Decision Support

Patient choice analysis and demand prediction for a health care diagnostics company

CrossMark

Yue Zhang[a], Liping Liang[b,*], Emma Liu[c], Chong Chen[d], Derek Atkins[e]

[a] *College of Business and Innovation, University of Toledo, USA*
[b] *Faculty of Business, Lingnan University, Hong Kong*
[c] *BC Cancer Agency, British Columbia, Canada*
[d] *Fraser Health Authority, British Columbia, Canada*
[e] *Sauder School of Business, The University of British Columbia, Canada*

A B S T R A C T

This paper describes a case study for a medical diagnostic laboratory service provider to model the behavior of patients when choosing a patient service centre for their medical tests and to estimate future demand volume. A tool developed based on our methodology allows the management of the diagnostic services to experiment with locations and capacities for locating or relocating service centres. In addition to the focal firm, the methodology considers the impact of decisions on another service provider and hospital laboratories located in the same area. The methodology identifies the most significant service centre attractiveness factors. Our models are validated from different perspectives and show good predictive capability. This case study is used to draw a number of lessons for applying these types of models to other similar services in order to assist other applications.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Medical diagnostic laboratory services play a crucial role in the detection, diagnosis, and treatment of disease for patients. An estimated 60–70 percent of all decisions regarding a patient's diagnosis and treatment, hospital admission, and discharge are based on laboratory test results (Mayo Clinic, 2010). The medical diagnostic laboratory service industry is a large industry. In the U.S., for example, diagnostic costs were approximately 10 percent of overall health care expenditures in the 1990s (Benge, Bodor, Younger, & Parl, 1997), and this industry recently had revenues of more than 50 billion dollars (Knowledge Source, 2010). British Columbia (BC), Canada, the location of this study, spent an estimated 457 million dollars on laboratory services in 2001–2002 (BC Ministry of Health, 2003).

Medical diagnostic services are offered to both outpatients and inpatients. Inpatients do not concern us here as in countries such as the U.S. and Canada, the vast majority of inpatient tests are done by hospital laboratories. We focus on outpatients. In BC, outpatient tests are conducted by either hospital test centres or private service test centres (commonly called Patient Service Centres or PSCs), in either case funded on a per test basis through the Medical Services Plan (MSP, http://www.health.gov.bc.ca/msp/). MSP is a government ad-

ministered health insurance plan, and enrolment with MSP is mandatory for all BC residents. Samples from the PSCs will typically be sent to centralized laboratories from which results are forwarded directly to physicians' offices or clinics.

With test revenues regulated, the profit of a private service provider depends on the volume of visits to its PSCs or its market share as well as on controlling costs. This brings us immediately to the core problem addressed in this paper. A private service provider in BC approached us for help in evaluating when and where to relocate a PSC, when to change the capacity of a PSC, and when to extend operating hours; all with the aim of increasing their share of volume of visits. The service provider would also like to know the impact on its PSCs when a competitor makes similar decisions. In the longer term, the service provider would like to experiment with different locations in the face of demographic trends, such as ageing and population growth.

The answer to any of these questions depends on our ability to predict which PSC a patient attends, which in turn means understanding what characteristics of a PSC are instrumental to this decision. For example, does a patient simply go to the nearest from their home, or from their physician, one adjacent to public transit, to a mall or having adequate parking, near other associated diagnostic facilities or PSCs at which they have not waited too long in the past?

Therefore, the primary problem posed was, "how to predict patient demand and market share for PSCs sufficiently well to enable

* Corresponding author. Tel.: +852 2616 8103; fax: +852 2892 -2442.
E-mail address: lipingliang@ln.edu.hk (L. Liang).

the evaluation of many managerial choices about location, capacity, and opening hours in terms of the impact on the entire system of PSCs including those of competitors." And to do this, we had the secondary problem, "how to predict patient choice of PSC based on characteristics of those PSCs."

In this paper, we shall describe the methodology that we employed and the results. Our key result is that the use of a probabilistic choice behavior model for patients is superior to other choice models while still keeping the methodology reasonably simple and portable. Although carried out with data pertaining to medical diagnostic facilities, we believe the method and results to be insightful for many other service sectors.

The remainder of the paper is organized as follows. We first present an overview of the relevant literature in Section 2. We then describe the background of the case study, our general methodology, and preliminary analysis in Sections 3, 4, and 5, respectively. In Section 6, we further discuss the explanatory variables included in the models, estimation, and model validation. We then provide a few key insights in Section 7 by comparing our model with three other simplified models. Section 8 concludes the paper.

2. Literature review

There is a rich literature in operations, economics, and marketing that studies the behavior of people choosing among a set of alternatives or predicts the flow of people visiting a set of locations.

Consumer shopping choice is one of the areas that are studied most. Huff (1964) developed an early gravity-type model that included distance to stores and size of stores as independent variables to estimate market shares of retail stores. Many extensions, such as the multiplicative competitive interaction (MCI) model (Nakanishi & Cooper, 1974), have been subsequently proposed and applied (see, e.g., Drezner & Drezner, 2002; Gautschi, 1981; Jain & Mahajan, 1979). These models are usually based on aggregate flows between consumer zones and stores. The flows are predicted as a function of a store's attractiveness factors, such as distance or travel time, store size, floor space, and accessibility. Other models of consumer shopping choice are based on disaggregate (individual) discrete choice models. They assume that the probability of people choosing a certain alternative is influenced by the attractiveness of that alternative. The most well-known models are probably the multinomial logit (MNL) model (McFadden, 1974) and its extensions (Bell & Lattin, 1998; Berry, 1994; Severin, Louviere, & Finn, 2001). As these models can also be applied at an aggregate level, they may be viewed as logical extensions of the Huff model as well.

