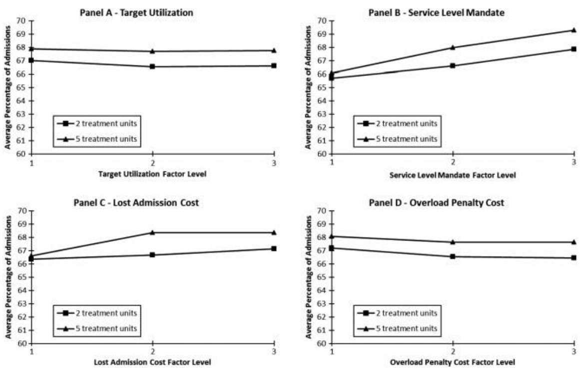

Fig. 7. Interaction plots for average admissions treated.

The further relaxation of utilization targets reduces utilization penalties sufficiently that the lower penalties along with lower lost admission costs are sufficient to compensate for higher treatment costs—hence, the proportion served goes up at the final level of utilization targets. Last, the proportion of admissions served as a function of the service level mandate. As expected, the service level rises with the service level mandate for both the two-center and five-center systems. Note however, that the five-center system serves more admissions than the two-center system which also corresponds to the fact that the five-center system is generally costlier.

5. Conclusions and future research

The research described in this paper has involved the development and application of a comprehensive model for location-allocation of specialized healthcare services. This model includes several additions to a standard healthcare location-allocation model including multiple acuity levels, multiple service level mandates by acuity, facility utilization targets by acuity, and a common resource constraint at each facility that limits the use of resources that are common to all levels of acuity (e.g., nurses and physicians).

We conducted an extensive managerial experiment in order to assess the impact of several factors on key goals of the system such as total cost and service level (proportion served). These factors include lost admission cost, service level mandate, target utilization percent, and overloading penalty cost. Comparative results are provided regarding the impact of these factors for a two-center (i.e., centralized) system versus a five-center (i.e., decentralized) system. An important finding is that a decentralized system is costlier than a centralized one but also serves a higher proportion of admissions. A finding that is not necessarily intuitive is that the proportion served first decreases and then increases when facility utilization targets are increased. A comprehensive set of results is provided to assist managers with fine-tuning the parameters of the system.

The model is applied in specific ways depending on the particular concerns of managers. As an illustration, if the healthcare manager aims to achieve the highest service level, Panel B of Fig. 1 shows that it makes little difference to total cost whether a centralized or decentralized system is used while the difference is substantial if a lower service level is desired (for instance, to reduce load on the system). In addition, if lost admissions cost is a particular source of concern, Panel B of Fig. 5 indicates that the centralized system results in higher lost admissions cost relative to the decentralized system. Combining the information from the two panels, the manager might favor the decentralized system particularly at the highest service level.

Our future research is focused on extending the model from a non-profit environment to a for-profit one. This of course involves the introduction of revenue and pricing considerations—in particular, the elasticity of demand becomes an important consideration. Another consideration is the pricing of non-medical services such as single-room treatment, cable television, and the like. Finally, lost admission costs are much more difficult to evaluate in a for-profit setting—typically, they have to be estimated from data and are not available as a fixed cost per admission. To conclude, we believe that an extension of our model to the for-profit environment is challenging but worthwhile in terms of the much broader range of applicability that it promises.

Appendix.

This Appendix provides details on the statistical distributions and functional forms that were used to generate data for the computational experiment. All data are representative of the operational data provided to us by our anonymous healthcare organization but should not be construed as fact. We use the notation N(μ,σ) to indicate a normally distributed random variable with mean, μ, and standard deviation, σ.

1. Annual number of admissions per district (varies normally between districts with mean and standard deviation as shown)
   - Acuity level 1: N(1000,100)
   - Acuity level 2: N(700,70)
   - Acuity level 3: N(400,40).
2. Capacity (i.e., number of admissions) per treatment unit for system with two open treatment units
   - Acuity level 1: N(25,000,5000)
   - Acuity level 2: N(17,500,5000)
   - Acuity level 3: N(10,000,2500).
3. Capacity (i.e., number of admissions) per treatment unit for system with five open treatment units
   - Acuity level 1: N(10,000,500)
   - Acuity level 2: N(7000,400)
   - Acuity level 3: N(4000,300).
4. Annual fixed treatment cost per treatment unit (in $)
   - Acuity level 1: $2000×Capacity0.95
   - Acuity level 2: $3000×Capacity0.95
   - Acuity level 3: $4000×Capacity0.95.
5. Variable treatment cost per admission (in $)
   - Acuity level 1: N(1000,100)
   - Acuity level 2: N(2000,200)
   - Acuity level 3: N(3000,300).
6. Hotel charge per day (in $)
   - N(70,20).
7. Length of stay (in days)
   - Acuity level 1: N(3,1)
   - Acuity level 2: N(7,2)
   - Acuity level 3: N(14,3).
8. Minimum staff
   - Acuity level 1: N(4,1)
   - Acuity level 2: N(6,1)
   - Acuity level 3: N(8,2).

