The Impact of Client Choice on Medical Facility Network Design under Competition

October 21, 2012

Yue Zhang
College of Business and Innovation, University of Toledo
2801 W. Bancroft Street, Toledo, Ohio, U.S.A., 43606
yue.zhang@utoledo.edu

Derek Atkins
Sauder School of Business, University of British Columbia
2053 Main Mall, Vancouver, British Columbia, Canada, V6T 1Z2
derek.atkins@sauder.ubc.ca
Abstract

This paper describes a methodology for designing a network of medical facilities under competition. We focus on “walk-in” medical facilities and assume that travel time and waiting time are the two main determinants for clients to choose where to receive the service. Subject to a total capacity limit, the problem is to optimize the location and capacity of each open facility so as to maximize total volume. We consider two alternative client choice models, one probabilistic-choice and the other optimal-choice. Both models are formulated as a mathematical program with equilibrium constraints. The lower level problem of both models is to determine the flows of clients to facilities, and the upper level is a facility location and capacity allocation problem. We propose a common solution methodology for both models. An exact allocation algorithm is developed to solve the lower level problem, and a greedy algorithm and a Tabu search procedure are developed to solve the upper level problem. Our computational experiments show that the Tabu search procedure is more effective. Finally, through a real-life case study, we discuss several interesting managerial insights about model selection and network design strategies.

Keywords: location, capacity, client choice, algorithm, equilibrium.
1 Introduction

Healthcare has become one of the largest and most important industries in North America and all over the world, especially due to the growth of ageing population. For instance, the U.S. spent more than $2 trillion or 16% of its GDP on healthcare (Catlin et al. 2008). Total employment in various health delivery settings in the U.S. is almost 14.4 million, including doctors, nurses, pharmacists, and administrators (Shi and Singh 2010).

Medical services in North America are provided by many separate entities, and medical facilities are largely owned and operated by the private sector (Shi and Singh 2010). For private healthcare providers, since they are reimbursed for their services from insurance plans, their primary goal is to maximize total volume. How to select the best locations of their facilities and how to allocate limited capacity so as to maximize total volume are their strategic decisions to make.

This paper presents a methodology for designing a network of medical facilities for a firm in a given area. In addition to the focal firm, there are existing facilities that belong to the competitors of the focal firm in the area. The focal firm attempts to optimize its facility network in order to maximize total volume to its own facilities or overall market share. Subject to a total capacity limit, the location and the capacity (number of servers) of each open facility are the main determinants of the configuration of a facility network.

In this paper, we focus on “walk-in” medical facilities (e.g., medical diagnostic laboratory) that clients are free to choose to visit, and assume that travel time to a facility and waiting time at the facility are the main attractiveness factors considered by clients. Note that, in this user-choice circumstance, it is important for a model to address how clients choose a facility to visit. In particular, we study the impact of client choice behavior on the configuration of a facility network. To this end, we present two alternative models: 1) probabilistic-choice model, a client may visit any facility with a certain probability, which increases with the attractiveness of the facility; 2) optimal-choice model, each client would go to the facility with the highest attractiveness.

Both models are formulated as a mathematical program with equilibrium constraints (MPEC), i.e., a bilevel nonlinear optimization model. The lower level problem of both
models is to determine the allocation of clients to facilities according to the client choice behavior, and the upper level is a facility location and capacity allocation problem. For the probabilistic-choice model, the lower level problem is formulated by a set of nonlinear equations, while it is formulated as a variational inequality for the optimal-choice model. To solve the problems efficiently, we propose a common solution methodology along with the bilevel framework for both models. Specifically, we develop an exact allocation algorithm to solve the lower level problem and a greedy algorithm and a Tabu search procedure to solve the upper level problem.

Note that this work is motivated by a real-life case study that we conducted for a private firm providing medical diagnostic laboratory services in British Columbia (BC), Canada, as described in Zhang et al. (2012b). In this paper, we use the models developed here to investigate this real-life case. A number of interesting findings and managerial insights are discussed. We believe that our methodology can also be applied to the problems of facility network design for other medical or more general services.

The main contribution of this research lies in four perspectives. First, we develop two generic models for the problem of medical facility network design under competition in the user-choice circumstance and formulate them as MPECs. Second, we propose a common solution methodology to solve the two problems and discuss mathematical properties of the lower level equilibrium problem. Third, the models for the lower level problem can be used alone to estimate parameter values or to predict the demand volume at a facility in a network. This could make a significant contribution to empirical studies in the service and healthcare industry, where waiting time is often considered a major attractiveness factor (Newman 1984 and Jan et al. 2000). In particular, our models do not require real waiting time data. Finally, by comparing the models applied to the real-life case, we discuss several key findings and managerial insights regarding model selection and facility network design as follows.

• We find that incorporating waiting time into the model can significantly improve the goodness-of-fit, and that waiting time is a better and more appropriate attractiveness factor to be considered than capacity.
Our analysis supports that the probabilistic-choice model in general fits real data better, but the optimal-choice model may perform closely at the facility level, i.e., fitting the demand volume at each facility.

Both models show that the optimal strategy for the stronger firm in a duopoly situation is to relocate its facilities and capacities closer to those of the weaker firm, leading to a more intensive face-to-face competition. In contrast, the optimal strategy of the weaker firm is to avoid such a face-to-face competition.

The remainder of the paper is organized as follows. The next section provides a brief literature review. Section 3 presents and formulates the two models. The solution methodology is described in Section 4. Section 5 presents computational results for the models. The case is studied in Section 6. In the final section, conclusions and future research directions are discussed.

2 Literature Review

There is rich literature in operations research on facility location (and capacity allocation) problems. In this section, we mainly focus on recent studies related to service or healthcare operations. Readers may also refer to Daskin and Dean (2004) for the review of location problems in healthcare as well as Berman and Krass (2002) and Marianov and Serra (2002) for more general review on public facility location problems.

Most of these studies can be classified into two categories: 1) monopoly model, there is only one firm operating all facilities in a studied area; 2) competitive model, there are several different firms in a studied area, each operating one or more facilities. In terms of modeling, a single decision-maker determines all the decisions for all facilities to maximize total profit (or coverage, social welfare, etc.) or to minimize total cost in monopoly models, while multiple decision makers determine the decisions for their own facilities in competitive models to maximize their own market share or profit.

From the perspective of client allocation, we can classify the studies into another two categories: 1) centralized model, a central decision-maker has the power to assign clients to
facilities; 2) user-choice model, clients are free to choose a facility to visit. Centralized models are common for monopoly cases. Within the category of user-choice models, we can further classify them into probabilistic-choice and optimal-choice models as defined earlier. In the classical facility location literature, optimal-choice models are more common. Huff (1963) developed the first probabilistic-choice model for spatial analysis with a gravity type. Many extensions, including the multiplicative competitive interaction (MCI) model (Nakanishi and Cooper 1974) and the multinomial logit (MNL) model (McFadden 1974), have been subsequently proposed and applied. Refer to Zhang et al. (2012b) for a more detailed review on these models.

