Working Paper 2015-08

# Patient's Choice of Hospital in Korean Inpatient Care Market

Daejung Kim

Korea Institute for Health and Social Affairs

Patient's Choice of Hospital in
Korean Inpatient Care Market

Daejung Kim, Associate Research Fellow



Korea Institute for Health and Social Affairs
Building D, 370 Sicheong-daero, Sejong city
30147 KOREA
http://www.kihasa.re.kr
ISBN: 978-89-6827-306-3 93510

# Contents

## List of Tables



## List of Figures

# 1 Introduction

# 1 Introduction

The supply of healthcare service in Korea relies on market mechanism given the fact that healthcare service is delivered mostly by private sector (in terms of beds, about 90% are supplied by private hospitals) and patients are free to choose hospitals without gate keeping. This engenders various consequences; the concentration of patients in metropolitan hospitals such as Big 5 hospitals in Seoul and medical arms race in hospital industry and financial difficulties for the rural hospitals, etc. This report will address the hospital choice in Korean market to analyze factors affecting patients choice and competition landscape.

A hospital could be regarded as a multi-product firm that provides a diverse range of services, from prescription and treatment to surgery and hospitalization, depending on the patient's sickness. The patient is the end user of the services provided by the hospital and selects the hospital based on subjective or objective factors. Subjective factors may include the independent perception of the hospital services, and objective factors may include the time taken to reach the hospital or distance to the hospital, as well as facilities, equipment, and human resources owned by the hospital.

However, the selection of medical service providers is unique compared to the selection of normal goods, in terms of the asymmetry of information. Normally, the patient may be able to evaluate factors that can be seen externally, such as the friendliness of the hospital staff, but it is difficult for the patient to accurately evaluate core factors such as the quality of medical technology because of asymmetry of information. As such, issues like supplier-induced demand may occur. Although the knowledge levels of patients have significantly improved recently, the characteristics of medical goods as experience goods or credence goods have not changed much.

This study primarily focuses on the estimation of factors affecting the choice of hospitals by the patient, based on Patient Survey data. Patient Survey data are collected by the Ministry of Health and Welfare and the Korea Institute for Health and Social Affairs since 1988, and it seeks to identify the significance placed on various factors governing the medical services demand structure in Korea as well as the competitive structure of Korean hospitals. A significant number of studies assert that key factors impacting hospital choice, including the quality of hospital services, the patient's place of residence, distance to the hospital, and the prices for the services, differ depending on the illness (e.g., Morill, Earickson and Ries 1970). This study targets the patient with cancer, the primary diagnosis code being ICD-10.

# 2 Previous literature on hospital choice models

# 2 Previous literature on hospital choice models

This study estimates the preference of patients by examining data on their hospital choice. There are very few literature which studies on the Korean healthcare market.  Kim.B (1990) estimated the logit model using a Korean province data. Several studies using non-Korean data have applied McFadden's (1973) findings on multinomial logit models, which can be estimated either by the maximum likelihood method or the minimum chi-square method, and the estimation models differ depending on the data format.

Folland (1983), Lee and Cohen (1985), and McGuirk and Porell (1984) used the minimum chi-square type method developed by Theil (1969) to derive their estimation results.

Folland (1983) applied independent variable data from surveys of non-profit hospitals in South Dakota from 1977 to the multi-attribute model to develop a market share model explaining inter-region patient flows. The results indicated that patients selected their hospitals based on their own individual characteristics, their physicians, as well as the characteristics of utilizable hospitals.

Lee and Cohen (1985) sought to develop a model of regional hospital utilization. They referred to the 1) discharged patient data of the PAS, 2) the hospital facilities database, 3) items from physician surveys conducted by the Ministry of Health, and 4) travel time data extracted from studies conducted by the Department of Planning for Rhode Island. Using the multinomial logit model, which allows easier calculations, they showed that travel time (accessibility) and the physician were the greatest influences in patients' choices, patients tended to choose hospitals with a wider range of services over those within the region, and the multinomial logit model explained patients' hospital choice patterns very well.

McGuirk and Porell (1984) sought to develop a spatial demand model of hospital choice by measuring the effects of time and distance on hospital utilization. They analyzed data on 184 residents of short-term acute care hospitals in Allegheny, Pennsylvania, from 1975 using the log-linear cross-product ratio estimation approach. The results indicated that distance (accessibility) and time factors strongly influenced hospital choice, and they differed by the hospital and the type of service provided.

Adams et al. (1991), Burns and Wholey (1992), Dranove, White, and Wu (1993), Garnick et al. (1989), Luft et al. (1990), and Phibbs et al. (1993) used the maximum likelihood method.

Adams et al. (1991) sought to study the hospital choices of rural Medicare beneficiaries. They used discharged Medicare patient records (age, DRG type, and address) from 1986, extracted from the MEDPAR file, and hospital characteristics (number of beds and scale of services). These data were analyzed using McFadden's (1973) conditional logit model. They found that Medicare beneficiaries tended to lean toward hospitals that provide a wide range of services and conduct research activities, and as the patient became older, the distance factor increasingly impeded hospital choice.

Burns and Wholey (1992) examined the effects of hospital choice factors and employed data on discharged patients (composed of physicians, hospitals, and patient characteristics) from the Arizona Department of Health Services, Phoenix, from 1989. They used the conditional logit model and found that the physician's disposition heavily impacted hospital choice, and that while quality of treatment and price differentials were influential, they could only explain relatively lower variations thereof.

Dranove, White, and Wu (1993) examined the determinants of hospital choice. They segmented the hospital market and conducted their study using the California OSHPD patient data from 1983 and 1989. The results of the study indicated that Medicaid beneficiaries had a tendency to select hospitals with lower prices and fewer services, privately insured patients selected hospitals with a wider range of services irrespective of the price, and Hispanic patients selected hospitals with lower prices. With regard to the distance, patients tended to select closer hospitals in general, and the market was segmented by race (Hispanics and Caucasians) and not by the clinical state of the patient.

