国際医療福祉大学審査学位論文(博士)

大学院医療福祉学研究科博士課程

A Grouping Method based on the N-dimension Euclidean Divergence using Cerebral Infarction Cases in China and Japan

平成29年度

保健医療学専攻　診療情報管理　分析学分野

診療情報管理・分析学領域

学籍番号：15S3040　　氏名：陳　迪

研究指導教員：　斎藤　恵一　准教授

副研究指導教員：小川　俊夫　准教授

山本　康弘　教授

陳　霞芬　教授

要旨

在院日数と医療費からなる2次元分布の類似度（ダイバージェンス）に着目した、処置，副傷病を考慮した新しいクラスタリング手法を試行し、日本と中国における病院の脳梗塞症例を用いて本法の適用性を検証した．決定木分析から両国に共通な在院日数への影響要因、リハビリテーション、肺炎、II型糖尿病を抽出し、症例を8グループに分類した。8グループに対して本法により集約されたクラスター数は、日本では4でDPCによる既存クラスター数の3とはやや異なるがほぼ同様の結果が示された。中国では6となり、在院日数と費用の分布をより精緻に反映している可能性が示唆された。なお、本研究では限定されたサンプルを用いた検証に留まっており、今後より大規模な評価と他の疾患への適用について検討する必要がある。処置，副傷病を考慮したグループに区分し、各グループの分布類似度からクラスターを構築し患者分類を精緻化する手法が有効であることが示唆された。

キーワード：2次元分布の類似度　在院日数　医療費　脳梗塞　患者分類の精緻化　中国 日本

Abstract

A new grouping method based on 2-dimension of divergence for length of stay (LOS) and medical charge with the consideration of treatment and comorbidities and complications (CC) has been proposed. This method was evaluated using the cases of cerebral infarction in Japan and China to verify the applicability in different settings.

The common factors between two countries, i.e., rehabilitation, pneumonia and diabetes type II, were extracted by decesion tree method. All samples were divided into eight groups using the common factors, then performed for the eight groups for case mix clustering, The Chinese samples were divided into 6 clusters whereas Japanese had 4 clusters. The Japanese clusters were similar to the existing CCPM clusters that has 3 clusters. Chinese clusters could also show the divergence of the distribution of LOS and medical charge for patients with cerebral infarction in China.

This research with the limitation of using sample data of cerebral infarction only suggests that it will be necessary to expand the data and target diseases for the further analysis. It is indicated that this methods will be applicable for improvement of patient classification.

Keywords: *Divergence of 2-dimensional distributions, Length of hospital stay, Hospital fee, Cerebral Infarction, China, Japan*

Contents

1 Introduction

Since the 1980s, due to the remarkable progress of medical technology and the diversity of healthcare needs with budget constraint in many countries, it becomes difficult to cover the medical expense with fee-for-service scheme (FFS). A new payment method, Diagnosis-Related Group/ Prospective Payment System (DRG/PPS), was introduced and adopted in the USA, Germany and other developed countries, that is a system to classify inpatient cases into one of originally 467 groups[1], referred to as "DRGs". The main purposes of the DRGs are to set up a health care standard and to identify the medical services provided by hospitals transparently.

The Diagnosis Procedure Combination/Per-Diem Payment System (DPC/PDPS), a Japanese-style DRG/PPS scheme, has introduced in Japan since early 2000s with concerns over healthcare costs, length of stay (LOS), and the healthcare needs. [2] Similar to the DRG/PPS, DPC is also a prospective payment system to classify the inpatient cases with coding. The unique part of this payment system is that per-diem and FFS payment schemes are integrated. Providers are paid a flat-rate prospective fee per day of inpatient hospital stay for certain DPC services and paid FFS for non-DPC services as well.

To evaluate the medical services and resources provided in hospitals more accurately, Comorbidity Complication Procedure Matrix (CCPM) was investigated in Japan as a comprehensive evaluation method with the consideration of severities.[3] The trial version of CCPM was introduced for the patients with cerebral infarction, pneumonia, and diabetes mellitus from 2016.

DRG/PPS has been imported into China from the early 2000s. However, DRGs have not been widely used in China, and it has not yet been adopted nationally.[4,5] It will be necessary to improve the usability of DRGs in China for promoting it to strengthen healthcare systems in future.

As DRGs/DPC system could not evaluate the differences in pathology and severity of the same disease adequately, CCPM as a new method based on the consideration of treatment and CC was developed for providing more accurate medical resources for the patients. A method to construct CCPM with Kullback-Leibler divergence had been conducted for Type 2 Diabetes Mellitus. [6,7]

In this study, a grouping method based on the N-dimension Euclidean divergence was going to proposed and examined. In addition, the data of cerebral infarction from medical institution was analyzed to verify the applicability of the method for improvement of DRGs or DPC systems in different setting.

1.1 The concept of case mix complexity

The term case mix complexity has been used to refer to an interrelated but distinct set of patient attributes which include severity of illness, prognosis, treatment difficulty, need for intervention and resource intensity. Each of these concepts has very precise meaning, which describes a particular aspect of a hospital's case mix.[8]

1. **Severity of illness.** Refers to the relative levels of loss of function and mortality that may be experienced by patients with a particular disease.[9]
2. **Prognosis.** Refers to the probable outcome of an illness including the likelihood of

improvement or deterioration in the severity of the illness, the likelihood for recurrence and the probable life span.

3. **Treatment difficulty.** Refer to the patient management problems that a particular illness presents to the health care provider. Such management problems are associated with illnesses without a clear pattern of symptoms, illnesses requiring sophisticated and technically difficult procedures and illnesses requiring close monitoring and supervision.
4. **Need for intervention.** Relates to the consequences in terms of severity of illness that lack of immediate or continuing care would produce.
5. **Resource Intensity.** Refers to the relative volume and types of diagnostic, therapeutic and bed services used in the management of a particular illness.

When clinicians use the notion of case mix complexity, they mean that the patients treated have a greater severity of illness, present greater treatment difficulty, have poorer prognoses and have a greater need for intervention. Thus, from a clinical perspective case mix complexity refers to the condition of the patients treated and the treatment difficulty associated with providing care. On the other hand, administrators and regulators usually use the concept of case complexity to indicate that the patients treated require more resources, which results in a higher cost of providing care. In addition, the purpose of the DRGs is to relate a hospital's case mix to the resource demands and associated costs experienced by the hospital. Therefore, a hospital having a more complex case mix from a DRG perspective means that the hospital treats patients who require more hospital resources.

1.2 Patient classification

Given that the purpose of the DRGs is to relate a hospital's case mix to its resource intensity, it was necessary to develop an operational means of determining the types of patients treated and relating each patient type to the resources they used. All patients are able to make unique groups with demographic, diagnostic and therapeutic attributes in common that determine their level of resource intensity. By developing clinically similar groups of patients with similar resource intensity, patients can be aggregated into meaningful patient classes. Moreover, if these patient classes covered the entire range of patients seen in an inpatient setting, then collectively they would constitute a patient classification scheme that would provide a means of establishing and measuring hospital case mix complexity. [10]The DRGs were therefore developed as a patient classification scheme consisting of classes of patients who were similar clinical conditions and consumption of hospital resources.

During the process of developing the DRG patient classification scheme, several alternative approaches to constructing the patient classes were investigated. [11] There was a tendency for their definitions to include an extensive set of specifications, requiring information, which might not always be collected through a hospital's medical information system. If the entire range of patients were classified in this manner, it would ultimately lead to thousands of DRGs, most of which described patients seen infrequently in a typical hospital. In addition, statistical algorithms applied to historical data would be useful to suggest ways of forming DRGs that were similar in terms of resource intensity. However, it was also discovered that statistical algorithms applied to this data in the absence of clinical input would not yield a satisfactory set of DRGs. The DRGs resulting from such a statistical approach, while similar in terms of resource intensity, would often contain patients with a diverse set of characteristics, which could not be interpreted from a clinical perspective. Thus, it became apparent that the development of the DRG patient classification scheme required that physician judgment, statistical analysis and verification with historical data be merged into a single process.