User-choice models must address how clients choose a facility to visit, i.e., how to define a utility function based on facility attractiveness factors. It is common to consider distance the only attractiveness factor, including Verter and Lapierre (2002), Wang et al. (2004), Berman et al. (2006), and Zhang et al. (2012a). In contrast, more recent models in the service and healthcare contexts normally incorporate waiting time as well, such as Marianov et al. (2008) and Zhang et al. (2010), as empirical analysis has shown that waiting time is a key factor to clients in these contexts (Newman 1984 and Jan et al. 2000). These models are typically formulated based on the MPEC structure, which is a common methodology for modeling equilibrium problems in operations research, economics, and engineering. Other facility location and/or capacity allocation models based on the MPEC structure include Chao et al. (2003), Marianov (2003), and Marianov et al. (2005).

According to these classifications, our paper addresses competitive and user-choice models based on the MPEC structure. In particular, our paper is most related to three earlier studies in the literature. Zhang et al. (2012a) compared two monopoly models (one probabilistic-choice and the other optimal-choice) for network design with the objective of maximizing total participation to a preventive healthcare program. They considered distance the only attractiveness factor and applied a service level constraint on mean waiting time to each open facility. As a result, their models were formulated as mixed integer programs (MIPs), which are much simpler than our MPECs; the solution methodology is completely different from ours. Zhang et al. (2010) developed an optimal-choice monopoly model based on the MPEC structure to maximize total participation, which is assumed to be elastic. In
contrast, our problem addresses a competitive environment and a fixed demand volume at each population node. Therefore, our formulation is different, especially the variational inequality formulation for the lower level problem. Further, their allocation algorithm to solve the lower level problem of the optimal-choice model is an approximation one when there are two servers or more at any facility, while we introduce a novel allocation algorithm, which can exactly solve the lower level problem of both the probabilistic-choice and optimal-choice models. Third, by comparing the two models via the case with real data, we discuss several key insights for model selection and network design, which were not investigated in their work. The MPEC model proposed by Marianov et al. (2008) is related to our probabilistic-choice model. However, they assumed that facility capacities are exogenous, but only considered facility locations as decisions. Our solution algorithms for both the lower and upper level problems are different from theirs. Moreover, their work does not investigate a real-life application or propose any managerial insights to practitioners.

3 Models

Let \( G = (N, L) \) be a network with a set of nodes \( N(|N| = n) \) and a set of links \( L \). The nodes represent the neighborhoods of a city or the population zones, and the links are the main transportation arteries. The fraction of clients residing at node \( i \) is denoted by \( h_i \). We assume that the number of clients who require the service over the entire network is Poisson distributed with a rate of \( \lambda \) per unit of time, and thus from each node \( i \) at a rate \( \lambda h_i \).

We assume that there is a finite set of potential locations \( M(|M| = m) \) in \( G \) for facilities of the focal firm, Let \( S \subset M \) be a set of facilities located by the focal firm. There is another set of existing facilities belonging to the competitors of the focal firm denoted by \( R(|R| = r) \). The travel time from node \( i \) to location \( j \) through the shortest path denoted by \( t_{ij} \). Suppose that there is a given client attraction value at location \( j \), \( j \in M \cup R \), denoted by \( u_j \), which includes main client attraction determinants at that location, such as convenience of parking, facility appearance, area security, etc.

We assume that there are \( s_{max} \) servers available for the focal firm in total, and that one or more servers can be allocated to each open facility. Denote by \( s_j \) the given number
of servers at the competitor’s facility \(j, j \in R\). Assume that each server is homogenous and can provide an average of \(\mu\) services per unit of time, and also that the service time is exponentially distributed.

Thus, each facility can be modeled as an M/M/c queuing system, where \(c\) denotes the number of servers at the facility. The mean waiting time in the system is denoted by \(W\), and the general formula for this is (Kleinrock 1975):

\[
W = \frac{C(c, u)}{c} \frac{1}{\mu(1 - \rho)} + \frac{1}{\mu},
\]

where

\[
u = \frac{\lambda}{\mu}, \quad \rho = \frac{\lambda}{c\mu}, \quad C(c, u) = \frac{1 - K(u)}{1 - \rho K(u)}, \quad K(u) = \sum_{l=0}^{c-1} \frac{u^l}{l!}.
\]

The functions \(C(c, u)\) and \(k(u)\) are just mathematical expressions to simplify (1).

To develop the mathematical models, we define three sets of decision variables:

\[
y_j = \begin{cases} 
1 & \text{if a facility is located at node } j, \ j \in M \\
0 & \text{otherwise};
\end{cases} \\
s_j = \text{number of servers at facility } j, \ j \in M; \\
x_{ij} = \text{number of clients (flow) from population node } i, \ i \in N, \text{ who visit facility } j, \ j \in M \cup R.
\]

Therefore, we have \(S = \{j : j \in M, y_j = 1\}\) and

\[
\sum_{j \in S \cup R} x_{ij} = \lambda h_i \quad i \in N.
\]

Denote by \(a_j\) the arrival rate of clients at facility \(j, j \in M \cup R\), and then we have

\[
a_j = \sum_{i \in N} x_{ij} \quad j \in S \cup R.
\]

We denote by \(U_{ij}\) the total utility that clients from node \(i\) receive the service at facility \(j\), which comprises of three components: 1) \(u_j\), a given constant client attraction at facility \(j\); 2) \(t_{ij}\), the travel time from node \(i\) to facility \(j\); and 3) \(\bar{W}(a_j, s_j)\), the mean system waiting
time at facility \( j \). We assume that the total utility from node \( i \) to facility \( j \) is a linear function of these three components, which is given by,

\[
U_{ij} = \beta_0 u_{ij} - \beta_1 t_{ij} - \beta_2 \bar{W}(a_j, s_j) \quad i \in N \quad j \in S \cup R.
\] (4)

Before we present the formulations of the probabilistic-choice and optimal-choice models in detail, let us first describe a key relationship between the arrival rate \( a_j \) and the mean system waiting time \( \bar{W}(a_j, s_j) \). As in both models, \( x_{ij} \) is assumed to depend on \( U_{ij} \), which depends on \( \bar{W}(a_j, s_j) \) according to expression (4); thus, \( a_j \) depends on \( \bar{W}(a_j, s_j) \). Therefore, we have to solve an equilibrium problem in order to determine \( x_{ij} \) and \( a_j \). This is referred to as a user-equilibrium problem in the operations research literature.

