定額支付的病例醫令項目之合理性研究

蔣肇慶、林熙禎
中央大學資訊管理系

摘要

全民健康保險制度是自給自足的社會保險制度。因而，對於醫療給付單位及醫療院所而言，在有限的醫療給付資源情形之下，必須建立醫療的收支平衡，方為持續經營之道。醫療給付單位為有效的控制醫療費用，實施總額預算制的醫療給付制度，其目的在協助醫療院所能有效的規劃醫療資源、控制醫療成本及提昇服務品質。在總額預算制度之定額支付的病例，醫令執行項目內容的問題，可以藉著醫令合理性予以適當的解決。因此，本研究提出一個 ART_N.V.計劃，藉由修改後之關連式規則與值比率的組合，將定額支付下之疾病醫令項目，實驗得出基本醫令群。此基本醫令群不但可以協助醫療給付單位，更精確的計算出病例的必須執行醫令項目及支付的費用；而且在醫療院所方面，可以節省不必要的處置與成本支出，提升營運的績效。此研究論證定額支付的病例之醫項目及支付的費用，作為目前及未來在定額支付之下病例開立醫令的參考。

關鍵字：基本醫令群、值比率、定額支付

Identifying Interesting Patterns in a Medical Database System ---Fixed Amount Payment's Order Items

Jaw-Ching Chiang、Shi-Jen Lin
Department of Information Management National Central University

Abstract

Maintaining a financial balance given limited medical payments is essential for health insurance payment units and hospitals. The Bureau of National Health Insurance (BNHI) implemented the Prospective Payment System of the Global Budget System to assist hospitals in planning and controlling medical care costs and service quality. Meanwhile, the BNHI also devised various plans for strengthening the operational utilization of medical resources. This study suggests the ART_N.V. scheme for computing Basic Order Groups (BOG) and assessing their costs/benefits in a Fixed Amount Payment System. The BOGs used to provide a contrast in the experiment are more savings in medical care costs than health insurance reporting payments. Therefore, the BOG can not only assist payment units in reducing costs, but also can assist hospitals in operating efficiently using the Fixed Amount Payment System. Consequently, this study demonstrates a standard of payment for current and future basic and suitable order items of Fixed Amount Payment system references.

Keywords: Basic Order Group, Value Rate, Fixed Amount Payment system

1. Introduction

The health insurance payment system in Taiwan in Taiwan was changed to the National Health Insurance System, from the previous Government Employees' Insurance, Labor Insurance, and Farmers' Insurance etc. on March 1, 1995. It is a social medical insurance system, thus it adopts a self-sufficient financial system. Therefore, to minimize payment expenditures and achieve a financial balance, the payment system for certain diseases has gradually changed. In July 1995, BNHI prescribed a case payment operation standard and the payment system for some cases was changed to 'Case Payment' from 'Fee for Service'. BNHI implemented the prospective payment system to assist all hospitals in planning and controlling their medical care costs, service quality and other medical resources. Furthermore, BNHI ordered that for each case payment, the disease order items[1] should comprise at least 65% of the basic required examinations and treatments (NHI, 2000).

The results of the Taiwan Public Health Report 2003 submitted by the Department of Health, Executive Yuan in 2003 demonstrated that healthcare expenditures exceed the premiums paid by the insured and the reserve funds were below the legal minimum threshold, and below the equivalent of one-month of medical expenditure in December 2001(DOH, 2003).

In this Case Payment System, hospitals must execute 65% basic required examinations and treatments. These changes can help hospitals to control medical resource costs and define standard order items. Under this condition, it is impossible to understand how '65% of basic required examinations and treatments' are calculated? Is the '35% of exanimations and treatments' unimportant?

This study focuses on the problem of 'which order items are basic required examinations and treatments'. Therefore, this study suggests the ART_N.V. scheme to solve this problem. This scheme employs the $aRule_D(T)$ algorithm (modified Apriori algorithm), Self-Organizing Map (SOM) and Value Rate method to analyze and generate Basic Order Groups (BOG).

This study employs the database of an academic hospital to generate BOGs, and compute the simulation fee (experiment contrast payment) of these BOGs. Meanwhile, this study employs the National Health Research Institute (NHRI) database to find health insurance reporting payments. The databases of both academic hospitals and NHRIs employ same cases of ICD-9-CM 571.4. '**Different Rate**' then is used to compare and analyze the difference between the BOG experiment contrast payments and the health insurance reporting payments. This study aims to find these BOGs, which can approach the refining order items.

The Via Experiment results demonstrate that the BOGs of the ART_N.V. scheme are

[1] Order item: Medication, laboratory examination, and any procedures prescribe by physician for patient.

available. Moreover, the experiment contrast BOGs save more medical care cost payments than health insurance reporting payment. Meanwhile, BOGs maintain service quality without reducing medical quality. Additionally, the basic and suitable order items of the BOG can be used by physicians to prescribe order references in the Fixed Amount Payment System.

The rest of this paper is organized as follows. Section 2 outlines the methodology and tools employed in this study. Section 3 then introduces the ART_N.V. scheme. Subsequently, experimental results are presented in section 4. Next, section 5 discusses the findings. Finally, section 6 summarizes the conclusions.

2. Method

Data Mining and Knowledge Discovery are processes for extracting the previously unknown; these processes involve comprehensible and actionable information from a large database, and are used to increase enterprise advantages and niches. Data Mining is a key focus of scholars and researchers, because it involves several important analytical methods, such as the Association rule(Agrawal & Srikant, 1994), Data Cube(Chaudhuri & Dayal, 1997), Clustering(McIlvance, 1983 ; Ng & Han, 1994), and so on. These methods can simulate patterns and models of real world situations. This technology has been applied commercially in various ways, including Marketing Research, Consumer shopping Behavior, Insurance Claims Analysis, and so on.

