

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2012.01379.x
METHODS ARTICLE

Development of Peer-Group-Classification Criteria for the Comparison of Cost Efficiency among General Hospitals under the Korean NHI Program

Hee-Chung Kang, Jae-Seok Hong, and Heon-jin Park

Objectives. To classify general hospitals into homogeneous systematic-risk groups in order to compare cost efficiency and propose peer-group-classification criteria.

Data Sources. Health care institution registration data and inpatient-episode-based claims data submitted by the Korea National Health Insurance system to the Health Insurance Review and Assessment Service from July 2007 to December 2009.

Study Design. Cluster analysis was performed to classify general hospitals into peer groups based on similarities in hospital characteristics, case mix complexity, and service-distribution characteristics. Classification criteria reflecting clustering were developed. To test whether the new peer groups better adjusted for differences in systematic risks among peer groups, we compared the R^2 statistics of the current and proposed peer groups according to total variations in medical costs per episode and case mix indices influencing the cost efficiency.

Data Collection. A total of 1,236,471 inpatient episodes were constructed for 222 general hospitals in 2008.

Principal Findings. New criteria were developed to classify general hospitals into three peer groups (large general hospitals, small and medium general hospitals treating severe cases, and small and medium general hospitals) according to size and case mix index.

Conclusions. This study provides information about using peer grouping to enhance fairness in the performance assessment of health care providers.

Key Words. Cost efficiency, peer group, cluster analysis

Policy makers in many health insurance systems experiencing rapidly escalating health care expenditures have focused on developing tools to hold physicians accountable for their decisions (Bindman 1999; Austin et al. 2004).

Several strategies have been developed, including utilization review and practice guidelines, to induce physicians to choose more rational, less expensive behavior (Bindman 1999). Many insurance programs have employed utilization review to foster more appropriate payment, but case-by-case utilization review is time consuming, and a more efficient methodology is required to complete the review of a large number of claims within the legally

mandated time frame. In this context, profiling is considered to constitute a more efficient mechanism for observing a physician's pattern of care (Bindman 1999; Romano 2004). Performance profiling involves comparing an individual physician's performance during a fixed period with normative standards or with the average performance level of comparable physicians during the same period (Bindman 1999; Smith 2000; Weiss and Wagner 2000; Austin et al. 2004). However, such a performance comparison between hospitals invites dispute due to the difficulties of defining comparable hospitals, particularly comparable general hospitals, which are multipurpose, multiproduct institutions providing ambulatory care in addition to inpatient services (Stefos, Lavalley, and Holden 1992; Sandy 1999; Pink et al. 2009). Contextual as well as patient characteristics must be considered when classifying peer groups of general hospitals to enable adjustment for the many systematic risks that are not easily manageable by hospitals themselves (Austin et al. 2004; Byrne et al. 2009).

The objective of this study was to classify general hospitals into peer groups so that differences in systematic risks can be adjusted and to propose peer-group-classification criteria for assigning newly established general hospitals or those with changing characteristics to appropriate peer groups.

METHODS

Study Setting

The Ministry of Health and Welfare (MOHW) of Korea oversees the National Health Insurance (NHI) Program. In terms of implementation, the National Health Insurance Corporation (NHIC) functions as the insurer, and the Health Insurance Review and Assessment Service (HIRA) conducts reviews and assessments of medical costs. The HIRA review process is designed to minimize the risk of payment for excessive or unnecessary patient care in a fee-for-service-based reimbursement system (HIRA 2010).

Address correspondence to Hee-Chung Kang, Ph.D., Review & Assessment Research Division, Health Insurance Review and Assessment Service, 1451-34, Peace BLDG 11F, Seocho3-Dong, Seocho-Gu, Seoul, 137-927, Korea; e-mail: kanghj71@gmail.com. Jae-Seok Hong, Ph.D., is with the Review & Assessment Research Division, Health Insurance Review and Assessment Service, Seoul, Korea. Heon-Jin Park, Ph.D., is with the Department of Statistics, Inha University, Incheon, Korea.

Since 1989, benefits under Korea's NHI program have been distributed in two steps. The first step involves primary care institutions (clinics, hospitals, general hospitals), and the second step involves the 44 tertiary general hospitals (NHIC 2011). When an insured individual (dependent) requests medical care at a tertiary general hospital, he or she must present a document issued by the referring physician. Health care institutions are classified into three groups by the number of beds they contain; clinics have fewer than 30 beds, hospitals have 30–99 beds, and general hospitals have more than 99 beds. The MOHW designates tertiary hospitals from general hospital applicants on the basis of whether they meet the standards for teaching hospitals and fulfill other criteria (NHIC 2007). Patients tend to be drawn to general hospitals because large institutions generate greater trust. This phenomenon has been identified as one of the contributors to increased medical costs (Lee and Park 2010).

The HIRA has operated the Comprehensive Management for Appropriate Medical Services System (CM System) overseeing clinics since 2003. The CM system is designed to encourage medical clinics showing extreme patterns of practice and patient care to change voluntarily through feedback provided by economic-performance profiles.

As medical costs continued to increase, the HIRA extended the implementation of the CM system to cover hospitals and, in July 2009, to cover general hospitals and tertiary general hospitals. The cost efficiency of a general hospital is measured relative to the average cost of an appropriate group of peer hospitals, defined only by the number of beds, into hospitals with ≥ 301 beds and those with ≤ 300 beds. Administrators of general hospitals have argued that a more refined classification of peer groups is required for fairer comparisons among general hospitals.