In the health care sector, the literature focuses primarily on patient hospital choice. While a few earlier works applied the MCI model (Erickson & Finkler, 1985; Folland, 1983; McGuirk & Porell, 1984), most of the studies were based on the MNL model to analyze the patient hospital choice at the individual level (Bronstein & Morrisey, 1990; Cohen & Lee, 1985; Roh, Lee, & Fottler, 2008; Sivey, 2012; Tai, Porell, & Adams, 2004). These studies usually considered both patient characteristics and hospital attractiveness factors. However, there appear to be little literature on medical diagnostic laboratory services and little guidance for governments, health authorities, or service providers who hope to improve patient satisfaction and revenues.

In the location literature, customer choice for alternative facilities is often considered in optimization models to determine the optimal facility locations or capacities. Traditional studies tend to simplify the customer choice behavior, and often assume that customers make their choices based on distance only and seek service from the closest facility (Berman, Krass, & Wang, 2006; Verter & Lapierre, 2002; Wang, Batta, & Rump, 2002). Now, gravity-type or discrete choice models have also been incorporated into location models (Aboolian, Berman, & Krass, 2007; Benati & Hansen, 2002; Haase & Muller, 2014; Mar-

ianov, Rios, & Icaza, 2008; Zhang, Berman, & Verter, 2012). Readers may refer to several recent review papers for facility location models (Boffey, ao, & Espejo, 2007; Daskin & Dean, 2004; Klose & Drexl, 2005; ReVelle & Eiselt, 2005).

3. Background

The private laboratory service provider for which we carried out the study is referred to as Firm A throughout the paper.[1] Firm A is located in the southwest of BC (henceforth called the service area). Operating 45 Patient Service Centres (PSCs), it serves 6,000 to 8,000 patients and performs 35,000 tests every day. The number of annual patient visits is around 2 million. Firm A requested assistance from the Centre for Operations Excellence at the Sauder School of Business in the University of British Columbia in understanding patient choice of PSC and developing tools to aid managerial decisions about location and capacity of PSCs.

There are two other main providers of these services in the service area. One is another private provider with 40 PSCs, which we shall refer to as Firm B. In addition, 24 hospitals in this area provide in-house laboratory services for both inpatients and outpatients.

The basic process is as follows. A patient is referred for a panel of tests by a physician, primarily a family physician. With a physician's requisition, a patient can visit any PSC for the service. Usually, no appointment is needed; most PSCs use first-come-first-served protocols. After waiting, a patient enters a medical (phlebotomy) station somewhat misleadingly called a "seat," where the test is taken. The number of "seats" effectively defines the capacity of the PSC. After collection, the sample is delivered to a central laboratory and results are electronically sent to physicians.

In Canada there is mandatory universal health insurance that covers most laboratory services. Insurance coverage is not an important factor in patient choice. As this is a government regulated industry, major operational decisions made by private service providers or hospitals in BC need approval by the BC Ministry of Health. For instance, opening a new PSC, moving an existing PSC to another site, or changing the number of "seats" at a PSC, all require submission of an application with adequate supportive evidence.

The key purpose of the study was to develop a practical and portable tool that can be used by service providers to predict demand or market share subsequent to any changes in the service facility network, such as addition, deletion, or move of a PSC.

4. Data and methodology

4.1. Data

We obtained the following data from Firm A:

- All referrals by physicians of patients to PSCs during 2004–2012, about 14 million records in total. The data includes the patient age, gender, postal code, the panel of tests requested, the date and time of the test, and the signing physician. We primarily used 2007, 2008, 2011, and 2012 data records.
- Information about PSC locations including hours of operation and capacities, meaning the number of "seats."
- Information about physicians, their clinical office address, and specializations.

No data were available about referrals completed at Firm B or the hospitals. However, their locations, hours of operation, and capacities were publicly known. Also, the total volume of visits in the service area was reported each year, including those of Firm A, Firm B, and

---

[1] The company wishes to remain anonymous, but a letter from the CEO affirming that the paper reflects the implementation of the work has been filed with the Editor and made available to referees.

hospitals. Moreover, we obtained geographic/demographic data of the service area from 2011 Canadian Census data and population projections data from BC Statistics.

4.2. Methodology

The objective of the study was to estimate future demand volume to each PSC by means of predicting the number of visits from each geographic region to each PSC. The smallest geographic unit in the Canadian Census data that is readily available is the dissemination area (DA); see Section 5.1 below for more description. Therefore, our task is to predict the number of visits from each DA to each PSC.

Our methodology is composed of two parts: a *demand* model and a *share* model. The former is to predict the total number of visits originating from each DA, while the latter is to predict the share or proportion of visits from a particular DA to a particular PSC.

Let I be the set of DAs, A be the set of Firm A's PSCs, B be Firm B's PSCs, and H the hospital PSCs. Let y_{ij} ($i \in I$, $j \in A \cup B \cup H$) denote the number of visits from DA i to PSC j. We have data only from Firm A, i.e., y_{ij} for $i \in I$ and $j \in A$, and the total number of visits in the service area $\sum_{i \in I} \sum_{j \in A \cup B \cup H} y_{ij}$. Let d_i ($i \in I$) be the total number of visits from DA i, and m_{ij} ($i \in I$, $j \in A \cup B \cup H$) the share or proportion of visits from DA i to PSC j. In general, neither d_i nor m_{ij} is known. Clearly these quantities can be expressed as:

$$y_{ij} = d_i m_{ij}, \quad i \in I, j \in A \cup B \cup H. \tag{1}$$

Our task is to predict the future values of d_i and m_{ij}.