This study wishes to analyze the medical cost of diagnosis order items and save medical resources. Consequently, the Association rule and Neural clustering are applied to solve the current problem. These methods are introduced below.

2.1 Association rule and its properties

The Association rule is intended to identify transaction items that imply the presence of other items in the same transaction. Support and confidence are two key parameters in the Association rule, and indicate the relative occurrence and strength, respectively, within the input data(IBM, 1999a). Therefore, mining results depend on controlling support and confidence values. Agrawal, the creator of Data Mining, attempts to associate the Apriori Algorithm to explore the correlation between each item and multiple relations, which are stored in a large volume data warehouse(Agrawal & Srikant, 1994).

The Apriori algorithm involves the following steps:

(1) Scanning all data warehouses to identify large itemsets that exceed the minimum support.

(2) Subsequently, this study employed these large itemsets to create these association rules.

Figure 1 illustrates the Apriori algorithm. The candidate C_k= apriori-gen(L_{k-1}) is created and presented using this process, as follows: $C_k = L_{k-1} * L_{k-1} = \{ X \cup X' \mid X, X' \in L_{k-1}, \mid X \cap X' \mid = k-2 \}$.

```
L1= { large 1-itemsets } ;
for (k=2; Lk-1 ≠ φ ; k++) do begin
    Ck=apriori-gen(Lk-1); //New candidate
    forall transactions t ∈ D do begin
        Ct= subset(Ck, t); // candidate contained in t
        forall candidate c ∈ Ct do
            c.count++;
    end
    Lk = { c∈ Ck | c.count≧minsup } ;
end
Answer = Uk Lk;
```

Figure 1 Apriori algorithm

Let **I** = $\{A_1, A_2, \dots A_m\}$ denote an itemset of **attributes** or **items**. A transaction T = $\{A_p, \dots A_q\}$ $A_i \in$ **I** for $1 \leqq p \leqq i \leqq \leqq q \leqq n$, and **itemset** are subsets of **I**. A transaction **T** is said to contain an itemset **X** if $\mathbf{X} \subseteq \mathbf{T}$. Let **D** = $\{T_1, T_2, \dots T_n\}$ denote a set of transactions and be called a **database**. The **support** $supp_D(X)$ of an itemset X in **D** is the fraction of the total number of transactions in D that contain $supp_D(\mathbf{X}) = \left|\{ \mathbf{T} \in \mathrm{D} \mid \mathbf{X} \subseteq \mathbf{T} \}\right| / \left|\mathbf{D}\right|$. Furthermore, the ratio of $supp_D(\mathbf{X} \cup \mathbf{Y}) / supp_D(\mathbf{X})$ is called **confidence** $conf_D(\mathbf{X}, \mathbf{Y})$ $X \Rightarrow Y$ in D, where $\mathbf{Y} \subset \mathbf{I}$, $\mathbf{X} \cap \mathbf{Y} = \Phi$, which indicates the probability that transaction **Y** occurs in D when transaction X occurs in D(Rothermel, 1997 ; IBM, 1999a ; IBM, 1999b).

2.2 Neural clustering and its properties

Neural Clustering is a computation system comprising artificial neuron networks which perform information processing. Neural Clustering possesses the advantages of learning ability, tolerance, parallel processing and trace memorizing(Kohonen, 1991 ; Lee & Gyvez, 1996 ; Chiou, 1998).

Neural Clustering detection can be guided through unsupervised Neural Networks, which form clusters on the vast amount of data without knowing the content of the output clusters and models. Neural Clustering, which employs a Kohonen Feature Map neural network, thus is employed in this study. Notably, Kohonen Feature Maps employ a process termed Self-Organizing Map (SOM) for grouping similar input records. Therefore, SOM creates several clusters(IBM, 1999b ; Kohonen, 1991).

Depending on the data cluster rule, clustering analysis can group similar data with high similarity inside and low similarity outside. Professor Ng suggested CLARANS clustering, which employs a random search and comparison method, and obtains a minimum clustering neighbor region (Ng, 1997).

The Clustering SOM is learned competitively and trained iteratively. The following steps (algorithm of ***Clustering$_{Dc}$ (Tc)***), explain the execution of the Clustering SOM method.

Step 1. A sample transaction vector T_{ck} is chosen randomly from database Dc. The distance between the random sample T_{ci} and every transaction then is calculated.

$$\left\|T_{ck} - T_{ci}\right\| = \sum_{j=1}^{m}\left[T_{ck}(m_{kj}) - T_{ci}(m_{ij})\right]^2$$

Step 2. Find the minimum distance of output transaction vector $T_{ci}(m_{ij})^*$, which is closest to T_c.

$$\left\|T_{ck} - T_{ci}{}^*\right\| = \min_{j=1}^{m}\left\{\sum_{j=1}^{m}\left[T_{ck}(m_{kj}) - T_{ci}(m_{ij})\right]^2\right\}$$

Step 3. Adjust the weight vectors. η is the learning rate coefficient, and $h_{bi}(t)$ is the neighbor kernel centered on the winner unit. $h_{bi}(t) = \exp(\,\|\, T_{ck} - (\,T_{ci}^*)\,\|\,/\,2\,\sigma^2(t)\,)$, with $\sigma^2(t)$ being as suitable decreasing monotonically with the function of time.

$$T_{ci}(t+1) = T_{ci}(t) + \eta . h_{bi}(t) . \left[T_{ck}(t) - T_{ci}(t)\right]$$

Step 4. Finally, increase time t and adjust learning rate η iteratively, until the weight vector value converges or the Euclidean distance becomes smooth.