Study Data and Measures

Data were obtained from NHI Claims Database and the NHI Institution Registration Data submitted to the HIRA by health care institutions. A total of 251 general hospitals except tertiary general hospitals provided health care under the NHI Program for the 12 months from January to December 2008. Their inpatient claims submitted to the HIRA from July 2007 to December 2009 were arranged according to admission date (from July 2007 to June 2009). Records were treated as belonging to the same episode if the time between the discharge date and the following admission date was 2 days or less and the second admission was associated with the same diagnosis-related

group (DRG). Inpatient episodes that began in 2008 were selected from the episode-based data for the analysis. Inpatient episodes are referred to as patients or episodes for simplicity. Additional costs were adjusted according to the standard cost of the NHI fee schedule. Outlier episodes by DRG were excluded, taking into account the effects of the small sample size, resulting in the exclusion of fewer than 10 inpatient episodes per DRG per general hospital (Pope and Kautter 2007).

The ratio of main-disease episodes to total number of episodes exceeded the criterion set by the MOHW for specialized hospitals in 29 general hospitals. As of 2010, hospitals could be designated as specialized in nine domains (joints, the cerebrovascular system, the colon, hands and feet, the cardiac system, alcohol, breasts, spinal issues, and burns). To qualify as specializing in spinal or alcohol-related diseases, a hospital must show that more than 66 percent of its cases involve one of these areas; the comparable requirement for the other domains is greater than 45 percent. These hospitals were excluded from the analysis as they were regarded as specialized settings for certain diseases, and their practice behavior differed from that of general hospitals. The final sample included 222 general hospitals and 1,236,471 inpatient episodes. The unit of analysis was a general hospital.

Peer groups, used to control for differences in systematic risks among hospitals, were defined on the basis of hospital characteristics, case mix complexity, and service-distribution characteristics, which have been studied as influences on the finances or clinical outcomes of hospitals (Ellis and McGuire 1988; Stefos, Lavalley, and Holden 1992; Byrne et al. 2009). The use of peer groups enabled us to categorize hospitals according to structural and patient characteristics to facilitate comparisons among similar institutions (Byrne et al. 2009).

Hospital characteristics consisted of type of ownership, teaching status, location, number of specialties, number of beds, number of medical staff, types of equipment, patient volume, and medical costs per inpatient episode. Outpatient episodes were added to the inpatient episodes to calculate patient volume. These variables have previously been used as hospital characteristics in analyses of sources of variations among hospitals in case mix complexity and performance indices (Ament, Kobrinski, and Wood 1981; Becker and Steinwald 1981; Rosko and Carpenter 1993; Jian et al. 2009).

Case mix complexity can be measured using scalar indices or information-theory indices (Park and Shin 2004). The former approach reduces the vector of the proportion of patients classified into diagnostic categories into a single-value index through multiplication by vector weights, which often

include the length of stay and costs or charges. The latter approach, first proposed by Evans and Walker, measures differences in the proportions of two sets of patients classified according to diagnostic categories (Klastorin and Watts 1980; Park and Shin 2004). This study used both the Case Mix Index (CMI) as a scalar index and the Professional Care Disease Index (PCDI) as an information-theory index.

The HIRA assigns an inpatient episode to a disease group using the *Korean Diagnostic Related Groupings* (KDRG) version 3.3, an inpatient case mix classification system. The CMI is calculated with the following equation. As the same average cost per DRG (C_i) is applied to the denominator (N_i) and the numerator (N_{hi}), the index value would be 1 if the case mix of a general hospital were equal to the average for its peer group. The DRG classification was developed so that resource consumption, as well as the clinical characteristics of the patients in a DRG, would be similar. Therefore, if the CMI of a general hospital were 1.2, the general hospital would be regarded as having a case mix spending cost that was 20 percent higher than that of the peer-group average.

$$\text{Case Mix Index (CMI)}_i = \frac{(\sum_i C_i \times N_{hi}) / (\sum_i N_{hi})}{(\sum_i C_i \times N_i) / (\sum_i N_i)},$$

where h : a general hospital in the country, i : by DRG, N_{hi} : number of inpatient episodes in the i th DRG of h general hospital, N_i : number of inpatient episodes in the i th DRG in all general hospitals, C_i : average medical costs per inpatient episode in the i th DRG in all general hospitals.

The PCDI is one of the accreditation criteria used by the MOHW to designate tertiary general hospitals in the NHI program. Tertiary hospitals are expected to treat severely ill patients who cannot be adequately diagnosed or treated at the first stage of the NHI health care delivery system (National Health Insurance Corporation (NHIC) 2011; Park and Shin 2004). The PCDI is the proportion of patients in the adjacent DRGs (ADRGs) that were approved by the MOHW as disease categories needing to be treated in tertiary hospitals, standardized by the proportion of these cases at the national level (Park and Shin 2004). The ADRG is the first four digits of the DRG code. In this study, we used this index as a second indicator to assess the case mix complexity of the study institutions (general hospitals not designated as tertiary institutions). Whereas the CMI represents the number of high-cost patients within the total population of patients, the PCDI indicates the proportion of all patients that pose a clinically high risk. Therefore, when the two indices are considered together, the case mix complexity per general hospital is more accurate.

$$\begin{aligned} &\text{Professional Care Disease Index (PCDI)}_i \\ &= \left(\frac{\sum_{j \in K} n_{ij}}{N_i} \right) / \left(\frac{\sum_{i \in H} \sum_{j \in K} n_{ij}}{\sum_{i \in H} N_i} \right), \end{aligned}$$

where H : set of all general hospitals in the nation; K : set of ADRGs that need to be treated in tertiary hospitals; i : subscript for H ; j : subscript for K ; n_{ij} : number of inpatient episodes in the j th ADRG of the i th general hospital; N_i : total number of inpatient episodes in the i th general hospital.

