Data Analysis for Healthcare: A Case Study in Blood Donation Center Analysis

Michael Chau
School of Business
The University of Hong Kong
mchau@business.hku.hk

Cheuk-Kwong Lee
Hong Kong Red Cross
Blood Transfusion Service
ckleea@ha.org.hk

Chi Wai Chan
School of Business
The University of Hong Kong
ktchan@gmail.com

Eddie Cheng
School of Business
The University of Hong Kong
hkueddie@graduate.hku.hk

ABSTRACT

With increasing emphasis on using data analysis in decision support in healthcare organizations, we describe an example in collecting and analyzing data for a health organization. In particular, we describe the process in identifying the best location for Hong Kong Red Cross Blood Transfusion Service to establish a new blood donation center based on past donation data. Our analysis is divided into two parts. The first part focuses on data analysis which studies the context of data itself. We have identified the relationship among donor's age, gender, blood type and frequency of donation, the relationship between donation frequencies and distance from the donation center, the characteristic of different donation center, and the trend of blood donation. The second part focuses on new center analysis which applies the findings derived from the first part in identifying the best location for establishing a new blood donation center.

Keywords

Healthcare, geographic analysis, data analysis, blood donation

1. INTRODUCTION

There is increasing emphasis in data analysis in healthcare organizations, which involves using data and the findings from data analysis to inform and support decision making processes. Due to the large amount of data, data management and analysis techniques are often useful for identifying results that reduce subjective biases and reveal patterns that may be previously unknown. Many healthcare organizations have used patient record management systems to support decision making (Heathfield & Louw, 1999; Abidi, 2001; Chen et al., 2003). The purpose of this paper is to present our data analysis in exploring information value in the decision support process of a healthcare organization. In particular, we present a case study in using data analysis to support establishing a new blood donation center for *Hong Kong Red Cross Blood Transfusion Service* (BTS) by analyzing donors' past donation records. We identify the relationship between donors' demographic information and their activeness in blood donation, the relationship of geographic distribution of blood donation centers and the frequencies of blood donations, and the characteristic of each blood donation center. Based on these findings, we identify the best location for a new blood donation center which can increase the usage of the new center.

There are a number of challenges in this data analysis for the BTS. First, although there is clear relationship between the geographic distribution of blood donation center and the likelihood of making blood donation, the degree of correlation is difficult to quantify. Second, there are many noises in the data which will affect the accuracy of the result. Third, fine-tuning the suggestion for best location is very challenging because of limited data available.

The rest of the paper is organized as follows. In the Section 2 we provide a brief review of some related work in data analysis for health organizations. Section 3 describes the data collected by the BTS which was used in our study. Section 4 presents the analysis methods and results. Lastly, we sum up the paper in Section 5 and suggest some future research directions.

2. LITERATURE REVIEW

It is very important for modern medical procedures to have sufficient supply of blood. In order to have a reliable supply of blood, increasing the number of first time donors and keeping them as repeat donors in subsequent time are critical. There are many studies that discuss influential factors in making blood donations, and we only focus on spatial location in this paper as we aim to identify the best location in our study.

Saberton (2009) investigated a large number of donors at the census tract level to determine the factors that influence the number of donors, such as proportion of young residents, proportion of immigrants, education level, and accessibility. Through statistical analysis of donor data, it was found that the results were consistent across geographical regions and urban sizes. There are also studies on the donation pattern of the Chinese first-time donors, which identified factors affecting donors' pattern in order to help develop effective marketing strategies (Yu et al., 2007; Lee et al., 2008). They presented a statistical model to identify those donors with potential to become committed donors and those committed donors who can donate more frequently.

The decision on the city or other area in which to locate a healthcare facility must be based on such factors as potential, growth, and competition (Winter 1985). Before the site can be selected, one must consider criteria for location. In retail decisions it is assumed that the closer a store is to consumers, the greater the chance of successful marketing to them (Nelson 1958). Therefore it would make sense for the healthcare facility to be located close to the users, especially those who potentially would make heavy use of the particular facility being proposed. Moreover, it would be logical to locate a facility close to those target customers who are far from other similar facilities.

