Kawasaki Journal of Medical Welfare Vol. 31, No. 2, 2026 61－79

Original Paper

Functional Classification of Medical Institutions Using Principal Component Analysis and Clustering of Performance Indicators Based on Hospital Bed Function Reports: A Case Study of Okayama Prefecture

Katsumi HONNO＊1

(Accepted October 29, 2025)

Key words: hospital bed function, principal component analysis, clustering, healthcare classification, regional medical planning, Japan

Abstract

This study aims to classify the functional types of medical institutions based on performance indicators reported under Japan's hospital bed function reporting system. Amid the growing need for objective assessment of actual care provision—especially given the limitations of self-reported classifications—this study adopts a data-driven approach using principal component analysis (PCA) and k-means clustering. Using open data from the 2022 hospital bed function reports in Okayama Prefecture, we conducted PCA on standardized performance indicators (Z-scores) and identified key components reflecting specific medical functions such as acute stroke care, cardiovascular surgery, and chronic inpatient care. Based on these components, a five-cluster classification was constructed, and significant intergroup differences were confirmed using the Kruskal-Wallis test and Holm-adjusted pairwise comparisons. The resulting clusters revealed both conventional functions and intermediate or specialized functional groups not captured by the existing four-function system. These findings suggest that performance-based classification offers a more nuanced and evidence-based understanding of healthcare delivery. The method is applicable to other regions and years, with potential implications for the redesign of medical planning and evaluation systems in line with Japan's evolving healthcare policy.

1. Introduction

In Japan, the rapid aging of the population has brought about significant changes in the structure of demand for medical care. While acute care has traditionally been the central focus, the need for convalescent and chronic care is projected to increase[1)]. In response, the Regional Medical Initiative (RMI) was launched in 2015, aiming to promote functional differentiation and coordination of hospital beds by 2025, across four categories of medical care: Advanced acute, acute, convalescent, and chronic[2,3)]. Optimizing the reallocation of regional medical resources and establishing care delivery systems that better meet patient

＊1 Department of Health Data Science, Faculty of Health and Welfare Services Administration
Kawasaki University of Medical Welfare, 288 Matsushima, Kurashiki, 701-0193, Japan
E-Mail: khn@mw.kawasaki-m.ac.jp

needs has become a pressing policy priority.

However, the current hospital bed function reporting system faces several structural limitations[4]. Under this system, medical institutions self-report the function of each ward. In reality, however, patients often transition from acute to convalescent stages during hospitalization, making fixed ward-based classifications insufficient[5]. Furthermore, it is common for patients at different stages of care to be housed in the same ward, limiting the accuracy and reliability of self-reported data[6]. As a result, gaps between actual care delivery and reported functional categories may arise, impeding the precise implementation of regional healthcare policy[7].

In response to these issues, recent years have seen growing attention to data-driven approaches for objectively assessing hospital functions[8]. Among them, statistical methods using performance indicators—such as the number of ambulance transports, surgeries performed, return-to-home rate, bed occupancy rate, number of discharges, percentage of full-time physicians, and nurse staffing ratios—have been shown to be effective[9,10]. These objective methods are expected to supplement the limitations of self-report-based systems and contribute to the more appropriate allocation of medical resources within regional health planning.

In fiscal year 2024, the Ministry of Health, Labour and Welfare established a Study Group on New Regional Healthcare Strategies, tasked with re-evaluating the long-term medical care delivery system through 2040[11]. This initiative emphasizes not only hospital bed coordination but also stronger linkages with outpatient, home-based, and long-term care services. It calls for a more flexible, sustainable care delivery system tailored to local characteristics—one that requires a more substantive, function-based understanding of healthcare institutions, rather than a superficial adjustment of bed numbers[11].

The aim of this study is to develop an objective, data-driven classification of healthcare institution functions—spanning acute to chronic care—based on publicly available medical performance indicators. Using principal component analysis (PCA) and k-means clustering, we conducted a simulation to identify typologies of medical institutions. Our goal is to provide foundational evidence to inform more effective regional medical resource reallocation within the framework of Japan's evolving regional healthcare policies.

2. Methods

2.1 Target facilities and dataset

The subjects of this study were medical institutions that reported their bed functions under the national Hospital Bed Function Reporting System, according to the four functional categories: "advanced acute," "acute," "recovery," and "chronic." The target area comprised five secondary medical care districts in Okayama Prefecture, and eligible facilities were identified using the 2022 reporting data[12].

The dataset consisted of medical performance indicators submitted by each institution as part of the bed function report. Key indicators (referred to as "X items") relevant to the purpose of this study were selected with reference to prior research[9,10]. These variables were standardized as Z-scores for further analysis. In selecting indicators, consistency with the medical function evaluation framework used in regional medical planning was considered. In addition to raw variables from the reporting system, derived variables computed from these were also included. A summary of the raw and derived indicators is presented in Table 1 (collectively referred to as the "dataset").

All data used in this study are publicly available from the Ministry of Health, Labour and Welfare and were confirmed to comply with the Act on the Protection of Personal Information and relevant ethical guidelines for research.

2.2 Analytical indicators and statistical methods

This study analyzed multiple performance indicators from the hospital bed function report. The analysis consisted of the following stages: principal component analysis (PCA), clustering, and inter-cluster statistical testing.

2.2.1 Identification of high-contribution indicators and principal component structure

PCA was performed on the standardized (Z-score) indicators across all institutions. The contribution

Table 1 List of medical performance indicators and variable definitions

Indicator description	Indicator description
X002_Open Data Medical Institution Code	X071_Emergency care admission
X003_Medical Institution Name	X072_Specified intensive care unit management fee
X005_Secondary Medical Area Code	X073_HCU inpatient care management fee
X006_Secondary Medical Area Name	X074_SCU inpatient care management fee
X009_Municipal Code	X075_Newborn specific intensive care unit management fee
X010_Municipal Name	X076_General perinatal specific intensive care unit management fee
X011_Physician Index	X077_Pediatric inpatient care management fee
X012_Full-time Physician Ratio	X078_Recovery Rehabilitation Unit
X013_All Physicians	X079_Regional Comprehensive Care Unit
X014_Dentist Number	X080_Basic Inpatient Care for Disabled
X016_Acute General Admission	X081_Palliative Care Unit
X017_Regional General Admission	X091_Basic Short Stay Surgery
X018_General Ward Special Admission	X095_Total Number of Operations
X019_General Hospital Beds	X096_Count of Delivery
X020_Basic Inpatient Fee for Convalescent Ward	X097_Total Number of General Anesthesia Operations
X024_Care Unit Beds	X098_Skin. Subcutaneous Surgery
X025_All Beds	X099_Musculoskeletal Surgery
X026_General Bed Ratio	X100_Nervous System. Cranial Surgery
X027_Bed Occupancy Rate	X101_Eye Surgery
X028_Average Length of Stay	X102_Ear, Nose & Throat Surgery
X029_Inpatients	X103_Facial. Neck Surgery
X030_Return to Home Rate	X104_Thoracic Surgery
X031_Total Number of Patients in Ward	X105_Cardiac. Vascular Surgery
X032_Transfers to Medical Facilities	X106_Abdominal Surgery
X033_Transfers to Nursing Homes, etc.	X107_Urinary Tract. Adrenal Surgery
X034_Number of Discharged Patients	X108_Genital_Surgery
X035_Number of All Nurses	X109_Dental_Surgery
X036_Other Nursing Staff	X110_Cardiopulmonary Bypass Surgery
X037_Nursing Need	X111_Thoracoscopic_Surgery
X041_CT	X112_Laparoscopic_Surgery
X042_MRI	X113_Endoscopic_Surgical_Aid_Instruments_for_Surgery
X045_Endoscopic Surgical Support Equipment	X114_Malignancy_Surgery
X051_Ratio of Rehabilitation Patients	X115_Histopathology_Patient_Sampling
X052_Average Rehab Units	X116_Rapid_Intraoperative_Histopathology_Sampling
X053_Number of All Rehab Staff	X117_Radiotherapy
X054_Rehab Performance Index	X118_Chemotherapy
X061_Number of Patients Seen on Holidays	X119_Cancer Patient Guidance and Management Fee
X062_Number of Patients Admitted on Holidays	X120_Local Continuous Infusion of Antineoplastic Drugs
X063_Night. Patients Seen after Hours	X122_Hyperacute Stroke Surcharge
X064_Nighttime. Number of Patients Admitted after Hours	X123_tPA Administration
X065_Number of Ambulance Admissions	X124_Cerebral Endovascular Surgery
	X125_Percutaneous Coronary Intervention (PCI)