Garnick et al. (1989) developed a measure that includes the type of the patient and the entire market area. They used discharged patient data of patients with heart catheters from the California Department of Health from 1983 and selected the non-linear choice model to conduct their investigation. The study estimated the conditional choice model with two techniques used previously to transform the non-linear choice model, and found that when there are many combinations of hospitals and locations with zero patients, the linear estimations were much more sensitive to the proportion of zeroes. Thus, they showed that the maximum likelihood estimation is appropriate when there are many zeroes.

Luft et al. (1990) measured the influence of treatment quality, price, ownership, and distance on hospital choice, and in doing so, used patient data from California from 1983. The study employed McFadden's conditional logit model and concluded that treatment quality had a significant influence on hospital choice.

Phibbs et al. (1993) evaluated whether there were differences between subgroups (high-risk/low-risk and private insurance/Medicare) regarding factors influencing hospital choice. The study used discharged patient information of the OSHPD of California from 1985 as well as the financial information of the hospital, and applied McFadden's conditional logit model. The study concluded that high-risk women considered quality of treatment more important in hospital choice compared to low-risk women, and beneficiaries of private insurance, compared to beneficiaries of Medicare, tended to select hospitals with a higher percentage of caesarean deliveries, on-site normal spontaneous vaginal delivery centers, hospitals that were harder to access, and private hospitals.

# 3 Data

A. Institutional Data

B. Discharged patient data

# 3 Data

The data used in this study is Korean Patient Survey data in 2010. Patient Survey is conducted at all medical institutions in the country for certain periods of time (a month) to identify trends in patient illnesses as well as the status of medical system utilization, health and medical facilities, and human resources. Doing so provides basic data for health and medical policies and statistics.[1] Patient Survey data from 2010 was sourced from 1,307 institutions in this analysis, and among the discharged patients utilizing these institutions, this study targeted 49,487 patients with cancer. “Patients with cancer” denotes patients with the core diagnosis code of ICD-10, which refers to neoplasms (C00~D48). The data used in the model were divided into institutional data, including the hospital characteristic variables, and the discharged patient data, including the patient characteristic variables, both of which are explained below.

1) Patient Survey in Korea was first conducted in 1953 as a non-scheduled census stemming from statistical research on disease/illness, and from 1988 onwards, it assumed the form of sample research, which was later recognized as the leading form of research pertaining to medical utilization in Korea. Until 2011, a total of 21 rounds of research had been conducted (Korea Institute for Health and Social Affairs, Patient Survey 2011, page 11).

## A. Institutional data

The institutional data for this research were sourced from a census or sampling of 9,259 medical institutions, that is, the total number of medical institutions in Korea. The census was conducted for general hospitals, hospitals, health centres, maternity centers, and clinic-level medical institutions. The details recorded in the institutional research include the name of the institution, location, type of registration, key medical devices, number of beds, number of personnel, and number of patients. The sample collection was web-based, wherein the head or personnel of the medical institution logged onto the Patient Survey management system and completed the data.

In order to determine the hospital choice model, the targets of this study were tertiary hospitals,[2] general hospitals,[3] and hospitals.[4] We sorted out biggest 5 hospitals(Big 5) from the

2) Tertiary hospitals are hospitals that conduct complicated treatments for serious illnesses. They are mandated to treat more than 20 medical specialties and must employ a specialist for each treatment specialty. Moreover, they must meet various standards set by the Ministry of Health and Welfare.

3) General hospitals with 100 to 300 beds must practice three specialties from among internal medicine, surgery, paediatrics, and obstetrics and gynaecology. In all, they must practice a total of at least seven specialties or more, including radiology, anaesthetics, laboratory medicine, and pathology, and they must employ a specialist for each treatment specialty. General hospitals with more than 300 beds must offer services in internal medicine, surgery, paediatrics, obstetrics and gynaecology, radiology, anaesthetics, mental health, and dentistry. The hospital must also practice either laboratory medicine or pathology, thus providing a total of nine treatment specialties, with a specialist for each one.

4) The term hospitals refers to medical institutions servicing patients requiring

tertiary hospitals in estimation to look more closely on them. Long-term care hospitals, dental hospitals, oriental medicine hospitals, health clinics, health centres, clinics, dental clinics, oriental medicine clinics, sub-health centres, primary health care posts, maternity centers, and mother and child health (MCH) centres were excluded from the study. Additionally, military hospitals, mental hospitals, rehabilitation hospitals, tuberculosis hospitals, long-term care hospitals for the aged with Alzheimer's disease, and leprosy hospitals were also excluded from the survey. Ultimately, 1,307 institutions were selected for research, including 44 tertiary hospitals, 254 general hospitals, and 1,009 hospitals. Regarding their classification by private or public sector, there were 19 public sector national hospitals, 17 public sector special hospitals, 41 public sector hospitals, 7 public sector miscellaneous hospitals, 469 private sector enterprises, and 754 private sector hospitals.

The number of beds for the institutions averaged at 200, with a standard deviation of 228. The number of beds by institution type averaged at 990 for tertiary hospitals, 362 for general hospitals, and 136 for hospitals. As factors relating to supply-induced demand may affect the patient's hospital choice process, this relationship can be identified by examining the availability of expensive medical devices, number of beds in the hospital

---

hospitalization and having more than 30 beds. They include hospitals, dental hospitals, oriental medicine hospitals, long-term care hospitals, mental hospitals, geriatric hospitals, and general hospitals.

s,[5] and the patient's choices.

The status of medical device ownership in the target institutions included the identification of computed tomography (CT) machines, magnetic resonance imaging (MRI) machines, and radiology devices. The results of the distribution of these key medical devices showed that 64.4% of the total of 1,307 hospitals selected for this study had more than one CT machine, 42.5% had more than one MRI machine, and 18.8% had more than one radiology device. The highest number of CT machines for a hospital was 10, and the corresponding number for MRI machines and radiology devices was 8 and 24 respectively.