In other words, the entire problem can be viewed as a bilevel problem based on the MPEC structure:

- **The lower level problem**: given a set of facility locations and the associated capacities, identify equilibrium flows of clients from population nodes to the facilities;

- **The upper level problem**: determine the best set of locations and the associated capacities.

### 3.1 The Probabilistic-Choice Model

In this section, we present our probabilistic-choice model based on the MNL formulation, to be consistent with Zhang et al. (2012b). Given \( S \) and \( s_j, j \in S \), the user-equilibrium can be determined by solving the following set of equations,

\[
x^*_{ij} = \lambda h_i \frac{\exp(\beta_0 u_{ij} - \beta_1 t_{ij} - \beta_2 \bar{W}(a^*_j, s_j))}{\sum_{k \in S \cup R} \exp(\beta_0 u_{ik} - \beta_1 t_{ik} - \beta_2 \bar{W}(a^*_k, s_k))} \quad i \in N \quad j \in S \cup R,
\] (5)

where \( x^*_{ij} \) and \( a^*_j \) denote the flow from node \( i \) to facility \( j \) and the arrival rate to facility \( j \) at equilibrium, respectively.

Theorem 1 proves the existence and uniqueness of the equilibrium flow \( x^*_{ij} \) once given \( y_j \) and \( s_j \).
Theorem 1. Given \( y_j \) and \( s_j \), there exists a unique equilibrium flow \( x_{ij}^* \) from node \( i \) to facility \( j \).

The proof for Theorem 1 is straightforward following Lemmas 1 and 2 in Marianov et al. (2008).

Then, the formulation for this competitive facility network design problem with the probabilistic-choice assumption is given by,

\[
\text{max} \sum_{i \in N} \sum_{j \in M} x_{ij} \tag{6}
\]

\[
\text{s.t.}
\]

\[
s_j \geq y_j \quad j \in M \tag{7}
\]

\[
\sum_{j \in M} s_j = s_{\text{max}} \tag{8}
\]

\[
\sum_{j \in M \cup R} x_{ij} = \lambda h_i \quad i \in N \tag{9}
\]

\[
y_j \in \{0, 1\}, \quad s_j = \text{Integer} \quad j \in M \tag{10}
\]

\[
x_{ij} \geq 0 \quad i \in N \quad j \in M \cup R \tag{11}
\]

\[
t'_{ij} = t_{ij} + B(1 - y_j) \quad i \in N \quad j \in M \tag{12}
\]

\[
x_{ij} = \lambda h_i \frac{\exp(\beta_0 u_j - \beta_1 t'_{ij} - \beta_2 W(a_j, s_j))}{\sum_{k \in M \cup R} \exp(\beta_0 u_k - \beta_1 t'_{ik} - \beta_2 W(a_k, s_k))} \quad i \in N \quad j \in M \cup R. \tag{13}
\]

The objective (6) is to maximize the total demand of the firm. Constraints (7) ensure that at least one server will be assigned to each open facility. Constraint (8) limits the total available servers to \( s_{\text{max}} \). Constraints (9) guarantee the total demand from each node. Constraints (13), where \( B \) represents a big number, stipulate that clients can obtain service
only from open facilities, i.e., the travel time to a location where there is no open facility is set to a big number. Constraints (12) and (13) together are equivalent to (5). Also, we note that \( y_j \) and \( s_j \) are the true decision variables, while \( x_{ij} \) is only an auxiliary decision variable.

3.2 The Optimal-Choice Model

Now, we present the optimal-choice model, which assumes that clients choose the facility with the highest utility. Denote by \( U_i \) the highest utility incurred by clients at node \( i \). i.e.,

\[
U_i = \max_{j \in S \cup R} U_{ij} \quad i \in N. \tag{14}
\]

As mentioned earlier, given facility locations \( y_j \) and the number of servers at each open facility \( s_j \), the lower level problem involves the clients’ facility choices so as to maximize their total utility. At equilibrium, no client wants to change her choice.

The equilibrium condition can be stated as: given \( S \) and \( s_j, j \in S \), for all pairs of \((i,j), i \in N, j \in S \cup R\),

\[
U^*_{ij} = \beta_0 u_j - \beta_1 t_{ij} - \beta_2 \bar{W}(a^*_j, s_j) \begin{cases} 
= U^*_i & \text{if } x^*_{ij} > 0 \\
\leq U^*_i & \text{if } x^*_{ij} = 0,
\end{cases} \tag{15}
\]

where \( U^*_{ij} \) and \( U^*_i \) denote the utility of clients at node \( i \) visiting facility \( j \) and the highest utility for clients at node \( i \) at equilibrium, respectively.

This equilibrium condition (15) states that if there is a flow of clients from node \( i \) to facility \( j \), then the total utility incurred by clients at node \( i \) to facility \( j \) must be equal to the highest one; otherwise, the total utility is less than or equal to the highest one.

To find \( x^*_{ij} \) and \( a^*_j \) in (15), we have to solve the following variational inequality problem,

\[
\sum_{i \in N} \sum_{j \in S \cup R} (U^*_{ij} - U^*_i) (x_{ij} - x^*_{ij}) \leq 0 \quad \forall x_{ij} \in X(y), \tag{16}
\]

where the feasible set \( X(y) \) is defined as,

\[
X(y) = \{ x_{ij} : x_{ij} \geq 0, \sum_{j \in S \cup R} x_{ij} = \lambda h_i, i \in N, j \in S \cup R; x_{ij} = 0, i \in N, j \in M \setminus S \}. \tag{17}
\]
Note that we expect that the stability of the queue \( a_j < s_j \mu, j \in S \cup R \), can be naturally satisfied in (16).

Theorem 2 proves that given \( y_j \) and \( s_j \), the solution to the variational inequality problem (16) is the equilibrium of this problem (15).

**Theorem 2.** Given \( y_j \) and \( s_j \), the flow pattern \( x_{ij}^* \) is in equilibrium if and only if it satisfies the variational inequality problem (16).

The proof for Theorem 2 is given in Appendix 1.

Further, Theorem 3 proves that solving (16) is equivalent to solving the following problem,

\[
\sum_{i \in N} \sum_{j \in S \cup R} U_{ij}^*(x_{ij} - x_{ij}^*) \leq 0.
\] (18)

**Theorem 3.** Solving (16) is equivalent to solving (18).

The proof for Theorem 3 is given in Appendix 1.

Theorem 4 proves the existence of the equilibrium flow \( x_{ij}^* \) once given \( y_j \) and \( s_j \).

**Theorem 4.** Given \( y_j \) and \( s_j \), there exists at least one equilibrium flow \( x_{ij}^* \) from node \( i \) to facility \( j \).

The proof for Theorem 4 is similar to that for Theorem 1.