Some clusters are obtained using this approach. Each cluster $Q_i = \{\, Q_{ip}, Q_{iA}{}^* = \langle m_{i1}{}^*, \ldots\, m_{im}{}^* \rangle \,\}$ has clustering population Q_{ip}, and attributes $\langle m_{i1}{}^*, \ldots\, m_{im}{}^* \rangle$, which arrange attributes based on the chi-square value

$$X^2 = \sum_{i=1}^{m}\left\{\left(T_{ck} - T_{ci}\right)^2 / T_{ci}\right\}.$$

2.3 APORES Model

The APORES model employed the Association Rule, Partially Ordered Set and Top methods to generate the Prototyping Basic Order Group. Subsequently, the model used Relative Strength to correct PBOG. Figure 2 shows the model.

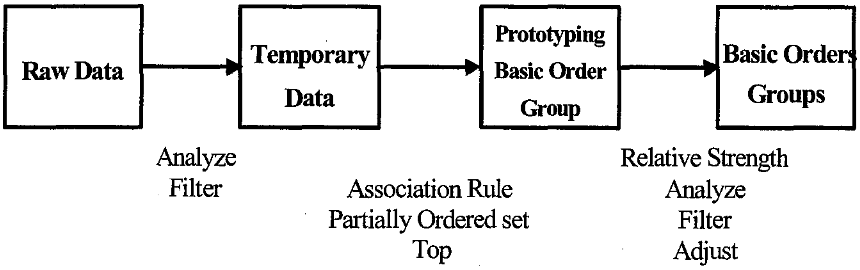

Figure 2 APORES model

Relative Strength (*rStrength (S, T)*) was the important viewpoint of the APORES model. It defined *rStrength (S, T)* as $|S \cap T| / |S \cup T|$, while S and T were Prototyping BOGs (PBOG). PBOGs were generated based on the modified association rule. Subsequently, these PBOGs were computed to create rStrength (S, T) in two and two together. If these values of rStrength (S, T) exceeded the threshold, PBOGs were called BOGs(Chiang & Lin, 2002).

In this model, these PBOGs were computed to create rStrength (S, T) in two and two together. This rStrength(S, T) computing was computed by PBOG themselves only. Therefore, 'rational and suitable' could be discussed continuously. Therefore, this study suggests applying the ART_N.V. scheme to solve this problem.

3. ART_N.V. Scheme

This study uses the ART_N.V. scheme, which is designed for creating BOGs and comprises the *aRuleD(T)*, TOP, Neural Clustering, Value Rate and Different Rate. Figure 3 illustrates the executive flow.

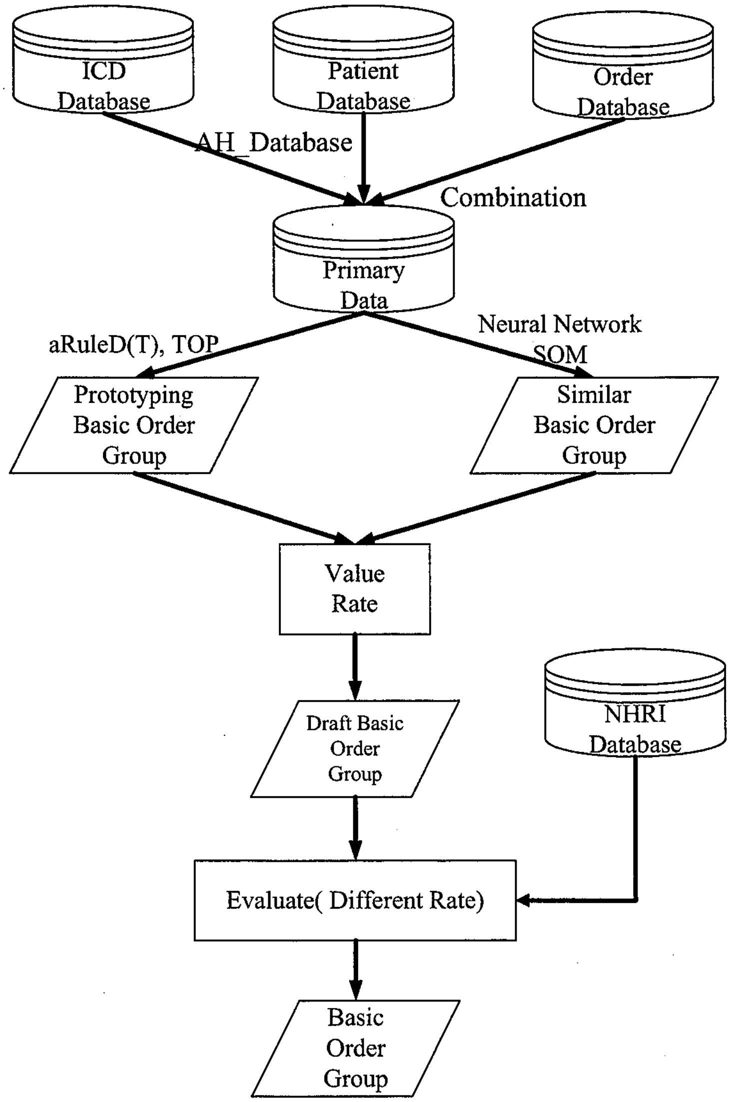

Figure 3 The Scheme of ART_N.V.

The ART_N.V. scheme involves the following steps.

Step 1: The primary data comprises ICD, Patient and Order databases.

Step 2: aRuleD(T) and TOP transform primary data to Prototyping BOGs.