〈Table 1〉 Number of key medical devices

| | CT machines | MRI machines | Radiology devices |
|---|---|---|---|
| 0 | 466<br>(35.6%) | 752<br>(57.5%) | 1,069<br>(81.7%) |
| 1 | 722 | 451 | 144 |
| 2 | 72 | 81 | 47 |
| 3 | 24 | 15 | 13 |
| 4 | 13 | 3 | 16 |
| 5 | 4 | 2 | 5 |
| 6 | 1 | 1 | 6 |
| 7 | 2 | 1 | 2 |

5) The relationship between the number of beds and the hospital utilization rate of the patient is often referred to as the medical supplier-induced demand or moral hazard. The Dartmouth Atlas of Healthcare in the United States indicates that the supply of beds was not uniformly influential in clinical decisions such as bed utilization. This research indicates that hospitalizations for hip fractures were almost irrelevant to the supply of beds ($R$-square = 0.08), but when the diagnosis pertained to surgery, the correlation was higher ($R$-square = 0.22). For general medical use, the number of beds explained the rate of hospitalization for 56% of cases, and the correlation thereof was very high ($R$-square = 0.56).

| | CT machines | MRI machines | Radiology devices |
|---|---|---|---|
| 8 | 1 | 1 | 1 |
| 9 | 0 | 0 | 1 |
| 10 | 1 | 0 | 1 |
| 13 | 0 | 0 | 1 |
| 24 | 0 | 0 | 1 |
| Total | 1048 | 701 | 466 |

The status of medical personnel at the target institutions can be examined by analyzing the number of physicians and nurses. The average values of the number of physicians and nurses were 33 (standard deviation = 104) and 73 (standard deviation = 186) respectively; however, the corresponding statistics by hospital type show that, on average, tertiary hospitals had 475.5 physicians and 793.0 nurses, general hospitals had 62.2 doctors and 164.3 nurses, while standard hospitals had 6.7 doctors and 19.8 nurses. The distribution of physicians by type can be found in Figure 2-5.

〈Table 2〉 Distribution of medical personnel by type

| Type | | Count | Average | Std. Dev. | Min. Value | Max. Value |
|---|---|---|---|---|---|---|
| Number of beds | Tertiary hospital | 44 | 990 | | | |
| | General hospital | 254 | 362 | | | |
| | Hospital | 1009 | 136 | | | |
| Number of physicians | Tertiary hospital | 44 | 475.55 | 277.1862 | 169 | 1413 |
| | General hospital | 254 | 62.24 | 71.27531 | 0 | 398 |
| | Hospital | 1009 | 6.73 | 5.256185 | 0 | 55 |
| Number of nurses | Tertiary hospital | 44 | 793.05 | 536.8864 | 312 | 2991 |
| | General hospital | 254 | 164.33 | 138.5075 | 2 | 742 |
| | Hospital | 1009 | 19.79 | 21.13032 | 0 | 170 |

## B. Discharged patient data

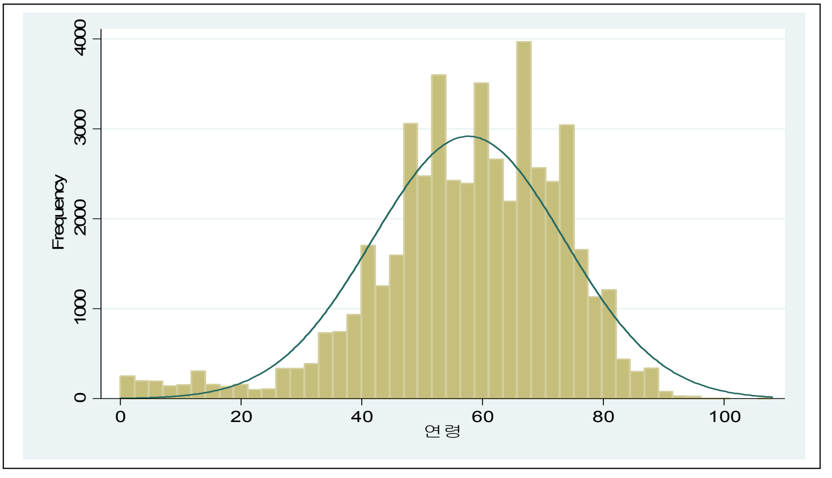
The patient discharge data were collected for a period of one month (31 days). The patient discharge sheet includes details about the treatment area, gender, age, address, diagnosis, reasons for the illness, name of the surgery, date of hospitalization, date of discharge, results of the treatment, discharge type, reasons for hospitalization, admission channel, and payment method for medical fees. Moreover, the distance and time taken to reach the hospital by patients in metropolitan areas, namely, the Gun and Gu levels, were investigated using Daum maps (http://map.daum.net/). If the addresses of the patient and the hospital were located within the same metropolitan area, the Gun and Gu levels as well as the time and distance were set to zero. This is an assumption necessitated by the limitations of the data, because ascertaining the exact addresses of the patients and hospitals was difficult.

The descriptive statistics show 52.9% and 47.1% of the total targeted discharge patients were men and women respectively. Of the total, 19.5% of patients were admitted after being referred by other institutions, and 81.5% were self-admitted. The average patient was 57.6 years old, the standard deviation being 15.88 years. The number of patients in the age groups below 30, 30~45, 45~65, and over 65 years was 2,587, 6,580, 23,099, and 17,221 respectively. The length of stay (LOS) averaged 10.28 days. The most popular method of payment was National Health Insurance (89.9%), followed by medical aids (7%), and out of pocket (1.8%). The time and distance from the hospital to the patient's place of residence averaged 74.3 minutes and 86.4 km, respectively. Interestingly, the distribution after 100 minutes shows a uniform trend.