The prediction of d_i, denoted by $\hat{d}_i$, in general would be of the following form:

$$\hat{d}_i = \sum_{k=1}^{K} \gamma_k g_{ik}, \quad i \in I.$$

Here the population has been grouped into K reasonably homogeneous groups sharing a common average number of annual visits per person γ_k. Such groups would typically be based on demographic characteristics, such as age, gender, etc. This grouping is context dependent, and when we discuss the actual data below, it will be seen that only three age groups were sufficient. In general, more groups are likely. Let g_{ik} denote the population of group k at DA i, for $1 \leq k \leq K$ and $i \in I$. Note that g_{ik} is known. We discuss both the choice of groups and the assumption that the γ_k's are common across DAs in more detail below. Given this and supposing for the moment that data concerning d_i are available, then the γ_k's can be estimated from a linear regression of the form:

$$d_i = \sum_{k=1}^{K} \gamma_k g_{ik} + \epsilon_i, \quad i \in I, \tag{2}$$

where ϵ_i denotes an independent normally distributed error term. The issue that data for d_i are not available is also addressed below.

For the share model, it was clear that no single factor made a PSC attractive and that patients were sufficiently diverse that there would always be some idiosyncratic reasons for their choice. Thus, we adopted the MNL model. Let u_{ij} denote the overall utility of PSC j for patients from DA i and x_{ijn} the value of the nth attractiveness factor of PSC j for patients from DA i. Based on the MNL model, u_{ij} can be expressed as:

$$u_{ij} = \sum_n \beta_n x_{ijn} + \varepsilon_{ij}, \quad i \in I, j \in A \cup B \cup H, \tag{3}$$

where β_n is the coefficient associated with the nth PSC attractiveness factor, to be estimated, and ε_{ij} as usual denotes an extreme value distributed error term representing the unidentified attractiveness.

Then the MNL model gives m_{ij} by:

$$m_{ij} = \frac{\exp(u_{ij})}{\sum_{l \in A \cup B \cup H} \exp(u_{il})}, \quad i \in I, j \in A \cup B \cup H. \tag{4}$$

An important assumption is that these coefficients β_n are identical for patients who visit the PSCs of Firm A, Firm B, or the hospitals. Management's opinion was that patients were reasonably homogenous in terms of attractiveness factors. A generalization to the case β_{nk} where attractiveness was only homogenous by population group k would be a straightforward extension if needed.

So far our two models have been general. Next, we will discuss our preliminary data analysis, which helped us select population groups and PSC attractiveness factors for the two regression models.

5. Preliminary analysis

5.1. Data aggregation

As mentioned earlier, DA is the smallest geographic unit in the Canadian Census data that is readily available. One DA may contain a few to nearly a hundred postal codes. We had postal codes for each patient, about 100,000 distinct codes for the service area. Data were aggregated from postal codes to DAs to ensure patient anonymity. Our network is composed of DAs (about 3,800) and PSCs (about 110 including those of Firm B and the hospitals).

5.2. Calculation of travel distance

Distance to a PSC is an obvious choice of an attractiveness factor. Using the latitude/longitude of each DA and PSC, the rectilinear (Manhattan) distance between the centroid of a DA and a PSC was used as the travel distance. The conversion of longitude and latitude to kilometers is 72.74 and 111.32 kilometers per degree respectively at the latitude of Vancouver, BC. We considered this a reasonably accurate measure, given the grid road system. An exception to this was an adjustment to deal with the major barrier to travel—crossing rivers by bridges. In this case distance was replaced by the sum of the Manhattan distance to each bridge crossed from the source and to the destination. In the case of multiple route choices, the shortest was taken.

5.3. Volume of visits

We observed that the growth in total volume of visits to Firm A during 2009–2011 was approximately 3 percent per year. This increase mainly results from both population increase as well as the population ageing, while the average number of visits per population per age does not change much. On average, one third of visits require fasting services.

Without data from Firm B and the hospitals, the total number of visits from each DA d_i was unknown even with knowing the population. Fortunately, there was a particular sub-region, containing about 10 percent of the service area population, notable for the almost complete absence of Firm B and hospital service locations. Let S be the set of DAs in this sub-region. With such a dominant market share in this sub-region, we could approximate the total number of visits from each DA d_i, $i \in S$, by $\sum_{j \in A} y_{ij}$, and then use this to estimate the average number of annual visits per person by population group γ_k. The details of the estimation will be discussed in Section 6.2.

Our preliminary analysis showed that the volume in this sub-region is influenced mostly by age and much less by gender. Fig. 1 shows the average number of visits per person with respect to age in 2008 and 2011. This allowed us to restrict our population analysis to age as the key predictor and drop gender.