Z.X indicates a standardized variable derived from each X indicator value.
Bold text denotes derived variables calculated based on raw data item values.

ratio and factor loadings for each principal component (PC1-PC10) were calculated. Principal components with a cumulative contribution ratio exceeding 70% were retained for analysis. For each variable, the average absolute contribution across the retained components and the standard deviation across clusters were computed. Based on these values, a statistical score was calculated, and the top eight indicators were selected as "high-contribution indicators."

2.2.2 Principal component scores by cluster

The principal component scores (PC1-PC10) derived from the PCA were used as input for k-means clustering to examine both 4-cluster and 5-cluster solutions. The mean score for each principal component was calculated by cluster, allowing visualization of their distributional structure. For each principal component, the most contributing indicator, its functional interpretation, and its tentative classification of medical function were also noted.

2.2.3 Median Z-scores by cluster for major indicators

For the eight high-contribution indicators, median Z-scores were computed for each cluster to facilitate visual comparison of quantitative trends.

2.2.4 Inter-cluster comparison using the kruskal-wallis test

To assess whether statistically significant differences existed among clusters for the eight major indicators, the non-parametric Kruskal-Wallis test was employed. This test is appropriate for comparing medians across groups and is robust to non-normality. The resulting p-values were presented alongside cluster-specific medians to clarify distributional patterns and inter-cluster differences.

2.2.5 Pairwise comparison with holm adjustment

For indicators showing significance in the Kruskal-Wallis test, pairwise comparisons between clusters were conducted using the Mann-Whitney U test. To control for Type I error in multiple comparisons, the Holm correction was applied. Adjusted p-values were used to determine statistical significance.

2.2.6 Functional labeling of clusters

Based on the cumulative results from Sections 2.1 to 2.5, each cluster was assigned a descriptive functional label. This labeling was based on patterns in principal component scores, the distribution of high-contribution indicators, and the results of statistical testing. Functional naming also considered representative variables, their principal component loadings, cluster-level medians, and inferred clinical roles.

Statistical significance was defined as $p < 0.05$, and all tests were two-tailed.

2.3 Analytical environment

All statistical analyses were performed using EZR (version 1.68)[13), a GUI-based software platform for medical statistics built on R (version 4.3.1 "Beagle Scouts"; R Foundation for Statistical Computing, 2023) and R Commander (version 2.9-1), developed by the Department of Hematology at Saitama Medical Center, Jichi Medical University[13)].

3. Results

The analytical procedures used in this study were summarized in Figure 1.

3.1 Principal component analysis (PCA)

PCA was conducted on Z-standardized medical performance indicators derived from hospital bed function reports. A total of 10 principal components (PC1-PC10) were extracted, cumulatively accounting for over 70% of the variance. For each indicator, its average absolute loading across components and the standard deviation across clusters were computed. Based on these "statistical scores," the top eight high-contribution indicators were identified (Table 2).

The indicator with the highest average contribution was Z.X125_Percutaneous Transluminal Coronary Intervention (mean contribution: 0.26), which exhibited the highest loading on PC1 (0.53), along with relatively high loadings on PC6 (0.43) and PC8 (0.43). Z.X124_Cerebral Endovascular Surgery (0.25) also loaded highest on PC1 (0.64), with considerable loadings on PC2 (0.50) and PC4 (0.34). Similarly, Z.X123_tPA

1. **Collection of medical performance data and indicator construction (hospital bed function reports)**

   ↓

2. **Standardization using Z-scores (analytical variables)**

   ↓

3. **Dimensionality reduction via principal component analysis (PCA) / Extraction of eight high-contribution indicators**

   ↓ → *Table 2: Medical Performance Indicators and Cumulative Contribution Rates from PCA*

4. **k-means clustering based on PCA Scores (4 vs. 5 clusters)**

   ↓ → *Table 3-1: Mean Principal Component Scores for Each Cluster (4 vs. 5)*

5. **k-means clustering based on key indicators (4 vs. 5 clusters)**

   ↓ → *Table 3-2: Cluster-wise Standardized Scores for Eight Major Indicators (4 vs. 5)*

6. **Kruskal-wallis test**

   ↓ → *Table 4 (Upper): Z-Score Distributions of Major Indicators by Cluster*

7. **Pairwise comparison with holm-adjustment (Mann-whitney U test)**

   ↓ → *Table 4 (Lower): Intergroup Differences for Eight Indicators (p-values after Correction)*

8. **Classification and interpretation of medical institution functions based on the results**

   → *Table 5: Evaluation of Naming Validity and Statistical Basis for the Five-Cluster Functional Classification*

Figure 1 Analytical flow

Administration and Z.X122_Hyperacute Stroke Surcharge (both 0.25) had strong loadings on PC1 (0.64), indicating a common reflection of acute stroke care functions.

On the other hand, Z.X110_Cardiopulmonary Bypass Surgery (0.25) exhibited substantial loadings not only on PC1 (0.58) but also on PC6 (0.47) and PC8 (0.46), demonstrating its multifaceted contribution to cardiovascular and surgical care.

The Z.X026_General Bed Ratio (0.25) contributed most to PC6 (0.47), indicating its distinctive role in capturing structural bed configuration. In addition, Z.X020_Basic Inpatient Fee for Convalescent Wards (0.25) and Z.X028_Average Length of Stay (0.24) were mainly associated with PC1, PC6, and PC8.

These results suggest that specific clusters of performance indicators—such as those related to cardiovascular treatment, stroke care, bed configuration, and hospitalization duration—are coherently organized within the principal component structure. This provides a meaningful foundation for subsequent functional clustering of medical institutions.

3.2 Cluster distributions of principal component scores

Using the PC1-PC10 scores derived from PCA, k-means clustering was applied to generate both four- and five-cluster solutions. The mean score for each cluster on each principal component was calculated to visualize institutional differentiation (Table 3-1).

For instance, in PC1, Cluster 4-4 had the highest score (31.26), followed by Cluster 4-3 (11.92), while Clusters 4-1 and 4-2 had notably lower scores (-1.51 and -0.63, respectively), indicating a clear polarity among institutions. A similar pattern was observed in the five-cluster model, with Clusters 5-5 and 5-4 scoring highly (31.26 and 11.92), whereas Clusters 5-2 (-2.48) and 5-1 (-1.29) scored low.

For PC2, Clusters 4-4 and 5-5 recorded high scores (5.57), while Clusters 4-3 and 5-3 had negative scores (-2.19 and -0.95), suggesting functional divergence across clusters.

