〈特別講演〉

Variation of Hospital Costs and Product Heterogeneity (Abstract)

Young Soo Shin, M.D., Dr. P.H.

College of Medicine, Seoul National University

Department of Preventive Medicine

The major objective of this research is to identify those hospital characteristics that best explain cost variation among hospitals and to formulate linear models that can predict hospital costs. Specific emphasis is placed on hospital output, that is, the identification of diagnosis related patient groups (DRGs) which are medically meaningful and demonstrate similar patterns of hospital resource consumption. A casemix index is developed based on the DRGs identified.

Considering the common problems encountered in previous hospital cost research, the following study requirements are estab-lished for fulfilling the objectives of this research:

1. Selection of hospitals that exercise similar medical and fiscal practices.
2. Identification of an appropriate data collection mechanism in which demographic and medical characteristics of individual patients as well as accurate and comparable cost information can be derived.
3. Development of a patient classification system in which all the patients treated in hospitals are able to be split into mutually exclusive categories with consistent and stable patterns of resource consumption.
4. Development of a cost finding mechanism through which patient groups' costs can be made comparable across hospitals.

A data set of Medicare patients prepared by the Social Security Administration was selected for the study analysis. The data set contained 27,229 record abstracts of Medicare patients discharged from all but one short-term general hospital in Connecticut during the period from January 1, 1971, to December 31, 1972. Each record abstract contained demographic and diagnostic information, as well as charges for specific medical services received. The "AUTOGRP System" was used to generate 198 DRGs in which the entire range of Medicare patients were split into mutually exclusive categories, each of which shows a consistent and stable pattern of resource consumption. The "Departmental Method" was used to generate cost information for the groups of Medicare patients that would be comparable across hospitals.

To fulfill the study objectives, an extensive analysis was conducted in the following areas:

1. Analysis of DRGs: in which the level of resource use of each DRG was determined, the length of stay or death rate of each DRG in relation to resource use was characterized, and underlying patterns of the relationships among DRG costs were explained.
2. Exploration of resource use profiles of hospitals; in which the magnitude of differences in the resource uses or death rates incurred in the treatment of Medicare patients among the study hospitals was explored.

3. Casemix analysis; in which four types of casemix-related indices were generated, and the significance of these indices in the explanation of hospital costs was examined.
4. Formulation of linear models to predict hospital costs of Medicare patients; in which nine independent variables (i. e., casemix index, hospital size, complexity of service, teaching activity, location, casemix-adjusted death rate index, occupancy rate, and casemix-adjusted length of stay index) were used for determining factors in hospital costs.

Results from the study analysis indicated that:

1. The system of 198 DRGs for Medicare patient classification was demonstrated not only as a strong tool for determining the pattern of hospital resource utilization of Medicare patients, but also for categorizing patients by their severity of illness.
2. The weighted mean total case cost (TOTC) of the study hospitals for Medicare patients during the study years was $1127.02 with a standard deviation of $117.20. The hospital with the highest average TOTC ($1538.15) was 2.08 times more expensive than the hospital with the lowest average TOTC ($743.45). The weighted mean per diem total cost (DTOC) of the study hospitals for Medicare patients during the sutdy years was $107.98 with a standard deviation of $15.18. The hospital with the highest average DTOC ($147.23) was 1.87 times more expensive than the hospital with the lowest average DTOC ($78.49).
3. The linear models for each of the six types of hospital costs were formulated using the casemix index and the eight other hospital variables as the determinants. These models explained variance to the extent of 68.7 percent of total case cost (TOTC), 63.5 percent of room and board cost (RMC), 66.2 percent of total ancillary service cost (TANC), 66.3 percent of per diem total cost (DTOC), 56.9 percent of per diem room and board cost (DRMC), and 65.5 percent of per diem ancillary service cost (DTANC). The casemix index alone explained approximately one half of interhospital cost variation; 59.1 percent for TOTC and 44.3 percent for DTOC. Thsee results demonstrate that the casemix index is the most importand determinant of interhospital cost variation

Future research and policy implications in regard to the results of this study is envisioned in the following three areas:

1. Utilization of casemix-related indices in the Medicare data systems.
2. Refinement of data for hospital cost evaluation.
3. Development of a system for reimbursement and cost control in hospitals.

1. Autogrp

Autogrp is an interactive computer system designed to facilitate rapid analysis of complex medical information. Autogrp allows the clinical or administrative expertise of the user to be combined with sophisticated computer techniques to permit rapid information retrieval, hypothesis testing, development of norms, and indentification of deviant cases. This interaction yields results of a uniquely high statistical and medical quality. Autogrp has been used to aid in understanding the process of patient care management in a variety of settings in order to enhance the effectiveness of

decision-making from both a medical and management point of view.

The AUTOGRP CLASSIFY algorithm is the means by which AUTOGRP suggests partitions of the data into groups which are distinct with respect to some dependent variable. Traditional clustering algorithms seek to form groups by minimizing some measure of distance between observations. Autogrp forms groups by minimizing the unexplained variance of a specified "dependent" variable.

Each observation in a given population of data will have a value of the independent variable X and a value of the dependent variable Y. If the independent variable is discrete, then for each distinct value the subset of observations which have that value is called a category. When the independent variable is continuous, a sequence of disjoint intervals is defined by the user, and all observations for which the independent variable value falls within a given interval are called a category. If there are N possible categories of the independent variable and there are Miobservations in the i^{th} category of the independent variable ($i \leqslant i \leqslant N$), the total sum of suqares (TSSQ) of th data with respect to the dependent variable is defined as:

$$TSSQ = \sum_{i=1}^{N} \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y})^2 \qquad (1)$$

in which

Y_{ij} = value of the dependent variable for the j^{th} observation in the i^{th} category of the independent variable

$\bar{Y}$ = mean value of the dependent variable for the entire date set.