Third, the service-distribution characteristics included the specialization status and the distribution of patients by disease type. The specialization status was measured using the Internal Herfindahl Index, a measure of the concentration of services in a health care institution that is derived from the Herfindahl-Hirschman Index and used to measure the diversity of procedures performed at a single hospital or within a region (Wachtel et al. 2010). The Internal Herfindahl Index (IHI) of i institution is calculated by the summation of the P_j^2 values.

$$\text{Internal Herfindahl Index (IHI)}_i = \sum (P_j^2),$$

where P_j = proportion of all the patients in i general hospital accounted for by the j th ADRG category.

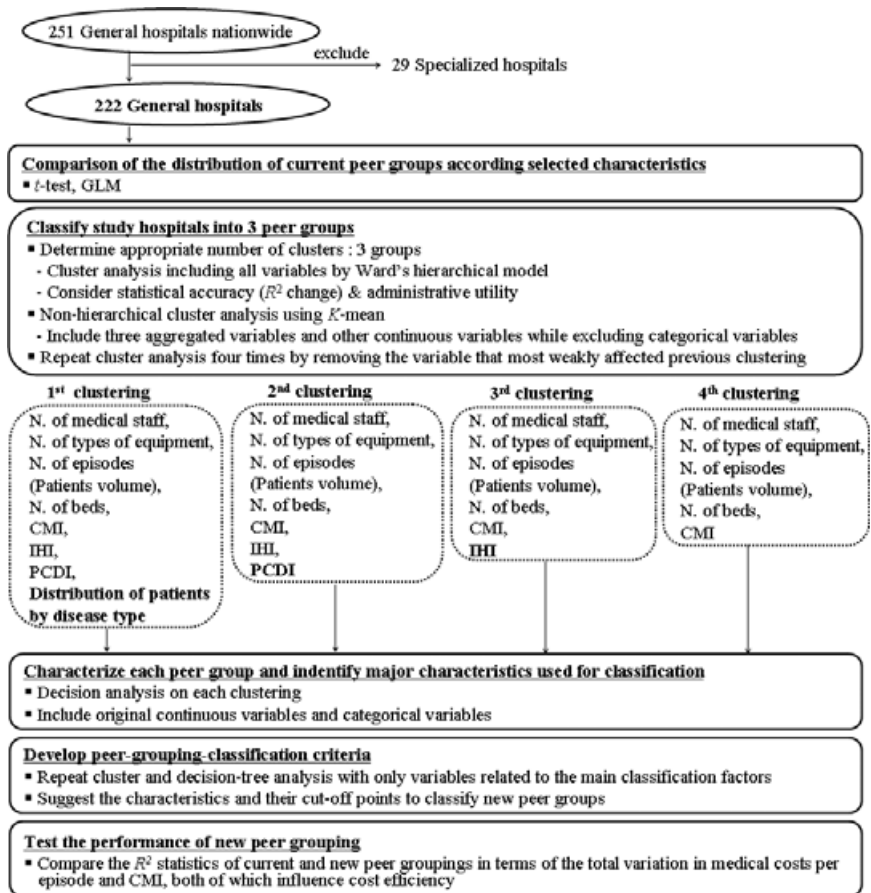
If few services are provided in a hospital, the concentration of patients in those services will be high (Lee and Chun 2008). In this study, the specialization status of the service was interpreted to be higher as the scope of inpatient episodes treated by the general hospital became narrower.

The proportion of the total number of inpatient episodes attributable to each disease was used as an indicator of the distribution of patients by disease type. This is one of the criteria used by the MOHW to designate specialized hospitals.

Statistical Analyses

Figure 1 presents the five steps involved in the analyses performed in this study. The first step was to compare the distributions of current peer groups (bed size: ≤ 300 , ≥ 301) by their characteristics (t -test, general linear model [GLM]). The second step was to classify general hospitals into peer groups using cluster analysis. Cluster analysis is a multivariate procedure that simultaneously considers all classification variables to arrange a sample of entities

Figure 1: Study Hospitals and Analytical Process



into distinct groups according to shared characteristics (Stefos, Lavallee, and Holden 1992). For the cluster analysis, all variables were standardized using PROC STANDARD, and nonhierarchical cluster analysis was conducted for the appropriate number of clusters, as determined by hierarchical analysis. PROC CLUSTER using Ward's method was used for hierarchical analysis, and PROC FASTCLUS using the *K*-mean was used for nonhierarchical analysis. Cluster analysis was performed for all variables with the exception of categorical variables such as ownership, teaching status, and location. Variables that were highly related according to correlation analyses were converted into

aggregated variables for performing cluster analysis. These included number of medical staff (number of physician specialists + number of nurses + number of medical technicians + number of specialties); number of types of equipment (number of types of diagnostic test equipment + number of types of radiological diagnostic and therapeutic equipment + number of types of physical therapy equipment + number of types of surgical and treatment-related equipment); and patient volume (number of inpatient episodes + number of outpatient episodes). Four cluster analyses were performed to classify hospitals into three peer groups by removing the variable that most weakly affected previous clustering. The first clustering was performed for three converted variables, number of beds, CMI, IHI, PCDI, and the distribution of patients by disease type. The second clustering was performed for the first clustering variables except the distribution of patients by disease type, which had most weakly affected the first cluster analysis. For the same reason, the PCDI in third clustering and the IHI in the fourth clustering were additionally excluded.

The third step was to characterize each peer group by identifying the common characteristics of the major classification factors shown in the iterative decision-tree analysis for the four cluster analyses.

The fourth step was to produce classification criteria through repeating the cluster and decision-tree analyses with only the variables related to the main factors identified in the prior step. To confirm that the classification criteria reflected cluster analyses, the consistency between the final peer grouping and each previous clustering was assessed with weighted *kappa* statistics.

The final step was to compare the R^2 statistics for the total variation in medical costs per episode and the CMI, both of which influence cost efficiency, to examine whether the proposed peer-grouping criteria better adjusted for systematic risks when measuring cost efficiency than did the current peer-grouping criteria (Lee 2007). *SAS 9.1* and *SAS Enterprise Minor* were used for this analysis.