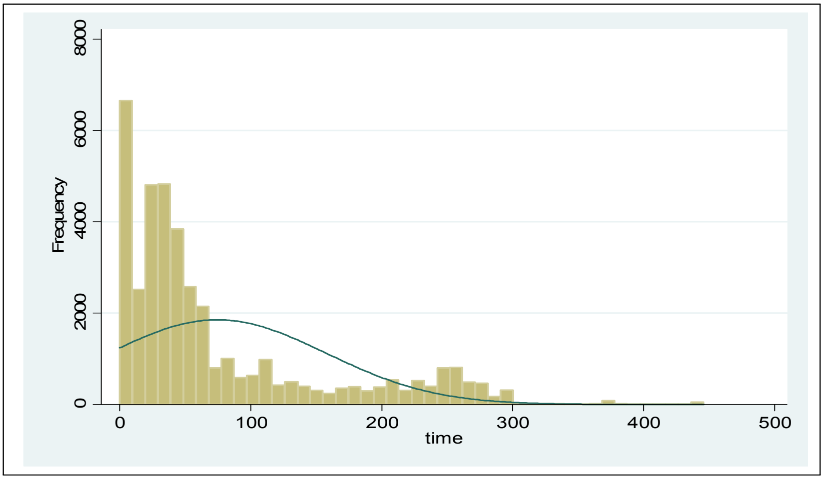
〈Table 3〉 Descriptive statistics of the discharged cancer patients sampled in this research

| Name of variable | Explanation | Average (standard deviation) |
|---|---|---|
| Male | Male = 1, Female = 0 | .5291447 (.49915) |
| Reference | Reference from other institution = 1, self-admitted = 0 | .1952255 (.39638) |
| dv5 | Age | 57.67256 (15.88713) |
| Age | Age category<br>(below 30 = 1, 30~45 = 2, 45~65 = 3, over 65 = 4) | 1 = 2,587<br>2 = 6,580<br>3 = 23,099<br>4 = 17,221 |
| dv18 | Length of stay in days | 10.28 days (24.30529) |
| dv17 | Payment method | Out of pocket = 913 (1.84%)<br>National Health Insurance = 44,481<br>Employment insurance = 80<br>Automobile insurance = 53<br>Medical aids = 3,459<br>Other = 501 |
| Time | Time taken to travel from the patient's residence to the hospital | 74.36952 (83.04087) |
| Distance | Distance from the hospital to the patient's residence | 86.48192 (130.463) |

[Figure 1] Age distribution of sampled cancer patients




[Figure 2] Distribution of time taken to travel from the patient's residence to the hospital



# 4 Hospital choice model of the sampled patients

# 4 Hospital choice model of the sampled patients

The multinomial choice model can be used to study the preference of the chosen subject (patient) who is faced with more than two options. Leading examples of the multinomial choice model are the multinomial logit model, developed by Theil (1969, 1970) and the conditional logit model introduced by McFadden (1974). Even though the multinomial logit model and the conditional logit model use different data, the multinomial logit model is categorized as a special type of conditional logit model. This model has been employed multiple times in the Korea context; however, some relevant points are provided below.

If $U_{ij}^*$ can be regarded as the utility level of patient $i$ selecting hospital $j$, the choice made by patient $i$ to use hospital $j$ ($Y_{ij}=1$) means that $U_{ij}^* = Max\{U_{i1}^*, U_{i2}^*, \cdots U_{im}^*\}$. The utility of patient $i$ choosing hospital $j$ is assumed to be a linear combination of patient and hospital characteristics; in other words,

$$U_{ij}^* = z_{ij}'\beta + \epsilon_{ij} \qquad \text{(Equation 1)}$$

McFadden (1973) showed that if $\epsilon_{ij}$ is a type I extreme value distribution, as in

$$F(\epsilon_{ij}) = \exp(-e^{-\epsilon_{ij}}) \qquad \text{(Equation 2)}$$

the probability that patient $i$ will choose hospital $j$ could be calculated as follows.

$$P_{ij} = prob(Y_{ij} = 1) = \frac{e^{z_{ij}'\beta}}{\sum_{j=1}^{J} e^{z_{ij}'\beta}} \quad \text{(Equation 3)}$$

This model is called the conditional logit model, and the explanatory variable $z_{ij}$ is a choice-specific variable that differs by the choice made; the variable does not take on another value depending on the patient only. If the explanatory variable is an individual-specific variable that differs by patient, then Equation 2-3 becomes

$$P_{ij} = prob(Y_{ij} = 1) = \frac{e^{z_i'\beta}}{\sum_{j=1}^{J} e^{z_i'\beta}} \quad \text{(Equation 4)}$$

These models are called multinomial logit models.

Models of these formats provide the desired value by estimating the coefficient that maximizes the log likelihood function. In other words,

$$\max_{\beta} \log L = \sum_{i=1}^{n} \sum_{j=1}^{J} d_{ij} \log Prob(Y_i = j) \quad \text{(Equation 5)}$$

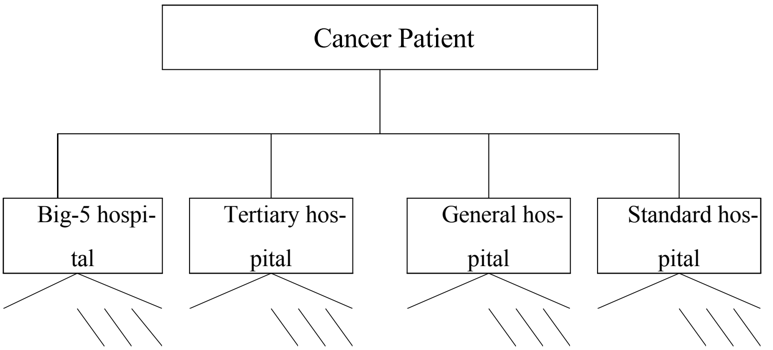
However, multinomial choice models, including conditional logit models, assume independence of irrelevant alternatives

(IIA). In other words, even if other choice alternatives are added, the odds ratio between the choices does not change. This is possible because of the assumption that the residual $\epsilon_{ij}$ is independent and homoscedastic. Hausman and McFadden (1984) developed an evaluation method to determine if the actual model satisfies IIA and suggested an appropriate statistical model for the test of IIA.

If the model fails the IIA test, the alternatives include the multinomial probit model and the nested logit model. The nested logit model can be used when the alternatives form a hierarchy.

If the hospital choice by the patient is assumed to take place in multiple stages, the nested logit model is the model of choice. Specifically, when the patient determines the hospit"al, he/she chooses the type of hospital to go to in the first stage (i.e., the standard, general, or tertiary hospital). Choosing the type of hospital is influenced by the patient's characteristics, such as his/her health status, gender, age, and area of residence, as well as the hospital's characteristics. In the second stage, the patient determines which hospital to go to in a segmented market. The evaluation criteria in this case would include factors associated with the hospital that would give the most satisfaction to the patient.