Step 3: Neural Clustering SOM transforms primary data to Similar BOGs.

Step 4: Acting on Value rate, draft BOGs are generated using these Prototyping BOGs and Similar BOGs.

Step 5: The inside of the draft BOGs elements is allowed to employ a reporting fee to calculate the payments of every draft BOGs.

Step 6: Finally, this study compares and tests statistical hypotheses between the experiment contrast payment for every draft BOGs and the health insurance reporting payments of each hospital level.

3.1 Prototyping Basic Order Group

This study employs the characteristics of the Apriori algorithm to identify frequent large itemsets. First, frequent and large itemsets(iFrequentLarge) are defined, and then the details of the scheme are illustrated.

[Definition]

Let **X, Y** $\subseteq$ **I** represent set of items from D. An itemset ***iFrequentLarge***(*frequent large itemset*) is one with support of no less than min_sup (threshold of support).

$$\boldsymbol{iFrequentLarge\ (X, Y, t)} = \{ (W, Z, t) \mid W \cup Z = \mathbf{X} \cup \mathbf{Y},\ W \cap Z = \phi,\ supp_D(\mathbf{X} \cup \mathbf{Y}) \geqq t \}.$$

Next, the Apriori algorithm is modified to identify and generate frequent large itemsets. The algorithm displayed in Fig. 4 provides an overview of the ***aRule$_D$(T)***.

```
Input: Database
Output: Frequent Large Itemsets
L1= { large 1-itemsets } ;
for (k=2; Lk-1 ≠ φ ; k++) do begin
    Ck=apriori-gen(Lk-1); //New candidate
    forall transactions t ∈ D do begin
        Ct= subset(Ck, t); // candidate contained in t
        forall candidate c ∈ Ct do
            c.count++;
    end
    Lk = { c∈ Ck | c.count≧tSupport } ;
    iFrequentLarge = U Lk;
end
```

Figure 4 *aRule$_D$(T)* algorithm

Subsequently, this study uses the concept of partially ordered sets (Birkhoff, 1967) to design the *TOP*. This study employs the *TOP* property for creating a framework of similar partially ordered sets using all frequent large itemsets. Therefore, the *TOP* can be used to establish a Prototyping Basic Order Group (PBOG) algorithm as shown in Fig. 5. The *TOP* is defined as follows.

[Definition]

Let Dp denote a partially ordered set of attributes$\leqq$, and moreover let Dp have no maximal element. If T exists there is no S in *Dp*, and $S \neq T$ and $T \subseteq S$, in which case T is called ***TOP***. It is represented as

$$\boldsymbol{TOP}(Dp, \leqq) = \{T \in Dp \mid \text{there is no } S \text{ in } Dp,\ S \neq T \text{ such that } T \subseteq S\}.$$

Input: Frequent Large Itemsets
Output: PBOG

```
1.  Lp = ⌈(Σ_{i=1}^{n} |T_i| / n) + 0.5⌉;  //  average size of order items of transactions
2.  POS = φ;     // partially ordered set
3.  PBOG = φ;   // prototyping basic order group
4.  for ( all iFrequentLarge and | iFrequentLarge |  ≦  Lp ) do begin{
5.        if ( conf_D(X, Y) > tConfidence) {
6.           POS = POS  ∪  iFrequentLarge ;
7.           break;
8.      }
9. }
10. PBOG = TOP(POS,  ≦);
```

Figure 5 PBOG algorithm

3.2 Similar Basic Order Group

Neural Clustering employs a *Kohonen Feature Map neural network*. Moreover, Kohonen Feature Maps use a process called self-organization map (SOM) to group similar input records together (Kohonen, 1991；IBM, 1999). This characteristic most frequently occurs in common, and groups the related records accordingly. Therefore, this characteristic is employed to generate Similar Basic Order Group(SBOG).

First, let Dc denote a set of transaction vectors $\{T_{c1}, \ldots T_{ci}, \ldots T_{cn}\}$, and let each m_{il}, $1 \leqq l \leqq m$ $m_{il} \in$ Tci, represent a set of binary digits of {0, 1}. The ***Vector Matrix*** is employed to transform the transaction of physical items into the transaction vector. The *Vector Matrix* is defined as follows.

[Definition]

Let **I** = $\{A_1, \ldots A_j \ldots A_m\}$ and $T_k = \{T_{k1}, \ldots T_{ki}, \ldots T_{kn}\}$ denote two sets, and moreover let $T_{ki} = \{A_p, \ldots, A_q\}$ $1 \leqq i \leqq n$, $1 \leqq p < q \leqq m$. Moreover, let $T_C = \{T_{C1}, \ldots T_{Ci}, \ldots T_{Cn}\}$ and $T_{Ci}=\{m_{i1}, m_{i2}, \ldots m_{im}\}$ denote a set of binary digits of {0, 1}. R can be represented as n * m matrix M(R) = [m_{ij}]

$$m_{ij} = \begin{cases} 1 & if\ A_j \in T_{ci} \quad 1 \le j \le m, 1 \le i \le n \\ 0 & otherwise \end{cases}$$

R = { for i= 1,..n (for j = 1,..m (if $A_j \in T_{ki}$ then $m_{ij} = 1$, otherwise $m_{ij}= 0$)) }

then M(R) is termed the ***Vector Matrix*** of R, and *Tci* is termed the ***Transaction Vector*** of T_{ki}.

Subsequently, the neural clustering clusters this transaction vector to enable the identification of different data types. Neural clustering partitions transaction vectors Q into clusters Q_i, i = $_{1, \ldots, n}$. Each Tci must belong to exactly one cluster Q_i. If the population of certain clusters exceeds the threshold, these clusters are termed Similar Basic Order Group (SBOG).