Given the N categories, the population is partitioned into G mutually exclusive "groups," so that all of the observations in a given category are in the same group. A group, then, is the union of specified categories.

Each of the G groups will have a "within group sum of squares" (WGSSQ) with respect to the dependent variavle.

$$WGSSQ(k) = \sum_{i \epsilon R_k} \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_k)^2 \qquad (2)$$

in which

WGSSQ (k) = WGSSQ of the k^{th} group

R_k = set of all categories of the independent variable in the k^{th} group.

Y_k = mean value of the dependent variable in the k^{th} group.

The total WGSSQ (TWGSSQ) for the G groups is given by

$$TWGSSQ(G) = \sum_{k=1}^{G} \sum_{i \epsilon R_k} \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_k)^2 \qquad (3)$$

For any G it is simple to show that TWGSSQ $(G + 1) \leqslant$ TWGSSQ(G). A criterion which is both intuitively and statistically appealing for the data into distinctive groups is the partitioning is the minimization of the TWGSSQ of the data. The TWGSSQ simply represents the sum of the total squared distance of each of the groups from its mean and as such is proportional to the variance of the data that is not explained by the independent variable. Thus, the minimization of the TWGSSQ of the data results in the minimization of the unexplained varience of the data. The whole population has an initial unexplained variance given by TSSQ. If the population is divided into two groups and the minimum TWGSSQ (2) is substantially less than the original TSSQ, then a better explanation of the variance in the data is obtained by treating the population as two separate groups. AUTOGRP allows the user to partition his data, using the minimization of the TWGSSQ criterion, into as many subgroups as desirable or meaningful.

2. Diagnosis Related Group (DRG)

Many of the health systems and institutions encountering management problems can benefit from a common approach to the identification of classes of consumers (patients) for whom consistent, stable, and reliable patterns of resource consumption can be predicted during any episode of care. The present practice of standard setting and implementation in the patient care monitoring mechanism has a major defect in the initial definition of patient groupings. In order to operate a patient care monitoring system, one must be able to identify, based on process measures, incidents of care for which patient management decisions appear to produce unexpected results in terms of utilization. If, for example, one employs bed-days consumed as one of measures, one must be able to predict the bed-days required by each patient. This means that some set of patient at tributes (such as diagnosis, age, surgical procedure, complications, and the like) must be identified sufficiently to allow stability in such predictions.

The Center for the Study of Health Services at Yale University has developed an interactive statistical analysis system, called AUTOGRP, which allows one to partition data (patient records in this case) so as to maximally explain variation in some dependent variable (e.g., bed-days) as a function of some set of independent variables (patient attributes).

The author using AUTOGRP at the Center for the Study of Health Services in Yale University, has developed 317 Diagnosis Related Groups (DRGs) in which entire ranges of patients treated in a hospital were split into mutually exclusive patient categories, each of which shows a consistent and stalbe pattern of hospital resource consumption. Since Medicare patients of age 65 and over are a part of the total hospital patients, the author has also developed 198 DRGs for the Medicare patients as a subset of the 317 DRGs. The 198 DRGs for Medicare patients was used as a patient classification scheme in the thesis.

3. Casemix Related Indices

1. Casemix Index (CM_{cost})

To identify singly the proportions of each DRG in a hospital and the corresponding costs, an average cost per case is calculated. It can be stated as follows:

$$AC_i = \Sigma_j P_{ij} C_{ij} \quad (4)$$

where

AC_i = average per diem cost per patient in hospital i

C_{ij} = average cost per diem of DRG j in hospital i

P_{ij} = proportion of patients in DRG j in hospital i

For each DRG, however, the average per diem cost in Connecticut was used as the reference cost. By introducing the reference cost of each DRG, equation (4) is transformed to

$$AC_i = \Sigma_j P_{ij} C_j \times (\Sigma_j P_{ij} C_{ij} [\Sigma_j P_{ij} C_j]^{-1}) \quad (5)$$

where

C_j = average per diem cost of DRG j in Connecticut (reference cost of DRG j).

Thus, the average cost per patient in hospital i (AC_i) is a multiplication of two components; the casemix index ($\Sigma_j P_{ij} C_j$) in hospital i, and the costliness ($\Sigma_j P_{ij} C_{ij} \cdot [\Sigma_j P_{ij} C_j]^{-1}$) of hospital i. The reference cost of each DRG (C.j) is the same and constant for all hospitals. Therefore, the casemix index (CM_{cost}: $\Sigma P_{ij} C_{.j}$) of a hospital is solely a function of the hospital's casemix.

2. Casemix Adjusted Length of Stay Index ($I_{LOS\ CM}$) and Casemix Adjusted Death Rate Index

(I or CM).

$$I_{LOS}^{CM_i} = \frac{LOS_i}{\widehat{LOS}_i}$$

$$I_{DR}^{CM_i} = \frac{DR_i}{\widehat{DR}_i}$$

where

LOSi = actual average LOS in hospital i
LOS_i = expected LOS by casemix in hospital i
DR_i = actual average DR in hospital i
DR_i = expected DR by casemix in hospital i

Acknowlegement

This research was conducted from December 1975 to January 1977 at the Center for the Study of Health Services of Institution for Social and Policy Studies in Yale University, where the author was appointed as a research associate. The research was submitted to the Department of Epidemiology and Public Health in Yale University in fulfillment of the requirements for the degree of Doctor of Public Health. The research was supported by the Social Security Administration, in U.S. Department of Health, Education, and Welfare under the Control Number 600-75-0180. The report was published by HEW, HEW Publication No. (SSA) 77-11722.