5.4. Pattern of visits

The travel distance to a PSC is certainly the first factor to consider, as it is reflected in the economics, marketing, and operations literature. Fig. 2 shows individual patient visits in 2008 and 2011 to their
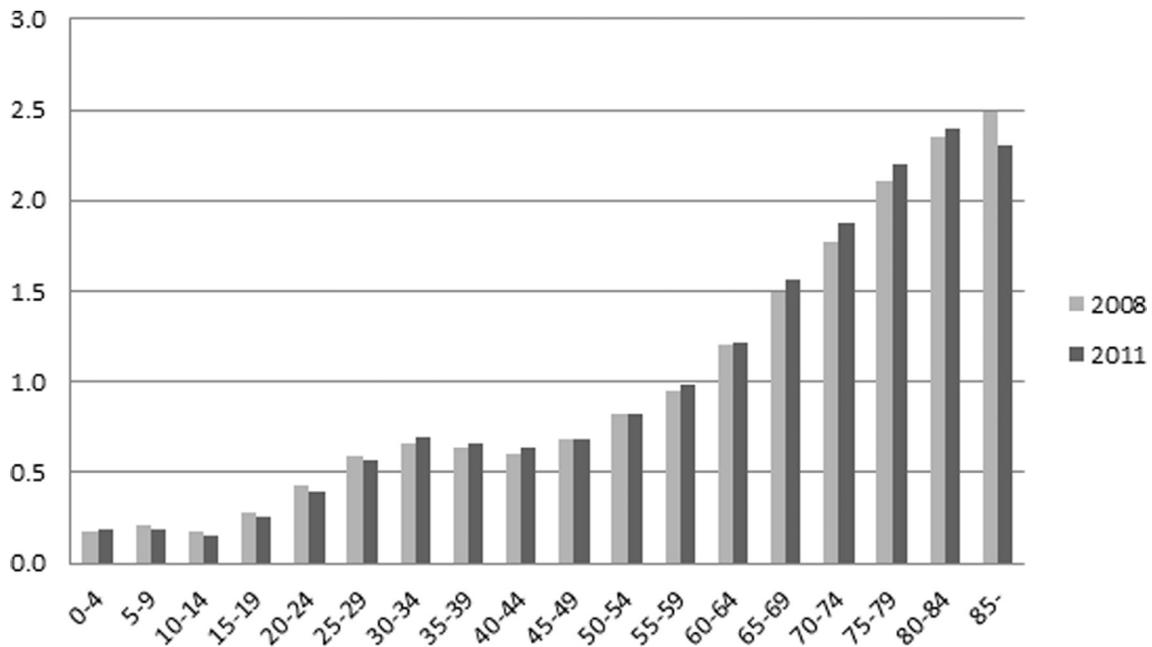
Fig. 1. Average number of annual visits per person with respect to age in 2008 and 2011.
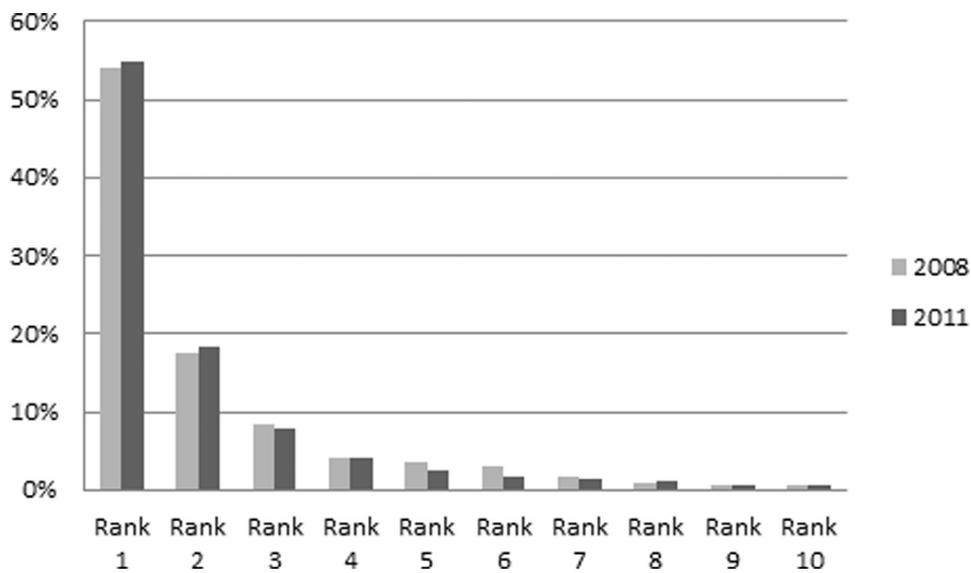

Fig. 2. Percentage of patient visits to the top ten closest PSCs.

top ten nearest PSCs. These account for 94 percent of all visits in 2008 and 93 percent of all visits in 2011. There are two clear messages: patients mainly choose a nearby location, but certainly not always. Thus, the travel distance is important but is not the only factor in patients' choice.

The second place to turn to understand patients' visits would be waiting times at PSCs. These data were not available and neither were there sufficient average data such as about the average number of "staffed seats" to allow for even an approximation. However, our work generated enough excitement among the executive team that equipment to count average waiting times were deployed in selected PSCs so that waiting times might be included in a subsequent upgrade. Instead we focused on what we did know and what was clearly observable to patients. At peak times, when all "seats" would be staffed, patients would experience the service rate, the rate the people waiting in front of them reduced. As the mix of tests stays reasonably constant a good surrogate for this is the maximum capacity,

the maximum number of "seats" in each PSC. The number of "seats" at a PSC varies between 2 and 9, with an average of 4.

Hours of operation vary considerably and affect both the attractiveness to a patient and the daily volume served by a PSC. Some PSCs open Monday to Friday, some open on Saturday; most PSCs open at 8am, but some as early as 6:30am; most PSCs open for a whole day, but a few close in afternoons (after 1pm).