[Figure 3] Concept tree of the patient's hospital choice model




If a two-stage choice model is assumed to form a tree structure with sub-level alternatives as twigs and higher level alternatives as branches, the probability of selecting branch $l$ and twig $j$ would be

$$prob[twig_j, branch_l) = p_{jl} = \frac{e^{x'_{jl}\beta + z'_l\gamma}}{\sum_{l=1}^{L}\sum_{j=1}^{J} e^{x'_{jl}\beta + z'_l\gamma}} \qquad \text{(Equation 6)}$$

If this is reorganized,

$$p_{jl} = p_{j|l}p_l = \left(\frac{e^{x'_{jl}\beta}}{\sum_{j=1}^{J_l} e^{x'_{jl}\beta}}\right)\left(\frac{e^{z'_l\gamma}}{\sum_{l=1}^{L} e^{z'_l\gamma}}\right)\frac{\left(\sum_{j=1}^{J_l} e^{x'_{jl}\beta}\right)\left(\sum_{l=1}^{L} e^{z'_l\gamma}\right)}{\left(\sum_{l=1}^{L}\sum_{j=1}^{J} e^{x'_{jl}\beta + z'_l\gamma}\right)}$$

(Equation 7)

If the inclusive value of choosing branch $l$ is assumed to be

$$I_l = \ln \sum_{j=1}^{J_l} e^{x'_{jl}\beta}$$

(Equation 8)

then

$$p_{j|l} = \frac{e^{x'_{jl}\beta}}{\sum_{j=1}^{J_l} e^{x'_{jl}\beta}}, \quad p_l = \frac{e^{z'_l\gamma + \tau_l I_l}}{\sum_{l=1}^{L} e^{z'_l\gamma + \tau_l I_l}}$$

There are two methods of estimating this model. The first method is the full information maximum likelihood method, and the probabilities of selecting a branch and of selecting a twig from that branch is transformed into a combined probability. The log likelihood thereof and the coefficient that maximizes this value are calculated. In other words,

$$\max_{\beta,\gamma,\tau} \ln L = \sum_{i=1}^{n} \ln \left(p_{j|l} \times p_l\right)$$

The second method is a two-stage conditional logit estimation. First, the twig within the branch is selected using a simple conditional logit model, and the value for $\beta$ is estimated. Second, the inclusive value $I_l$ is calculated, and when a branch is chosen, the choice variables of $z_i$ and $I_i$ are used to estimate $\gamma$ and $\tau$. Normally, the first method (full information maximum likelihood estimation) is more effective; however, for studies such as ours where there are multiple alternatives, the two-stage estimation method is used. While the two-stage estimation is not as effective as the first method, the estimated values have been proven to match (McFadden 1984).

The following process was used to estimate the actual model. First, the higher-order choice unit of branches was set as the type of hospital. During the research period, the number of discharged patients using the big-5 hospitals, tertiary hospitals except the big-5 hospitals, general hospitals, and standard hospitals were 14,372, 16,984, 15,347, and 2,784 respectively.

〈Table 5〉 Number of patients in the sample by type of medical institution

| Type | Frequency | Percent |
|---|---|---|
| Big-5 hospitals | 14,372 | 29.04 |
| Tertiary hospitals excluding the big-5 | 16,984 | 34.32 |
| General hospitals | 15,347 | 31.01 |
| Standard hospitals | 2,784 | 5.63 |
| Total | 49,487 | 100 |

# 5 Results of the estimation

# 5 Results of the estimation

This study applied the nested logit model using the two-stage estimation method. The first stage involved category selection for the hospitals (big-5 hospitals, tertiary hospitals, general hospitals, and standard hospitals), and the estimation for this stage was completed using the multinomial logit model. The second stage, which was completed using the conditional logit model, involved the choices within each type of institution.

As choice-specific regressors are used as explanatory variables for conditional logit models, the effects reflecting characteristics of individual patients, such as age and gender coefficients, cannot be estimated. Thus, one may use regression methods to analyze patient-specific effects only after defining the interaction variables with alternative-specific regressors. This study used the variables formed by interacting the number of beds and distance with the age of the patient in the regression model. Thus, the results indicate the influence of patient age on hospital choice, depending on the number of beds and distance between the patient's residence and the hospital.

At first glance, it is difficult to interpret the coefficient $\hat{\beta}$ estimated using the conditional logit model. Two methods can be used to do so: the first uses the odds ratio, and the second, marginal effects. The odds indicate the ratio ($\frac{p}{1-p}$) of an event happening ($p$) to not happening ($1-p$). Furthermore, the odds ratio can also be depicted as $\frac{p_1/(1-p_1)}{p_2/(1-p_2)}$, and it can be calculated by using the coefficients of the conditional logit functions. The odds ratio indicates the effect taking place when only one variable changes while the other explanatory variables are controlled.

$$\widehat{odds\,ratio} = \exp(\hat{\beta}_i)$$

The marginal effect of the explanatory variable on the probability of choice is calculated using the following equation.

$$\begin{cases} \dfrac{d\hat{p}_k}{dx_k} = \hat{p}_k(1-\hat{p}_k)\hat{\beta} \\ \dfrac{d\hat{p}_k}{dx_j} = -\hat{p}_k\hat{p}_j\hat{\beta} \end{cases}$$

When $\hat{\beta}$ is positive, the own effect indicates a positive effect ($\hat{p}_k(1-\hat{p}_k)\hat{\beta} > 0$), and the cross effect shows a negative effect.

The conditional logit model cannot estimate the coefficients of patient age or gender, as the explanatory variables are alternative-specific regressors. To understand patient-specific effects, the interaction variables with alternative-specific regressors need to be defined prior to interpretation.

The explainability of the model at the twig level is indicated by the pseudo $R$-square values suggested by McFadden, as seen below.

$$Pseudo\,R-square : R^2 = 1 - \frac{Log\,L_1}{Log\,L_0}$$

Here, $L_0$ indicates the likelihood value in the intercept only model, when there are no explanatory variables, and $L_1$ is the likelihood value when the goodness of fit is reflected.