1.3 Basic characteristics of the DRG patient classification scheme

Given the limitation of existing patient classification schemes and the experience of attempting to develop DRGs with physician panels and statistical analysis, it was concluded that, in order for the DRG patient classification scheme to be practical and meaningful, it should have the following characteristics: [12]

1. The patient characteristics used in the definition of the DRGs should be limited to information routinely collected on hospital abstract systems.
2. There should be a manageable number of DRGs, which encompass all patients seen on an inpatient basis.
3. Each DRG should contain patients with a similar pattern of resource intensity.
4. Each DRG should contain patients who are similar from a clinical perspective.

Restricting the patient characteristics used in the definition of the DRGs to those readily available insured that the DRGs could be extensively applied. Currently, the patient information routinely collected includes age, principal diagnosis, secondary diagnoses and the surgical procedures performed.

Limiting the number of DRGs to manageable numbers insures that for most of the DRGs, a typical hospital will have enough experience to allow meaningful comparative analysis to be performed. If there were only a few patients in each DRG, it would be difficult to detect patterns in case mix complexity.

1.4 The development of DRGs

The design and development of DRGs began in the late 1960s at Yale University. The initial motivation was to create an effective framework for monitoring the utilization of services in a hospital setting. The first large-scale application of DRGs was conducted in the late 1970s by the State of New Jersey in its hospital prospective payment system (PPS). In 1984, a DRG-based PPS was implemented for the Medicare program. Subsequently, a number of states and large payers implemented DRG-based PPS for non-Medicare patients. In addition, DRGs have been used as the basis for global budget allocation and payment in several countries in Western and Eastern Europe as well as Australia.[13]

The initial DRG system developed by Yale was intended to describe all types of patients seen in an acute care hospital. There was an inherent problem, however, in that the database used for its development attempted to be representative of a cross section of community hospitals. This ensured there would not be sufficient case volume of complex low-volume pediatric and neonatal conditions to detect certain problems or to develop solutions. Of note, freestanding acute children's hospitals were not included in the Yale study database.[14]

The initial generation of DRG systems provided only modest differentiation for severity within a DRG category. Certain of the DRG categories were split into two categories based on the presence or absence of a secondary diagnosis from a list of comorbidities and complications (CC) conditions that included approximately 3,000 of more than 12,000 International Classification of Diseases, 9th Revision Clinical Modification (ICD-9-CM) codes. The CC was a simple yes/no split. No differentiation was made as to whether certain of the CC diagnoses represented more extreme conditions or whether the patient had multiple CC diagnoses.[15]

The effort to develop a more advanced severity adjustment methodology began in the mid-to-late 1980s when US Health Care Financing Administration (HCFA) funded a 2-year study project conducted by Yale University. This project produced the first "Refined" Diagnosis-Related Group (RDRG) system, which is a DRG system with multiple CC (or severity) levels within each DRG category. Nearly all the DRG categories were given either three or four severity subclasses (mild, moderate, major, extreme) based on the presence of certain secondary diagnoses.

The All Patients DRGs (AP-DRGs) implemented by New York State in 1988 designated a subset of secondary diagnoses as major CCs. These diagnoses were similar to those classified as catastrophic by the initial RDRGs. To avoid significantly increasing the number of AP-DRG categories, AP-DRG major CC categories were formed for groups of surgical AP-DRGs and medical AP-DRGs in a body system.

The severity-adjusted version of DRGs that has come into widest acceptance and use in the 1990s is the All Patient Refined DRGs (APR-DRGs), first introduced in 1991. As of June 1998, there were above 1400 hospitals and other organizations using the APR-DRGs. This included 17 state health departments and data commissions using the APR-DRGS for comparative profiling of hospitals.

The APR-DRGs are developed and updated through the combined research activities of 3M Health Information Systems and the National Association of Children's Hospitals and Related Institutions (NACHRI). The APR-DRGs are different from other DRG systems in a number of respects including: 1) definitions for DRG categories; 2) revisions to surgical hierarchies; 3) updates to diagnoses on the CC list; 4) assignment of all diagnoses to one of four CC (or severity) levels; 5) severity subclass algorithm that takes into account the interactive effect of multiple secondary diagnoses; 6) a severity subclass methodology specifically developed for neonatal patients; and 7) a separate subclass methodology for risk of mortality.[16]

1.5 The development of DPC in Japan

In 2003, the Japanese government began the implementation of the Diagnostic Procedure Combination/Per-Diem Payment System (DPC/PDPS) for reimbursements to acute care hospitals under the public medical insurance scheme. The DPC/PDPS is similar to the US prospective payment system with diagnosis-related groups (DRG/PPS), and was implemented with the aim of reducing length of stay (LOS) without decreasing quality of care. [17,18] However, the current DPC/PDPS has not been able to achieve appropriate reimbursements for medical resource due to an inadequate consideration of patient severity and comorbidities, despite the wide variations inpatient severity in Japan.[19,20]

As the DPC produces a tiered tree structure with the condition name as the top layer, the level of one node (parent: branch condition) impacts all of the child nodes, and thus affects the structure of the diagnostic categories. Further, when there are a large number of branch conditions, there was the problem that the leaf nodes located on the tips increase too much. On the other hand, the elaboration of DPC is expected, the comorbidity complication procedure matrix (CCPM) has been investigated as a new method depending on severities, which were considered as the combination of operations, treatments and complications and other factors. The CCP matrix would enable analysts to account for variations in patient severity and comorbidities when investigating the use of health care resources. In addition, the Ministry of Health, Labour and Welfare (MHLW) is considering the integration of a "Comorbidity Complication Procedure" (CCP) matrix into the current DPC/PDPS. The systematic methods of CCPM construction have been examined recently. Moreover, the trial introduction of CCPM for cerebral infarction, pneumonia, and diabetes mellitus was conducted from 2016.

1.6 The development of DRG in China

Health system reform has been promoted[21] in order to optimize the medical resources in China, and trials have been made for the introduction of Chinese

DRG/PPS from the early days, but it has not yet been implemented nationwide.[22] It is thought that the movement towards the introduction will be accelerated in the future, and it is expected that efficient construction of DRG will be an urgent issue.

Although DRG/PPS has been imported into China from the early 2000s, DRGs have not been widely used in China[23,24], and it has not yet been adopted nationally. It will be necessary to improve the usability of DRGs in China for promoting it to strengthen health system in future.

Introducing DRGs would potentially switch China's current per-head costing system to a cost classification system that categorizes patients of diagnoses based on distinct groupings, for the purpose of reimbursing hospitals or each case in a given category with a fixed fee regardless of the actual costs incurred. [25]

If implemented, DRGs would lead to a complete shift in the approach of drug pricing and procurement managers at public hospitals, as pharmaceuticals effectively become a cost, rather than a revenue stream, potentially limiting waste and improving management in China.[26]

2 Purpose of the Research

The purpose of this study is to develop a grouping method of DRGs based on N-dimension divergence, and to evaluate this method using the cases of cerebral infarctions in Japan and China to verify the applicability in different settings. Based on the analyses, I would like to discuss the applicability of this method for the improvement of DRGs.

3 Methods

The processes of grouping method of DRGs based on N-dimension divergence were shown in **Figure 3-1**. Firstly, I select the data according to the inclusion and exclusion criteria, and then extract the factors related to the LOS by applying the decision tree method for case grouping. The difference of the distribution of LOS and medical charge between divided groups are measured, and finally the groups with the similarity were allocated into the same cluster to get the clustering result of case mix grouping.

One of the advantages of this method is that classification could be constructed using both LOS and medical charge while considering the health conditions. Formerly, patient classification is usually developed focus on either the distribution of medical resources or LOS.