PC6 exhibited a distinct separation, with Clusters 4-3 and 5-4 scoring positively (3.88), while Clusters 4-4 and 5-5 scored negatively (-2.87). Since PC6 was strongly associated with cardiovascular indicators such as cardiopulmonary bypass and percutaneous coronary interventions, this suggests that these clusters differ in terms of cardiac and surgical capabilities.

These trends demonstrate that principal component scores by cluster effectively capture variations in institutional function. In particular, PC1 and PC6 showed pronounced inter-cluster differences, indicating

Table 2 Principal component analysis (PCA) of medical performance indicators and cumulative contribution rate

Variable Description	Mean contribution (absolute value)	Inter-cluster variability (standard deviation)	Statistical score	Principal component (maximum contribution)	Maximum loading (signed)	Absolute value	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Z.X125_Percutaneous Coronary Intervention (PCI)	0.2647	0.1716	1.5424	PC1	0.5294	0.5294	0.5294	0.4065	0.1870	0.2140	0.0578	0.4250	0.1552	0.4251	0.0231	0.2241
Z.X124_Cerebral Endovascular Surgery	0.2546	0.1930	1.3193	PC1	0.6352	0.6352	0.6352	0.5031	0.1561	0.3401	0.2213	0.2909	0.0257	0.1800	0.0694	0.1245
Z.X123_tPA Administration	0.2546	0.1930	1.3193	PC1	0.6352	0.6352	0.6352	0.5031	0.1561	0.3401	0.2213	0.2909	0.0257	0.1800	0.0694	0.1245
Z.X122_Hyperacute Stroke Surcharge	0.2546	0.1930	1.3193	PC1	0.6352	0.6352	0.6352	0.5031	0.1561	0.3401	0.2213	0.2909	0.0257	0.1800	0.0694	0.1245
Z.X110_Cardiopulmonary Bypass Surgery	0.2532	0.1909	1.3267	PC1	0.5823	0.5823	0.5823	0.1030	0.2841	0.1353	0.0684	0.4674	0.1974	0.4645	0.0250	0.2048
Z.X026_General Bed Ratio	0.2514	0.1930	1.3029	PC6	0.4693	0.4693	0.3140	0.4385	0.4896	0.2704	0.1035	0.4693	0.2062	0.0794	0.0377	0.1770
Z.X020_Basic Inpatient Fee for Convalescent Ward	0.2464	0.1693	1.4554	PC1	0.4953	0.4953	0.4953	0.3383	0.1727	0.3162	0.2079	0.3255	0.0506	0.2737	0.0596	0.2076
Z.X028_Average Length of Stay	0.2448	0.1698	1.4420	PC1	0.4892	0.4892	0.4892	0.3380	0.1691	0.3140	0.1999	0.3187	0.0546	0.2746	0.0587	0.2060

Table 3-1 Cluster-wise mean scores of principal components and model-based distribution comparison (4 vs. 5 clusters): functional characteristics of clusters based on PCA Results

CLASS	4				5					Representative variable		Tentative medical function	Reasons for contribution
Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5				
PC1	-1.5063	-0.6344	11.9248	31.2646	-1.2853	-2.4784	-0.6344	11.9248	31.2646	Z.X124_Cerebral Endovascular Surgery	Z.X123_tPA Administration	Advanced Acute Care (Stroke) / Hyperacute Stroke Management	Characterized primarily by "cerebral endovascular surgery" and "tPA administration," reflecting capabilities for advanced and hyperacute stroke care. Indicates the institution's capacity for endovascular stroke treatment and early-stage thrombolytic therapy (tPA).
PC2	0.1887	-0.9466	-2.1894	5.5736	0.0857	0.6419	-0.9466	-2.1894	5.5736	Z.X124_Cerebral Endovascular Surgery	Z.X123_tPA Administration	Advanced Acute Care (Stroke) / Hyperacute Stroke Management	Characterized primarily by "cerebral endovascular surgery" and "tPA administration," reflecting capabilities for advanced and hyperacute stroke care. Indicates the institution's capacity for endovascular stroke treatment and early-stage thrombolytic therapy (tPA).
PC3	0.3394	-0.8139	-1.2206	-3.1314	0.1576	1.1395	-0.8139	-1.2206	-3.1314	Z.X110_Cardiopulmonary Bypass Surgery	Z.X026_General Bed Ratio	Cardiovascular Surgery / Advanced Critical Care / Balanced Acute-to-Recovery Care	Characterized by "cardiopulmonary bypass surgery" and "general bed ratio," reflecting cardiovascular surgical capabilities and tertiary emergency care. Also indicates a functional balance shaped by bed composition.
PC4	-0.6634	3.0206	2.0364	-4.0258	-0.2768	-2.3645	3.0206	2.0364	-4.0258	Z.X026_General Bed Ratio	Z.X020_Basic Inpatient Fee for Convalescent Ward	Balanced Acute-to-Recovery Care / Chronic and Long-Term Care	Characterized by "general bed ratio" and "basic inpatient fee for convalescent wards," reflecting a balanced structure from acute to recovery phase and chronic, long-term care. Suggests a functional structure based on bed mix and indicators of chronic or long-stay care.
PC5	0.3988	-2.3250	2.7834	-3.0274	0.8669	-1.6608	-2.3250	2.7834	-3.0274	Z.X026_General Bed Ratio	Z.X020_Basic Inpatient Fee for Convalescent Ward	Balanced Acute-to-Recovery Care / Chronic and Long-Term Care	Characterized by "general bed ratio" and "basic inpatient fee for convalescent wards," reflecting a balanced structure from acute to recovery phase and chronic, long-term care. Suggests a functional structure based on bed mix and indicators of chronic or long-stay care.
PC6	-0.0936	-0.3680	3.8827	-2.8685	-0.4099	1.2983	-0.3680	3.8827	-2.8685	Z.X110_Cardiopulmonary Bypass Surgery	Z.X125_Percutaneous Coronary Intervention (PCI)	Cardiovascular Surgery / Advanced Critical Care / Advanced Acute Care (Cardiology)	Characterized by "cardiopulmonary bypass surgery" and "percutaneous coronary intervention (PCI)," indicating advanced cardiovascular surgery and critical care capacity. Reflects emergency cardiac intervention for conditions such as myocardial infarction.
PC7	0.0709	-0.0778	-0.9614	0.2869	0.1663	-0.3487	-0.0778	-0.9614	0.2869	Z.X028_Average Length of Stay	Z.X026_General Bed Ratio	Chronic / Community-based / Balanced Acute-to-Recovery Care	Characterized by "average length of stay" and "general bed ratio," reflecting chronic or community-based care with balanced acute-to-recovery functions. Indicates a tendency toward long-term hospitalization and a functionally balanced bed structure.
PC8	-0.0199	-0.0097	0.5499	-0.4922	-0.1155	0.4009	-0.0097	0.5499	-0.4922	Z.X110_Cardiopulmonary Bypass Surgery	Z.X125_Percutaneous Coronary Intervention (PCI)	Cardiovascular Surgery / Advanced Critical Care / Advanced Acute Care (Cardiology)	Characterized by "cardiopulmonary bypass surgery" and "percutaneous coronary intervention (PCI)," indicating advanced cardiovascular surgery and critical care capacity. Reflects emergency cardiac intervention for conditions such as myocardial infarction.
PC9	-0.0389	0.0874	0.3994	-0.2001	-0.0721	0.1070	0.0874	0.3994	-0.2001	Z.X028_Average Length of Stay	Z.X124_Cerebral Endovascular Surgery	Chronic / Community-based / Advanced Acute Care (Stroke)	Characterized by "average length of stay" and "cerebral endovascular surgery," representing chronic and community-based care with acute stroke treatment capability. Reflects both a tendency for long hospital stays and advanced stroke care skills.
PC10	-0.0179	0.2798	-1.1171	1.1065	0.0698	-0.4040	0.2798	-1.1171	1.1065	Z.X028_Average Length of Stay	Z.X125_Percutaneous Coronary Intervention (PCI)	Chronic / Community-based / Advanced Acute Care (Cardiology)	Characterized by "average length of stay" and "percutaneous coronary intervention (PCI)," indicating community-based and chronic care alongside advanced cardiovascular intervention capacity. Reflects both long-stay orientation and the ability to treat acute cardiac events.