9. Additional staff per patient
    Acuity level 1: N(0.01,0.003)
    Acuity level 2: N(0.02,0.003)
    Acuity level 3: N(0.03,0.003).
10. Annual salary per individual staff (in $)
    Acuity level 1: N(50,000,10,000)
    Acuity level 2: N(60,000,10,000)
    Acuity level 3: N(70,000,10,000).
11. Maximum service distance limit for two treatment unit system (in miles)
    Acuity level 1: 425
    Acuity level 2: 475
    Acuity level 3: 525.
12. Maximum service distance limit for five treatment unit system (in miles)
    Acuity level 1: 250
    Acuity level 2: 300
    Acuity level 3: 350.
13. Transportation cost per mile (in $)
    N(1,0.5).
14. Common capacity available for treatment
    N(1.5,0.1)×Sum of individual Acuity level capacities.
15. Common capacity utilization per admission (as a %)
    Acuity level 1: N(1.5,0.1)×0.33
    Acuity level 2: N(1.5,0.1)×0.334
    Acuity level 3: N(1.5,0.1)×0.336.
16. Admission retention rates for Acuity level 1
    0–70 miles: 80%
    71–125 miles: 70%
    126–600 miles: 60%.
17. Admission retention rates for Acuity level 2
    0–100 miles: 85%
    101–200 miles: 75%
    126–600 miles: 65%.
18. Admission retention rates for Acuity level 3
    0–150 miles: 90%
    151–250 miles: 70%
    251–600 miles: 70%.

References

[1] M. Côté, S. Syam, W. Vogel, D. Cowper. **A mixed integer programming model to locate traumatic brain injury units in the Department of Veterans Affairs: a case study.** *Health Care Management Science*, 10 (2007), pp. 253-267

[2] R. Love, J. Morris, G. Wesolowsky. **Facilities Location: Models and Methods.** North-Holland, New York (1988)

[3] G. Cornuéjols, G.L. Nemhauser, L.A. Wolsey. **The uncapacitated facility location problem.** P.B. Mirchandani, R.L. Francis (Eds.), *Discrete Location Theory*, Wiley-Interscience, New York (1990)

[4] A. Mehrez. **On the implementation of quantitative facility location models: the case of a hospital in a rural region.** *Journal of the Operational Research Society*, 47 (1996), pp. 612-625

[5] P.R. Harper, A.K. Shahani, J.E. Gallagher, C. Bowie. **Planning health services with explicit geographical considerations: a stochastic location-allocation approach.** *Omega*, 33 (2005), pp. 141-152

[6] W. Price, M. Turcotte. **Locating a blood bank.** *Interfaces*, 16 (1986), pp. 17-26

[7] S.B. Schweikhart, V.L. Smith-Daniels. **Location and service mix decisions for a managed health care network.** *Socio-Economic Planning Sciences*, 27 (1993), pp. 289-302

[8] C.C. Branas, E.J. MacKenzie, C.C. ReVelle. **A trauma resource allocation model for ambulances and hospitals.** *Health Services Research*, 35 (2000), pp. 489-507

[9] J.E. Stahl, N. Kong, S.M. Shechter, A.J. Schaefer, M.S. Roberts. **A methodological framework for optimally reorganizing liver transplant regions.** *Medical Decision Making*, 25 (2005), pp. 35-46

[10] S.S. Syam, M.J. Côté. **A location-allocation model for service providers with application to not-for-profit healthcare organizations.** *Omega*, 38 (2010), pp. 157-166

[11] M.S. Daskin, L.K. Dean. **Location of health care facilities.** M.L. Brandeau, F. Sainfort, W.P. Pierskalla (Eds.), *Operations Research and Health Care*, Kluwer Academic Press, Boston (2004)

[12] D. Erlenkotter. **A dual-based procedure for uncapacitated facility location.** *Operations Research*, 26 (1978), pp. 992-1009

[13] M.R. Garey, D.S. Johnson. **Computers and Intractability: A Guide to the Theory of NP-Completeness.** Freeman, San Francisco (1979)

[14] S. Syam. **A model and methodologies for the location problem with logistical components.** *Computers and Operations Research*, 29 (2002), pp. 1173-1193

[15] V.R. Muppani (Muppant), G.K. Adil. **Efficient formation of storage classes for warehouse storage location assignment: a simulated annealing approach.** *Omega*, 36 (2008), pp. 609-618

[16] T. Crainic, L. Delorme. **Dual-ascent procedures for multicommodity location-allocation problems with balancing requirements.** *Transportation Science*, 27 (1993), pp. 90-101

[17] IBM. www-01.ibm.com/software/integration/optimization/cplex-optimizer, 2011.

[1]Tel.: +1 414 288 5462; fax: +1 414 288 5754.