Cowper (1984) described a framework for analyzing the primary healthcare-seeking patterns of consumers for primary care providers in rural northern California. A cost-efficient principle of spatial interaction (Smith 1978) was combined with consumer search to provide an appealing theoretical rationale for the spatial interaction model. The model predicts the proportion of visits made from each origin captured by each destination or care site on the basis of (a) the characteristics of providers and facilities at the care site in question and at competing care sites and (b) the travel time between the origin and alternative sites in the market.

However, there is no existing paper discussing the correlation between spatial distance and the incentive for blood donors. In this paper, we describe the process in identifying the best location for the BTS to establish a new blood donation center to increase the incentive for donors in making blood donations.

3. DATA

The dataset provided by the BTS contains 775,690 records for blood donation which included information about the ID of donors which is the unique identifier for each donor, their demographic information, information about each blood collection, and the information about the nature of sites for making donations. The characteristics of data are summarized in the following table.

Items	*Description*
Dataset	There are totally 775,690 data from the dataset.
ID	All donors have their own unique ID.
Gender	Binary option of Male or Female.
Date of Birth	The birth day of donors.
Data Collection Date	The blood collection date.
Address	The residential address of the donor
Site and Venue	The site of donation.
Blood Type	Eight distinctive blood types: A, B, AB, O (Positive and Negative).

Table 1. Characteristics of the Dataset

In our data analysis, we have taken some extra environmental factors into consideration, such as the latitude and longitude of the address and sites according to the global positioning system from Google Map, and the population of the districts from the Census and Statistics Department of the Hong Kong government. The purpose of this is to find out the relationship between geographical distribution of the blood donation centers, population, and the activeness of blood donation. We also

group the addresses into 18 districts as defined by the Hong Kong government (see Figure 1). We further group the 18 districts into 5 metro-districts according to their corresponding blood donation ratios in the 5 centers.

1	Islands	10	Kowloon City
2	Kwai Tsing	11	Kwun Tong
3	North	12	Sham Shui Po
4	Sai Kung	13	Wong Tai Sin
5	Sha Tin	14	Yau Tsim Mong
6	Tai Po	15	Central and Western
7	Tsuen Wan	16	Wan Chai
8	Tuen Mun	17	Eastern
9	Yuen Long	18	Southern

Center	Center Location	Mainly Serve
Tsuen Wan	7	2,7,8,9
Sha Tin	5	5,6,3
Kwun Tong*	11	4,13,11
Yau Ma Tei	14	10,12,13,14
Central and Western	15	15
Wan Chai	18	16,17,18

Figure 1. Geographic Information of Hong Kong and Blood Donation Centers

4. ANALYSIS

We first studied the context of data. We identified the relationship among donors' age, gender, blood type and frequency of donations, the characteristics of different donation centers, the trend of blood donations from year 2004 to 2007, and the relationship between donation activeness and distance from the donation center.

Relationship between Age, Gender, and Frequency of Blood Donations

We aim to find out the average frequency of donations, the relationship between age and frequency of donation, the relationship between gender and frequency of donation, and the integrated relationship between age, gender and frequency of donation. By categorizing all records based on the 18 districts, the average of total blood donation per total number of donors is 2.23, this means each donor makes blood donation 2.23 times in average in the past four years. For restricting the venue of blood donation to donation centers only, the number changes to 2.79. This means people in the centers make blood donation at higher frequency compared with the average. There is a 19.44% increase in blood donation frequency in the centers when compared to other mobile donation units.

From Figure 2, we can see the relationship between age and frequency of donation. The curve peaks at age 19, and the number drops sharply from age 19 to age 23, followed by steady decrease beyond age 23.

Out of the 775,690 times of blood donations, there are 415,246 blood donations made by male, and 360,444 made by female. In other words, male shares 54% of total blood donation.

From Figure 3, which combined the data of age and gender, we can observe that young female (age 16-25) make more blood donations than young male. However, male donates more frequently than female starting from age 26 and the difference keeps increasing from age 26 to age 42, and afterward it goes downs steadily.