RESULTS

Comparison of Characteristics between Current Peer Groups

When the current peer-group characteristics were compared, no significant differences in ownership were evident (Table 1). A greater number of medical schools were present in general hospitals with ≥ 301 beds. General hospitals with ≥ 301 beds were also characterized by larger medical staffs, more

Table 1: Descriptive Statistics of Study Hospitals

Variables	Current Peer Group		X^2/t Value
	No. of Beds \leq 300 ($N = 120$)	No. of Beds \geq 301 ($N = 102$)	
Hospital characteristics			
Ownership			
Private	102 (85.0)	92 (90.2)	1.3
Public	18 (15.0)	10 (9.8)	
Teaching status			
Medical school	9 (7.5)	26 (25.5)	13.4**
Other	111 (92.5)	76 (74.5)	
Location			
Metropolitan	99 (82.5)	74 (72.6)	13.0**
Municipality	9 (7.5)	24 (23.5)	
County	12 (10.0)	4 (3.9)	
Composition of medical staff			
Number of physician specialists	19.5 \pm 10.6	64.1 \pm 41.1	-10.6**
Number of nurses	73.8 \pm 36.3	206.2 \pm 103.5	-12.3**
Number of medical technicians	18.7 \pm 7.8	52.2 \pm 28.2	-11.6**
Number of specialties	11.7 \pm 2.7	19.1 \pm 3.3	-18.1**
Number of beds	210.6 \pm 55.4	494.5 \pm 136.1	-19.7**
Types of equipment by field			
Diagnostic tests	31.1 \pm 7.4	45.2 \pm 1.5	-12.1**
Radiological diagnostic & therapy	11.1 \pm 2.3	16.1 \pm 3.4	-12.7**
Physical therapy equipment	14.2 \pm 3.2	20.4 \pm 4.1	-12.5**
Surgery & treatment equipment	8.5 \pm 2.5	12.7 \pm 3.4	-10.2**
Patient volume			
Number of inpatient episodes	2,441 \pm 1,604	9,249 \pm 7,676	-8.8**
Number of outpatient episodes (10,000 cases)	145 \pm 72	433 \pm 285	-9.9**
Medical costs per inpatient episode (100,000 KRW)	9.5 \pm 2.6	13.5 \pm 3.4	-9.7**
Case mix complexity			
Case mix index	0.65 \pm 0.17	0.86 \pm 0.22	-8.1**
Professional care disease index	0.08 \pm 0.19	0.49 \pm 0.56	-7.0**
Service-distribution characteristics			
Specialization status (Internal Herfindahl Index)	0.19 \pm 0.05	0.14 \pm 0.04	8.2**
Distribution of patients by disease type (%)			
Cerebrovascular	8.0 \pm 5.0	9.9 \pm 5.5	-2.7*
Cardiac	3.0 \pm 5.2	6.3 \pm 5.5	-4.5**
General surgical (breast)	2.5 \pm 3.3	2.7 \pm 2.3	-0.6
Joint and spinal	17.5 \pm 9.9	11.9 \pm 6.8	4.9**
Colon	21.3 \pm 8.4	19.2 \pm 6.5	2.0*
Pediatric	21.9 \pm 15.2	22.8 \pm 11.5	-0.4
Obstetrics and gynecology	2.3 \pm 6.3	4.9 \pm 4.9	-3.4*

* $p < .05$,** $p < .001$.

specialties, and a greater number of beds. The CMI was higher in the general hospitals with ≥ 301 beds than in those with ≤ 300 beds, and the PCDI was higher in general hospitals with ≥ 301 beds.

Regarding service distribution, the specialization status was higher in general hospitals with ≤ 300 beds. With respect to patient proportions by disease type, no significant differences in general surgical and pediatric patients were evident, but lower proportions of joint, spinal, and colon patients were admitted to general hospitals with ≥ 301 beds.

The distributions of most of the characteristics assessed were wider in general hospitals with ≥ 301 beds than in general hospitals with ≤ 300 beds. The distribution of the two groups was not distinctive and only partially overlapped.