Therefore, this study employs both *Vector Matrix* and Clustering SOM to create SBOGs. Figure 6 illustrates the SBOG algorithm. To confirm the correctness of PBOGs, this study employed **Value Rate** to be corrected. Meanwhile, the **Value Rate** is introduced in the following section.

Input: Database
Output: SBOG
1. Lp = $\lceil (\Sigma_{i=1}^{n} |T_i| / n) + 0.5 \rceil$; // average size of order items of transactions
2. tPopulation; // threshold of population
3. SBOG = ϕ ; // similar basic order group
4. $Q_i = \phi$; // cluster
5. **for** (**all** transaction *T*, $T \in$ D) **do begin**{
6. Tc = ***Transaction Vector*** (T_k) ; // from ***Vector Matrix***
7. }
8. ***Clustering***$_{Dc}$ ***(Tc)***; // SOM
9. Q_{iA} = ***Transaction Vector***$^{-1}$(Q_{iA}*); // inverse function of Vector Matrix
10. **for** (**all** Q_{ci}) **do begin**{
11. **if** (population(Q_{ip}) $\geqq$ tPopulation) {
12. SBOG = SBOG $\cup$ { $\{A_{i1}, \ldots, A_{i(Lp)}\}$ };
13. break;
14. }
15. }

Figure 6 SBOG algorithm

3.3 Value Rate

PBOGs and SBOGs were obtained from *aRule$_D$(T), TOP, Vector Matrix* and Clustering SOM respectively. To confirm the accuracy of PBOGs, this study employs SBOGs to confirm their ***Value Rates*** and generate draft BOGs.

First, this study transforms PBOGs from the *Vector Matrix* into a format of '0' or '1'. Since every PBOGs and SBOGs is considered a vector, PBOGs and SBOGs are vectors, and they shot out from (0, 0) in the n-dimension vector space. Meanwhile, Boolean Logic 'equivalence' is applied to the same sub-vectors of the PBOGs and SBOGs. Similarity increases if the number of equal sub-vectors of PBOGs and SBOGs exceeds a certain threshold.

Next, draft BOGs are created from the function of $\mathbf{Val}(\vec{P}\ ,\ \vec{Q})$(definitions are as follows) of PBOGs and SBOGs. Figure 7 displays draft algorithm for generating BOGs.

Subsequently, draft BOGs are assessed based on fee difference. Therefore, two formulae are defined for computing the 'Experiment Contrast Payment' and 'Health Insurance Reporting Payment'.

Experiment Contrast Payment =

Σ Health Insurance Payment at each level hospital (inside order items of BOG) +

diagnostic fee (at each level hospital).

Health Insurance Reporting payment =
The average reporting payments of each department at each level hospital were 571.4 code.

Finally, to clarify the fee difference between the Experiment Contrast Payment and Health Insurance Reporting Payment, the following formula is used to calculate the 'Different Rate'. The Different Rate is used to test the statistical hypotheses, and to confirm the correctness and suitability of BOGs.

Different Rate = (Health Insurance Reporting Payment −

Experiment Contrast Payment) ÷

Experiment Contrast Payment * 100%

[Definition]

Let $\vec{P} = (P_1, ..., P_j, ...P_n)$ represent a ***Transaction Vector.*** P_j denotes the decreasing jth rank in all items by frequency, where P_j represents the last item with a value of '1'. Additionally, j has a ***length*** of $\vec{P}$, and is represented as ***len***($\vec{P}$).

□

[Definition]

Let $\vec{P}$ 、 $\vec{Q}$ be the *Transaction Vector* and let n = max(*len*($\vec{P}$), *len*($\vec{Q}$)). Mark '$\odot$' represent the Boolean function '*equivalence*'. The '| $\vec{P} \odot \vec{Q}$ | / ***(n+1)***' is called the ***Value Rate*** between $\vec{P}$ and $\vec{Q}$. This study presents ***Val*** ***(***$\vec{P}$, $\vec{Q}$ ***)***. If ***Val(***$\vec{P}$, $\vec{Q}$ ***)*** $\geqq$ threshold and ***Val(***$\vec{P}$, $\vec{Q}$ ***)*** approaches 1 then $\vec{P}$ and $\vec{Q}$ are much similar.

```
Input: PBOGs(), SBOGs()
Output: Draft BOGs
tValuerate     // threshold of value rate
dBOG;     // draft basic order group
\vec{P_i} =  Transaction Vector ( P_i );
\vec{S_j} =  Transaction Vector ( S_j );
for ( i = 1; m; i ++) {
    k= 0;
    for ( j = 1; n; j ++) {
       if Val(\vec{P_i}, \vec{S_j}) ≧ tValuerate
          k++;
       else break;
    }
    if  k ≧ int( n/2 + 0.5 )
       dBOG  =dBOG ∪ { P_i };
}
```

Figure 7 Value Rate Algorithm

4. Experimental Evaluation

To assess scheme correctness, this study prepared some Academic Hospital databases and National Health Research Institute databases. This conduct the experiment using DB2(IBM, 1995), Intelligent Miner (IBM, 1996) and Excel(Microsoft, 2000) run on a personal computer.

4.1 Materials

Code 571.4 cases of ICD-9-CM were selected from 19,974,470 outpatients in some Academic Hospital, from January to December 2001. The code of 571.4 cases, which is 'chronic hepatitis B', must be traced over a long time, and has a stable illness state. The 571.4 cases belongs to series of Internal Medicine field. Therefore, it is an ideal material choice.