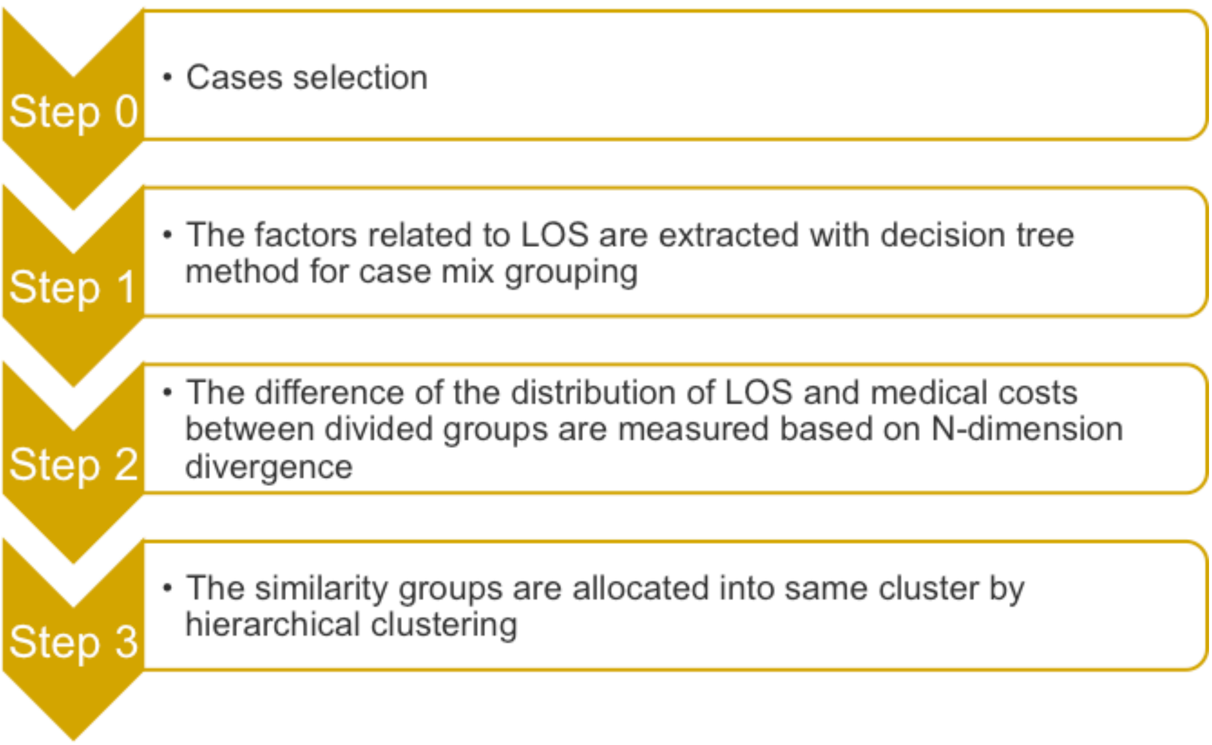


Figure 3-1 Processes of classifying method of DRGs based on N-dimension divergence

To achieve the purpose of research, firstly the grouping method described above was applied into the Japanese data, and comparing the results to the official published CCPM for cerebral infarction, when the comparing result showed that the method is available for the Japanese data, then the same method was applied into the Chinese data, finally the applicability of this grouping method was validated in Japan and China.

DPC/PDPS is a per-diem payment system, which calculate the medical charge every day according to the payment standard from LOS I to LOS III, but the DRG/PPS is per case payment system, which calculate the medical charge together when the patient discharge from the hospital.

Although the DPC/PDPS is different payment system from DRG/PPS, in this study, the total length of stay (LOS) and total medical charge were chosen as the dependent variables for analyzing. When the LOS and medical charge were added from the inpatient day to the discharge day in DPC, the distribution of DPC is similar with the distribution of DRG, so the research results were not influenced by the different of DRG and DPC in this study.

3.1 Data source and sample inclusion/exclusion criteria

The medical records of inpatients with cerebral infarction (ICD-10: I63) as the principal discharge diagnosis were extracted both in Japan and China. Personal information was anonymous and incapable of being connected. The Chinese data were the DRG data with 929 inpatients between January 1st 2015 and December 31st 2015 in a rehabilitation institute of China. The Japanese data were the DPC data with 918 inpatients between July 1st 2010 and February 29th 2012 discharged from 5 general hospitals in Japan.

Inclusion criteria of this study were patients with cerebral infarction as the main diagnosis, and patients covered by a medical insurance. Exclusion criterion was the patients who had been performed a surgical operation. All of the analyses were performed using SPSS ver. 22 (IBM, NYC USA).

3.2 Data collection and definitions

The extracted medical records of the inpatients with cerebral infarction contained various information such as (1) Basic information such as type of medical insurance, gender, age, admission date, discharge date, length of stay (LOS) and major diagnosis with ICDs; (2) detailed information related to cerebral infarction such as complications and comorbidities (CC) with ICDs; (3) Medical expenses information for drugs, type of bed, imaging, rehabilitation, laboratory analysis, sanitary materials, diagnostics and nursing care, which were used for this analysis.

3.3 Statistical analysis

Numeric variables were reported as mean $\pm$ standard deviation (SD), median/25th-75th percentile, and categorical variables were described as both numbers and percentages. Data filtering was performed using Microsoft Excel 2011 and analysis were carried out using IBM SPSS Statistics ver. 22.

3.3.1 Variables selection method

The improvement of DRGs/DPC is need to integrate the severity of illness (CC) for more accurate reflection of given health care services and related medical resources utilization, thus we use the decision tree method[27] with Chi-square Automatic Interaction Detector (CHAID) [28]of which applications[29,30] were reported in medical data mining, to extract variables of clinical characteristics such as treatment and comorbidities and complications for the case mix grouping.

The decision tree was used to extract the variables associated with LOS. The factor in the first branch of the decision tree was considered as the most relevant explanatory variable, then excluded it and conducted decision tree analysis again using the remaining items. The analysis was repetitious performed until a decision tree could not be created. Finally, all the factors related to the dependent variable LOS were extracted.

3.3.2 Euclidean divergence and Euclidean divergence matrix

The similairty between distribution between groups was usually measured by divergence, and the Euclidean divergence is one of the commonly used and applied to many research fields. [31,32]

The Euclidean divergence is calculated by the same formula as the ordinary straight-line distance between two points in Euclidean space, and the N-dimension Euclidean divergence is calculated by the Euclidean divergence matrix to measure the difference of the relative frequency distribution during the N-dimension space.

If $\mathbf{p} = (p_1, p_2,..., p_n)$ and $\mathbf{q} = (q_1, q_2,..., q_n)$ are two points in Euclidean n-space, of which p_i and q_i are the i-th relative frequency, then the distance (d) from **p** to **q**, or from **q** to **p** is given by the formula:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

3.3.3 Grouping method

The clustering of the DRG groups was conducted by the hierarchical clustering method[33] using SPSS ver. 22, and the group average method[34] was used for calculating the distance of each group and cosines of the distance between each group.

In order to decide which clusters should be combined, or where a cluster should be spilt, a measure of dissimilarity between sets of observations is required. In most methods of hierarchical clustering, this is achieved by use of an divergence metrix (a measure of divergence between groups of distribution).

3.3.4 Applicability of method

The grouping method was applied to both the cases of China and Japan to verify the applicability in different settings. And the Japanese result was compared to the trial version of CCPM for cerebral infarction published in Japan at April 2016.

3.4 Ethical considerations

This research is supported by International University of Health and Welfare (IUHW) and China Rehabilitation Research Center (CRRC) with the approvals of the ethics committees in two institutions (Approval number of IUHW were 15-Io-14 and 17-Ig-43; Approval number of CRRC was 2017-020-1). Personal information was anonymous and incapable of being connected.

4 Results

4.1 Demographic, clinic characteristics of the subjects

There are two groups of study subjects included in the study. The Demographic and clinical characteristics of the study subjects in China and Japan was shown in **Table 4-1**.

Through this study, the gender status of study subjects in China (Male/Female: 62.7%/37.3%) and Japan (Male/Female: 59.8%/40.2%) were almost same. The age of admission (means: 66.7±13.0 years old, median: 68.0 year old) in China was slightly shorter then Japan (means: 72.3±12.4 years old, median: 74.0 years old). LOS of inpatients with cerebral infarction (means: 25.1±16.9 days, median: 23.0 days) in China was longer than the LOS (means: 14.0±11.7 days, median: 11.0 days) in Japan.