In the case of the big-5 hospitals, the value was 0.0594, whereas the corresponding values for the tertiary hospitals, general hospitals, and standard hospitals were 0.0192, 0.04613, and 0.4787, respectively. Tentatively, the explainability of the variables selected in the Patient choice Model for the big-5 and tertiary hospitals is lower compared to the cases of the general and standard hospitals. As low pseudo R-square values indicate misspecification issues in the model arising from causes such as omitted variables, the interpretation becomes uncertain. For example, if the problem arises on account of omitted variables, adding variables to increase model explainability may change the coefficient values of the existing variables as well as their signs.

〈Table 6〉 Estimation results of the hospital choice model for cancer patients

(coefficient values)

| | Big-5 hospitals | Tertiary hospitals | General hospitals | Standard hospitals |
|---|---|---|---|---|
| _Ikv2_2 | (omitted) | -0.4280219*** | -0.7609365*** | (omitted) |
| _Ikv2_3 | (omitted) | (omitted) | -0.5039556*** | 1.46921 |
| _Ikv2_4 | (omitted) | (omitted) | -2.500343*** | 0.4200025** |
| _Ikv2_5 | 7.019748 | -0.2656112*** | -0.9168416*** | 2.038991*** |
| _Ikv2_6 | (omitted) | (omitted) | -1.321587*** | 1.43955* |
| bed | -0.0230457 | -0.0010517*** | 0.0019224*** | 0.9855461*** |
| mri | 5.362792 | -0.2454507*** | 0.4184815*** | 0.5618782*** |
| inpatient | (omitted) | 0.0023031*** | 0.0014905*** | 0.9963498*** |
| doct | 0.0299806 | -0.0006487*** | -0.0111907*** | 1.072556*** |
| nurse | (omitted) | 0.0005968*** | 0.0046413*** | 1.005206*** |
| quality | (omitted) | 0.004349** | 0.0692484*** | 2.647576 |
| bedage2 | 0.0007275*** | 0.0001631** | -0.0004997* | 1.014517*** |
| bedage3 | 0.0006308*** | 0.0000729 | -0.000548** | 1.014832*** |
| bedage4 | 0.0006068*** | -0.0001633** | -0.000911*** | 1.015778*** |
| distanceage2 | 0.0065333 | 0.0000273 | -0.0014516** | 1.005604 |
| distanceage3 | -0.0029684 | 0.0003608 | -0.002029*** | 1.009523* |
| distanceage4 | -0.0166816*** | -0.0002662 | -0.0074514*** | 1.001108 |
| time | -0.0142986*** | -0.002338*** | -0.0470272*** | 0.9547931*** |
| distance | -0.018741*** | 0.0021337*** | 0.0046773*** | 0.9932771 |
| Pseudo $R$-square | 0.0594 | 0.0193 | 0.4613 | 0.4787 |

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

〈Table 7〉 Estimation results of the hospital choice model for cancer patients (odds ratio)

| Variable | Big-5 hospitals | Tertiary hospitals | General hospitals | Standard hospitals |
|---|---|---|---|---|
| _Ikv2_2 | (omitted) | 0.6517972*** | 0.4672287*** | (omitted) |
| _Ikv2_3 | (omitted) | (omitted) | 0.6041362*** | 1.46921 |
| _Ikv2_4 | (omitted) | (omitted) | 0.0820568*** | 0.4200025** |
| _Ikv2_5 | 1118.504 | 0.7667372*** | 0.3997797*** | 2.038991*** |
| _Ikv2_6 | (omitted) | (omitted) | 0.2667118*** | 1.43955* |
| bed | 0.9772179 | 0.9989489*** | 1.001924*** | 0.9855461*** |
| mri | 213.3197 | 0.7823519*** | 1.519652*** | 0.5618782*** |
| inpatient | (omitted) | 1.002306*** | 1.001492*** | 0.9963498*** |
| doct | 1.030435 | 0.9993515*** | 0.9888717**( | 1.072556*** |
| nurse | (omitted) | 1.000597*** | 1.004652*** | 1.005206*** |
| quality | (omitted) | 1.004358** | 1.071702*** | 2.647576 |
| bedage2 | 1.000728*** | 1.000163** | 0.9995005* | 1.014517*** |
| bedage3 | 1.000631*** | 1.000073 | 0.9994522** | 1.014832*** |
| bedage4 | 1.000607*** | 0.9998367** | 0.9990894*** | 1.015778*** |
| distanceage2 | 1.006555 | 1.000027 | 0.9985494** | 1.005604 |
| distanceage3 | 0.997036 | 1.000361 | 0.9979731*** | 1.009523* |
| distanceage4 | 0.9834568*** | 0.9997338 | 0.9925763*** | 1.001108 |
| time | 0.9858031*** | 0.9976647*** | 0.9540614*** | 0.9547931*** |
| distance | 0.9814335*** | 1.002136*** | 1.004688*** | 0.9932771 |
| Pseudo $R$-square | 0.0594 | 0.0193 | 0.4613 | 0.4787 |

Note: *** p 〈 0.01, ** $p$ 〈 0.05, * $p$ 〈 0.10.

## 1) Big-5 hospitals

The estimation results for the big-5 hospitals show a negative value for the coefficient on number of beds (bed), but the coefficients on the number of MRI machines (mri) and doctors (doct) show positive values in hospital choice. However, the coefficients of these variables were not statistically significant. The interaction variables between age and number of beds (bedage), between age and distance, and between the patient's residence and the hospital, as well as the time and distance variables, were significant.

For patients older than 30 years of age, the probability of choosing the hospital increased when the number of beds was higher, compared to patients below 30 years of age (agecategory1). For patients over 65 years of age, the probability of choosing the hospital decreased when the distance to the hospital increased, compared to patients below 30 years of age (agecategory1). The time and distance variables were both negative, as expected. In other words, as the distance or time between the patient's residence and the hospital increased, the probability of selecting the hospital decreased.