Table 3-2 Cluster-wise average Z-scores of the eight key indicators and model-based comparison

CLASS	4				5				
Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Z.X125_Percutaneous Coronary Intervention (PCI)	-0.0596	-0.1417	0.2818	2.5735	-0.0409	-0.1417	-0.1417	0.2818	2.5735
Z.X124_Cerebral Endovascular Surgery	-0.0842	-0.0842	-0.0842	3.8739	-0.0842	-0.0842	-0.0842	-0.0842	3.8739
Z.X123_tPA Administration	-0.0842	-0.0842	-0.0842	3.8739	-0.0842	-0.0842	-0.0842	-0.0842	3.8739
Z.X122_Hyperacute Stroke Surcharge	-0.0842	-0.0842	-0.0842	3.8739	-0.0842	-0.0842	-0.0842	-0.0842	3.8739
Z.X110_Cardiopulmonary Bypass Surgery	-0.0607	-0.1463	0.0102	3.2823	-0.0412	-0.1463	-0.1463	0.0102	3.2823
Z.X026_General Bed Ratio	-0.1248	0.2345	0.8087	0.8087	0.2613	-1.8237	0.2345	0.8087	0.8087
Z.X020_Basic Inpatient Fee for Convalescent Ward	0.0740	-0.0309	-0.7278	-0.7278	-0.2306	1.4142	-0.0309	-0.7278	-0.7278
Z.X028_Average Length of Stay	0.0952	-0.2355	-0.4816	-0.4977	-0.1508	1.1776	-0.2355	-0.4816	-0.4977

their importance in functional classification.

3.3 Cluster-wise distribution of eight key medical performance indicators

For the eight high-contribution indicators identified via PCA, the median Z-scores were calculated for each cluster to compare the central tendencies in institutional performance (Table 3-2). This allowed for assessment not only of overall PCA structure, but also of indicator-specific profiles.

3.3.1 Acute stroke care indicators

The indicators Z.X124_Cerebral Endovascular Surgery, Z.X123_tPA Administration, and Z.X122_ Hyperacute Stroke Surcharge showed extremely high median Z-scores (3.87) in Clusters 4-4 and 5-5. All other clusters had consistently low medians (-0.08), clearly distinguishing these clusters as having specialized capacity in acute stroke care.

3.3.2 Cardiovascular and cardiac surgery indicators

Z.X110_Cardiopulmonary Bypass Surgery and Z.X125_Percutaneous Coronary Intervention were also concentrated in Clusters 4-4 and 5-5, with medians of 3.28 and 2.57, respectively. In contrast, other clusters had near-zero or negative medians for these indicators (e.g., -0.15 and -0.14 in Cluster 4-2), revealing their uneven distribution and supporting the distinction of cardiovascular functions.

3.3.3 Bed configuration and length of stay

Z.X026_General Bed Ratio scored consistently high (0.81) in Clusters 4-3 and 4-4, indicating stronger general bed availability. Notably, Cluster 5-2 showed a markedly low value (-1.82), suggesting limited general bed allocation. Conversely, Z.X020_Basic Inpatient Fee for Convalescent Wards was highest in Cluster 5-2 (1.41), identifying this cluster as indicative of chronic care function.

Likewise, Z.X028_Average Length of Stay was high in Cluster 5-2 (1.18) and low in Clusters 4-4 and 5-5 (-0.50), suggesting that this metric serves as a strong marker of chronic and long-term care orientation.

3.4 Validation of the five-cluster structure and identification of functional differences

To evaluate the validity of the five-cluster structure constructed using the eight high-contribution indicators extracted from PCA, Kruskal-Wallis tests were conducted for each indicator to assess intergroup differences. For indicators exhibiting statistical significance, pairwise comparisons using the Mann-Whitney U test were conducted with Holm's correction for multiple testing (Table 4).

The Kruskal-Wallis test revealed significant differences across all eight major indicators. For instance, Z.X125_Percutaneous Coronary Intervention yielded a p-value of 8.03×10^{-10}. Similarly, the three stroke-related indicators—Z.X124_Cerebral Endovascular Surgery, Z.X123_tPA Administration, and Z.X122_ Hyperacute Stroke Surcharge—each showed a p-value of 2.46×10^{-9}. Other indicators such as Z.X110_ Cardiopulmonary Bypass Surgery ($p = 3.08 \times 10^{-10}$), Z.X026_General Bed Ratio ($p = 7.91 \times 10^{-12}$), Z.X020_ Basic Hospitalization Fee for Convalescent Ward ($p = 6.85 \times 10^{-9}$), and Z.X028_Average Length of Stay ($p = 2.30 \times 10^{-10}$) also demonstrated statistically significant inter-cluster variation. These results indicate systematic and robust differences in the distribution of performance indicators among the five clusters.

Holm-adjusted pairwise comparisons further supported the presence of inter-cluster differences, especially for cardiovascular and cardiac surgical functions. For example, comparisons for Z.X110_Cardiopulmonary Bypass Surgery revealed significant differences between Cluster 5 and Cluster 1 ($p = 1.01 \times 10^{-8}$) and between Cluster 5 and Cluster 2 ($p = 0.00$).

In contrast, although the three stroke-related indicators were statistically significant in the overall Kruskal-Wallis test, many of the pairwise comparisons yielded p-values of 1.00 after Holm correction. This suggests that these indicators had highly skewed distributions—with extreme values concentrated in a limited number of clusters (primarily Cluster 5), and uniformly low values in others—resulting in fewer statistically significant pairwise differences.

Notable differences were also observed for indicators associated with chronic care and long-term hospitalization. For example, Z.X020_Basic Inpatient Fee for Convalescent Wards showed significant differences between Cluster 2 and Cluster 3 ($p = 0.00$), and between Cluster 2 and Cluster 4 ($p = 0.00$). For Z.X028_Average Length of Stay, significant differences were also observed: $p = 0.00$ between Cluster 2 and

Table 4　Kruskal-wallis test and holm-adjusted pairwise comparisons for eight key indicators

Cluster	Z.X125_Percutaneous coronary intervention (PCI)	Z.X124_Cerebral endovascular surgery	Z.X123_tPA Administration	Z.X122_Hyperacute stroke surcharge	Z.X110_Cardiopulmonary bypass surgery	Z.X026_General bed ratio	Z.X020_Basic inpatient fee for convalescent ward	Z.X028_Average length of stay
Kruskal-Wallis p-value	8.02998E-10	2.46285E-09	2.46285E-09	2.46285E-09	3.08474E-10	7.90751E-12	6.8547E-09	2.30371E-10
Cluster 1 (Median)	-0.1417	-0.0842	-0.0842	-0.0842	-0.1463	0.8087	-0.7278	-0.3524
Cluster 2 (Median)	-0.1417	-0.0842	-0.0842	-0.0842	-0.1463	-1.9043	1.2174	1.1068
Cluster 3 (Median)	-0.1417	-0.0842	-0.0842	-0.0842	-0.1463	0.8087	-0.7278	-0.3503
Cluster 4 (Median)	-0.1417	-0.0842	-0.0842	-0.0842	-0.1463	0.8087	-0.7278	-0.4802
Cluster 5 (Median)	-0.0114	-0.0842	-0.0842	-0.0842	3.6054	0.8087	-0.7278	-0.4926