Fig. 3(a) and (b) display the percentage of patient visits on days within a week and hours within a day based on *all* 45 PSCs. From the first plot, we do not observe significant difference in daily volumes from Monday to Friday. Saturday volume is similar to that of weekdays for the PSCs open on Saturday (27 out of 45 PSCs), thus opening on Saturday may proportionally increase a PSC's weekly volume. The second plot shows that more patients visit a PSC in the morning than in the afternoon. This is in part because many tests require fasting. These important differences should be included in the model.
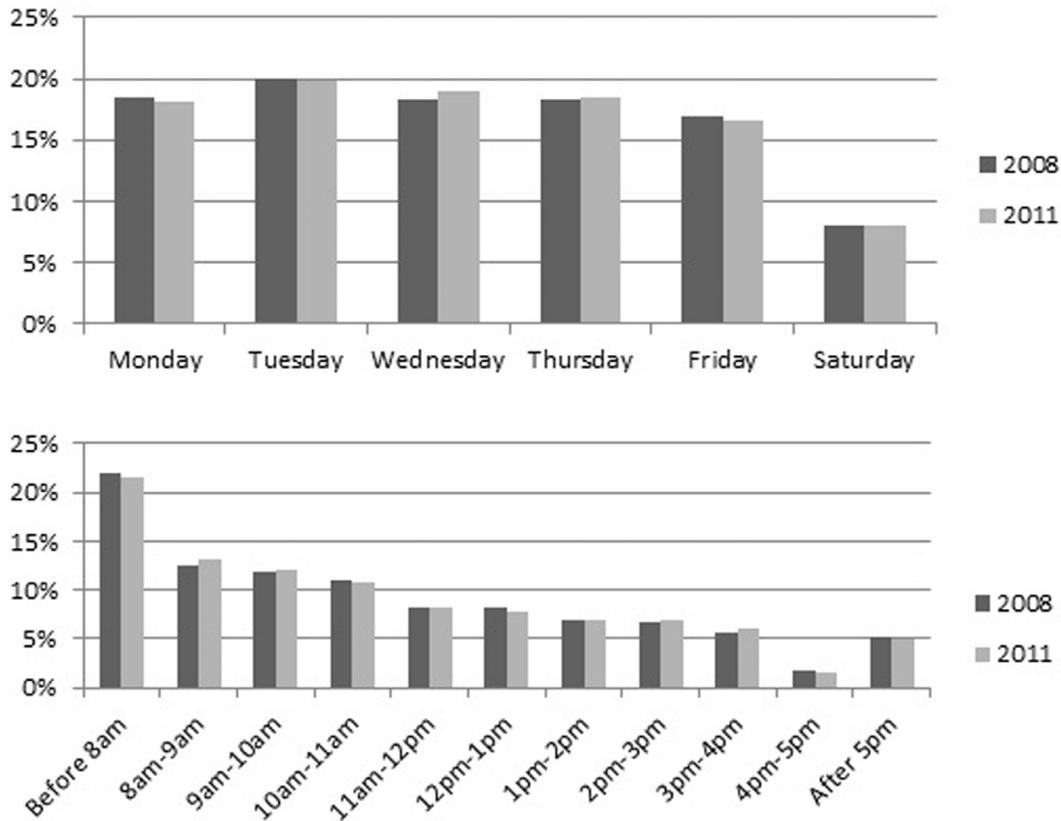
Fig. 3. Percentage of patient visits (a) on different days within a week and (b) during different hours within a day over all PSCs.
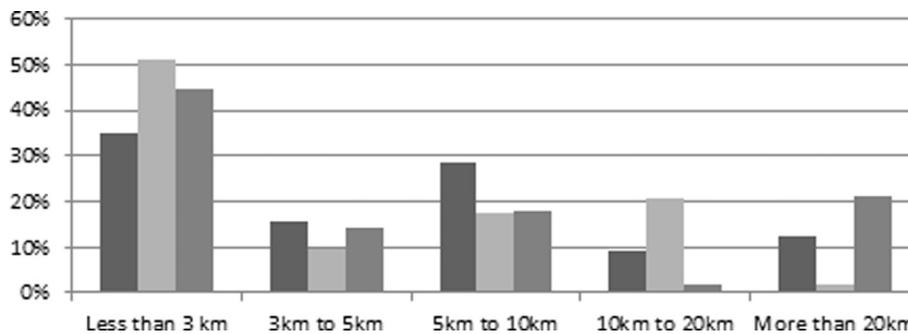

Fig. 4. The proportion of physician requisitions originating within distance radii and used at three different PSCs (Based on 2011 data; the three PSCs are represented by different colors).

5.5. Physicians' impact

Patients are usually referred to a PSC by family physicians. Our data showed that 60 percent of physicians were general practitioners (GPs) rather than specialists and GPs accounted for nearly 80 percent of referrals in 2007. GPs are a key driver of this industry.

The second important characteristic of physicians is their proximity to PSCs. Fig. 4 shows the proportion of physician requisitions originating within distance radii for three typical PSCs in 2011. Clearly, physician-PSC proximity is important in deciding where a patient goes for testing. In the next section, we will include the number of GPs within a certain radius of a PSC in the model as a proxy measure to represent the impact of physicians' proximity.

5.6. Test prices

In Canada most diagnostic testing is available without charge when required by physicians, and where there are charges, the cost is regulated and so constant. Thus we had no attractiveness factor associated with test charges, but in other cases this can be easily included.

6. Explanatory variables, estimation, and validation

6.1. Explanatory variables

Table 1 presents all the explanatory variables with the descriptions that were eventually included in the two regression models. As mentioned in Section 5, the total number of visits from DAs primarily depends on age. After trying different combinations of age groups, we chose three age groups (as explanatory variables in the demand model): below 50, 50–65, and above 65. Then, the demand is a linear function of the population of the three age groups. Using more age groups did not significantly improve the model.

For the share model, we tested a list of PSC attractiveness factors (explanatory variables), which showed relevance in the preliminary analysis. The MNL estimation result then showed what factors

Table 1
Explanatory variables.

Variable	Description
$Pop\_Under50_i$	Population at DA i under the age of 50
$Pop\_50\text{-}65_i$	Population at DA i between 50 and 65
Pop_Over65_i	Population at DA i over the age of 65
$Distance_{ij}$	Travel distance from DA i to PSC j
$Seat\_Hour_j$	Number of seats $\times$ weekly hours of operation of PSC j
$Open\_Early_j$	$Open\_Early_j = 1$, if PSC j is open before 8 a.m.
$X\text{-}Ray_j$	$X\text{-}Ray_j = 1$, if there is X-ray service within 500 meters of PSC j
$Transport_j$	$Transport_j = 1$, if there is convenient public transport at PSC j
$Open\_Saturday_j$	$Open\_Saturday_j = 1$, if PSC j is open on Saturdays
GP_j	Number of GPs within 3 kilometers of PSC j

best explain patients' PSC choices. Table 1 describes only the most significant PSC attractiveness factors. The capacity of a PSC comes from both the number of seats and the hours of operation, which were highly correlated. Therefore, we combined them into a single variable, $Seat\_Hour$, representing the weekly capacity. Prior patient surveys suggested that patients prefer a PSC with nearby auxiliary medical services, of which an X-ray service is the best representative. Proximity to public transport was thought to be important for this population. Opening early was considered to be important mainly because of fasting. Opening on Saturdays was considered an attractiveness factor because it offers convenience, as evidenced by Fig. 3(a). Finally, the number of GPs nearby was included to reflect their impact. A number of other factors were initially included but were not significant, such as opening late, proximity to shopping malls, and parking.