Unlike the previous case, the uniqueness of the equilibrium for the optimal-choice model may not hold. One simple example with multiple equilibrium solutions is a network with two client nodes and two facilities, where all the parameters are symmetric. In practice, however, almost all of the cases that we face have precisely one equilibrium.

Then, the bilevel formulation of the optimal-choice model follows that for the probabilistic-choice model with replacing constraints (13) by

\[
\sum_{i \in N} \sum_{j \in M \cup R} U_{ij}(x_{ij}' - x_{ij}) \leq 0 \quad \forall x_{ij}' \in \{x_{ij} : (9) \text{ and } (11) \text{ hold}\}.
\] (19)

Note that constraints (12) and (19) together are equivalent to (16). Again, the stability of the queue \( a_j < s_j \mu, j \in M \), can be naturally satisfied.
4 Solution Methodology

Since both models are highly nonlinear and the decision variables of the upper level problem are binary or integer, the entire problem is extremely difficult to solve. Thus, the focus of the study is on developing an effective and efficient heuristic.

Our solution methodology is along with the bilevel framework. We develop an allocation algorithm presented in Section 4.1 for the lower level problem of both models, i.e., given a set of facility locations and the associated capacities, the allocation algorithm is to identify equilibrium flows of clients to facilities. For the upper level problem of both models, we develop a greedy algorithm and a Tabu search procedure presented in Section 4.2 to determine the best set of locations and the associated capacities. In other words, in the entire bilevel framework, the allocation algorithm serves as a sub-routine for the greedy algorithm and the Tabu search procedure.

4.1 Allocation Algorithm

In this section, we propose a common exact allocation algorithm that can be applied to both the probabilistic-choice and optimal-choice models. In other words, it can find the user-equilibrium for given $S$ and $s_j, j \in S$, by solving both the set of nonlinear equations (5) and the variational inequality (16).

Our algorithm is an iterative method. Let $k$ be the iteration index and $K$ be a maximum iteration number. Let $\epsilon$ be an error tolerance parameter, $\theta^k, k = 1, \ldots, K$, be a step-size parameter at iteration $k$, and $\delta$ be a step-size adjusting parameter. Both $\theta^k, k = 1, \ldots, K$, and $\delta$ are between zero and one. Also, to improve the efficiency of the algorithm, we let $\bar{W}_{\text{max}}$ be the maximum allowable mean waiting time at each facility, and $a_{\text{max}}(s_j)$ be the maximum arrival rate at facility $j$ given $s_j$ to satisfy $\bar{W}_{\text{max}}$. That is, given $s_j$ and $\bar{W}_{\text{max}}$, $a_{\text{max}}(s_j)$ can be calculated inversely from expression (1).

Allocation Algorithm

Step 0: Choose appropriate values for $\epsilon, \theta^k, \delta, K, \text{ and } \bar{W}_{\text{max}}$; calculate $a_{\text{max}}(s_j), j \in S \cup R$; set $k = 0$; set $x_{ij}^0 = \frac{\lambda_i}{|S \cup R|}, i \in N, j \in S \cup R$. 

11
Step 1: Set $k = k + 1$; calculate $a_j, j \in S \cup R$, from expression (3); calculate $W(a_j, s_j), j \in S \cup R$, from expression (1); calculate $U_{ij}, i \in N, j \in S \cup R$, from expression (4); find $\overline{U}_i, i \in N$.

Step 2: For the probabilistic-choice model, set

$$x'_{ij} = \lambda h_i \exp(U_{ij}) \sum_{k \in S \cup R} \exp(U_{ik}), i \in N, j \in S \cup R;$$

for the optimal-choice model, set

$$x'_{ij} = \begin{cases} 
\lambda h_i & \text{if } U_{ij} = \overline{U}_i \\
0 & \text{if } U_{ij} > \overline{U}_i.
\end{cases}$$

Step 3: Define $d_{ij} = x'_{ij} - x_{ij}^{k-1}, i \in N, j \in S \cup R$, as a search direction.

Step 4: Set $x_{ij}^k = x_{ij}^{k-1} + \theta^k d_{ij}, i \in N, j \in S \cup R$.

Step 5: Calculate $a_j, j \in S \cup R$, from expression (3); if $a_j > a_{\text{max}}(s_j), j \in S \cup R$, set $\theta^k = \delta \theta^k$, and go to Step 4.

Step 6: If $|x_{ij}^k - x_{ij}^{k-1}| \leq \epsilon, i \in N, j \in S \cup R$, or $k \geq K$, set $x_{ij} = x_{ij}^k, i \in N, j \in S \cup R$, and then stop; otherwise, go to Step 1.

In each iteration, the mechanism of the algorithm is to find a new search direction for $x_{ij}$ in Step 3 and then update $x_{ij}$ in Steps 4 and 5. The process continues until reaching the stopping criteria. A main difference between this algorithm and some other methods in the literature for solving a set of nonlinear equations or a nonlinear program is that, in each iteration, this algorithm does not compute the optimal step-size; instead, we set the step-size in each iteration $\theta^k$ in advance and may adjust it by multiplying it by $\delta$ if needed. There are many different ways to set $\theta^k$. In general, $\theta^k$ should decrease in $k$ to make the algorithm converge. In particular, we set $\theta^k = \phi^{k-1}$, where $\phi$ is a constant value between zero and one. Other formats of $\theta^k$ may be applicable too.
### Table 1: The Travel Time Matrix and the Fractions

<table>
<thead>
<tr>
<th>Facility</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>( h )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.405</td>
<td>0.683</td>
<td>0.285</td>
<td>1.008</td>
<td>0.875</td>
<td>0.061</td>
</tr>
<tr>
<td>2</td>
<td>0.840</td>
<td>0.425</td>
<td>0.775</td>
<td>0.175</td>
<td>0.408</td>
<td>0.152</td>
</tr>
<tr>
<td>3</td>
<td>1.038</td>
<td>0.380</td>
<td>0.630</td>
<td>0.600</td>
<td>0.520</td>
<td>0.071</td>
</tr>
<tr>
<td>4</td>
<td>0.843</td>
<td>0.530</td>
<td>0.415</td>
<td>0.435</td>
<td>1.070</td>
<td>0.066</td>
</tr>
<tr>
<td>5</td>
<td>1.068</td>
<td>0.433</td>
<td>0.700</td>
<td>0.595</td>
<td>1.038</td>
<td>0.095</td>
</tr>
<tr>
<td>6</td>
<td>0.310</td>
<td>0.615</td>
<td>0.940</td>
<td>0.533</td>
<td>0.643</td>
<td>0.022</td>
</tr>
<tr>
<td>7</td>
<td>0.688</td>
<td>0.468</td>
<td>0.653</td>
<td>0.475</td>
<td>0.818</td>
<td>0.023</td>
</tr>
<tr>
<td>8</td>
<td>0.193</td>
<td>0.278</td>
<td>1.053</td>
<td>0.253</td>
<td>0.625</td>
<td>0.177</td>
</tr>
<tr>
<td>9</td>
<td>0.198</td>
<td>0.613</td>
<td>0.153</td>
<td>0.780</td>
<td>0.270</td>
<td>0.160</td>
</tr>
<tr>
<td>10</td>
<td>0.945</td>
<td>0.438</td>
<td>0.993</td>
<td>0.698</td>
<td>0.223</td>
<td>0.170</td>
</tr>
</tbody>
</table>