Holm-adjusted p-value / Mann-whitney U test

Cluster Comparison Pairs	Z.X125_Percutaneous coronary intervention (PCI)	Z.X124_Cerebral endovascular surgery	Z.X123_tPA Administration	Z.X122_Hyperacute stroke surcharge	Z.X110_Cardiopulmonary bypass surgery	Z.X026_General bed ratio	Z.X020_Basic inpatient fee for convalescent ward	Z.X028_Average length of stay
1 vs 2	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000***	0.0000***	0.0000***
1 vs 3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3667
1 vs 4	0.0000***	1.0000	1.0000	1.0000	0.0008***	0.1788	0.1982	0.0015**
1 vs 5	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.6114	0.6437	0.0316*
2 vs 3	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000***	0.0011**	0.0000***
2 vs 4	0.0139*	1.0000	1.0000	1.0000	0.0877	0.0000***	0.0009***	0.0007***
2 vs 5	0.0019**	0.1134	0.1134	0.1134	0.0019**	0.0013**	0.0491	0.0316*
3 vs 4	0.0082**	1.0000	1.0000	1.0000	0.0645	0.3039	0.4162	0.0001***
3 vs 5	0.0007***	0.0728	0.0728	0.0728	0.0008***	0.6572	0.7771	0.0038**
4 vs 5	1.0000	1.0000	1.0000	1.0000	0.6315	1.0000	1.0000	0.3667

*p<0.05:**

*p<0.01:***

*p<0.001:****

Cluster 4, and p = 0.03 between Cluster 2 and Cluster 5.

These findings demonstrate that the five-cluster structure derived from PCA and k-means clustering effectively reflects statistically significant functional distinctions among medical institutions. In particular, the quantitative identification of clusters specialized in cardiovascular, stroke, and chronic care functions highlights the analytical validity of the classification model and its potential utility for informing future regional healthcare planning and resource allocation.

3.5 Final refinement and naming of functional clusters

Based on the results of the PCA and clustering analyses, each cluster was descriptively labeled according to its functional characteristics, as summarized in Table 5. Naming was based on the distribution of median Z-scores, the contribution of key indicators, and statistically significant performance differences.

Cluster 1 (88 facilities) was labeled General Acute Care. No extreme values were observed in its Z-score distribution, indicating a balanced profile without specific specialization. These institutions appeared to provide standard acute care without a pronounced functional tilt. The classification was deemed appropriate.

Cluster 2 (20 facilities) was labeled Chronic/Convalescent Care. High Z-scores in Average Length of Stay and Basic Inpatient Fee for Convalescent Wards clearly indicated specialization in long-term and chronic care. These indicators showed statistically significant differences even after multiple testing correction, supporting the classification's validity as highly appropriate.

Cluster 3 (23 facilities) was designated the Balanced Recovery-Phase Type. Although indicators related to acute care were generally low, a high General Bed Ratio suggested moderate inpatient function, potentially within a community-based or transitional care context. This classification was also considered appropriate.

Cluster 4 (7 facilities) was tentatively labeled Community-Based Comprehensive Care. This group exhibited no striking Z-score deviations and showed proximity to other clusters in the PCA distribution, implying unclear structural uniqueness. While statistical distinctions were modest, the classification was evaluated as partially appropriate, with the possibility that this group represents a functionally mixed model addressing broad community needs.

Cluster 5 (3 facilities) was labeled Advanced acute/Lifesaving Type. This group showed overwhelmingly high Z-scores for advanced procedures such as Cardiopulmonary Bypass, PCI, and tPA Administration, along with pronounced principal component scores. Statistically significant differences were found for multiple indicators, suggesting specialization in tertiary emergency care and Advanced acute services. This classification was considered highly appropriate.

In summary, the cluster classification was derived from both the structural patterns of performance indicators and the results of multivariate statistical analysis, enabling the differentiation of medical institution types with a substantial degree of validity. In particular, Clusters 2 and 5 demonstrated clearly specialized functional roles, affirming the usefulness of PCA and clustering as tools for objective, data-driven functional classification in healthcare system planning.

4. Discussion

4.1 Overview and significance of the study

This study aimed to empirically classify medical institution functions and visualize their structural characteristics through the statistical analysis of medical performance indicators reported in the hospital bed function system. A set of performance indicators, selected with reference to prior studies[9,10], was summarized using PCA, and clustering was subsequently performed using the k-means method based on the resulting principal component scores. As a result, five statistically supported functional groups were identified, revealing diverse configurations of medical institutions that differed from those defined by the existing institutional framework.

In Japan's regional medical planning, the current classification system categorizes beds into four functional types—advanced acute, acute, recovery, and chronic—based on ward-level self-reporting by institutions. However, this approach does not necessarily reflect the actual clinical activities or resource

Table 5 Evaluation of appropriateness and statistical basis for proposed functional labels in the five-cluster classification

Cluster	Functional label / medical function name	Number of facilities	Evaluation of appropriateness	Basis / Rationale	Description of characteristics
Cluster 1	General Acute Care Type	88	Appropriate	The median Z-scores are moderate, and the cluster can be interpreted as a balanced type.	Exhibits an average Z-score distribution, suggesting an intermediate function between acute and chronic care.
Cluster 2	Chronic and Long-term Care Type	20	Highly Appropriate	High values in length of stay and convalescent ward surcharge clearly indicate chronic care functions.	High values for convalescent ward surcharge and length of stay indicate a strong tendency toward chronic and long-term care.
Cluster 3	Balanced Recovery-phase Care Type	23	Appropriate	Although acute care indicators are low, the high general bed ratio suggests a community-based care orientation.	Although acute care indicators are low, the high general bed ratio suggests responsiveness to community-based care.
Cluster 4	Community-based Integrated Care Type	7	Partially Appropriate	This cluster is in close proximity to other clusters, making its differentiation somewhat ambiguous.	No pronounced deviations observed overall, suggesting a potential role in comprehensive community-based care.
Cluster 5	Advanced Acute and Lifesaving Care Type	3	Highly Appropriate	Advanced acute care procedures (e.g., PCI and cardiopulmonary bypass) are statistically and markedly prominent.	Extremely high values in cardiopulmonary bypass, PCI, and tPA administration indicate advanced emergency and acute care functions.

use, and discrepancies between institutional declarations and practice have been frequently pointed out[5]. In contrast, the present study is significant in that it demonstrates the feasibility of a performance-based classification scheme using multivariate analysis, thus offering a more realistic and objective perspective.

Moreover, the "Summary of a New Regional Medical Plan" released by the Ministry of Health, Labour and Welfare in 2024 highlights a policy transition from static bed management to securing flexible medical functions, emphasizing the need for objective and evidence-based functional classification[14]. The analytical framework proposed in this study is responsive to such policy directions and has the potential to enhance the precision and practical applicability of regional medical resource planning.

Importantly, the five functional clusters derived in this study both encompass the existing four categories and reveal intermediate or specialized types not explicitly defined in the current system—such as the "Advanced acute/lifesaving type" and "chronic/long-term care type." While these facility types have traditionally been overlooked under the administrative classification, the findings underscore their distinct roles in the local healthcare ecosystem. As such, this study contributes valuable insights for future revisions of functional evaluation indicators and the development of more flexible and sustainable healthcare systems.