6.2. Estimation

The data described in Section 4.1 were used to estimate the values of the coefficients associated with the explanatory variables in the regression models. In general, we used 2011 data for the estimation, while the data of year 2012 was used as hold-back data to validate the model as discussed in the next section.

As described in Section 5.3, we do not know the total number of visits from each DA d_i, because of no available data from Firm B and the hospitals; however, there was a particular sub-region of the service area with the absence of Firm B and hospital service locations. Therefore, we approximated d_i by $\sum_{j \in A} y_{ij}$, $i \in S$, and ran the linear regression demand model based on the DAs in this sub-region. The results of the regression gave values of the coefficients γ_k for the three age groups shown in Table 2. For example, the population over age 65 averaged 3.2 tests per year compared with 0.52 tests per year for the population under 50's. The coefficient of determination R^2 for the linear regression demand model is 0.87.

However, we were aware that this sub-region had more patients traveling out of the sub-region for testing than patients travelling inwards. Thus, we suspected that the coefficients would be underestimated. Applying these coefficients to the whole population in the service area, we obtained the predicted total number of visits.

Table 2
Estimation results.

Parameter	Variable	Estimate	Std. error	t_value	p_value
γ_1	$Pop\_Under50_i$	0.52	1.40E-01	3.85	<.05
γ_2	$Pop\_50\text{-}65_i$	1.72	5.70E-01	3.01	<.05
γ_3	Pop_Over65_i	3.2	2.90E-01	11.03	<.01
β_1	$Distance_{ij}$	−0.306	2.00E-04	−1387.15	<.0001
β_2	$Seat\_Hour_j$	0.0019	9.01E-06	210.27	<.0001
β_3	$Open\_Early_j$	0.25	2.90E-03	86.58	<.0001
β_4	$X\text{-}Ray_j$	0.015	2.10E-03	6.99	<.0005
β_5	$Transport_j$	0.037	2.50E-03	14.60	<.0001
β_6	$Open\_Saturday_j$	0.4003	2.70E-03	148.66	<.0001
β_7	GP_j	0.00015	1.13E-05	13.22	<.0001

Comparing it to the actual total number of visits reported (including those of Firm B and hospitals), we indeed observed an underestimation. We therefore inflated all the three coefficients by the same multiplier (1.37) to match with the actual total demand.

For the *share* model, we estimated the coefficients β_n associated with the explanatory variables using equations (3) and (4) for all the PSCs of Firm A (i.e., $j \in A$, the data of Firm A for the entire service area), based on the expression:

$$m_{ij} = \frac{\exp\left(\sum_n \beta_n x_{ijn} + \varepsilon_{ij}\right)}{\sum_{l \in A} \exp\left(\sum_n \beta_n x_{iln} + \varepsilon_{il}\right)}, \quad i \in I, j \in A. \tag{5}$$

We used the PHREG Procedure in SAS for the MNL estimation (Kuhfeld, 2010). The likelihood function of the MNL model has the same form as a survival-analysis model fit by the PHREG Procedure.

From Table 2, we can observe that older people had more visits (as expected). Although all the PSC factors are significant, $Distance_{ij}$ is especially so and $Seat\_Hour_j$ is second. $Open\_Early_j$ and $Open\_Saturday_j$ are almost at the same significance level.

As below, the estimated coefficients β_n from Table 2, and our knowledge of x_{ijn}, $j \in A \cup B \cup H$, the attractiveness factors for all the PSCs, give us the market share and patient flow estimates:

$$\hat{m}_{ij} = \frac{\exp\left(\sum_n \beta_n x_{ijn}\right)}{\sum_{l \in A \cup B \cup H} \exp\left(\sum_n \beta_n x_{iln}\right)}, \qquad i \in I, j \in A \cup B \cup H,$$

$$\hat{y}_{ij} = \hat{m}_{ij} \hat{d}_i, \qquad i \in I, j \in A \cup B \cup H.$$

6.3. Validation

We validated the models at three levels. First, we compared the predicted number of visits $\hat{y}_{ij}$ from each DA i to each Firm A's PSC j using the coefficients estimated from the 2011 data to the actual value y_{ij} from the hold-back data of year 2012. We used a pseudo-R^2 as a measure for goodness-of-fit, which is defined as:

$$1 - \frac{\sum_{i \in I} \sum_{j \in A} (y_{ij} - \hat{y}_{ij})^2}{\sum_{i \in I} \sum_{j \in A} (y_{ij} - \bar{y})^2},$$

where $\bar{y}$ denotes the average number of visits from any DA i to a PSC j, $i \in I, j \in A$. In this case, the pseudo-R^2 at the DA-PSC level is 0.80.

Second, we compared the predicted *total* number of visits to each Firm A's PSC (denoted by $\hat{Y}_j$, $j \in A$) to the actual value Y_j. We used another pseudo-R^2 as a measure for goodness-of-fit, which is defined as:

$$1 - \frac{\sum_{j \in A} (Y_j - \hat{Y}_j)^2}{\sum_{j \in A} (Y_j - \bar{Y})^2},$$

where $\bar{Y}$ denotes the average total number of visits to a Firm A's PSC j, $j \in A$. This pseudo-R^2 at the PSC level is 0.87.

Fig. 5 shows the actual and predicted volumes for each PSC. A perfect fit would place each PSC exactly on the 45 degree line. A PSC above the line indicates overestimation, while that under the line indicates underestimation. This graph was most valuable getting "buy-in" from management. The visualization allowed the management to check the PSCs individually and identify reasons why an overestimation or underestimation might have been expected. For example, the upmost point in Fig. 5 is a PSC with 9 "seats," but known by management to be a "poor performer" due to a poor layout. The "seats" are not effectively used, leading to an overestimation for this PSC.