### 4.1.1 A Small-sized Example

In this section, we use a small-sized example to compare the allocation results of the probabilistic-choice and optimal-choice models. A similar example has been used in Zhang et al. (2012a). Suppose there are 5 open facilities and 10 population nodes. The travel time matrix in hours and the fractions \((h_i)\) are given in Table 1, and \(\lambda = 10\) clients/hour. There is only one server at each open facility with \(\mu = 4\) clients/hour. The parameters used in both models include \(u_j = 0, j = 1, \ldots, 5\), \(\beta_1 = 1\) and \(\beta_2 = 1\).

Tables 2 and 3 show the allocation results of the optimal-choice and probabilistic-choice models, respectively, using the above allocation algorithm. It is clear that the allocation results are different. Based on the optimal-choice model, clients from the same node usually visit a single facility. If they visit two or more facilities, the utilities to the multiple facilities are actually identical. Therefore, if the focus of a real project is on the flows of clients to facilities, it is critical to conduct a thorough investigation of client choice behavior prior to choosing a model.

Another interesting difference between the two allocation results is that the variation in the numbers of clients to the facilities is less in the probabilistic-choice model ([1.864, 2.127]...
Table 2: The Allocation Result of the Optimal-Choice Model

<table>
<thead>
<tr>
<th>Facility</th>
<th>Node</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.610</td>
<td>0.610</td>
<td></td>
<td></td>
<td></td>
<td>0.610</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>1.520</td>
<td></td>
<td></td>
<td></td>
<td>1.520</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.710</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.045</td>
<td>0.615</td>
<td></td>
<td></td>
<td></td>
<td>0.660</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.950</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.220</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.230</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1.770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.770</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.054</td>
<td>1.547</td>
<td>0.019</td>
<td></td>
<td></td>
<td>1.620</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>1.710</td>
<td></td>
<td></td>
<td>1.710</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2.044</td>
<td>1.890</td>
<td>2.202</td>
<td>2.135</td>
<td>1.729</td>
<td>10.000</td>
</tr>
</tbody>
</table>

clients/hour) than in the optimal-choice model ([1.729, 2.202] clients/hour). This observation is consistent with that in Zhang et al. (2012a). They have conducted more rigorous analysis to verify this observation. The main reason for this is that, since clients at each node split to all the facilities in the probabilistic-choice model, the numbers of clients to the facilities are more smooth. It is also interesting to note that the variations themselves as well as their difference are smaller than those based on the two models in Zhang et al. (2012a). In other words, the allocation results between our probabilistic-choice and optimal-choice models are closer than the allocation results between theirs. We believe that it is mainly because waiting time, considered as an attractiveness factor, works as a control to the arrival rate to each facility in our models.

4.2 Location Algorithm

We develop two heuristics to solve the upper level problem for both models, a greedy algorithm and a tabu search procedure (Glover 1986).

As denoted earlier, $s_{max}$ is the maximum available servers for the focal firm, and $M$ is the
set of potential locations for facilities of the focal firm. For simplicity of the presentation, we call any location in \( M \) as a facility. If there is at least one server at the facility, it is open; otherwise, it is closed.

The greedy algorithm starts from the set of all the facilities with no servers and adds one server to a facility in each iteration until \( s_{\text{max}} \) servers are added. In each iteration, the algorithm selects the facility whose addition causes the greatest improvement in the objective function.

As the greedy algorithm may not be able to find the global optimum, we develop a tabu search procedure. Tabu search is one of the most successful meta-heuristics, and it is intended not to get stuck at a local optimum and explore other regions of the solution space.

The Tabu search heuristic starts from a given initial solution with \( s_{\text{max}} \) servers. An example of generating an initial solution is to randomly allocate all the servers to any facilities in \( M \). Then, each iteration of the Tabu search focuses on a “neighborhood” of the current solution. We define two types of neighborhood moves: “Remove” and “Add”. “Remove” results from removing a server from an open facility, which leads to the least decrease in the
overall market share; “Add” results from adding a server to one of the facilities (including the closed ones), which leads to the greatest improvement in the overall market share. In each iteration, “Remove” and “Add” are executed successively. The procedure repeats until no solution that improves the objective function can be identified within a given number of iterations denoted by $N_{ite}$. Eventually, the heuristic outputs the best feasible solution found so far.

When moves are selected, tabu restrictions are used to prevent moving back to previously investigated solutions. In this paper, we define a tabu list in which each value is associated with a facility to represent its tabu status. Once a server is removed or added to a facility, the facility is classified as tabu with a length equal to $T_{len}$, which represents the number of iterations in which the node typically will not be selected for removing or adding. However, even for a tabu node, it can still be selected for adding, if an aspiration criteria is satisfied. We use the typical criteria which states that if a move produces a feasible solution which is better than the best known feasible solution, then the tabu status is disregarded and the move is executed.

5 Computational Experiments

The aim of this section is to examine performances of our solution algorithms. In the experiments, the number of potential locations for the focal firm $m$ is set to 10 and 20, while the number of population nodes $n$ is set to 100, 200, and 400. In total, there are 6 problem sets. In each problem set, 10 instances are generated. For each instance, the following parameter values are used:

- Problem parameters
  - the number of available servers $s_{max} = 2m$;
  - the number of existing facilities of competitors $r = m/2$;
  - the number of servers at each facility of competitors is randomly generated between 2 and 6;
- the demand rate at each zone $\lambda h_i$ is randomly generated in the interval $[0, 10m/n]$ clients per hour;
- the travel time $t_{ij}$ is randomly generated in the interval $[0, 1]$ hour;
- the service rate at each facility $\mu = 4$ clients/hour;
- the constant client attraction $u_j = 0$;
- the sensitivity to the travel time $\beta_1 = 1$;
- the sensitivity to the waiting time $\beta_2 = 1$.

• Allocation algorithm parameters
  - the error tolerance $\epsilon = 0.001$;
  - the step-size parameter $\phi = 0.95$;
  - the step-size adjusting parameter $\delta = 0.9$;
  - the maximum allowable mean waiting time $\bar{W}_{\text{max}} = 4$ hours.