4.2 Consistency and divergence from existing classification systems

The five functional clusters identified in this study generally aligned with the four categories defined in the hospital bed function reporting system—namely, Advanced acute, acute, recovery, and chronic care. However, the analysis also visualized functionally intermediate or overlapping groups not captured within the current institutional framework.

For instance, facilities in Cluster 5, designated as the Advanced acute/lifesaving type, showed statistically significant concentrations of advanced procedures such as cardiopulmonary bypass, percutaneous coronary intervention (PCI), and tPA administration. These institutions appear to serve tertiary emergency care roles within the high-acuity spectrum. While broadly consistent with the "Advanced acute" category, this cluster adds granularity by distinguishing a core subset of institutions with exceptional intervention capabilities—essentially a "high-functioning stratum" within the Advanced acute class.

Similarly, Cluster 2, classified as chronic/convalescent care, showed high performance in indicators such as length of stay and basic convalescent ward fees. These findings support its correspondence with the chronic care category. However, by using institutional-level performance data rather than ward-level declarations, the analysis revealed variability within this group, highlighting facility types specializing in long-term inpatient care that might otherwise be administratively aggregated under a single label.

More notably, Clusters 1 and 3 exhibited intermediate functional profiles not clearly represented in the system. Cluster 1, the general acute care type, displayed moderate performance across both acute and recovery-phase indicators, forming a balanced, non-specialized group within the broader acute care domain. Cluster 3, the balanced recovery-phase type, featured low acute care intensity but high general bed ratios, a pattern often observed in mid-sized, community-based hospitals. These clusters demonstrated that performance-based classification can capture functional variation not delineated by existing categories.

Cluster 4, though statistically less distinct, was located near the centroid of the principal component space, suggesting a functionally integrative profile. This community-based comprehensive care type may represent multipurpose facilities supporting continuity of care, including integration with nursing care services. Although these institutions are often ambiguously categorized in the system, they may play critical roles in community-based care networks and warrant further attention in future system reforms.

In sum, while the current ward-based classification system provides a general directional framework, the data-driven functional groups extracted in this study underscore both the limitations and the potential of institutional classification. In particular, performance-based classification at the institutional level may complement existing systems by revealing nuanced, overlapping, or intermediate care functions—offering a flexible approach to regional healthcare management (appendix).

4.3 Analytical strengths and limitations

This study employed a combined PCA and k-means clustering approach to summarize and classify medical performance indicators derived from the hospital bed function reports. PCA enabled dimensionality reduction based on covariance structure, extracting 10 principal components (PC1-PC10) that explained approximately 70% of the variance. These were then used as input for clustering.

Importantly, the first six principal components (PC1-PC6) showed clear structural alignment with major medical functions such as acute stroke care, cardiovascular interventions, and chronic care. This indicates that the use of PCA as a preprocessing step in k-means clustering provides a robust analytical framework for objectively grouping institutions, while avoiding overfitting or instability due to high dimensionality and multicollinearity.

Nonetheless, the methodology has several limitations. PCA, being a linear transformation technique, may not fully capture complex nonlinear relationships among variables. Moreover, interpretation of the resulting components and their labeling inevitably involve some degree of subjectivity, requiring domain expertise for meaningful classification[15].

The k-means algorithm also has methodological constraints. It requires the number of clusters to be specified a priori, and although this can be guided by techniques such as the elbow method or silhouette analysis, it ultimately involves researcher judgment. Furthermore, because k-means relies on Euclidean distance, results can be sensitive to variable distributions and scaling. The impact of outliers and the appropriateness of standardization therefore warrant careful evaluation[16].

In terms of variable selection, the performance indicators used were limited to those available in the hospital bed function report, potentially omitting important domains such as outpatient services or community-based nursing care. While Holm correction was applied to mitigate type I error in multiple comparisons, this may have reduced sensitivity to practically relevant but statistically subtle differences. Thus, interpretation should consider practical effect size alongside statistical significance.

Additionally, the analysis was confined to secondary medical regions within Okayama Prefecture. The demographic and geographic characteristics of the area may have influenced the clustering outcomes, limiting external generalizability. Future work should extend to multiple regions and time periods, incorporating longitudinal data and multidisciplinary variables, including home care and long-term care data, to enhance robustness and policy relevance.

Taken together, the PCA-clustering approach used in this study provides a statistically grounded method for classifying medical institution functions based on empirical data. However, careful attention must be paid to the analytic assumptions, variable coverage, statistical corrections, and regional context when interpreting the results. Ongoing methodological refinement and broader data integration will be essential for translating these insights into practical healthcare system reform.

4.4 Future applications and policy implications

The five medical function clusters identified in this study represent a structural classification that transcends the uniform functional categories defined by the existing system and more accurately reflects the actual medical services being delivered. Current regional medical care plans are grounded in four standardized functional classifications—"Advanced acute," "acute," "recovery," and "chronic"—which are institutionally defined and often do not fully capture the diversity and complexity of real-world care provision. In contrast, the present study applied multivariate analysis based on empirical performance indicators, independently of the notification-based system, thereby uncovering functionally intermediate or specialized groups that are not explicitly addressed in the current policy framework.

Particularly noteworthy is the identification of distinct functional types—such as the "advanced acute/lifesaving" and "chronic/convalescent" clusters—which have traditionally been ambiguously treated in institutional classifications. These functional groups were clearly differentiated in the distribution of principal component scores and Z-scores. Such performance-based classifications are consistent with the policy direction outlined in the "New Regional Medical Vision"[11], which emphasizes the visualization of medical functions and the enhancement of regional collaboration. The findings of this study can serve as objective, data-driven evidence to support coordination meetings and the formulation of future regional healthcare plans.

Moreover, the identification of intermediate function groups—for example, Cluster 1 ("General Acute Care") and Cluster 3 ("Balanced Recovery-Phase Care")—offers a novel lens through which to reassess medical institutions that do not clearly align with the four system-defined categories. Although these institutions have not been assigned explicit functional roles within the existing framework, they appear to play a substantive part in sustaining regional healthcare capacity. As such, the proposed classification scheme may inform policy discussions related to the development of function-specific evaluation criteria and the refinement of support mechanisms for under-recognized facility types.

Additionally, the analytical framework employed in this study is grounded in publicly available open data from hospital bed function reports, making it generalizable across both regions and time periods. By extending this method to different geographic areas and subsequent fiscal years, it will be possible to visualize inter-regional disparities in healthcare functions and monitor dynamic shifts in healthcare delivery systems over time. In particular, the use of time-series data to trace functional transitions and structural realignments at the institutional level could yield critical insights for the strategic reallocation of healthcare resources in response to evolving demographic and policy needs.

In this regard, the present study serves as an initial step toward the practical application of performance-based, data-driven functional assessment in health system design. The proposed methodology not only offers a flexible alternative to the current institutional framework but also provides a conceptual foundation for enhancing the responsiveness and sustainability of regional healthcare governance.

5. Conclusion

This study employed multivariate statistical techniques to classify the functions of medical institutions using performance indicators derived from hospital bed function reports. The analysis revealed a diverse and complex landscape of medical care functions that are not fully captured by the current institutional framework. By combining dimensionality reduction through principal component analysis (PCA) with clustering via the k-means algorithm, we objectively identified functionally distinct groups of medical institutions. This approach provides valuable foundational evidence for the reorganization of hospital beds and the development of integrated care systems within regional medical planning.