Table 4-1 Demographic and clinical characteristics of the study subjects in China and Japan

Country		China				Japan			
Characteristics		Mean ± SD or number	Median or percentage	Percentile		Mean ± SD or number	Median or percentage	Percentile	
				25th	75th			25th	75th
Gender	Male	582	62.7%			549	59.8%		
	Female	347	37.3%			369	40.2%		
Age		66.7±13.0	68.0	57.0	77.0	72.3±12.4	74.0	65.0	81.0
LOS		25.1±16.9	23.0	14.0	31.0	14.0±11.7	11.0	5.0	19.0
Medical charge		$4,675.6 ±$7,409.7	$3,534.5	$2,321.0	$4,420.3	$5,531.0 ±$4,134.2	$4,675.8	$2,249.64	$7,467.6

To comparing the demographic and clinical characterizes of the study subjects in China and Japan, the tests for normality (Shapiro-Wilk test) was conducted in the distribution of age, LOS and medical charge. The test results show that all the distributions of age, LOS and medical charge were non-normality, then the Mann-Whitney test were applied for comparing the different of age, LOS and medical charge in Japan and China, and the test results show that there are significant different in age ($p < 0.05$), LOS ($p < 0.05$) and medical charge ($p < 0.05$) between Japan and China.

Pearson's chi-squared test was conducted for comparing the gender status in Japan and China, and the result shows that there is not significant different in gender proportion ($p > 0.05$) between two countries.

4.2 Extracted variables for case mix grouping

The factors closely associated with LOS were extracted. For Chinese data, the variables are rehabilitation, speech disability, peripheral neuropathy, pneumonia, diabetes type II. On the other hand, the variables for Japanese data are rehabilitation, dysphagia, pneumonia, paralysis, peripheral neuropathy, high blood pressure, speech disability, hyperlipidemia and diabetes type II.

As using all the factors for case mix grouping, the number of cases included in each group will be too few to create relative frequency distribution, so in this research, for conducting a comprehensive research on the grouping results of cerebral infarction in China and Japan under the same situation of DPC/PDPS scheme, the three common variables as rehabilitation, pneumonia and diabetes type II were selected. Then all the Chinese and Japanese cases were assigned into eight groups by the three extracted variables as shown in **Table 4-2**.

Table 4-2 Case mix grouping for cerebral infarction in China and Japan

<table>
<tr><td colspan="6">Treatment or Comorbidities and Complications</td><td rowspan="2">Groups</td></tr>
<tr><td colspan="2">1st Factor</td><td colspan="2">2nd Factor</td><td colspan="2">3rd Factor</td></tr>
<tr><td rowspan="8">Rehabilitation</td><td rowspan="4">No</td><td rowspan="8">Pneumonia</td><td rowspan="2">No</td><td rowspan="8">Diabetes type II</td><td>No</td><td>G1</td></tr>
<tr><td>Yes</td><td>G2</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G3</td></tr>
<tr><td>Yes</td><td>G4</td></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td>G5</td></tr>
<tr><td>Yes</td><td>G6</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G7</td></tr>
<tr><td>Yes</td><td>G8</td></tr>
</table>

4.3 Distribution of LOS and medical charge

Under the DPC/PDPS scheme, the LOS was known as categorized into four periods, but the original data type of LOS and medical charge are continuous variables. In this research, the LOS and medical charge were considered as a whole index of medical resources utilization, thus it is necessary to transfer the data type into categorical variables to get the distribution of LOS and medical charge together.

4.3.1 Period of LOS distribution

The LOS distribution were categorized into four groups which are Period I, Period II, Period III and Period IV in reference to Japan's DPC/PDPS [35]. Period I is defined as the first day to the days of 25th percentile of LOS. Period II is defined as the day after Period I to the day of average LOS. Period III is defined as the day after the

Period II to the an 30 integral multiple day equal or above the average LOS plus twice the standard deviation of LOS, and after the Period III is Period IV. In this study, the distribution of Period IV was excluded as the frequency is too few.

As shown in **Table 4-1**, for the Chinese data, the average LOS is 25.1 days, the standard deviation (SD) is 16.9 days, the 25th percentile of LOS is 14.0 days, so according to the standard described above, the Period I is from 1 to 14 days, the Period II is from 15 to 25 days, the Period III is from 26 to 60 days, and the Period IV is 60 days above. The period of LOS distribution for cerebral infarction in China was shown in **Table 4-3.**

Table 4-3 Period of LOS distribution for cerebral infarction in China

<table>
<tr><td colspan="6" rowspan="2">Treatment or Comorbidities and Complications</td><td rowspan="3">Groups</td><td colspan="3">Period of LOS</td></tr>
<tr><td>1-14 days</td><td>15-25 days</td><td>26-60 days</td></tr>
<tr><td colspan="2">1st Factor</td><td colspan="2">2nd Factor</td><td colspan="2">3rd Factor</td><td>Freq.</td><td>Freq.</td><td>Freq.</td></tr>
<tr><td rowspan="8">Rehabilitation</td><td rowspan="4">No</td><td rowspan="8">Pneumonia</td><td rowspan="2">No</td><td rowspan="8">Diabetes type II</td><td>No</td><td>G1</td><td>103</td><td>61</td><td>19</td></tr>
<tr><td>Yes</td><td>G2</td><td>14</td><td>11</td><td>2</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G3</td><td>12</td><td>7</td><td>4</td></tr>
<tr><td>Yes</td><td>G4</td><td>5</td><td>2</td><td>2</td></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td>G5</td><td>91</td><td>146</td><td>328</td></tr>
<tr><td>Yes</td><td>G6</td><td>10</td><td>7</td><td>8</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G7</td><td>2</td><td>10</td><td>8</td></tr>
<tr><td>Yes</td><td>G8</td><td>1</td><td>2</td><td>4</td></tr>
</table>

Abbreviations: Freq., Frequency

Meanwhile for the Japan, the average LOS is 14.0 days, the standard deviation is 11.7 days, the 25th percentile of LOS is 5.0 days, thus the Period I is from 1 to 5 days, the Period II is from 6 to 14 days, the Period III is from 15 to 60 days, and the Period IV is 60 days above. The period of LOS distribution for cerebral infarction in Japan was shown in **Table 4-4**.

Table 4-4 Period of LOS distribution for cerebral infarction in Japan

<table>
<tr><th colspan="6">Treatment or Comorbidities and Complications</th><th rowspan="2">Groups</th><th colspan="3">Period of LOS</th></tr>
<tr><th colspan="6"></th><th>1-5 days</th><th>6-14 days</th><th>15-60 days</th></tr>
<tr><th colspan="2">1st Factor</th><th colspan="2">2nd Factor</th><th colspan="2">3rd Factor</th><th></th><th>Freq.</th><th>Freq.</th><th>Freq.</th></tr>
<tr><td rowspan="8">Rehabilitation</td><td rowspan="4">No</td><td rowspan="8">Pneumonia</td><td rowspan="2">No</td><td rowspan="8">Diabetes type II</td><td>No</td><td>G1</td><td>177</td><td>234</td><td>135</td></tr>
<tr><td>Yes</td><td>G2</td><td>49</td><td>63</td><td>50</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G3</td><td>5</td><td>4</td><td>13</td></tr>
<tr><td>Yes</td><td>G4</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td>G5</td><td>3</td><td>35</td><td>86</td></tr>
<tr><td>Yes</td><td>G6</td><td>1</td><td>8</td><td>32</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G7</td><td>0</td><td>1</td><td>7</td></tr>
<tr><td>Yes</td><td>G8</td><td>0</td><td>0</td><td>3</td></tr>
</table>

Abbreviations: Freq., Frequency

4.3.2 Categories of medical charge distribution

For simplified classifying, all the cost distribution was categorized by the 25th percentile into 3 groups, Category I was defined as 0 to the 25th percentile of medical charge, Category II was defined as the points after Category I to 2 times of the 25th percentile, and Category III was defined of the points above Category II.

As also shown in **Table 4-1**, for the Chinese data, the average medical charge is 31,170.94 CNY ($4,675.64, with an exchange rate for 1 USD = 6.5 CNY in 2015), the standard deviation (SD) is 49,397.64 CNY ($7,409.7), the 25th percentile of medical charge is 15,473.55 CNY ($2,321.0), thus the category I of medical charge is from 0 to 15,000 CNY, the category II is from 15,000 to 30,000 CNY, the category III is above 30,000 CNY.