Clustering and Cluster Characterization

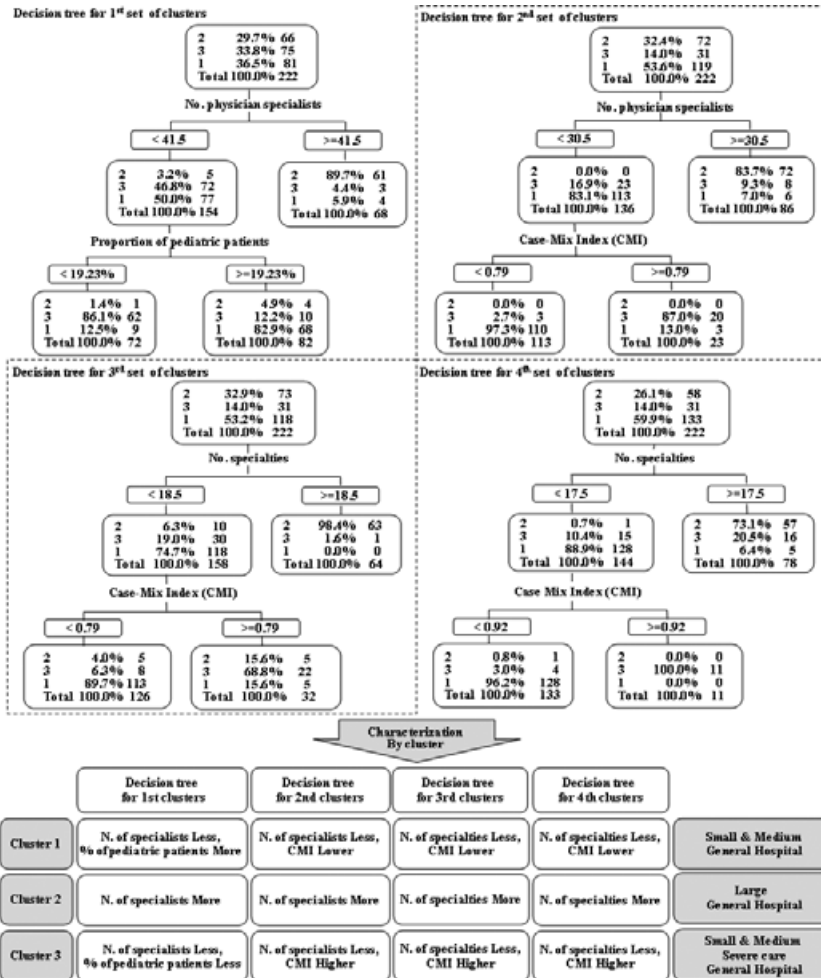
Three clusters emerged from cluster and decision-tree analyses. Cluster 1 included small and medium-sized general hospitals; cluster 2 included large general hospitals; and cluster 3 included small and medium-sized general hospitals that treated severe cases (Figure 2). Hospitals in cluster 1 had fewer specialists and specialties and lower proportions of pediatric patients or CMIs. Cluster 2 hospitals had more specialists and specialties and higher proportions of pediatric patients or CMIs. Hospitals in cluster 3 had fewer specialists and specialties, but higher proportions of pediatric patients or CMIs.

The first classification variable in the decision tree for the first clustering was the number of physician specialists, and the second classification variable, which divided the group with fewer than 41.5 physician specialists into two subgroups, was the proportion of pediatric patients; the second classification variable in the second, third, and fourth decision trees was the CMI.

The CMI of general hospitals, rather than the proportion of patients with each disease, was used to further classify hospitals because the CMI includes a measure of the complexity of the case mix of all patients.

When we compared the classification consistency using weighted *kappa* statistics, the level of consistency of the first clustering (0.3–0.4), which included the proportions of patients by disease type, was low with each subsequent clustering. The second, third, and fourth clustering showed high levels of consistency with one another (0.7–0.8), with the exception of the analysis of the proportion of patients by disease.

Figure 2: Characterization of Each Cluster Using Cluster and Decision-Tree Analyses



Development of the Classification Criteria Used to Define Peer Groups

Following the characterization of clusters, we used the size of the hospital as the first factor for defining peer groups and its case mix complexity as the second factor.

The new classification criteria were developed in two steps. The first step involved a cluster analysis to classify general hospitals into two groups using size-related variables (i.e., number of specialists, nurses, beds, and specialties),

which were converted to a new variable, size. The correlation coefficient among the four size-related variables was as high as 0.8, and thus all variables were converted into one principal component value to perform the cluster analysis. In the principal component analysis, Prin1 explained 85 percent of the total variance, and the principal component score for Prin1 was identical in all variables. Therefore, a size variable was produced to standardize and summarize the four variables with the same weight. The decision tree by Prin1 and that by the new variable of size produced identical results.

We conducted cluster and decision-tree analyses using size to determine the point at which to divide groups. The reference point for size was 8.2, which divided general hospitals into 78 large general hospitals and 144 small and medium-sized general hospitals.

$$\text{Size}_i = N_{pi}/\sigma_p + N_{ni}/\sigma_n + N_{bi}/\sigma_b + N_{si}/\sigma_s,$$

where N_{pi} : number of physician specialist at general hospital i , N_{ni} : number of nurse at general hospital i , N_{bi} : number of bed at general hospital i , N_{si} : number of specialty at general hospital i , σ_p , σ_n , σ_b , σ_s : standard deviation of N_{pi} , N_{ni} , N_{bi} , N_{si} .

The second step involved classifying general hospitals smaller than a certain size (8.2) into two groups using the CMI. The distribution of the CMIs in the small and medium-sized general hospitals ($n = 144$) was positively skewed. The frequency of the distribution dramatically decreased at $\text{CMI} > 0.7$, revealing outliers among the hospitals. Thus, we divided the small and medium-sized general hospitals into two groups at the 75th percentile ($\text{CMI} = 0.73$).

When we assessed the consistency of the groupings according to size and CMI with that of all but the first previous clustering results, including the proportion of pediatric patients, the weighted *Kappa* statistics were 0.77, 0.74, and 0.63 for the second, third, and fourth clusterings, respectively. The highest level of classification consistency was for the second clustering, which involved the largest number of variables. This demonstrated that the classification criteria reflected the previous clustering.

Comparison of Characteristics among Proposed Peer Groups

Finally, the proposed new peer groups were compared in Table 2. Cluster 1, small and medium-sized general hospitals, had the lowest number of beds and specialists and the lowest CMI. Cluster 2, large general hospitals, had larger staffs, greater numbers of specialists and beds, and the highest CMIs. Cluster