The five functional clusters extracted in this study were generally aligned with the existing four-function framework (advanced acute, acute, recovery, and chronic), but also illuminated intermediate and specialized functional groups—such as advanced acute/lifesaving, chronic/convalescent, and balanced recovery-phase types—that are not explicitly defined in the current system. These groups reflect the actual structure of medical service delivery in the region and offer new perspectives for the reallocation of medical resources and the future refinement of institutional evaluation frameworks.

Moreover, the classification of medical functions based on empirical performance indicators provides a practical complement to the limitations of self-reported systems. As such, it can be leveraged as quantitative evidence in policymaking and represents a meaningful step toward implementing data-driven regional healthcare policy. The methodology is sufficiently versatile to allow for application across different regions and time periods, facilitating broader comparisons and the tracking of functional transitions over time.

However, this study is not without limitations. The findings are influenced by several analytical assumptions, including the selection of variables, interpretation of principal components, and determination of the optimal number of clusters. Future research should incorporate more advanced techniques, such as methods for capturing nonlinear relationships, integration of outpatient and home care data, and time-series analyses using multi-year datasets.

In summary, this study offers new insights into the classification and understanding of medical institution functions based on performance data. The findings have important implications for evidence-based policymaking and provide a foundation for the flexible and strategic design of regional healthcare systems in the future.

Acknowledgments

We would like to express our sincere gratitude to Ms. Naomi Yamaguchi of the Department of Medical Information Management, National Hospital Organization Kobe Medical Center, for her valuable advice and support in the course of this study.

Ethical considerations

This study exclusively used publicly available open data released by the Ministry of Health, Labour and Welfare. As the data contain no personally identifiable information, the study complies with the ethical guidelines for medical and health research involving human subjects and does not require approval by an ethics committee. Additionally, the use of such open data does not violate any provisions of the Act on the Protection of Personal Information.

Conflict of interest

The author declares that there is no conflict of interest.

References

1. Ministry of Health, Labour and Welfare : *Demographic trends and medical demand toward 2040*. https://www.mhlw.go.jp/content/10802000/001090217.pdf, [2022]. (March 30, 2025)
2. Ministry of Health, Labour and Welfare : *Status of discussions on a new regional medical care plan (report)*. https://www.mhlw.go.jp/content/10801000/001300089.pdf, [2024]. (March 30, 2025)
3. Ministry of Health, Labour and Welfare : *About the new regional medical care plan*. https://www.mhlw.go.jp/content/10800000/001309842.pdf, [2023]. (March 30, 2025)
4. Matsuda S : *Study on understanding and visualizing the actual state of medical function differentiation based on hospital bed function reports*. https://mhlw-grants.niph.go.jp/system/files/2019/193011/201922014A_upload/201922014A202005201611394730005.pdf, [2020]. (March 30, 2025)
5. Fushimi K : The future of regional medical care planning from the perspective of public hospitals responsible for (highly) acute care. *Hospital*, 81(1), 32-36, 2022. (In Japanese)
6. Honno K : Hospital management based on data analysis in regional medical plans — Vol. 11: Specific methods of regional diagnosis (Part 2: Functional analysis of the Sapporo medical area). *Hospital Management Compass*, 271, 50-55, 2025. (In Japanese)
7. Ministry of Health, Labour and Welfare : *Implementation of the FY2023 hospital bed function report*. https://www.mhlw.go.jp/content/10800000/001100422.pdf, [2023]. (March 30, 2025)
8. Ministry of Health, Labour and Welfare, Shimizu Y : *Current status and challenges of secondary use of medical data in Japan*. https://www.mhlw.go.jp/content/10808000/001206322.pdf, [2023]. (March 30, 2025)
9. Masuzawa N and Matsuda S : Analysis of hospitals that selected Advanced acute and acute phases in the Kanagawa Prefecture hospital bed function report. *Yokohama Medical Journal*, 73, 507-516, 2022. (In Japanese)
10. Matsuda S : An attempt to classify hospitals using DPC open data and hospital bed function reporting open data. *Hospital*, 80(8), 739-742, 2021. (In Japanese)
11. Ministry of Health, Labour and Welfare : *Progress in discussions on a new regional medical care plan*. https://www.mhlw.go.jp/stf/shingi2/0000161127_00057.html, [2024]. (April 5, 2025)
12. Ministry of Health, Labour and Welfare : *FY2023 hospital bed function report: Published data*. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/open_data_00016.html, [2023]. (April 5, 2025)
13. Kanda Y : Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplantation*, 48, 452-458, 2013.
14. Ministry of Health, Labour and Welfare : *Summary of the new regional medical care plan (report)*. https://www.mhlw.go.jp/content/10800000/001357306.pdf, [2024]. (April 5, 2025)

15. Lever J, Krzywinski M and Altman N : Principal component analysis. *Nature Methods*, 14, 641-642, 2017, https://doi.org/10.1038/nmeth.4346.
16. Hastie T, Tibshirani R and Friedman J : *The Elements of Statistical Learning* (Kurihara S Trans.), Kyoritsu Shuppan, Tokyo, 2009. (Original work published 2009) (In Japanese)