First, raw data from the outpatient database of the Academic Hospital, which was a combination of the ICD database, order database and patient database, was searched. This database was called the AH_database, and produced PBOGs and SBOGs. Subsequently, screened suspect records and eliminated unreasonable records were created using primary data from the AH_database. These data included physician name, patient age, department, patient sex, ICD-9-CM code and the first 20 order items (only 1.80% patients had over 18 order items). The departments selected for study here included 'Medical(Internal Medical)', 'Medical Digestive' and 'Family', along with the 'All patients' category. The 'All patients' category is indicated with the code ICD-9_CM 571.4. 'Medical Digestive' denotes a medical specialization in digestion. Therefore, the 'Medical', 'Medical Digestive' and 'Family' departments are subsets of 'All patients'. Thus, the set of relations is 'All patients' $\supset$ 'Medical' $\supset$ 'Medical Digestive' and 'All patients' $\supset$ 'Family'.

Subsequently, the primary data was transformed into a suitable format for IBM DB2 and the Intelligent Miner. The primary data then employs the $aRule_D(T)$ method and the *TOP* algorithm, as well as the tool of Intelligent Miner and Excel, to compute PBOGs. Finally, PBOGs are transformed into the format data '0' or '1' using *Vector Matrix*. Simultaneously, SBOGs are created using SOM and *Vector Matrix* based on the primary data. PBOGs and SBOGs act on the *Value Rate* to generate draft BOGs.

So far this study has been unable to confirm the rationality and suitability of draft BOGs. Therefore, this study obtained the NHRI raw database, which combined the 'Hospital Basic File' and 'Output-Patient Order File' into two files, which included hospital level, patient age, department, ICD-9-CM code, total amount, and so on. It is called the NHRI_database. Accordingly, this study employs both the AH_database and the NHRI_database to compute the average BOGs. Furthermore, the NHRI_database is employed for reference and to provide comparison with the AH_database.

Finally, this study suggests 'Different Rate' to assess the suitable and rational draft BOGs.

4.2 Result

The AH_database and NHRI_database yielded some interesting information. In the AH_database, physicians prescribe the average order items, including 'All patients', 'Medical', 'Medical Digestive' and 'Family department', which were 6.10, 6.34, 6.59 and 5.93 (Table 4.1). However, the averages of order items can be used to deduce these 'registration' and 'diagnostic' two order items. Subsequently, the percentages of 'All patients', 'Medical', 'Medical Digestive' and 'Family' department patients for whom at least one of the above order items were prescribed were 94.99%, 90.52%, 96.21%, and 90.58%, respectively. However, 1.80%, 1.59%, 1.79% and 0.48% of patients were prescribed 18 order items (Table 1), which appears to offer an advantage.

Table 1 Percentage of Number of Order Items

No. of Order Items	1	2	3	4	5	6	7	8
All Patient	94.99%	87.71%	79.42%	69.86%	58.09%	48.53%	39.52%	32.95%
Medical	90.52%	83.69%	76.10%	66.97%	55.36%	46.03%	37.33%	30.99%
Medical Digestive	96.21%	93.75%	87.10%	77.78%	64.04%	53.37%	43.01%	35.48%
Family	90.58%	83.63%	71.27%	59.75%	49.19%	38.15%	27.50%	20.46%

9	10	11	12	13	14	15	16	17	18
27.63%	22.82%	18.16%	14.59%	11.65%	8.97%	6.57%	4.75%	3.35%	1.80%
25.92%	21.38%	17.02%	13.74%	11.06%	8.51%	6.18%	4.42%	3.03%	1.59%
29.55%	24.27%	19.38%	15.76%	12.78%	9.79%	6.95%	5.06%	3.42%	1.79%
15.22%	11.32%	8.28%	6.09%	4.38%	2.95%	1.71%	1.14%	0.86%	0.48%

Physicians prescribed 739 different order items, including 571.4 *'chronic hepatitis B'* cases. Notably, 555 order items had a frequency of below 0.2%. According to the order item frequency, the 100th order item frequency (or probability) was 0.56%. That is, these physicians prescribe this order item to just 5.6 patients for every 1000 cases of this disease. Therefore, these order items are only employed occasionally.

Table 2 lists the order items [09026CZP], [09025CZP], and so on, as well as their frequency. The top three order items, namely [009026CZP], [09025CZP] and [05209A00], have occurrence frequencies of 62.49%, 61.99% and 45.23%, respectively.

Table 2 The Relative Occurrence of Order Items

Order Item	09026CZP1st	09025CZP2nd	05209A003rd	27049C0P4th	PRO4FE095th	......
Frequency	62.49%	61.99%	45.23%	36.45%	28.63%	

....	09040CZP10th	......	27036A0P 15th	......	14051C0P 20th	....
	17.97%		10.56%		7.63%	

....	09031CZP 25th	....	27034AZP30th	......	ACZ4HD0250th	....	GLI4LB09100th
	6.48%		4.06%		1.76%		0.56%

Liver trouble ranks among the top ten diseases in Taiwan in terms of mortality(NHI, 2000). The number of carriers of Liver viral type B hepatitis in Taiwan is the highest in the world. Figure 8 illustrates the age distribution. From the figure, the 'Medical', 'Medical Digestive' and 'Family' departments display similar age distribution curves, although patients distinguish three different departments. Meanwhile, the figure displays that people can easily suffer liver trouble between the ages of 25 and 70, regardless of whether they are treated in the Medical, Medical Digestive or Family departments.