Table 2: Comparison of Characteristics among Proposed Peer Groups

<i>Variables</i>	<i>Large General Hospitals (N = 78)</i>	<i>Small and Medium General Hospitals with Severe Cases (N = 36)</i>	<i>Small and Medium General Hospitals (N = 108)</i>	<i>X²/F</i>
Hospital characteristics				
Ownership				
Private	73 (93.6)	94 (87.0)	27 (75.0)	7.7*
Public	5 (6.4)	9 (13.0)	9 (25.0)	
Teaching status				
Medical school	28 (35.9)	6 (5.6)	1 (2.8)	36.9***
Other	50 (64.1)	102 (94.4)	35 (97.2)	
Location				
Metropolitan area	57 (73.1)	88 (81.5)	28 (77.8)	16.7*
Municipality	20 (25.6)	8 (7.4)	5 (13.9)	
County	1 (1.2)	12 (11.1)	3 (8.3)	
Composition of medical staff				
Number of physicians (specialists)	77.6 ± 38.6	18.5 ± 7.2	22.9 ± 8.1	152.7***
Number of nurses	240.2 ± 94.9	72.7 ± 31.8	91.6 ± 38.9	172.8***
Number of medical technicians	60.6 ± 26.9	18.6 ± 7.0	23.1 ± 9.9	144.2***
Number of specialties	20.6 ± 2.1	13.1 ± 2.6	11.8 ± 2.6	300.8***
Number of beds	527.1 ± 136.3	272.9 ± 78.6	229.3 ± 87.3	187.5***
Types of equipment by field				
Diagnostic tests	48.9 ± 7.3	30.6 ± 6.8	33.9 ± 7.1	158.8***
Radiological diagnostic and therapy	17.2 ± 3.0	11.2 ± 2.3	11.8 ± 2.0	136.0***
Physical therapy equipment	21.6 ± 3.8	14.4 ± 3.2	15.5 ± 3.2	104.4***
Surgical and treatment equipment	13.9 ± 2.8	8.2 ± 2.4	9.3 ± 1.9	127.5***
Patient volume (10,000 cases)				
Number of inpatient episodes	11.2 ± 7.6	2.5 ± 1.6	2.6 ± 2.1	86.1***
Number of outpatient episodes	517.3 ± 271.3	144.1 ± 68.7	156.6 ± 90.5	117.6***

continued

Table 2. Continued

Variables	Large General Hospitals (N = 78)	Small and Medium General Hospitals with Severe Cases (N = 36)	Small and Medium General Hospitals (N = 108)	X ² /F
Medical costs per inpatient episode (100,000 KRW)	14.1 ± 3.3	13.5 ± 2.9	8.6 ± 1.4	130.0***
Case mix complexity				
Case mix index	0.90 ± 0.20	0.90 ± 0.18	0.58 ± 0.09	117.1***
Professional care disease index	0.63 ± 0.58	0.16 ± 0.25	0.04 ± 0.12	59.3***
Service-distribution characteristics				
Specialization status (Internal Herfindahl index)	0.12 ± 0.03	0.19 ± .05	0.19 ± 0.05	61.7***
Distribution of patients by disease type (%)				
Cerebrovascular	9.6 ± 5.0	7.6 ± 4.6	11.0 ± 6.9	6.94*
Cardiac	6.9 ± 5.1	2.4 ± 4.8	5.5 ± 6.4	17.3***
General surgical (breast)	3.1 ± 2.3	2.5 ± 3.4	1.8 ± 1.6	2.6
Joint and spinal	10.6 ± 5.5	16.3 ± 9.3	19.9 ± 10.3	18.3***
Colon	18.5 ± 6.0	21.8 ± 8.2	20.1 ± 8.5	4.2*
Pediatric	22.7 ± 10.4	24.3 ± 15.1	15.3 ± 13.2	6.2*
Obstetrics and gynecology	6.3 ± 4.9	2.5 ± 6.6	0.3 ± 0.9	18.0***

* $p < .05$,
 ** $p < .001$,
 *** $p < .0001$.

Table 3: R^{2*} Statistics: Comparison of Medical Costs Per Episode and CMI between Proposed and Current Peer Groups (Unit: %)

N = 222	Proposed Peer Groups	Current Peer Groups
Medical costs per episode [†]	65.4	51.1
CMI	56.6	42.2

* Adjusted R² value from the simple regression analysis with cost efficiency as a dependent variable and peer groups as an independent variable.

[†] Log-transformed medical costs per episode.

3, small- and medium-sized hospitals treating severe cases, was similar to the small- and medium-sized group in terms of number of specialists and beds, but it was higher than that group in terms of CMIs (Table 2).

The R^2 for the medical costs per episode and the CMI for the proposed peer-group classification were 65.4 and 56.6 percent, respectively, both of which were higher than those (51.1 and 42.2 percent) for the current peer-group classification (Table 3).

DISCUSSION

This study presents criteria to classify general hospitals into three peer groups (group 1: large-sized general hospitals [size ≥ 8.2], group 2: small- and medium-sized general hospitals with severe cases [size < 8.2 , CMI ≥ 0.73], group 3: small- and medium-sized general hospitals [size < 8.2 , CMI < 0.73]). According to consistency testing, the proposed classification criteria reflected the results of the cluster analysis well. The R^2 statistics for the variation in costs per episode and the CMI, both of which influence the efficiency index, were shown to increase substantially in the peer groups classified according to the proposed grouping criteria (R^2 : 65.4 and 56.6 for cost per episode and CMI, respectively) compared with those in the groups classified according to the current grouping criteria, which use number of beds (R^2 : 51.1, 42.2 for cost per episode and CMI, respectively). Indeed, when the R^2 of the effect of peer-group classification on total variation was higher, the variation within the group decreased, whereas that among the groups increased.

Structural differences in staff, facilities, equipment, and case mix can lead to variations among hospitals in systematic risks, which subsequently lead to differences in the patient-care costs and therefore in cost efficiency (Ellis and McGuire 1988).

The issue of fairness in comparisons of the performance of hospitals was explored by Jencks et al. (1984) and Ellis and McGuire (1988). They demonstrated that the systematic risks, which health care providers cannot manage independently, should be taken into account when comparing performance. Therefore, when the definition of a peer group incorporates adjustments for systematic risks, the fairness of comparisons is enhanced.