Our third validation was at a more aggregate level. In another particular sub-region of the service area, actual market shares of Firms A and B were available (but that of the hospitals was not). The actual market share of Firm A is 74 percent, while our predicted value is 76 percent.

In summary, at all of these three levels, the prediction was accurate enough to give management confidence to employ the model in making facility locating related decisions.
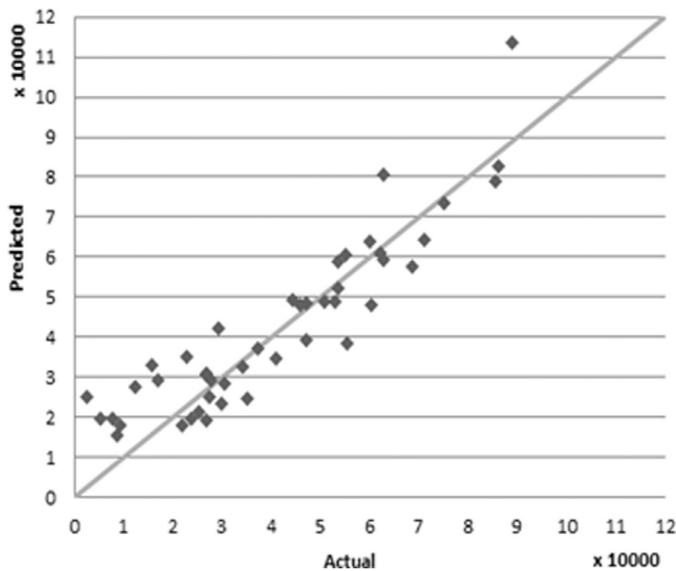
Fig. 5. Comparison of predicted and actual PSC volumes in 2012.

7. Discussion

In this section, we discuss some managerial insights based on the study. We believe that these are important to practitioners, administrators, and researchers in the medical diagnostic laboratory service industry and possibly in other service sectors.

Our share model was formulated as an MNL model, which is based on random utility theory. In contrast, in the location literature mentioned before, it is common to assume that people seek services from the closest open facility (Berman et al., 2006; Verter & Lapierre, 2002; Wang et al., 2002). This assumption makes the analysis easier. Thus, the question is whether the additional complexity associated with using MNL is worth the effort?

A second question is about the value of information. Having identified distance and capacity as the two most significant factors, what is the loss if we drop all other factors?

To investigate these two issues, we compare three simplified models against the MNL share model above, which is now referred to as the "*Base*" model. We keep the *demand* model fixed, hence the number of visits from each DA is fixed. In the "*Closest*" model, we assume patients visit the closest PSC. In the "*Distance*" model, we still use the MNL formulation but include only distance as an attractiveness factor. Thus, as in Fig. 4, patients might go to the second, third, etc. ranked PSC. Similarly, in the "*D&C*" (Distance and Capacity) model, we incorporated only these two factors in the MNL model. For the latter two models, we re-estimated the values of the coefficients for each case separately.

The comparison was in terms of two measures: the sum of squared errors (SSE) between the actual and the estimated numbers of visits from each DA to each Firm A's PSC (at the DA-PSC level), and the SSE between the actual and the estimated total numbers of visits to each Firm A's PSC (at the PSC level).

Table 3 summarizes the results for models using 2012 data. For the measures on both levels, the "*Closest* " model is a poor fit, thus simply choosing the closest PSC is not appropriate for the medical

diagnostic laboratory service industry. The "*Distance*" model improves the fit with respect to both measures. Comparing the "*Distance* " and "*D&C*" models, it is clear that the SSE at the PSC level can be improved substantially by incorporating the capacity factor in the model. In addition, we observe a smaller further improvement from the "*Base*" model over the "*D&C* " model. This finding indicates that, when there is limited time or effort, collecting data on the other factors may not be necessary in practice for this industry.

Table 3 also demonstrates that the choice of a model may depend on the purpose of the study. For example, if the purpose of the study is only to examine the pattern of visits, the "*Distance*" model is probably enough, as the SSE at the DA-PSC level is sufficiently low. In contrast, if the study is to accurately predict the volume to a PSC, like ours, at least the "*D&C*" model should be used.

We believe that some of these findings may apply to the service sector in general. For example, our study demonstrates that, although distance is shown to be the most important one, other factors, especially waiting time, might not be ignored. This finding empirically supports recent analytical studies (Marianov et al., 2008; Zhang, Berman, Marcotte, & Verter, 2010; Zhang, Berman, & Verter, 2009), which assume that both travel time and waiting time are the primary facility attractiveness factors.

8. Conclusion

We have employed the MNL methodology to explore patient choice behavior and to predict future volume of visits for a medical diagnostic laboratory service provider. Our methodology requires no specialized software, employing only standard "off the shelf" statistical and spreadsheet tools. It assists the management with their strategic decisions on the location, resource planning, and forward planning, and it has been shown to be appropriate and simple to use. Despite the lack of direct data for them, other competing service providers can be taken into account. Letting the data "speak" to identify and quantify the most significant facility attractiveness factors is a critical adjunct to customer surveys. Both have their place and we observed the management conversation comparing these data sources to be a rich and informing one. A comparison of this model with three simpler models demonstrated the importance of using the correct patient choice behavior for the application involved.

Our model is easy to use and easy to maintain. Re-estimation when more recent data are available requires someone able to use a statistical package, but in the long intervals between such re-estimation, a very simple spreadsheet can calculate the predictions essentially instantly. This makes what-if experiments with changing the PSC locations and the attractiveness factors available to any manager's desktop. No model can replace the experience of the management on the ground, but we believe this type of model can be of great assistance in many similar service delivery systems.

References

Aboolian, R., Berman, O., & Krass, D. (2007). Competitive facility location and design problem. *European Journal of Operational Research, 182*(1), 40–62.