Figure 2. Age of Donation

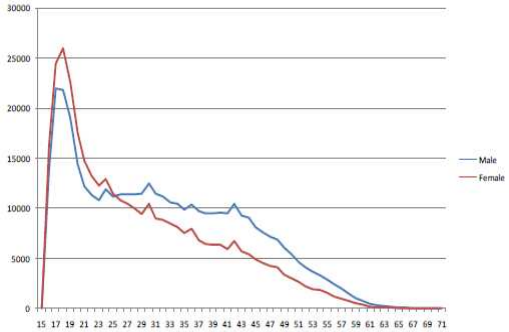

Figure 3. Gender and Age of Donation

Trend of Blood Donations

We aim to find out the changes of number of blood donation from year to year, and estimate the growth of the blood donation in the coming years. Therefore, the data derived from this trend can benefit a more accurate forecast of the number of donations in different districts in the future, and this leads to better planning for the new blood donation center.

For the total donation, from year 2004 to 2006, blood donation frequency is slowing down but there is a sharp increase in year 2007. There is no pattern that we can find from the data because of the short period of data provided in the dataset. For the changes in the number of total donors in the same period of time, the number is unstable every year and we cannot find any pattern either. Both the donation frequency and the total number of donors sharply increase in year 2007. We suggest that this sharp increase is contributed by the reallocation of the Mongkok Blood donation center.

While categorizing the donation into the 18 districts, we can derive more meaningful information for our new center analysis. For all blood donations, we find that the North district has the highest absolute increase and percentage increase in the 4 year period. Sham Shui Po and Sai Kung are in the second place and third place, respectively. Table 2 shows a summary of these data.

	Total Frequency of Donations				Total Frequency at Donations Made at Centers			
District	**2004**	**2007**	**Growth Rate**	**Growth Number**	**2004**	**2007**	**Growth Rate**	**Growth Number**
Central and Western	7394	7373	-0.28%	-21	4712	4954	5.14%	242
Eastern	16909	16878	-0.18%	-31	10456	10890	4.15%	434
Islands	2897	3154	8.87%	257	1722	1913	11.09%	191
Kowloon City	8442	8601	1.88%	159	4867	5235	7.56%	368
Kwai Tsing	15179	15425	1.62%	246	9558	10848	13.50%	**1290**
Kwun Tong	15613	16370	4.85%	757	9071	10242	12.91%	**1171**
North	7781	8819	**13.34%**	**1038**	3765	4698	**24.78%**	933
Sai Kung	11275	12290	**9.00%**	**1015**	6039	7002	15.95%	963
Sha Tin	20187	20119	-0.34%	-68	12048	12937	7.38%	889
Sham Shui Po	9065	10040	**10.76%**	**975**	5788	6787	**17.26%**	999
Southern	7151	7510	5.02%	359	4546	5067	11.46%	521
Tai Po	8582	9088	5.90%	506	4196	4868	16.02%	672
Tsuen Wan	8798	9396	6.80%	598	6153	6943	12.84%	790
Tuen Mun	15973	16237	1.65%	264	8534	9467	10.93%	933
Wan Chai	3806	3784	-0.58%	-22	2635	2647	0.46%	12
Wong Tai Sin	12834	13378	4.24%	544	7392	8711	**17.84%**	**1319**
Yau Tsim Mong	6962	7306	4.94%	344	4239	4660	9.93%	421
Yuen Long	13513	14396	6.53%	883	6711	7377	9.92%	666

Table 2: Summary of Growth Rate of Blood Donation for the 18 Districts

Geographic Distribution of Center and Blood Donation Frequency

This part focuses on site of donation and we aim to find out the unique characteristics of each blood donation center in terms of the attractiveness to blood donors. There are five blood donation centers in our dataset, namely Mongkok Center, Tsuen Wan Center, Causeway Bay Center, Shatin Center, and Central Center, listed in decreasing order of total blood donation. Table 3 summarizes the percentage of blood donation frequency of different residential address to its highest frequency blood donation center. Most of the following data analysis is based on this grouping which represents five metro-districts in Hong Kong.