On the Japan side, the average medical charge is 614,550 JPY ($5,530.95, with an exchange rate for 1USD = 111.1 JPY in 2011), the standard deviation is 459,360 JPY ($4,134.24), the 25th percentile of medical charge is 249,960 JPY, so the Category I is from 0 to 250,000 JPY, the Category II is from 250,000 to 500,000 JPY, and the Category III is above 500,000 JPY.

4.4 The distribution of LOS and medical charge

As the Euclidean divergence matric without the applying limitation of KLD, so all the groups could be included in this study, and for further comparison study of CCPM for cerebral infarction between China and Japan. The common factors were selected, which are rehabilitation, pneumonia and diabetes type II. The distribution of LOS and medical charge in China and Japan were shown in **Table 4-5** and **Table 4-6**.

Table 4-5 Distribution of LOS and Medical charge in China

Treatment or Comorbidities and Complications						Groups	LOS		Medical charge	
1st Factor		2nd Factor		3rd Factor			Avg.	SD	Avg.	SD
Rehabilitation	No	Pneumonia	No	Diabetes type II	No	G1	15.1	8.7	$2,829.5	$2,256.5
					Yes	G2	14.1	8.0	$2,638.1	$1,669.4
			Yes		No	G3	14.7	8.8	$3,596.0	$3,265.4
					Yes	G4	16.1	8.8	$6,005.9	$3,554.8
	Yes		No		No	G5	26.6	10.9	$4,173.7	$3,811.9
					Yes	G6	19.0	10.1	$3,757.3	$2,138.8
			Yes		No	G7	24.6	11.6	$7,776.9	$7,062.9
					Yes	G8	25.0	10.4	$8,538.4	$5,802.6

Abbreviations: Avg., average; SD, Standard Deviation.

Table 4-6 Distribution of LOS and medical charge in Japan

<table>
<tr><td colspan="6" rowspan="2">Treatment or Comorbidities and Complications</td><td rowspan="3">Groups</td><td colspan="2">LOS</td><td colspan="2">Medical charge</td></tr>
<tr><td rowspan="2">AVG</td><td rowspan="2">SD</td><td rowspan="2">AVG</td><td rowspan="2">SD</td></tr>
<tr><td colspan="2">1st Factor</td><td colspan="2">2nd Factor</td><td colspan="2">3rd Factor</td></tr>
<tr><td rowspan="8">Rehabilitation</td><td rowspan="4">No</td><td rowspan="8">Pneumonia</td><td rowspan="2">No</td><td rowspan="8">Diabetes type II</td><td>No</td><td>G1</td><td>11.4</td><td>9.8</td><td>$4,733.5</td><td>$3,819.2</td></tr>
<tr><td>Yes</td><td>G2</td><td>13.2</td><td>11.6</td><td>$5,483.8</td><td>$4,423.4</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G3</td><td>19.1</td><td>15.0</td><td>$7,707.6</td><td>$5,637.6</td></tr>
<tr><td>Yes</td><td>G4</td><td>39.0</td><td>N/A</td><td>$12,772.4</td><td>N/A</td></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td>G5</td><td>19.5</td><td>9.0</td><td>$6,875.4</td><td>$2,435.9</td></tr>
<tr><td>Yes</td><td>G6</td><td>22.8</td><td>10.6</td><td>$8,134.9</td><td>$3,561.9</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G7</td><td>30.4</td><td>14.3</td><td>$9,804.6</td><td>$4,004.5</td></tr>
<tr><td>Yes</td><td>G8</td><td>33.7</td><td>17.7</td><td>$10,786.6</td><td>$4,080.2</td></tr>
</table>

Abbreviations: Avg., average; SD, Standard Deviation.

4.4.1 Euclidean divergence matric of LOS and medical charge in China

The Euclidean divergence matric was calculated with the Chinese data to show the separation degree (Euclidean divergence) between each groups labeled from G1 to G8 by the distribution of LOS and medical charge, the result was shown in **Table 4-7**.

Table 4-7 Euclidean divergence matric of LOS and medical charge in China

Q/P	G1	G2	G3	G4	G5	G6	G7	G8
G1	0.00	0.12	0.15	0.64	0.48	0.31	0.49	0.66
G2	0.12	0.00	0.25	0.67	0.52	0.31	0.51	0.63
G3	0.15	0.25	0.00	0.67	0.46	0.31	0.43	0.68
G4	0.64	0.67	0.67	0.00	0.71	0.57	0.60	0.65
G5	0.48	0.52	0.46	0.71	0.00	0.46	0.44	0.60
G6	0.31	0.31	0.31	0.57	0.46	0.00	0.34	0.38
G7	0.49	0.51	0.43	0.60	0.44	0.34	0.00	0.50
G8	0.66	0.63	0.68	0.65	0.60	0.38	0.50	0.00

4.4.2 Euclidean divergence matric of LOS and medical charge in Japan

The Euclidean matric was conducted on the Japanese data to calculate the separation degree (Euclidean divergence) between each groups labeled from G1 to G8 by the distribution of LOS and medical charge as shown in **Table 4-8**.

Table 4-8 Euclidean divergence matric of LOS and medical charge in Japan

Q/P	G1	G2	G3	G4	G5	G6	G7	G8
G1	0.00	0.08	0.39	0.86	0.53	0.62	0.72	0.86
G2	0.08	0.00	0.33	0.79	0.46	0.55	0.66	0.79
G3	0.39	0.33	0.00	0.49	0.24	0.28	0.38	0.49
G4	0.86	0.79	0.49	0.00	0.37	0.26	0.18	0.00
G5	0.53	0.46	0.24	0.37	0.00	0.11	0.24	0.37
G6	0.62	0.55	0.28	0.26	0.11	0.00	0.16	0.26
G7	0.72	0.66	0.38	0.18	0.24	0.16	0.00	0.18
G8	0.86	0.79	0.49	0.00	0.37	0.26	0.18	0.00

4.5 Clustering results

4.5.1 Clustering for Euclidean divergence matric in China

The clustering method was performed by SPSS with Euclidean divergence matric to classify the similarity of distribution of LOS and medical charge into the same cluster. The clustering result for China was described in **Table 4-9**.

Table 4-9 Clustering results of Euclidean divergence matric in China

Groups	7 Clusters	6 Clusters	5 Clusters	4 Clusters	3 Clusters	2 Clusters
1:G1	1	1	1	1	1	1
2:G2	1	1	1	1	1	1
3:G3	2	1	1	1	1	1
4:G4	3	2	2	2	2	2
5:G5	4	3	3	3	3	2
6:G6	5	4	4	4	3	2
7:G7	6	5	4	4	3	2
8:G8	7	6	5	4	3	2

4.5.2 Clustering for Euclidean divergence matric in Japan

The clustering method was performed by SPSS with Euclidean divergence matric to classify the similarity of distribution of LOS and medical charge into the same cluster. The clustering result for Japan was described in **Table 4-10**.

Table 4-10 Clustering results of Euclidean divergence matric in Japan

Groups	7 Clusters	6 Clusters	5 Clusters	4 Clusters	3 Clusters	2 Clusters
1:G1	1	1	1	1	1	1
2:G2	2	1	1	1	1	1
3:G3	3	2	2	2	2	1
4:G4	4	3	3	3	3	2
5:G5	5	4	4	4	3	2
6:G6	6	5	5	4	3	2
7:G7	7	6	3	3	3	2
8:G8	4	3	3	3	3	2

5 Discussion

This study presented a grouping method based on the N-dimension Euclidean divergence of LOS and medical charge distribution for improvement of DRGs and DPC systems by involving the severity of illness using the cerebral infarction cases.

5.1 The case mix grouping

For further refinement of current DRGs and DPC systems and development of more accurate reflection of medical resources and provided health care services, it is needed to involve the comprehensive variables as much as possible into the case mix grouping. On the other hand, we could not use all of them, as it will results in an numerous DRGs and DPC classifications, so that it is necessary to keep the balance of the number of case mix grouping at a manageable level for practical applying. It is necessary to identify the key factors which including the severity of illness with comorbidities and complications (CC) and reflecting healthcare resource utilizations.