Bell, D., & Lattin, J. M. (1998). Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer EDLP. *Marketing Science, 17*(1), 66–88.

Benati, S., & Hansen, P. (2002). The maximum capture problem with random utilities. *European Journal of Operational Research, 143*(3), 518–530.

Table 3
Model comparison.

SSE	Closest	Distance	D&C	Base
DA-PSC level	6.9×10^8	1.5×10^8	1.1×10^8	1.0×10^8
PSC level	1.8×10^{10}	1.1×10^{10}	2.8×10^9	2.2×10^9

Benge, H., Bodor, G. S., Younger, W. A., & Parl, F. F. (1997). Impact of managed care on the economics of laboratory operation in an academic medical center. *Archives of Pathology and Laboratory Medicine, 121*, 689–694.

Berman, O., Krass, D., & Wang, J. (2006). Locating service facilities to reduce lost demand. *IIE Transactions, 38*, 933–946.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *RAND Journal of Economics, 25*, 242–262.

Boffey, B., ao, R. G., & Espejo, L. (2007). A review of congestion models in the location of facilities with immobile servers. *European Journal of Operational Research, 178*(3), 643–662.

Bronstein, J. M., & Morrisey, M. (1990). Determinants of rural travel distance for obstetrics care. *Medical Care, 28*, 853–865.

Mayo Clinic (2010). Laboratory sciences career overview. http://www.mayo.edu/mshs/lab-career.html (URL last accessed on July 31, 2015.).

Cohen, M. A., & Lee, H. L. (1985). The determinants of spatial distribution of hospital utilization in a region. *Medical Care, 23*, 27–37.

Daskin, M. S., & Dean, L. K. (2004). Location of health care facilities. In M. L. Brandeau, F. Sainfort, & W. P. Pierskalla (Eds.), *Operations Research and Health Care: A Handbook of Methods and Applications, Kluwers International Series* (pp. 43–76). Kluwer Academic Publishers Group.

Drezner, T., & Drezner, Z. (2002). Validating the gravity-based competitive location model using inferred attractiveness. *Annals of Operations Research, 111*, 227–237.

Erickson, G. M., & Finkler, S. A. (1985). Determinants of market share for a hospital's services. *Medical Care, 23*, 1003–1018.

Folland, S. T. (1983). Predicting hospital market shares. *Inquiry, 20*(1), 34–44.

Gautschi, D. A. (1981). Specification of patronage models for retail center choice. *Journal of Marketing Research, 18*(2), 162–174.

Haase, K., & Muller, S. (2014). A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research, 232*(3), 689–691.

BC Ministry of Health (2003). BC laboratory services review. http://www.health.gov.bc.ca/library/publications/year/2003/lab_review.pdf (URL last accessed on July 31, 2015.).

Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing, 28*(3), 34–38.

Jain, J. K., & Mahajan, V. (1979). Evaluating the competitive environment in retailing using multiplicative competitive interactive models. *Research in Marketing, 2*, 217–235.

Klose, A., & Drexl, A. (2005). Facility location models for distribution system design. *European Journal of Operational Research, 162*(1), 4–29.

Kuhfeld, W. F. (2010). *Multinomial Logit, Discrete Choice Modeling - An Introduction to Designing Choice Experiments, and Collecting, Processing, and Analyzing Choice Data with SAS*. SAS Institute Inc.

Marianov, V., Rios, M., & Icaza, M. J. (2008). Facility location for market capture when users rank facilities by shorter travel and waiting times. *European Journal of Operational Research, 191*(1), 32–44.

McFadden, D. (1974). Conditional logit analysis of quantitative choice behavior. In P. Zarembkar (Ed.), *Frontiers in Economics*. New York: Academic Press.

McGuirk, M. A., & Porell, F. W. (1984). Spatial patterns of hospital utilization: The impact of distance and time. *Inquiry, 21*(1), 84–95.

Nakanishi, M., & Cooper, L. G. (1974). Parameter estimation for a multiplicative competitive interaction model: Least squares approach. *Journal of Marketing Research, 11*(3), 303–311.

ReVelle, C. S., & Eiselt, H. A. (2005). Location analysis: A synthesis and survey. *European Journal of Operational Research, 165*(1), 1–19.

Roh, C. Y., Lee, K. H., & Fottler, M. (2008). Determinants of hospital choice of rural hospital patients: The impact of networks, service scopes, and market competition. *Journal of Medical System, 32*(2), 343–352.

Severin, V., Louviere, J. J., & Finn, A. (2001). The stability of retail shopping choices over time and across countries. *Journal of Retailing, 77*(2), 185–202.

Sivey, P. (2012). The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics, 21*(4), 444–456.

Knowledge Source (2010). Clinical laboratory testing market overview. Online, http://knowsource.ecnext.com/coms2/summary_0233-4761_ITM (URL last accessed on July 31, 2015.).

Tai, W. T., Porell, F. W., & Adams, E. K. (2004). Hospital choice of rural medicare beneficiaries: Patient, hospital attributes, and the patient-physician relationship. *Health Services Research, 39*, 1903–1922.

Verter, V., & Lapierre, S. D. (2002). Location of preventive health care facilities. *Annals of Operations Research, 110*, 123–132.

Wang, Q., Batta, R., & Rump, C. M. (2002). Algorithms for a facility location problem with stochastic customer demand and immobile servers. *Annals of Operations Research, 111*, 17–34.

Zhang, Y., Berman, O., Marcotte, P., & Verter, V. (2010). A bilevel model for preventive healthcare facility network design with congestion. *IIE Transactions, 42*(12), 865–880.

Zhang, Y., Berman, O., & Verter, V. (2009). Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research, 198*(3), 922–935.

Zhang, Y., Berman, O., & Verter, V. (2012). The impact of client choice on preventive healthcare facility network design. *OR Spectrum, 34*, 349–370.