If the interpretation results are analyzed using the odds ratio, a one-minute increase in time resulted in an odds ratio of $\exp(-0.0142986)=0.9858$ for choosing the hospital. In other words, when the time increases by 1 minute, the odds of selecting the hospital were found to be 0.98, assuming the choice is made under normal conditions. If the time taken increases by 10 minutes, the odds ratio was $\exp(-0.142986)=0.4240$, or 42.4% of the odds prior to the increase in time by 10 minutes. As such, the probability of hospital choice decreased to 42.4% for a 10-minute increase in time. If the time increases by 100 minutes, the odds ratio became $\exp(-1.42986)=0.2393$; that is, the probability decreased to 23.9%.

### 2) Tertiary hospitals

Differences by type of ownership were significant for the probability of choosing tertiary hospitals. The results calculated using a dummy variable show that the odds of choosing a public-sector special hospital (_Ikv2_2) were 65.1% compared to the odds of selecting public-sector national hospitals, which served as the baseline variables. The odds of selecting private-sector hospitals (_Ikv2_5) were 76.6% of the odds of the baseline variables, namely, public-sector national hospitals. In other words, all other conditions being constant, the probability of choosing public-sector special hospitals and private-sector hospitals decreased.

When the number of beds increased by one unit, the odds ratio was 0.9989. When the number of beds increased by 10 and 100, the odds ratio decreased to 0.9895 and 0.90048 respectively. In other words, the odds of selecting a hospital when the number of beds increases by 100 will decrease by approximately 10% compared to the previous odds. The number of MRI machines (mri) and doctors (doct) also affected hospital choice negatively, but the intuitive interpretation of these results was difficult. An increase in the number of nurses had a positive influence on hospital choice.

As explained previously, unintuitive coefficient values could result from misspecification issues arising from problems such as omitted variables.

### 3) General hospitals

The pseudo *R*-square values of the patient choice models of general hospitals and hospitals are 0.4613 and 0.4787, respectively, showing great improvements over the patient choice models of the big-5 and tertiary hospitals. This implies that when patients choose big-5 and tertiary hospitals, they may be influenced by variables other than the explanatory variables set in this study.

The probability of the patient choosing a general hospital increased as the number of beds (bed), MRIs (mri), and nurses (nurse) increased. When the number of beds increased by 1 unit, the odds ratio for hospital choice changed to 1.001924. The odds ratio for MRI was 1.519, indicating that when one MRI machine was added, the odds of hospital choice increased by approximately 51.9% compared to the odds of not having that extra MRI machine. The coefficient of the number of doctors (doct) did not have a positive influence on hospital choice. However, the odds ratio was found to be 1.004652 for the number of nurses. If the number of nurses increased by 10, the odds ratio increased to $\exp(0.046431) = 1.047507$. Assuming that the variable for distance affects hospital choice depending on the patient's age, the latter was analyzed as an interaction variable with distance and number of beds. If the number of doctors per 100 beds were thought to indicate the quality of medical treatment, providing an additional doctor per 100 beds was found to increase the likelihood of hospital choice by 107.17%.

### 4) Standard hospitals

Standard hospitals having between 30 and 100 beds tend to compete for patients. The number of beds and MRI machines did not positively influence hospital choice by the patient. However, the number of doctors and nurses increased the odds of hospital choice. The number of doctors per 100 beds did not significantly influence the patient's hospital choice. Apart from these observations, as the distance between the patient's residence and hospital and the time taken to reach the hospital increased, the odds of the patient selecting the hospital decreased, as expected. However, the results for the distance were not largely significant, which was attributed to the failure of the distance variable to reflect the actual distance, as the distance taken to reach the hospital by a patient in a metropolitan area was recorded in terms of Gun and Gu levels.

This study assumes that the competitive structure of the standard hospitals differs from those of the general and tertiary hospitals. While the coefficients on the number of beds, MRI machines, doctors per 100 beds, and nurses and the variable for time increased the odds of hospital choice for general hospitals, in the case of standard hospitals, the coefficients on number of doctors and nurses, and the variables for time and distance were found to be significant, unlike those on number of doctors per 100 beds.

### 5) Choice Model between Hospitals

The first-stage choice model in this study estimated choice between hospitals. The multinomial logit model was used to determine the variables that influenced the patients' choices between the big-5, tertiary, general, and standard hospitals. The multinomial logit model is used when the explanatory variable differs by the observed values, such as the age and gender of the patient. The probability of the patient $i$ choosing institution type $j$ is as follows.

$$P_{ij} = prob(Y_{ij} = 1) = \frac{e^{x_i'\beta}}{\sum_{j=1}^{J} e^{x_i'\beta}}$$

When $J$ is assumed to be the base alternative, the above equation can be re-organized as follows.

$$\frac{P_{ij}}{P_{iJ}} = e^{x_i'\beta}$$

This equation helps prove that the index value of the coefficient of $\beta$ (estimated by the model) is the odds ratio for the base alternative.

First, the big-5 and standard hospitals were compared as follows. Table 8 shows that when the number of doctors per 100 beds (quality) increased by one unit, the patient was 1.65 times more likely to choose a big-5 hospital over a standard hospital. When the length of the discharged patient's stay (dv18) increased by 1 day, the odds of selecting a big-5 hospital over standard hospital were found to increase by 1.00219 times. When the variable for time increased by 1 minute, the odds of the patient selecting a big-5 hospital over a standard hospital decreased by 0.9009 times.