• Tabu search parameters
  - the stopping criteria $N_{\text{ite}} = 50$;
  - the Tabu length $T_{\text{len}} = 5$.

The algorithms are coded in VC++ 2010. All runs are performed on a workstation with 3.0 GHz Intel Core(TM)2 Quad CPU and 16 GB of RAM. The greedy algorithm and tabu search procedure are compared with regards to two performance measures:

- Deviation: the average relative deviation (%) of objective value obtained by the corresponding algorithm from the best solution obtained by the two algorithms;

- CPU Time: the average CPU running time in seconds.

Tables 4 and 5 reports the results for the probabilistic-choice and optimal-choice models, respectively. Both tables demonstrate that the tabu search procedure finds a better solution for most of the problem instances, and thus it is a preferred algorithm to use. It is actually
Table 4: Computational performance of our solution algorithms for the probabilistic-choice model

<table>
<thead>
<tr>
<th>(m)</th>
<th>(n)</th>
<th>Deviation</th>
<th>CPU Time (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>1.16%</td>
<td>29</td>
</tr>
<tr>
<td>200</td>
<td>1.84%</td>
<td>46</td>
<td>1464</td>
</tr>
<tr>
<td>400</td>
<td>0.93%</td>
<td>74</td>
<td>2192</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>1.04%</td>
<td>119</td>
</tr>
<tr>
<td>200</td>
<td>1.16%</td>
<td>186</td>
<td>7796</td>
</tr>
<tr>
<td>400</td>
<td>0.81%</td>
<td>268</td>
<td>10681</td>
</tr>
</tbody>
</table>

Table 5: Computational performance of our solution algorithms for the optimal-choice model

<table>
<thead>
<tr>
<th>(m)</th>
<th>(n)</th>
<th>Deviation</th>
<th>CPU Time (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>0.45%</td>
<td>34</td>
</tr>
<tr>
<td>200</td>
<td>0.38%</td>
<td>57</td>
<td>1542</td>
</tr>
<tr>
<td>400</td>
<td>0.23%</td>
<td>79</td>
<td>2641</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>1.24%</td>
<td>133</td>
</tr>
<tr>
<td>200</td>
<td>0.86%</td>
<td>211</td>
<td>8595</td>
</tr>
<tr>
<td>400</td>
<td>1.09%</td>
<td>322</td>
<td>11628</td>
</tr>
</tbody>
</table>

applied to the case study presented in the next section. The cost for the performance of the tabu search procedure is its longer CPU running time. Between the two models, we observe that the optimal-choice model typically requires a longer CPU running time, since its MPEC formulation seems to be more complicated. This observation is different from that shown in Zhang et al. (2012a), as their two models are MIPs and the probabilistic-choice model is more complicated.
6  A Case Study

As mentioned earlier, this research is motivated by a real-life case study that we conducted for a private firm providing medical diagnostic laboratory services in BC, Canada. Zhang et al. (2012b) presented the background of the case study and main findings in detail. Here, we only briefly describe its background. The focal firm operates 45 patient service facilities, and the number of annual patient visits is around 1.9 million. It is referred to as Firm A in that paper and also here. There are two other main providers of these services in the service area. One is another private firm, referred to as Firm B, with 40 facilities. In addition, 23 hospitals in this area provide in-house laboratory services for both inpatients and outpatients.

Each facility has a number of medical stations called “seats”. Each seat can serve one patient simultaneously, and thus the number of seats effectively defines the capacity of the facility. It is a government-regulated industry in BC. The number of seats at each facility is issued by the government. It is normally difficult to apply for new seats; however, it is relatively easier to apply for transferring seats from one facility to another.

Zhang et al. (2012b) introduced a methodology to explore patient choice behavior and to predict future volume of visits. Based on the real data, they demonstrated that probabilistic-choice is more appropriate in this industry. They also found that the two most significant facility attractiveness factors are distance and capacity. They considered capacity rather than waiting time as a factor, mainly because the actual waiting time data was not available. Their model was primarily used for demand prediction, and they did not investigate what the optimal facility locations and capacities are, though this issue may be explored via scenario analysis.

In this section, we intend to use the methodology introduced above to further study this case with respect to three purposes. First, it is interesting to examine whether incorporating waiting time as a factor rather than capacity would improve model fitting. Second, we attempt to study the impact of client choice on facility network design through the comparison of the two models. Third, we are interested in drawing managerial insights about facility network design strategies under competition.

Zhang et al. (2012b) considered a network composed of about 3,700 population nodes
(called dissemination areas) and about 110 facilities (including those of Firm A, Firm B, and the hospitals). In this section, to reduce the complexity of the problem, we focus only on a subarea with approximately 500 population nodes, which is also distant from other subareas. In this subarea, there are 8 Firm A’s facilities, 3 Firm B’s facilities, and one hospital. In total, Firm A has 41 seats, Firm B has 10 seats, and the hospital has 4 seats.

6.1 Parameter Estimation and Allocation

First of all, we need to estimate the values of the model parameters. Zhang et al. (2012b) estimated the number of annual visits from each population node based on age-specific population sizes. Given the average annual operating hours, we convert the number of annual visits to the arrival rate per hour for each node, i.e., $\lambda h_i$. Zhang et al. (2012b) calculated the travel distance between each node and each facility using the rectilinear distance. We assume a constant travel speed of 5 kilometers per hour, so that the distance is converted to the travel time $t_{ij}$. The average service time is approximately 15 minutes, and thus the service rate $\mu$ is set to 4 patients per hour. Then, both travel time and waiting time are measured in hours.

As discussed in Zhang et al. (2012b), the services provided at all the facilities do not have much difference, and other attractiveness factors are not as significant as distance and capacity or waiting time. Hence, we assume that $u_j$ is identical and thus can be eliminated. Therefore, the most critical parameters to be estimated are $\beta_1$ and $\beta_2$.

Note that the actual number of annual visits from each population node to each Firm A’s facility is known, denoted by $\tilde{x}_{ij}$. Given a pair of $\beta_1$ and $\beta_2$, we can compute the number of visits per hour using the algorithm in Section 4.1. Given the annual operating hours for each Firm A’s facility, the number of visits per hour can then be converted to the number of visits per year, denoted by $\hat{x}_{ij}$. Thus, to find the values of $\beta_1$ and $\beta_2$ that fit the actual data best, we run a two-dimensional grid search with the objective of minimizing:

$$\sum_{i \in N} \sum_{j \in \tilde{S}} (\tilde{x}_{ij} - \hat{x}_{ij})^2,$$

where $\tilde{S}$ denotes the current set of Firm A’s facilities. In other words, our allocation algorithm
could be helpful for empirical studies that focus on the service and healthcare industry, where waiting time is often considered a major attractiveness factor, to estimate parameter values or to predict the demand volume at a facility in a network. Although a better estimation method is worth further investigation, this is out of the scope of this paper.