To adjust for differences among hospitals in systematic risks, Stefos, Lavallee, and Holden (1992) defined peer groups using cluster analysis so that the various characteristics of the hospitals in a group tended to be similar. Zodet and Clark (1996) classified hospitals in Michigan into 13 groups that displayed similar characteristics. A recent study suggested a new method for measuring the similarity of the characteristics of hospitals using Euclidean distance to define peer groups of public health care institutions (Byrne et al. 2009).

These studies addressed classification methods using cluster analysis, but they did not present a method for developing classification criteria that reflect the results of clustering, which health insurance administrators could use to assign a newly established general hospital or an established hospital whose characteristics have changed to a peer group within a reasonable period of time. The current study could therefore increase the application of cluster analysis.

This study also reflects all the factors leading to the variations among the hospitals examined in previous studies in terms of cost efficiency, including human resources, facilities, equipment, case mix per disease, CMI, and specialization index (Ament, Kobrinski, and Wood 1981; Becker and Steinwald 1981; Rosko and Carpenter 1993).

When the performance of a health care provider is compared with the average performance of the appropriate peer group, the definition of the peer group may also influence the accuracy of the cost-efficiency index. The Medicare Resource Use Reports Plan also includes the selection of peer groups as among the main tasks to be completed prior to the calculation of the efficiency index (MaCurdy et al. 2008).

Many health care systems encourage health care facilities that show extreme practice patterns to reach the peer-group-specific benchmark. However, the definition of comparable hospitals has always proven controversial (Stefos, Lavalley, and Holden 1992; Pink et al. 2009). The appropriate benchmark may vary depending on the interests of the stakeholders. Health care institutions would prefer to be compared within peer groups whose members possess similar characteristics. In contrast, health care consumers would ask for a comparison of performance with the nationwide average, as this information can be used to select health care institutions from which services can be purchased (Austin et al. 2004; Romano 2004). Health insurance administrators must consider both interests in determining the appropriate benchmark level if the comparison is to be meaningful enough to promote cost efficiency.

Pink et al. (2009) discussed a peer-group effect whereby the proportion of health care providers meeting the benchmark varies by peer group, indicating that the ability to reach a certain level differs among peer groups. It was therefore claimed that a benchmark should represent a high level of financial performance regardless of the factors that influence the ability of a hospital to reach the benchmark. That is, if hospitals were compared only with others in their peer group, the overall efficiency of health care institutions may decrease further. However, it is difficult to ignore the assertion

that hospitals have immutable systematic characteristics and that these characteristics may have a negative effect on both financial performance and quality of care (Ellis and McGuire 1988; Austin et al. 2004; Romano 2004; Byrne et al. 2009).

Potential users of our methodology may ask which variables should be part of a risk-adjustment system and which should be used to define a peer group. We would respond that peer grouping constitutes an additional option for determining the benchmark used for comparisons because hospitals outside the peer group would not be included in the comparative analysis. In this study, we included each hospital's CMI, a measure of hospital-wide case complexity, as a variable for classifying hospitals into peer groups.

In Korea, the HIRA regularly defines peer groups in terms of number of beds and compares the risk-adjusted performance (using DRGs) of hospitals within each peer group.

This study emphasized the need to adjust for a hospital's patient level of risk when comparing performance and further suggested the desirability of including a hospital-wide case-complexity measure in the definition of peer groups to enhance the fairness of comparison when peer grouping is necessary.

The present study may have some limitations. First, the adjustment for systematic risks among the peer groups may have been inadequate. Indeed, the first clustering of the cluster analysis, in which the proportion of patients by disease type was considered, differed from the other three clusterings. This indicates that the proposed peer-group classification cannot consider all features of a hospital. However, because specialized hospitals were excluded from this study, we determined that additional considerations regarding the proportion of patients according to disease type were unnecessary. Efforts to overcome such limitations through focusing the scope of the evaluation on specific disease types are required. In addition, it is important to re-evaluate the appropriateness of peer-group assignments and to redefine peer groups periodically.

Second, determining the appropriate number of clusters is difficult; no statistical tests currently exist to confirm that group numbers are optimal. In this study, a two-stage cluster analysis was performed (Stefos, Lavalley, and Holden 1992). As a first stage, Ward's hierarchical method was used to select the appropriate number of groups. The changes in R^2 and pseudo- t^2 values were analyzed according to changes in the number of clusters, and the results suggested two to four clusters. We set three groups as the optimum number of clusters, as the number of groups should be minimized to the extent that it satisfies both statistical accuracy, after adjusting for systematic risk, and administrative utility.

Third, as no information regarding patient addresses was available in the HIRA claims data, we could not include the competition level of the hospital location as a variable. When discussing the causes of variation in health care utilization by location, the medical supply and market aspects that cause inappropriate and excessive utilization of health care services have been of interest (Goodman and Green 1996; Do 2007). However, the CMI may already reflect the structural or external circumstances of a general hospital, and thus the exclusion of such variables would not constitute a major bias in the study results (Becker and Steinwald 1981).

Continued efforts to improve the fairness and accuracy of peer-group comparisons are required to increase the receptivity of health care providers to performance profiling. In this context, this study is expected to provide useful information for defining and managing peer groups in the assessment of health care performance.