Appendix Table: Medical institutions classified by cluster assignment

X002_Open data medical institution code.	X003_Medical institution name	Cluster
3333010079	Public Interest Incorporated Foundation Okayama Health Promotion Foundation Affiliated Hospital	1
3333010106	Sakakibara Heart Disease Center Hospital	1
3333010062	Fukuwatari Hospital, National Health Insurance, Kumeminami Town, Okayama	1
3333010003	Ohmoto Hospital	1
3333010034	Kajiki Hospital, Medical Corporation Sankikai	1
3333010013	Asahigawasou Medical and Rehabilitation Center	1
3333010080	Saiwaicho Memorial Hospital	1
3333010083	National Hospital Organization Okayama Kanagawa Hospital	1
3333010058	Okayama Memorial Hospital	1
3333010095	Kosei Hospital, Social Medical Corporation	1
3333010128	NASVA Okayama Rehabilitation Center	1
3333010067	Okayama Eastern Neurosurgical Hospital	1
3333010065	Okayama Daiichi Hospital	1
3333010009	Perinate Mother and Child Hospital	1
3333010050	Okayama Asahi Higashi Hospital, Sōfūkai Foundation	1
3333010036	Sanyo Hospital, Ryoyūkai Medical Corporation	1
3333010040	Nagashima Hospital, Chōkōkai Medical Corporation	1
3333010039	Okayama Saidaiji Hospital, Seizen-kai Medical Corporation	1
3333010031	Niwa Hospital, Kokutaikai Medical Corporation	1
3333010092	Ukita Obstetrics and Gynecology Hospital	1
3333010096	Okamura Isshindō Hospital, Social Medical Corporation	1
3333010100	Hospital Attached to Shigei Medical Research Institute	1
3333010087	Satō Hospital	1
3333010006	Senoo Hospital, Okayama City	1
3333010007	Central City Hospital	1
3333010129	Okayama Rosai Hospital, Japan Organization of Occupational Health and Safety	1
3333010076	Tamano Mitsui Hospital	1
3333010120	Nakaya Surgical Hospital	1
3333010119	Onishi Hospital	1
3333010019	Yura Hospital, Aizen-kai Medical Corporation	1
3333010116	Tamano Branch, Okayama Red Cross Hospital	1
3333010133	Bizen Municipal Bizen Hospital, National Health Insurance	1
3333010132	Bizen Municipal Hinase Hospital, National Health Insurance	1
3333010118	Soka Hospital	1
3333010131	Bizen Municipal Yoshinaga Hospital, National Health Insurance	1
3333010085	Oku Komyoen, National Sanatorium	1
3333010084	Nagashima Aiseien, National Sanatorium	1
3333010026	Kitagawa Hospital, Kiten-kai Medical Corporation	1
3333010044	Taira Hospital, Taira Medical Corporation	1
3333010130	Kibi Kogen Rehabilitation Center, Japan Organization of Occupational Health and Safety	1
3333020083	Mizushima Kyodo Hospital	1
3333020005	Chikuba Surgical, Gastroenterological, and Proctological Hospital	1
3333020057	Tamashima Kyodo Hospital	1
3333020058	Kurashiki Central Hospital Riverside, Ohara Memorial Kurashiki Central Medical Organization	1
3333020022	Kurashiki City Hospital, Kyoai-kai Medical Corporation	1
3333020025	Akamatsu Hospital, Shiseikai Medical Corporation	1
3333020090	Takeda Hospital	1
3333020032	Kurashiki Kita Hospital, Showa-kai Medical Corporation	1
3333020028	Kojima Seikou Hospital, Goseikai Medical Corporation	1
3333020050	Kurashiki Sweet Hospital, Wakakai Medical Corporation	1
3333020008	Mabi Memorial Hospital	1
3333020080	Kurashiki Municipal Citizen's Hospital	1
3333020011	Matsuda Hospital, Tenwakai Medical Corporation	1
3333020021	Prime Hospital Tamashima, Kashinkai Medical Corporation	1
3333020039	Watanabe Gastroenterology Surgical Hospital, Sōseikai Medical Corporation	1
3333020030	Kasaoka Daiichi Hospital, Seiwakai Medical Corporation	1
3333020055	Kasaoka Municipal Hospital	1
3333020049	Kasaoka Central Hospital, Midori Cross Medical Corporation	1
3333020052	Ibara Municipal Hospital	1
3333020026	Kan Hospital, Medical Corporation	1
3333020013	Oda Hospital, Oda-ujikai Medical Corporation	1
3333020033	Morishita Hospital, Jintoku-kai Medical Corporation	1
3333020046	Yakushiji Jikei Hospital, Yakushiji Medical Corporation	1
3333020031	Konko Hospital, Dojin-kai Medical Corporation	1
3333020017	Miwa Memorial Hospital, Miwa Medical Corporation	1
3333020088	Minami-Okayama Medical Center, National Hospital Organization	1
3333020093	Yakage National Health Insurance Hospital	1
3333030003	Osugi Hospital, Keishinkai Medical Corporation	1
3333030008	Nariwa Hospital, Takahashi National Health Insurance	1
3333030007	Takahashi Central Hospital, Seiryo-kai Medical Corporation	1

X002_Open data medical institution code.	X003_Medical institution name	Cluster
3333030004	Watanabe Hospital, Shisei-kai Medical Corporation	1
3333030005	Hasegawa Memorial Hospital, Junwa-kai Medical Corporation	1
3333030006	Niimi Central Hospital, Shinsei-kai Medical Corporation	1
3333030011	Ota Hospital	1
3333040005	Katsuyama Hospital, Mikamukai Medical Corporation	1
3333040004	Kondo Hospital, Keiwa-kai Medical Corporation	1
3333040009	Yubara Onsen Hospital, Maniwa City National Health Insurance	1
3333040008	Kaneda Hospital, Ryokusokai Medical Corporation	1
3333040003	Ochiai General Hospital, Iguchi-kai Medical Corporation	1
3333050022	Nihonbara Hospital, Seifukai Medical Corporation	1
3333050014	Nakajima Hospital, Wafukai Medical Corporation	1
3333050025	Ishikawa Hospital	1
3333050028	Tsuyama Memorial Hospital	1
3333050006	Tajiri Hospital, Sansuikai Medical Corporation	1
3333050033	Ohara Hospital, Mimasaka City	1
3333050016	Yoshino Hospital, Kyoai-kai Foundation	1
3333050019	Kagamino National Health Insurance Hospital	1
3333010145	Okayama Saiseikai Outpatient Center Hospital	1
3333010126	Dojin Hospital	2
3333010048	General Foundation Junpukai Long Life Hospital	2
3333010025	Kitamura Hospital	2
3333010074	Miyamoto Orthopedic Hospital	2
3333010066	Okayama East Chuo Hospital	2
3333010138	Nakano Branch, Okayama Saidaiji Hospital	2
3333010137	Hayashi Hospital	2
3333010121	Nakajima Hospital	2
3333010027	Tamano Chuo Hospital, Tamano Medical Corporation	2
3333010108	Setouchi Memorial Hospital	2
3333010073	Kibi Kogen Lumiere Hospital	2
3333020068	Shibata Hospital	2
3333020051	Shimotsui Hospital, Jakufūkai Medical Corporation	2
3333020064	Tamashima Hospital, Kōjinkai Foundation	2
3333020098	Nagano Hospital, Gyodo-kai Medical Corporation	2
3333020060	Kunisada Hospital	2
3333020086	Torigoe Hospital	2
3333050009	Otani Hospital, Seiken-kai Medical Corporation	2
3333050012	Mimasaka Central Hospital, Bifukai Medical Corporation	2
3333050005	Sakubara Hospital, Sanken-kai Medical Corporation	2
3333010097	Social Medical Corporation Kouninkai Okayama Central Hospital	3
3333010088	Saiseikai Kibi Hospital	3
3333010136	Ryuso Orthopedic Hospital	3
3333010059	Okayama Kyōritsu Hospital	3
3333010049	Okayama Rehabilitation Hospital, Sōfūkai Foundation	3
3333010099	Okayama Hakuai Hospital, Okayama Hakuai Welfare Association	3
3333010035	Fujita Hospital, Fujita Medical Corporation	3
3333010060	Okayama Kōnan Hospital	3
3333010117	Tamano Municipal Hospital, Tamano Medical Center (Independent Administrative Institution)	3
3333010109	Setouchi City Hospital	3
3333010139	Akaiwa Medical Association Hospital	3
3333020079	Kurashiki Rehabilitation Hospital	3
3333020002	Coop Rehabilitation Hospital	3
3333020034	Mizushima Daiichi Hospital, Suiseikai Medical Corporation	3
3333020040	Shigei Hospital, Sowa-kai Medical Corporation	3
3333020053	Kurashiki Daiichi Hospital, Seiwa-kai Medical Corporation	3
3333020029	Tamashima Central Hospital, Shinpūkai Medical Corporation	3
3333020082	Kurashiki Heisei Hospital	3
3333020019	Kojima Central Hospital, Isei-kai Medical Corporation	3
3333020072	Mizushima Central Hospital	3
3333020037	Kurashiki Memorial Hospital, Seiwa-kai Medical Corporation	3
3333050029	Tsuyama Daiichi Hospital, Hirano Dojin-kai Medical Corporation	3
3333050001	Sato Memorial Hospital	3
3333010115	Okayama Red Cross Hospital	4
3333010127	National Hospital Organization Okayama Medical Center	4
3333010113	Kawasaki Medical School General Medical Center	4
3333010061	Okayama Saiseikai General Hospital	4
3333010114	Okayama Municipal Citizen's Hospital	4
3333020077	Kawasaki Medical School Hospital	4
3333050017	Tsuyama Central Hospital, Tsuyama Jifukai Foundation	4
3333010063	Okayama University Hospital	5
3333020059	Kurashiki Central Hospital, Ohara Memorial Kurashiki Central Medical Organization	5
3333020081	Kurashiki Seijinbyō Center (Adult Disease Center)	5