For the probabilistic-choice model, the range for the two-dimensional grid search on $\beta_1$ and $\beta_2$ is set from 0 to 5 with the increment of 0.1. We found that expression (20) is minimized when $\beta_1 = 2.7$ and $\beta_2 = 1.7$. Figure 1 describes the curve of the expression with respect to the two parameters. It is clear that the impact of travel time is greater than that of waiting time, which is consistent with the finding in Zhang et al. (2012b).

For the optimal-choice model, since only the ratio of $\beta_2$ over $\beta_1$ matters, we attempted to run a one-dimensional grid search on the ratio with the range from 0.1 to 10. However, we found that expression (20) decreases with respect to the ratio in this range. The main reason for the decrease can be explained as follows. The real data shows the probabilistic-choice behavior, i.e., the flows from a population node to the facilities are relatively smooth. In order to fit the data, the optimal-choice model is forced to increase the ratio and thus reduce the impact of travel time continuously in this range, so as to make the flows smooth. We believe that the values out of this range are not reasonable any more in practice. Thus, in this case, we apply the same values obtained above to the optimal-choice model.

Now, we compare the prediction results based on our models and those in Zhang et al. (2012b) to draw managerial insights. They primarily used a probabilistic-choice model with a list of six attractiveness factors, including distance and capacity. In addition, they compared three simplified models to this “Base” model: 1) the “Closest” model: patients visit the closest facility; 2) the “Distance” model: a probabilistic-choice model with distance as the only attractiveness factor; and 3) the “Distance and Capacity” model: a probabilistic-choice model with only these two attractiveness factors. They compared these models in terms of two goodness-of-fit measures: 1) the correlation between the actual number of visits from node $i$ to facility $j$, $j \in S$, and the predicted number (at the node-facility level); 2) the correlation between the actual total number of visits to facility $j$, $j \in S$, and the predicted number (at the facility level).

The comparison here includes our probabilistic-choice and optimal-choice models as well
as their four models. These six models are applied to this subarea, and the same two goodness-of-fit measures are used. Table 5 summarizes the results of these six models. The first two rows present the results of the “Base” and “Distance and Capacity” models in Zhang et al. (2012b). The fourth and sixth rows present the results of our probabilistic-choice and optimal-choice models with $\beta_1 = 2.7$ and $\beta_2 = 1.7$. The third row presents the results of the probabilistic-choice model with $\beta_2 = 0$, i.e., the “Distance” model. The fifth row presents the results of the optimal-choice model with $\beta_2 = 0$, i.e., the “Closest” model.

Among these models, the “Base” model fits best, because it includes six attractiveness factors. The probabilistic-choice model is the second best model, better than the “Distance and Capacity”. This finding suggests that it is superior to consider waiting time instead of capacity as the attractiveness factor. Intuitively, from the perspective of clients, waiting time is also a more reasonable factor than capacity to be considered to influence their choice decisions.

The “Distance” model is worse than the above three, because only distance (travel time) is considered, even though the fitting may be improved by using a better value of $\beta_1$. In particular, the correlation at the facility level is low, due to the lack of consideration on waiting time or capacity as a control. This result is consistent with the earlier study. Unsurprisingly, the “Closest” model is the poorest one. However, it is surprising to see that the optimal-choice model achieves a very high correlation at the facility level, even though the correlation at the node-facility level is not good. This finding again verifies the importance of incorporating waiting time in the model as a control. It also suggests that, if the focus of a study is on the facility level, the optimal-choice model taking waiting time into account might be acceptable to use in some situations. This comparison is insightful for future analytical studies to choose a proper client choice model.

6.2 Facility Network Design

In this section, we present main results and sights about the facility network design based on our models and the above parameter values. The current facility network with the population density of this area is shown in Figure 2; the number beside each facility represents its number of seats (capacity).
Table 6: Model Comparison for Fitting

<table>
<thead>
<tr>
<th>Study</th>
<th>Model</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>Node-Facility Level</th>
<th>Facility Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. (2012b)</td>
<td>Base</td>
<td></td>
<td></td>
<td>0.888</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>Distance and Capacity</td>
<td></td>
<td></td>
<td>0.880</td>
<td>0.945</td>
</tr>
<tr>
<td>Current paper</td>
<td>Distance</td>
<td>2.7</td>
<td>0</td>
<td>0.801</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td>Probabilistic-Choice</td>
<td>2.7</td>
<td>1.7</td>
<td>0.884</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>Closest</td>
<td>2.7</td>
<td>0</td>
<td>0.627</td>
<td>-0.097</td>
</tr>
<tr>
<td></td>
<td>Optimal-Choice</td>
<td>2.7</td>
<td>1.7</td>
<td>0.685</td>
<td>0.981</td>
</tr>
</tbody>
</table>
First of all, let us again focus on Firm A. Clearly, its current network is not optimal. For instance, there are two separate facilities close to the hospital. Why not merge them so as to benefit from capacity pooling? There are historical issues for the current network. For example, the facilities may be purchased by the firm from different owners at different times. This is why the firm needs to reorganize its facility network.

Since it is difficult to apply for new seats, we assume that the total number of seats of Firm A is fixed. Nevertheless, the firm can open a new facility, close an existing facility, move an existing facility, and transfer seats from one facility to another. In addition to the current locations, we chose 12 other sites in this area as the potential facility locations. These 12 sites are typically located at main intersection points of the road network, close to large population communities. Also, at least some physicians have to be located around these sites, according to the government regulation. To be consistent with our models, we do not consider costs associated with opening, closing, moving facilities, or transferring seats.

Given Firm B’s facilities and the hospital fixed, what is the optimal facility network of Firm A to maximize its market share? Figures 3 and 4 display the optimal facility networks
for Firm A derived from the optimal-choice and probabilistic-choice models, respectively. In terms of improvement, both models show that by redesigning the facility network accordingly, Firm A’s overall market share increases by more than 2%.

Figure 3 shows that Firm A still operates 8 facilities; however, the locations or capacities of some facilities change. As expected, the two facilities originally close to the hospital merge together now; as a result, one seat is saved. More importantly, it is clear to see that the overall strategy for Firm A is to relocate its facilities and capacities from east to west, closer to Firm B’s facilities. In particular, one Firm A’s facility is located on top of one Firm B’s. This kind of “back-to-back” strategy has been derived from earlier studies in economics and operations research (e.g., Hotelling 1929). Interestingly, Firm A’s facilities that are next to Form B’s ones or the hospital have at least one more seat. This is sort of the “undercutting” strategy especially studied in the location and pricing literature. Note that, unlike the location and pricing models, Firm B would not be undercut in this case due to the property of the user-equilibrium.