District id	District	Total Frequency in District	Frequency in Nearest Center	Frequency in Nearest Center/ Total Frequency in District (%)
Mongkok Center				
11	Kwun Tong	37628	**23271**	61.84%
13	Wong Tai Sin	31603	**21379**	67.65%
12	Sham Shui Po	24661	**15753**	63.88%
4	Sai Kung	25299	**13735**	54.29%
10	Kowloon City	19898	**13292**	66.80%
14	Yau Tsim Mong	17112	**12666**	74.02%
Tsuen Wan Center				
2	Kwai Tsing	40267	**24953**	61.97%
7	Tsuen Wan	25612	**19753**	77.12%
8	Tuen Mun	35550	**19736**	55.52%
9	Yuen Long	27836	**13688**	49.17%
Causeway Bay Center				
16	Eastern	41820	**27963**	66.87%
17	Southern	18630	**12676**	68.04%
18	Wan Chai	10381	**7469**	71.95%
Shatin Center				
5	Sha Tin	49104	**29252**	59.57%
3	North	16651	**6453**	38.75%
6	Tai Po	17613	**6238**	35.42%
Central Center				
15	Central and Western	19164	**11089**	57.86%
1	Island	7292	**2492**	34.17%

Table 3. Summary of Blood Donation Frequency for different metro-districts in Hong Kong

In this section, we aim to find out the correlation between the numbers of blood donation to the distance from donor's residential address to the site of donation center. After that, we aim to identify the best location to establish a new center which gives maximum increase in blood donation. The process of the data analysis for studying the Geographic Distribution of Center and Blood Donation Frequencies as follows:

1. Plot the locus defined by Google Map onto the map of Hong Kong. Group all the donors and their blood donations into five metro-districts according to Table 3.

2. Select donations made in center only, because only these are relevant to our problem. We also remove cross metro-district donations, in which donors make blood donation not in their residential metro-district. Cross metro-district donations are irrelevant to the building of new center, since they make donation independent to distance between residential address and site of donation. The total donation made outside of donors' home metro-district is 39.53% out of all donations made at the centers.

3. For each metro-district, we calculate the distance between donors' residential address to blood donation center and count the number of donation for each location. We then can plot the regression line with distance on x axis, and numbers of blood donation on y axis. We find the slope of the regression line which indicates the correlation between Geographic Distribution of Center and Blood Donation Frequencies (Figure 4).

4. For abnormal high numbers of blood donation which indicate possible sites for new blood donation center (As blue circle in Figure 4), we hypothetically put a new center there and study the expected number of blood donation increased according to the derived correlation between Geographic Distribution of Center and Blood Donation Frequencies for that particular metro-district.

5. Compare all possible sites for all five metro-districts, followed by adjusting the figure with expected growth rate, and then we rank the selected sites in term of expected increased numbers of blood donation in year 2010.

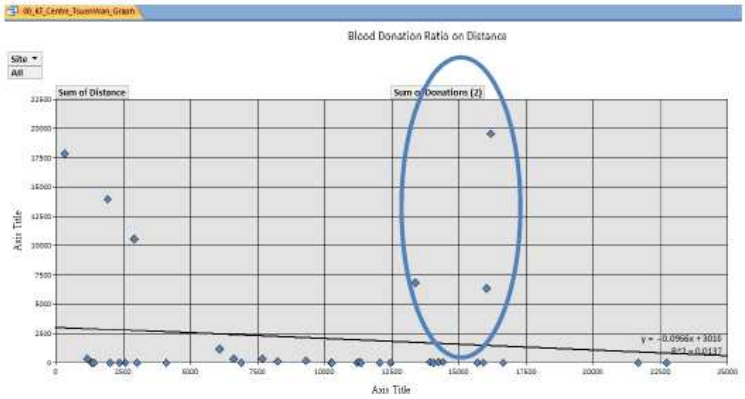

Figure 4. Correlation Analysis

From the data derived at the district level, Yuen Long, Tin Shui Wai and Tuen Mun (TYT) are the best locations for building a new center with an average increase of 14,279 donations, while Tseng Kwan O is the second best choice with an average increase of 13,268 donations. The result is summarized in Table 4.