Regardless of analyzed data types and numbers, the decision tree method was used for this study to extract the factors related to the LOS with all the collected variables. To keep all the distribution of LOS and medical charge were available in each group for further study, three common factors were selected, which are rehabilitation, pneumonia and diabetes type II, for case mix grouping both in China and Japan.

5.2 The divergence of distribution LOS and medical charge

Length of stay and medical charge are known as the classic indexes of medical resources for given health care services[36], so when we want to calculate the divergence of medical resources utilization between each groups, it is necessary to consider LOS and medical charge as whole together.[37,38]

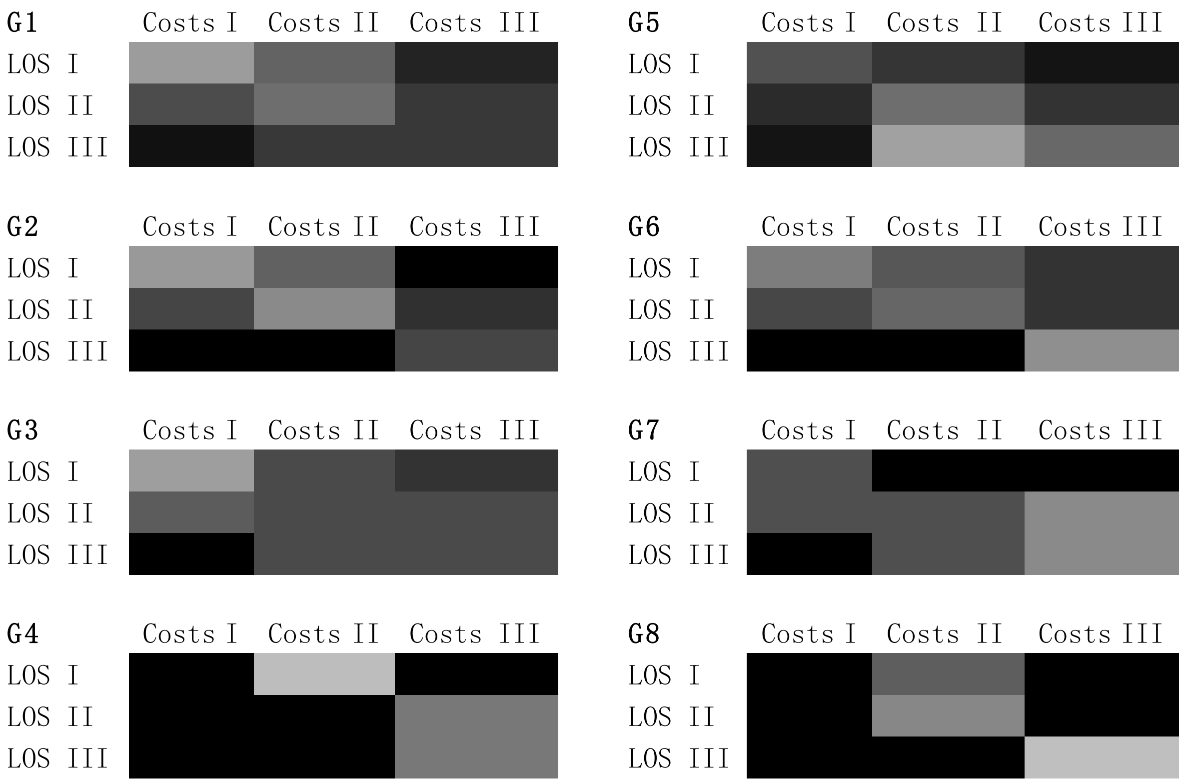
Depending upon the application involved, there are some commonly used distances for describing the divergence between sets of observations, such as Euclidean divergence and Kullback-Leibler divergence which is a measure of how one probability distribution diverges from a second, expected probability distribution.[39,40]

In this study, it is needed to consider the distribution of LOS and medical charge entirely and to make all the distribution of them in each divided group available for getting more accurate reflection between medical resources utilization and given health care services. Thus the Euclidean divergence was preferred to apply other than Kullback-Leibler divergence with the applying limitation of existing zero distribution.

5.3 The clustering of case mix grouping

The clustering of case mix grouping was performed by the hierarchical clustering based on the divergence of distribution of LOS and medical charge between eight groups in China and Japan.

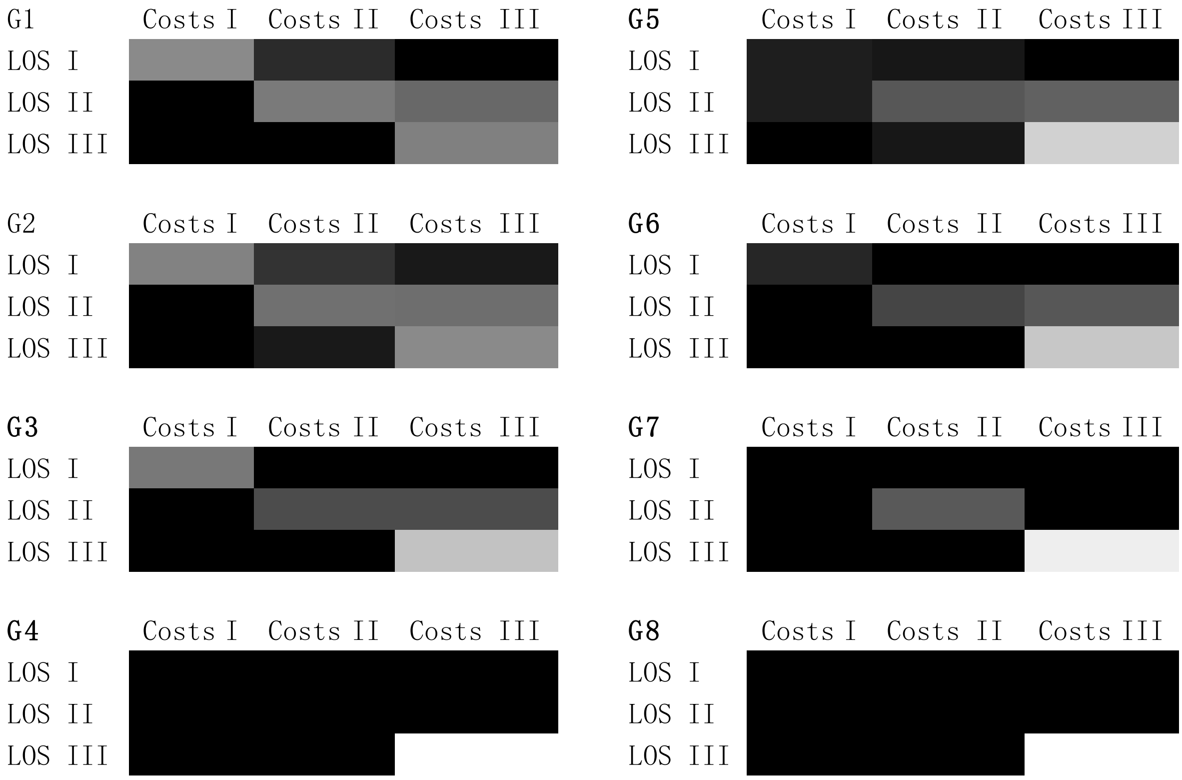
According to the Euclidean divergence matrix of LOS and medical charge in China and Japan shown in **Table 4-7** and **Table 4-8**. The results indicatred six clustering (two clusters to seven clusters) of case mix grouping both in China and Japan. However, the results of two clusters are too centralized, while on the contract aspect, the results of seven clusters are too decentralized. Therefore I used the results of three to six clusters for the further consideration.