A major difference of Figure 4 from Figure 3 is that Firm A operates 13 facilities. Almost all Firm A’s facilities are close to high population communities. In other words, they are closer to the population and thus the travel time is reduced. Because of the greater number of facilities, the capacities at the facilities are normally fewer. One of the main reasons for the differences is no consideration on costs. Hence, this solution may not be practical. We have modified the probabilistic-choice model by constraining the number of operating facilities to 8. Then, the derived facility network is close to that shown in Figure 3.

In spite of these differences, the overall strategy for Firm A derived from both models is the same, i.e., to relocate its facilities and capacities from east to west, closer to or even surrounding Firm B’s facilities, leading to a more intensive face-to-face competition.

On the other hand, since Firm B is a weaker player in this area, what is the strategy for this firm to maximize its market share by optimizing its facility network? Is it the same as that for Firm A, i.e., should Firm B relocate its facilities and capacities closer to Firm A’s?

We applied our models to Firm B, fixing Firm A’s facilities and the hospital. Figures 5 and 6 display the optimal facility networks for Firm B derived from the optimal-choice and probabilistic-choice models, respectively. In terms of the facility network, both figures
demonstrate that Firm B still operates 3 facilities with 3, 3, and 4 seats, respectively, but the locations are different. Again, although the locations are different, we believe that the strategies derived from the two models for Firm B are similar. That is, instead of a face-to-face competition, Firm B should relocate its facilities and capacities farther away from Firm A’s facilities; more importantly, these locations should be close to relatively high population communities that are uncovered by Firm A’s facilities yet.

In terms of improvement, the optimal-choice model shows a 6% increase in the market share, while the other model only shows a less than 2% increase. Since the probabilistic-choice model is more appropriate in this case, the difference implies that using the optimal-choice model may significantly overestimate the improvement. Thus, even though the optimal-choice model might be acceptable to use in some situations, we would recommend using the probabilistic-choice model in practice. Applications need to be proceeded after a thorough empirical investigation.

The above analysis is under the condition that one of the firms remains fixed. One may ask what the Nash equilibrium facility network would be for the case. A question even prior to this is whether such equilibrium exists. To investigate these issues, we run our models in an iterative way given that the hospital is fixed. In one iteration, the models apply to Firm A, and in the next iteration, the models apply to Firm B. We found that, unless charging costs associated with relocating a facility or a seat, the equilibrium does not exist in this case. In other words, the firms change their facility locations or capacities dynamically, leading to a loop. This kind of system dynamics is sometimes realistic for a duopoly situation with one strong player and one weak player. In contrast, if costs are charged, it is easy to imagine that the equilibrium always exists. However, the equilibrium then mainly depends on the values of the costs as well as that which firm moves first.

7 Conclusion

This paper describes a methodology for designing a network of medical facilities under competition. We focus on “walk-in” medical facilities and assume that travel time and waiting time are the two main determinants for clients who choose where to receive the service. Sub-
Figure 3: Optimal facility network for Firm A based on the optimal-choice model

Figure 4: Optimal facility network for Firm A based on the probabilistic-choice model
Figure 5: Optimal facility network for Firm B based on the optimal-choice model

Figure 6: Optimal facility network for Firm B based on the probabilistic-choice model
ject to a total capacity limit, the problem is to optimize the location and capacity of each open facility so as to maximize total volume. To incorporate the client choice behavior, we consider two different models, one probabilistic-choice and the other optimal-choice. Both models are formulated as MPECs. The lower level problem is to determine the flows of clients to facilities, and the upper level is a facility location and capacity allocation problem. We propose a common solution methodology along with the bilevel framework for both models. An allocation algorithm is developed to solve the lower level problem, and a greedy algorithm and a Tabu search procedure are developed to solve the upper level problem. Our computational experiments show that the Tabu search procedure is more effective.

Based on the models and the solution methodology, we investigate a real-life case study and derive several interesting managerial insights about model selection and network design strategies. In particular, we compared a few client choice models in terms of their goodness-of-fit. We found that incorporating waiting time as a client choice determinant instead of capacity would improve the goodness-of-fit. When waiting time is considered, the optimal-choice model may fit the data and predict the demand volume at the facility level as accurately as the probabilistic-choice model. This comparison is insightful for future analytical studies to choose a proper client choice model. Furthermore, we observed that the network design strategies derived from both models are similar. In a duopoly situation, the stronger player should relocate its facilities and capacities closer to those of the weaker player, leading to a more intensive face-to-face competition; on the other hand, the weaker player should relocate its facilities and capacities farther away from those of the stronger player. We also observed that a Nash equilibrium network may not exist in this duopoly case, unless charging costs associated with relocating facilities or capacities.

In the future, our models can be further extended in a few ways. Instead of a competitive environment, we address a centralized problem in a parallel study, where a single decision maker, such as the government, can determine the locations and capacities of all the facilities on a network to maximize total social welfare. In this centralized environment, it is interesting to compare the probabilistic-choice and optimal-choice models to the system-optimal model where each client is assigned to a facility. Moreover, a game-theoretical framework for this facility location and capacity allocation problem with duopoly deserves further in-
vestigation.

References


Appendix 1: Proofs for Theorems

Theorem 2

Proof: First, it is shown that if $x_{ij}^*$ satisfies (15) then it also satisfies (16).

From (15), one has that

$$ (U_{ij}^* - U_i^*)^* x_{ij}^* = 0. \quad (21) $$

Note that given $s_j$, for a fixed pair $(i, j), i \in N, j \in S \cup R$, one must have that

$$ (U_{ij}^* - U_i^*)(x_{ij} - x_{ij}^*) \leq 0 \quad (22) $$

for any nonnegative $x_{ij}$. Hence, summing over all pairs, one then has (16).

Now, it is shown that if $x_{ij}^*$ satisfies (16) then it also satisfies (15). For any pair $(k, l)$, let $x_{ij} = x_{ij}^*, \forall (i, j) \neq (k, l), i \in N, j \in S \cup R$, and then (22) simplifies to

$$ (U_{kl}^* - U_k^*)(x_{kl} - x_{kl}^*) \leq 0, \quad (23) $$

from which (15) follows for this $(k, l)$ and consequently for every pair. □

Theorem 3

Proof: To prove Theorem 3, one may simply prove that

$$ \sum_{i \in N} \sum_{j \in S \cup R} U_i^* (x_{ij} - x_{ij}^*) = 0. \quad (24) $$

From (2), one has

$$ \sum_{j \in S \cup R} x_{ij} = \sum_{j \in S \cup R} x_{ij}^* = \lambda h_i. \quad (25) $$

Therefore, one must have (24). □