	Causeway Bay	Central	Shatin	Tsuen Wan	Mongkok
Metro Donations	48,108	13,581	41,943	78,130	100,096
Metro Population	1,016,800	386,700	1,180,800	1,844,800	2,424,900
Ratio (per 4 Yr)	4.731%	3.512%	3.552%	4.235%	4.128%
Ratio (per Yr)	1.183%	0.878%	0.888%	1.059%	1.032%
Location	***Chai Wan***	***Tung Chung***	***Fan Ling***	***TYT***	***Tseung Kwan O***
Total Increase	18,851	7,884	32,481	57,117	53,071
Yearly Average	4,713	1,971	8,120	14,279	13,268
Growth Rate	1.11%	2.29%	4.04%	3.72%	6.03%
Annualized Growth Rate	0.37%	0.76%	1.33%	1.22%	1.97%
Expected Increase	**4,765**	**2,016**	**8,448**	**14,810**	**14,068**

Table 4. Summarized Data for Establishing New Blood Donation Center

5. CONCLUSION

Using data analysis methods, we have identified the relationship between donors' age and gender and their frequency of donations, the relationship between donation frequencies and distance from the donation center, the characteristics of different donation centers, and the trend of blood donation. In addition, using a spatial analysis method, we identify the best location for establishing a new blood donation center.

Through this case study, we show an example of using data analysis for decision support in a healthcare organization. We hope the study can encourage other similar studies in data analysis for healthcare organizations. In our future work, we will use more data analysis and data mining techniques on the data as well as use our analysis methods for other organizations.

ACKNOWLEDGMENT

We thank the Hong Kong Red Cross Blood Transfusion Service for providing the data used in this study, and Alice Lee of the University of Hong Kong for her help in editing the paper.

REFERENCES

1. Abidi, S. S. R. (2001). "Knowledge Management in Healthcare: Towards 'Knowledge-driven' Decision-support Services," *International Journal of Medical Informatics* 63, 5-18.
2. Chen, H., Lally, A. M., Zhu, B., and Chau, M. (2003). "HelpfulMed: Intelligent Searching for Medical Information over the Internet," *Journal of the American Society for Information Science and Technology* 54(7), 683-694.
3. Cowper, P. A. (1984) *A Spatial Analysis of Primary Health Care Markets in Rural Areas*, Ph.D. thesis, University of California, Davis.
4. Heathfield, H. and Louw, G. (1999). "New Challenges for Clinical Informatics: Knowledge Management Tools," *Health Informatics Journal*, 5(2), 67-73.
5. Lee, C. K., Hong, J., and Hung, A. T. F. (2008). "An Update of Blood Donor Recruitment and Retention in Hong Kong," *Asian Journal Transfusion Science* 2(2): 47–50.
6. Nelson, R. (1958). *The Selection of Retail Locations*, New York. F W Dodge.
7. Saberton, P. J., Paez, A., Newbold, K. B., and Heddle, N. M. (2009). "Geographical Variations in the Correlates of Blood Donor Turnout Rates: An Investigation of Canadian Metropolitan Areas," *International Journal of Health Geographics,* 2009.
8. Smith, T. E. (1978). "A Cost-Efficiency Principle of Spatial Interaction Behavior," *Regional Science and Urban Economics 8*, 313-37.
9. Winter, F. W. (1985). "Location Decisions Made Easier," *Medical Group Management,* 32 (March-April), 38-47.
10. Yu, P. L. H., and Chung, K. H., Lin, C. K., Chan, J. S. K., and Lee, C. K. (2007). "Predicting Potential Drop-out and Future Commitment for First-time Donors Based on First 1.5-year Donation Patterns: The Case in Hong Kong Chinese Donors," *Vox Sanguinis* 93, pp. 57–63.