CONCLUSIONS

This study demonstrated the procedure for performing cluster analyses of the characteristics of general hospitals that provide health care under the Korea NHI program so that the systematic risks within a peer group are similar and peer-group-classification criteria can be developed. Our results suggest that cluster analysis may be a useful method for classifying multipurpose, multi-product hospitals into peer groups because it is a multivariate procedure that simultaneously considers all variables affecting performance. This study contributes to increasing the fairness and accuracy of performance comparisons among health care providers and also expands the utilization of statistical analysis through the development of classification criteria reflecting the results of cluster analyses.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors thank the Health Insurance Review and Assessment Service (HIRA) of Korea. This study was conducted using data from the Korean National Health Insurance Claims Database of the HIRA.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Ament, R. P., E. J. Kobrinski, and W. R. Wood. 1981. "Case Mix Complexity Differences between Teaching and Nonteaching Hospitals." *Journal of Medical Education* 56 (11): 894–903.
- Austin, P. C., D. A. Alter, G. M. Anderson, and J. V. Tu. 2004. "Impact of the Choice of Benchmark on the Conclusions of Hospital Report Cards." *American Heart Journal* 148 (6): 1041–6.
- Becker, E. R., and B. Steinwald. 1981. "Determinants of Hospital Case Mix Complexity." *Health Service Research* 16 (4): 439–58.
- Bindman, A. B. 1999. "Can Physician Profiles Be Trusted?" *Journal of American Medical Association* 281 (22): 2142–3.
- Byrne, M. M., C. N. Daw, H. A. Nelson, T. H. Urech, K. Pietz, and L. A. Petersen. 2009. "Method to Develop Health Care Peer Groups for Quality and Financial Comparisons across Hospitals." *Health Services Research* 44 (2): 577–92.
- Do, Y. K. 2007. "Research on Geographic Variations in Health Services Utilization in the United States: A Critical Review and Implication." *Korean Journal of Health Policy and Administration* 17 (1): 91–124.
- Ellis, R. P., and T. G. McGuire. 1988. "Insurance Principles and the Design of a Prospective Payment System." *Journal of Health Economics* 7 (3): 215–37.
- Goodman, D. C., and G. R. Green. 1996. "Assessment Tools: Small Area Analysis." *American Journal of Medical Quality* 11 (1): s12–4.
- Health Insurance Review and Assessment Service (HIRA). 2010. "Major Activities of HIRA: Review" [accessed on August 8, 2010]. Available at http://www.hira.or.kr/cms/rb/rbb_english/13/13_01/01/review.html.
- Jencks, S. F., A. Dobson, P. Willis, and P. H. Feinstein. 1984. "Evaluating and Improving the Measurement of Hospital Case Mix." *Health Care Financing Review* (Annual Supplement, November): 1–11.
- Jian, W., Y. Huang, M. Hu, and X. Zhang. 2009. "Performance Evaluation of Inpatient Service in Beijing: A Horizontal Comparison with Risk Adjustment Based on Diagnosis Related Groups." *BMC Health Services Research* 9 (72): DOI: 10.1186/1472-6963-9-72.
- Lee, K. H. 2007. "The Effects of Case Mix on Hospital Costs and Revenues for Medicare Patients in California." *Journal of Med Systems* 31 (4): 254–62.
- Lee, K. S., and K. H. Chun. 2008. "Analyzing the Specialization Status of Hospital's Services in Korea." *Korean Journal of Health Policy and Administration* 18 (2): 67–85.
- Lee, S. J., and J. Y. Park. 2010. "Changing Trends in Daegu and Gyeongbuk-Based Patients' Use of Health Facilities in Seoul." *Korean Journal of Health Policy and Administration* 20 (4): 19–44.
- MaCurdy, T., N. Theobald, J. Kerwin, and K. Ueda. 2008. "Prototype Medicare Utilization Report Based on Episode Groupers" [accessed on August 10, 2010]. Available at <https://www.cms.gov/reports/downloads/MaCurdy2.pdf>.
- National Health Insurance Corporation (NHIC). 2007. *National Health Insurance Program of Korea: Healthcare Delivery System*. Seoul: National Health Insurance Corporation [Korean].

- National Health Insurance Corporation (NHIC). 2011. "NHI Program" [accessed on May 10, 2011]. Available at <http://www.nhic.or.kr/english/insurance/insurance01.html>.
- Park, H., and Y. Shin. 2004. "Measuring Case-Mix Complexity of Tertiary Care Hospitals Using DRGs." *Healthcare Management Science* 7 (1): 51–61.
- Pink, G. H., G. M. Holmes, R. T. Slifkin, and R. E. Thompson. 2009. "Developing Financial Benchmarks for Critical Access Hospitals." *Healthcare Financing Review* 30 (3): 55–69.
- Pope, G. C., and J. Kautter. 2007. "Profiling Efficiency and Quality of Physician Organizations in Medicare." *Healthcare Financing Review* 29 (1): 31–43.
- Romano, P. S. 2004. "Peer Group Benchmarks Are Not Appropriate for Healthcare Quality Report Cards." *American Heart Journal* 148 (6): 921–3.
- Rosko, M. D., and C. E. Carpenter. 1993. "Development of a Scalar Hospital-Specific Severity of Illness Measure." *Journal of Medical Systems* 17 (1): 25–36.
- Sandy, L. G. 1999. "The Future of Physician Profiling." *Journal of Ambulatory Care Management* 22 (3): 11–6.
- Smith, W. R. 2000. "Evidence for the Effectiveness of Techniques to Change Physician Behavior." *Chest* 118 (2 suppl): 8s–17s.
- Stefos, T., N. Lavallee, and F. Holden. 1992. "Fairness in Prospective Payment: A Clustering Approach." *Health Services Research* 28 (2): 239–61.
- Wachtel, R. E., F. Dexter, B. Barry, and C. Applegeet. 2010. "Use of State Discharge Abstract Data to Identify Hospitals Performing Similar Types of Operative Procedures." *Economics, Education, and Policy* 110 (4): 1146–54.
- Weiss, K. B., and R. Wagner. 2000. "Performance Measurement through Audit, Feedback, and Profiling as Tools for Improving Clinical Care." *Chest* 118 (2 suppl): 53s–8s.
- Zodet, M. W., and J. D. Clark. 1996. "Creation of Hospital Peer Groups." *Clinical Performance and Quality Health Care* 4 (1): 51–7.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.