Note: In this image, black denotes a relative frequency of 0 and white is maximal relative frequency

Figure 5-1 Heat map of distribution of LOS and medical charge in China

During the results of three to six clusters, For the Chinese data, **Table 4-7** shows that the divergence of distribution of LOS and medical charge between G5 and G6 (0.46), G6 and G7(0.34), G7 and G8 (0.50) were relative large, but the divergence between G1 and G2(0.12), G2 and G3(0.25) were relative small. Figure 5-1 also shown that the patterns of distribution of LOS and medical charge between G1,G2 and G3 are similar, but the patterns are different between G4, G5, G6, G7 and G8, thus it is preferred to consider that the result of 6 clusters is suitable for the case mix grouping in China.



Note: In this image, black denotes a relative frequency of 0 and white is maximal relative frequency

Figure 5-2 Heat map of distribution of LOS and Medical charge in Japan

For the Japanese data, **Table 4-8** shows that the divergence of distribution of LOS and medical charge between G1 and G2 (0.08), G5 and G6 (0.11) , G7 and G8 (0.18) are relative small, and divergence between G2 and G3 (0.33), G3 and G4 (0.49), G4 and G5 (0.37) are relative large. Figure 5-2 also shown that the similar patterns of distribution of LOS and medical charge were existed between G1 and G2, G5 and G6, G4,G7 and G8 accordingly. thus it is preferred to consider the result of 4 clusters is suitable for the case mix grouping in Japan.

5.4 The case mix grouping in Japan

The trial version of Comorbidity Complication Procedure Matrix (CCPM) for cerebral infarction was published by MHLW in 2016. For comparison the result of Japanese data and the CCPM grouping, both results showed into **Table 5-1** according to inclusion/exclusion criteria.

According to the four clusters result in **Table 4-10**, the first cluster of CCPM for cerebral infarction in Japan includes G1 and G2 which indicates the patients without rehabilitation and with mild CC, the second cluster includes G3 witch indicates that the patients without rehabilitation but with CC of pneumonia, the third cluster includes G4, G7 and G8 which indicates that patients with treatment of disability or with severe CC, at last the fourth cluster includes G5 and G6 which indicates that the patients with treatment of disability and with moderate CC.

Table 5-1 Clustering of case mix grouping for cerebral infarction in Japan

Treatment and Comorbidities and Complications						Groups	Japan (4 clusters)		
							CCPM GROUPING	LOS	COST
1st Factor		2nd Factor		3rd Factor				Avg.	Avg.
Rehabilitation	No	Pneumonia	No	Diabetes type II	No	G1	01	11	$4,733.5
Rehabilitation	No	Pneumonia	No	Diabetes type II	Yes	G2	01	13	$5,483.8
Rehabilitation	No	Pneumonia	Yes	Diabetes type II	No	G3	03	19	$7,707.6
Rehabilitation	No	Pneumonia	Yes	Diabetes type II	Yes	G4	03	39	$12,772.4
Rehabilitation	Yes	Pneumonia	No	Diabetes type II	No	G5	03	19	$6,875.4
Rehabilitation	Yes	Pneumonia	No	Diabetes type II	Yes	G6	03	23	$8,134.9
Rehabilitation	Yes	Pneumonia	Yes	Diabetes type II	No	G7	05	30	$9,804.6
Rehabilitation	Yes	Pneumonia	Yes	Diabetes type II	Yes	G8	05	34	$10,786.6

Abbreviations: Avg., average.

Table 5-1 shows that the clustering results of Japanese data is similar with the official published CCPM, but it suggested that the G4 stands for the patients without the treatment of rehabilitation but with severe CC could be clustered into the 05 cluster in CCPM for cerebral infarction.

5.5 The case mix grouping in China

According to the six clusters result in **Table 5-2**, the first cluster of CCPM for cerebral infarction in China includes G1, G2, G3 which indicates the patients without rehabilitation and with mild CC. The second cluster includes G4 indicates that the patients without rehabilitation but with severe CC. The third cluster includes G5 which indicates that patients with treatment of disability but without CC. The fourth cluster includes G6 which indicates that the patients with treatment of disability and with moderate CC. The fifth cluster includes G7 which indicates that the patients with treatment of disability and with severe CC, and at last the sixth cluster G8 indicates that the patients with treatment of disability and with severe CC.

Table 5-2 Clustering of case mix grouping for cerebral infarction in China

<table>
<tr><td colspan="6" rowspan="2">Treatment and Comorbidities and Complications</td><td rowspan="3">Groups</td><td colspan="2">China (5 clusters)</td></tr>
<tr><td>LOS</td><td>COST</td></tr>
<tr><td colspan="2">1st Factor</td><td colspan="2">2nd Factor</td><td colspan="2">3rd Factor</td><td>Avg.</td><td>Avg.</td></tr>
<tr><td rowspan="8">Rehabilitation</td><td rowspan="4">No</td><td rowspan="8">Pneumonia</td><td rowspan="2">No</td><td rowspan="8">Diabetes type II</td><td>No</td><td>G1</td><td>15</td><td>$2,829.5</td></tr>
<tr><td>Yes</td><td>G2</td><td>14</td><td>$2,638.1</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G3</td><td>15</td><td>$3,596.0</td></tr>
<tr><td>Yes</td><td>G4</td><td>16</td><td>$6,005.9</td></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td>G5</td><td>27</td><td>$4,173.7</td></tr>
<tr><td>Yes</td><td>G6</td><td>19</td><td>$3,757.3</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td>G7</td><td>25</td><td>$7,776.9</td></tr>
<tr><td>Yes</td><td>G8</td><td>25</td><td>$8,538.4</td></tr>
</table>

Abbreviations: Avg., average.

Comparing the clustering results between China and Japan, there are different in clustering results between China and Japan, the G5 (the cases only with rehabilitation) was classified into an individual cluster for Chinese data, but G4 (without rehabilitation and diabetes type II), G7 (with rehabilitation and diabetes type II), G8 were classified into the same cluster for Japanese data.

As one of the reasons is that Chinese data were collected from the rehabilitation institute and the Japanese data were mainly collected from the general hospitals, so the rehabilitation is major treatment for the Chinese data, while it is just a common treatment in general hospitals for the Japanese cases.

6 Study Limitation

This study has several limitations. As this study was conducted with data from a single rehabilitation institute in China, findings might have been different in other parts of the country or in other types of health institute.

The suggested sample number was calculated by the following formula [41]

$$n = \frac{N}{\frac{N-1}{p(1-p)}\left(\frac{L}{2k}\right)^2 + 1}$$

N stands for the total number of survey subjects, n stands for the suggested sample number, p stands for probability (generally as 0.5), k stands for degree of reliability (generally as 99%), L stands for allowable error (generally as 0.05).

There were about 6 million patients with cerebral infraction in China in 2015[42]. Meanwhile there were about 1.12 million patients with cerebral infarction in Japan in 2011[43]. According the result of this formula, both of suggested cases with cerebral infarction in Japan and China were about 1,500.

In addition, the Chinese data was limited to the period in 2015, a relatively short period of the data collection. Moreover, although surgery is very important treatment for patient classification the selected cases in this study are almost without surgery, and the sample number is less than one thousand. It may not be enough for study all the situation of patients with cerebral infarction. Therefore, this study was focus on the specific field of the patients with cerebral infarction, and the data for demographic and clinic characteristics were limited such as without surgery and with health insurance. If it is possible to use more data with richer clinic characteristics, it will show more accurate estimates which will be able to compare the results between China and Japan.

7 Conclusion

A grouping method based on the N-dimension Euclidean divergence using the cerebral infarction cases was discussed in this study. This method was focused on the distribution of LOS and cost with other variables such as treatment related to severity and CC. It assigned the groups with similarity of distribution of medical resources such as LOS and cost into the same cluster, which with the similar medical resources. The method has been proved to be suitable and verified by analyzing and comparing the research subjects in China and Japan.

Though the clustering results of Japanese data is similar with the official published CCPM, the present study suggested that the G4 stands for the patients without the treatment of rehabilitation but with severe CC could be clustered into the 05 cluster in CCPM for cerebral infarction.

Although the present study indicated that the method could be used in China, the clustering results need to be further discussed in clinical practice, and it is necessary to extend it into other diseases. The clustering results may be used as a guideline for DRGs development both in China and Japan.