〈Table 8〉 Probability of choosing hospitals at the branch level

| Hospital | Variable | Coefficient value (β) | Odds ratio |
|---|---|---|---|
| Big-5 hospitals | male | -0.0170545 | 0.9830901 |
| | quality | 0.5014233*** | 1.65107*** |
| | dv18 | 0.0021163 | 1.002119 |
| | inclusivevalue | -4.431634*** | 0.011895*** |
| | time | -0.1043479*** | 0.9009118*** |
| | _cons | -6.336029*** | 0.0017713*** |
| Tertiary hospitals excluding the big-5 hospitals | male | 0.5290063*** | 1.697245*** |
| | quality | 0.2502949*** | 1.284404*** |
| | dv18 | 0.0096558*** | 1.009703*** |
| | inclusivevalue | 1.493528*** | 4.452778*** |
| | time | 0.029115*** | 1.029543*** |
| | _cons | -13.52607*** | 1.34E-06*** |
| Standard hospitals | male | 0.2344617*** | 1.264228*** |
| | quality | 0.0832992*** | 1.086867*** |
| | dv18 | 0.0085741*** | 1.008611*** |
| | inclusivevalue | 1.165377*** | 3.207131*** |
| | time | 0.0069979*** | 1.007022*** |
| | _cons | -3.779233*** | 0.0228402*** |

Note: The base alternative is the standard hospital. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Next, the tertiary and standard hospitals were compared by examining the choice variables for a total of 39 tertiary hospitals (excluding the big-5 hospitals). The results show an increase in the odds ratio by 69.7% for men. This indicates that the percentage of male patients would be higher in tertiary hospitals compared to standard hospitals. When the length of stay of the discharged patient increased by 1 day, the odds ratio of the tertiary hospital was not much different compared to that for the standard hospital (1.009 times). When the time taken increased by 1 minute, the odds of selecting a tertiary hospital over a standard hospital increased by 1.029 times. This indicates that the tertiary hospitals (excluding the big-5 hospitals) are more sensitive to the distance between the patient's residence and the hospital, compared to the big-5 hospitals.

A similar phenomenon was observed for standard and tertiary hospitals. The odds of male patients selecting a standard hospital were 1.2642 times higher. The number of doctors per 100 beds (the variable measuring quality) recorded an increase of 1.086 times, while the length of stay of the discharged patient increased 1.0086 times, and the variable for time showed an increase of 1.0070 times.

# 6 Conclusion

# 6 Conclusion

The purpose of this study was to identify the patient and hospital characteristic variables that influence patients' hospital choice by the type of hospital and preferences between hospital types in a segmented market. This study analyzed 1,307 tertiary, general, and standard hospitals in Korea, and 49,487 cancer patients were surveyed over a period of one month as specified by the institution. This study assumed that patients' hospital choice is structured in a two-stage choice process. The first stage involves choosing the type of hospital, and the second stage concerns selecting a hospital from within a particular group of hospitals. The two-stage estimation method of the nested logit model was used to conduct the study.

The main results of the analysis are as follows. First, the estimation results indicated that the explanatory power of the model toward the choice for tertiary hospitals between hospital types was low, but that of the general and standard hospitals was high. This implies that different variables affect hospital choice within the group of tertiary hospitals. It then appears that additional explanatory variables (not considered in this study) are needed to analyze the factors affecting hospital choice within tertiary hospitals.

Secondly, when the patient selects between hospital types, the probability of selecting a tertiary and general hospital was similar when compared with the base alternatives of standard hospitals. However, there were numerous differences when the big-5 hospitals were compared. When the variable denoting time increased, the odds ratio for the big-5 hospitals decreased while those for the tertiary and general hospitals increased.

Third, in the case of general hospitals, the probability of selecting hospitals that were further away decreased when the age of the patients increased; however, this result differed for hospitals. Normally, with all other conditions remaining unchanged, higher age implies physical weakness, and it may be more difficult to move to hospitals that are further away; moving long distances for treatment may become less likely. These results align with those of Hogan (1988).

Fourth, an increase in the variable denoting time, which indicates the time spent travelling between the patient's home and the hospital, resulted in lower odds ratios regardless of hospital type. For the variable denoting distance, the odds ratio decreased for the big-5 and standard hospitals and increased for the tertiary and general hospitals, which is not consistent. If the time taken for the patient to arrive at the hospital is considered to be more important than the actual distance, it is possible that distance is not an important variable. This study classified hospitals into the big-5, tertiary, general, and standard hospitals, and assumed that the competitive structure within these types of medical institutions varies on account of differing competitive factors. The most important reason for making such an assumption lies in the fact that the determinants affecting a patient's hospital choice are likely to differ by hospital type. In reality, the results of the study show that the patient choice models for the big-5, tertiary, general, and standard hospitals have different signs and coefficient values for the same variable.

# References


Burns, L. R., & Wholey, D. R. (1992). The impact of physician characteristics in conditional choice models for hospital care. Journal of health economics, 11(1), 43-62.

Dranove, D., Shanley, M. and White, W.D(1993) “Price and concentration in hospitl markets: The switch from patient-driven to payer-competitioin”, Journal of Law and Economics, 36:179-204

Erikson G.M., A. Finkler(1985). Determinants of Market Share for a Hospital’s Services. Medical Care 23:1003-18.

Folland, S. T.(1983). Predicting Hospital Market Shares. Inquiry 20, no. 1 : 34-44.

Garnick, D. W., E. Lichtenberg, C. S. Phibbs, H. S. Luft, D. J. Peltzman, and S. J. McPhee.(1989). The Sensitivity of Conditional Choice Models for Hospital Care to Estimation Technique. Journal of Health Economics 8, no. 4 (1989): 377-97.

Kim, B.(1990), “A logit analysis of hospital choice behavior in Chollabukdo province of Korea”. Social Science and Medicine 30:1119-1129

Lee, H. L., and M. A. Cohen.(1985). A Multinomial Logit Model for the Spatial Distribution of Hospital Utilization. Journal ofBusiness and Economic Statistics 3, no. 2 (1985): 159-68

Luft HS, Garnick DW, Mark DH, Peltzman DJ, Phibbs CS, Lichtenberg E, McPhee SJ.(1990). Does quality influence choice of hospital?, Journal of the American Medical Association, 263(21): 2899-2906

McGuirk, M. A., & Porell, F. W. (1984). Spatial patterns of hospital utilization: the impact of distance and time. Inquiry, 84-95.

Phibbs, C. and J. Robinson (1993) ‘A Variable-Radius Measure of Local

Hospital Market Structure', Health Services Research, 28, 313-324.

Porell, F. W., & Adams, E. K. (1995). Hospital choice models: a review and assessment of their utility for policy impact analysis. Medical Care Research and Review, 52(2), 158-195.

Theil, H. (1969). A multinomial extension of the linear logit model. International Economic Review, 10(3), 251-259.