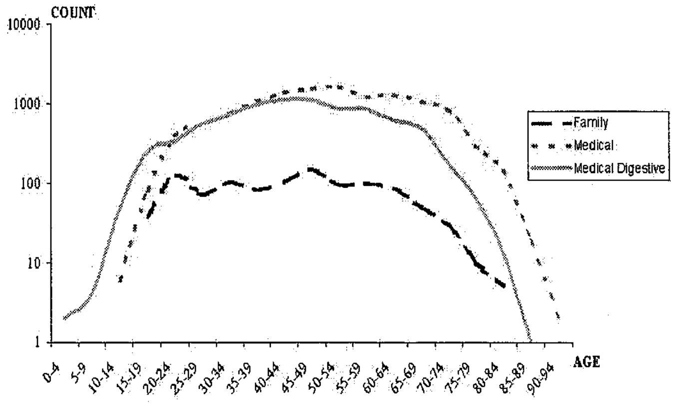

Figure 8 The Distribution of Age

Thirty-three PBOGs were produced using the *aRule$_D$(T)* and *TOP* methods (Table 3). The 'All patients', 'Medical', 'Medical Digestive' and 'Family' departments contained nine, nine,

nine and six PBOGs respectively. Next, this study employs Neural Clustering SOM to group similar order items. Table 4 lists the outcomes, from which only the first ten order items were selected. Because the frequency of the tenth order item is only 17.97% and PBOGs have a maximum of seven order items, this study chose ten order items related to SBOGs. Only the top five groups were selected, because the populations of the other groups are below the initially defined threshold. That is, if the order items of group S1 are similar, then the 15.18% transactions are grouped, and '15.18%' is called population. These five groups of order items are termed SBOG.

Table 3 Prototyping Basic Order Group

All Patients	P1	[09040CZP]+[09027BZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27049C0P]+ [09033CZP]
	P2	[09040CZP]+[09027BZP]+[09029CZP]+[09025CZP]+[09026CZP]+[27049C0P]+ [09033CZP]
	P3	[09040CZP]+[09038CZP]+[09027BZP]+[09025CZP]+[09026CZP]+[27049C0P]+ [09033CZP]
	P4	[09025CZP]+[09026CZP]+[09004CZP]
	P5	[09025CZP]+[09026CZP]+[09001CZP]
	P6	[09040CZP]+[09038CZP]+[09025CZP]+[09026CZP]+[27035A0P]
	P7	[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27035A0P]
	P8	[09025CZP]+[09026CZP]+[27036A0P]+[27049C0P]
	P9	[09038CZP]+[09029CZP]+[09025CZP]+[09026CZP]+[27035A0P]
Medical	P10	[09040CZP]+[09027BZP]+[09030CZP]+[27049C0P]+ [09033CZP]
	P11	[09040CZP]+[09027BZP]+[09029CZP]+[27049C0P]+ [09033CZP]
	P12	[09038CZP]+[09027BZP]+[09029CZP]+[09030CZP]+[27049C0P]+ [09033CZP]
	P13	[09040CZP]+[09038CZP]+[09027BZP]+[27049C0P]+ [09033CZP]
	P14	[09025CZP]+[09026CZP]+[09004CZP]
	P15	[09025CZP]+[09026CZP]+[09001CZP]
	P16	[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27035A0P]
	P17	[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27036A0P]
	P18	[09025CZP]+[09026CZP]+[27036A0P]+[27049C0P]
Medical Digestive	P19	[09038CZP]+[09027BZP]+[09030CZP]+[14051C0P]+ [27033AZP]
	P20	[09027BZP]+[09025CZP]+[27049C0P]+[14051C0P]+ [27033AZP]
	P21	[09027BZP]+[09029CZP]+[27049C0P]+[14051C0P]+ [27033AZP]
	P22	[09025CZP]+[09026CZP]+[09004CZP]
	P23	[09025CZP]+[09026CZP]+[09001CZP]
	P24	[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27035A0P]+[27049C0P]
	P25	[09027BZP]+[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27049C0P]+[14051C0P]
	P26	[09038CZP]+[09027BZP]+[09025CZP]+[09026CZP]+[27049C0P]+[14051C0P]
	P27	[09038CZP]+[09029CZP]+[09030CZP]+[09025CZP]+[09026CZP]+[27035A0P]

Family	P28	[09025CZP]+[09004CZP]+ [09001CZP]
	P29	[09026CZP]+[27049C0P]+ [09025CZP]
	P30	[09026CZP]+[09001CZP]+ [09025CZP]
	P31	[09025CZP]+[09026CZP]+[09030CZP]
	P32	[09025CZP]+[09026CZP]+[09029CZP]
	P33	[09025CZP]+[09026CZP]+[09004CZP]

Table 4 Similar Basic Order Group

Group	Population	Similar Basic Order Group
S1	29.28%	[09026CZP]+[09025CZP]+[05209A00]+[27049A0P]+[PRO4FE09]+[09029CZP] +[09030CZP]+[09038CZP]+[09040CZP]+[09027BZP]
S2	11.55%	[09026CZP]+[09025CZP]+[05209A00]+[27049A0P]+[PRO4FE09]+[09029CZP] +[09030CZP]+[09038CZP]+[19001CGA]+[09040CZP]
S3	9.99%	[09033CZP]+[09027BZP]+[09040CZP]+[09038CZP]+[09030CZP]+[09029CZP] +[19001CGA]+[27036Z0P]+[PRO4FE09]+[27034AZP]
S4	8.86%	[09026CZP]+[09025CZP]+[05209A00]+[27049A0P]+[PRO4FE09]+[09029CZP] +[09030CZP]+[09038CZP]+[09040AZP]+[09033CZP]
S5	8.39%	[09026CZP]+[09025CZP]+[05209A00]+[27049A0P]+[PRO4FE09]+[09029CZP] +[09030CZP]+[09038CZP]+[00157A00]+[ZEF4AM06]

PBOGs were identified as P1, P2, …P33, and as S1, S2, …S5 in the SBOGs. Table 5 lists the Value Rate. The tValuerate is defined as 0.76 in this system of Value Rate Algorithm(Fig. 7). Moreover, the 'n' is 5 in the Value Rate Algorithm because the system contains five SBOGs. Therefore, each value 'k' of PBOG exceeds 3 = int (n/2 + 0.5) = int (5/2 + 0.5), and thus PBOG is termed draft BOG. Therefore, the values 'k' of P3, P12, P24 and P27 are 3, 3, 4 and 4, respectively. Accordingly, P3, P12, P24 and P27 are said to be draft BOGs.