Acknowledgements

I would like to express my deep gratitude to associate professor Keiichi Saito and associate professor Toshio Ogawa, my research supervisors, for their patient guidance, and useful critiques of this research work.

I would also like to thank professor Yamamoto and Chen and other professors and teachers in faculty of medical information management of International University of Health and Welfare and the colleagues in China Rehabilitation Research Center for their advice and assistance in keeping my progress on schedule.

Finally, I would like express profound gratitude to my family members for their constant support and enthusiastic encouragement to accomplish the study.

References

[1] Iglehart J. Prospective payment panel faces key decisions as new DRG system

[2] SAKOI, M. The Making of DPC (Diagnosis Procedure Combination): The difference between DRG and DPC. Journal of the National Institute of Public Health. 2014;63,488-501.

[3]Shinya Matusda, Kenji Fujimori, Kazuaki Kuwabara, Kohichi B Ishikawa, Kiyohide Fushimi, Diagnosis Procedure Combination as an Infrastructure for the Clinical Study Development of CCPM: A revision proposal for a Japanese case-mix classification system that more closely reflects severity. Asian Pacific Journal of Disease Management. 2011;5(4),81-87

[4] Wang, Z., Liu, R., Li, P., Jiang, C., & Hao, M. How to Make Diagnosis Related Groups Payment More Feasible in Developing Countries- A Case Study in Shanghai, China. Iran J Public Health. 2014;43(5), 572-578.

[5] Wang, Z., Liu, R., Li, P., & Jiang, C. Exploring the transition to DRGs in Developing Countries: A case study in Shanghai, China. Pak J Med Sci. 2014; 30(2), 250-255.

[6] SAITO Keiichi, FUSHIMI Kiyohide. A Method to Construct CCP Matrices with Kullback-Leibler Divergence, and an Application to Type 2 Diabetes Mellitus. BMFSA. 2014;(27), 131-132.

[7] SAITO Keiichi, FUSHIMI Kiyohide. A Method to Design CCP Matrix with Kullback-Leibler Divergence, and an Application to Type 2 Diabetes Mellitus. The 34th Joint Conference on Medical Informatics. 2014; 142-143.

[8] Definition and Uses of Health Insurance Prospective Payment System Codes (HIPPS Codes), CMS Division of Institutional Claims Processing, Centers for Medicare and Medicaid Services, 2008.3.17

[9] Averill RF. The evolution of case-mix measurement using DRGs: past, present and future. Stud Health Technol Inform. 1994;14:75-83.

[10] Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis related groups. Medical Care. 1980;18(2):1–53

[11] Mayes, Rick. The Origins, Development, and Passage of Medicare's Revolutionary Prospective Payment System. Journal of the History of Medicine and Allied Sciences. Oxford: Oxford University Press. 2007; 62 (1): 21–55.

[12] De Marco MF, Lorenzoni L, Addari P, Nante N. [Evaluation of the capacity of the APR-DRG classification system to predict hospital mortality]. Epidemiologia e prevenzione. 2002;26(4):183-90.

[13] Averill R, Muldoon J, Vertrees J, Goldfield N, et al. The evolution of case mix measurement using diagnosis-related groups (DRGs). Physician Profiling and Risk Adjustment. Goldfield: Aspen Publishers, 1999:43-52.

[14] Muldoon J. Pediatrics and DRG case mix classification. Physician Profiling and Risk Adjustment.Goldfield N, Boland P, eds. 1996;24:252–270

[15]Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. Medical care. 1980;18(2 Suppl):iii, 1-53.

[16] Averill, R. F., J. H. Muldoon, J. C. Vertrees, et al. The Evolution of Casemix Measurement Using Diagnosis Related Groups (DRGs) 3M HIS Working Paper. Wallinford, CT: 3M Health Information Systems, 1997

[17]Hamada H, Sekimoto M, Imanaka Y. Effects of the per diem prospective payment system with DRG-like grouping system (DPC/PDPS) on resource usage and healthcare quality in japan. Health Policy. 2012;107: 194–201.

[18]Davis CK, Rhodes DJ. The impact of DRGs on the cost and quality of health care in the united states. Health Policy. 1988;9: 117–131.

[19]Uematsu H, Hashimoto H, Iwamoto T, Horiguchi H, Yasunaga H. Impact of guideline-concordant microbiological testing on outcomes of pneumonia. Int J Qual Health Care. 2014; 26: 100–107.

[20] Hamada, H., Sekimoto, M., & Imanaka, Y. Effects of the per diem prospective payment system with DRG-like grouping system (DPC/PDPS) on resource usage and healthcare quality in Japan. Health Policy. 2012;107(2-3), 194-201.

[21] Meng Q. The impact of provider payment reforms on cost containment. Chinese Health Economics Research. 2002; 9: 18–20 (in Chinese).

[22] Wu A, Li Y et al. DRG-based payment reform for urban health insurance scheme. Chinese Journal of Health Economics. 2004;9: 38–39 (in Chinese).

[23] Wei-Yan Jian, Ming Lu, Tao Cui, Mu Hu, : Evaluating performance of local case-mix system by international comparison: a case study in Beijing, China. The international Journal of Health Planning and Managemen. 2011;26(4),471-481.

[24] Karen Eggleston, Li Ling, Meng Qingyue, magnus Lindelow & Adam Wagstaff, :Health service delivery in China: A literature review. Health Economics. 2008; 17, 149-165.

[25] Steven, P. D. m. S.
China has started introducing Diagnosis Related Groups (DRGs) pricing system. https://www.linkedin.com/pulse/china-has-started-introducing-diagnosis-related-groups-steven. 2017.2.17

[26] Yang Z, Zhao Z et al. The impact of DRG-based user charge on expenditure control. Chinese Journal of Health Economics.2001;4: 25–26.

[27] J. R. Quinlan, Induction of decision trees. Machine Learning. 1986;1(1),81-106.

[28] Kass, G.V. An exploratory technique for investigating large quantities of categorical data. Applied Statistics. 1980;29(2),119-127.

[29] M. Abe, H. Toyama, K. Saito :Evaluation of DPC classification by decision tree method –Analysis and discussion nabout pneumonia. BMFSA. 2014;16(1),7-13.

[30] C. Sakamoto, H. Toyama, K. Saito. :Decision tree analysis of factors relating ot lung cancers using DPC data. BMFSA. 2014;16(1), 1-6.

[31] T. Saito., H. Yadohisa, Data analysis of asymmetric structures, Marcel Dekker, New York, 2004:103-105.

[32] T. Saito. Analysis of asymmetric proximity matrix by a model of distance and additive terms. Behaviormetrika. 1991;29:45-60.

[33] S. C. Johnso. Hierarchical clustering schemes. Psychometrika. 1967;32(3): 241-254.

[34] G. Punj, D. W. Stewart. Cluster analysis in marketing research: Review and suggestions for application. J. of Marketing Research. 1983;20(2):134-148.

[35] Ishii, M. DRG/PPS and DPC/PDPS as Prospective Payment Systems. Japan Med Assoc J. 2012; 55(4): 279-291.

[36] Ma, Y., Wang, X. M., et al. Evaluation of admission characteristics, hospital length of stay and costs for cerebral infarction in a medium-sized city in China. Eur J Neurol. 2010;17(10): 1270-1276.

[37] Li, Y., Liu, H., Wang, J., et al. Variable lengths of stay among ischemic stroke subtypes in Chinese general teaching hospitals. PLoS One. 2012; 7(9):34-35

[38] Wei, J. W., Heeley, E. L., Jan, et al. Variations and determinants of hospital costs for acute stroke in China. PLoS One. 2010; 5(9):45-46.

[39] Kullback, S.; Leibler, R.A. On information and sufficiency. Annals of Mathematical Statistics. 1951,22 (1): 79–86.

[40] P. Lindstrom, B. M. Namee, S. J. Delany. Drift detection using uncertainty distribution divergence. Evolving System. 2013;4(1):13-25.

[41] http://blog.thetheorier.com/entry/number-of-sample

[42] http://news.163.com/15/0523/07/AQ9I8TTO00014Q4P.html

[43] http://www.seikatsusyukanbyo.com/statistics/2016/009093.php