Table 5 Each Prototyping Basic Order Group Value Rate

$Val(\vec{P_i}, \vec{S_j})$	P1	P2	**P3**	P4	P5	P6	P7	P8	P9
S1	**0.813**	**0.813**	**0.813**	0.688	0.688	0.688	0.750	0.719	0.750
S2	0.750	0.750	0.750	0.688	0.688	0.750	0.750	0.719	0.750
S3	0.719	0.750	**0.781**	0.563	0.625	0.625	0.625	0.594	0.625
S4	**0.813**	**0.813**	**0.813**	0.688	0.688	0.750	0.750	0.719	0.750
S5	0.688	0.688	0.688	0.688	0.688	0.688	0.750	0.719	0.750

	P10	P11	P12	P13	P14	P15	P16	P17
S1	0.750	0.750	**0.781**	0.750	0.688	0.688	0.750	0.750
S2	0.688	0.688	0.719	0.688	0.688	0.688	0.750	0.750
S3	**0.813**	**0.813**	**0.844**	**0.813**	0.563	0.563	0.625	0.625
S4	0.750	0.750	**0.781**	0.750	0.688	0.688	0.750	0.750
S5	0.625	0.625	0.719	0.625	0.688	0.688	0.750	0.750

	P18	P19	P20	P21	P22	P23	**P24**	P25
S1	0.719	0.688	0.688	0.688	0.688	0.688	**0.781**	**0.813**
S2	0.719	0.625	0.625	0.625	0.688	0.688	**0.781**	**0.781**
S3	0.594	**0.813**	0.750	**0.813**	0.563	0.563	0.656	0.750
S4	0.719	0.688	0.688	0.688	0.688	0.688	**0.813**	0.750
S5	0.719	0.625	0.625	0.625	0.688	0.688	**0.781**	0.750

	P26	**P27**	P28	P29	P30	P31	P32	P33
S1	**0.781**	**0.781**	0.625	0.750	0.688	0.750	**0.781**	0.719
S2	0.750	**0.781**	0.625	0.750	0.688	0.750	0.750	0.688
S3	0.750	0.656	0.563	0.625	0.563	0.625	0.656	0.563
S4	0.719	**0.781**	0.625	0.750	0.688	0.750	0.750	0.688
S5	0.719	**0.781**	0.625	0.750	0.688	0.750	0.750	0.688

In these draft BOGs, BOG P3, BOG P12 and BOG P24, P27 belong to the 'All Patients', 'Medical' and 'Medical Digestive' departments, respectively. To confirm the suitability and rationality of these draft BOGs, this study compared the 'Different Rate' of every draft BOGs between the AH_database and the NHRI_database. BOG comprises some order items. Meanwhile, some order items which cannot be performed at the Clinic can be provided at the

District Hospital, Regional Hospital and Academic Hospital. Therefore, some draft BOGs can be done by the District Hospital, Regional Hospital and Academic Hospital, but these draft BOGs cannot be done in a Clinic. Draft BOG P24 can be done in the Regional Hospital and Academic Hospital, but cannot be done in a District Hospital or Clinic.

Table 6 represents the order items that are provided the status of draft BOGs. If a draft BOG can be done at some level hospital then mark '○' , otherwise mark '※'. This study computes the Experiment Contrast Payment of every draft BOG. Table 7 displays this result. It is marked '※'to indicate that it has not yet been carried out. Similarly, Table 8 lists the Health Insurance Reporting Payment of draft BOGs in each hospital level. Finally, this study computes 'Different Rate' within the common items of two tables. Table 9 lists the calculation results of the 'Different Rate'.

To summarize, these P3, P12, P24 and P27 draft the Difference Rates of BOGs, namely 5.98%, 12.76%, -2.41% and 44.89%, which appear to fit the needs of this study, with the exception of P24. As for whether or not it is a suitable and rational approach, this applies a statistical hypotheses test. This study uses SPSS for Windows 8.0(SPSS, 1997) to execute the t-test of the statistical hypothesis.

The Mean, standard deviation and standard error mean are 15.3050, 20.6763 and 10.3381. Moreover, the confidence interval is between −17.5956 and 48.2056. Meanwhile, the p-value 0.235 exceeds 0.05. We accepts the null hypothesis (Different Rate $\geqq$ 0.0%) in the t-test of the statistical hypothesis of $\alpha = 0.05$.

Table 6 The provider of BOGs

Level \ BOGs	P3	P12	P24	P27
Academic Hospital	○	○	○	○
Regional Hospital	○	○	○	○
District Hospital	○	○	※	※
Clinic	※	※	※	※

Table 7 Experiment Contrast Payment of Every BOGs in Each Level of Hospital

Department		All Patients	Medical	Medical Digestive	
BOGs		P3	P12	P24	P27
Experiment Contrast Payment	Academic Hospital	903	993	1103	743
	Regional Hospital	903	993	1103	743
	District Hospital	902	992	※	※
	Clinic	※	※	※	※

Table 8 Health Insurance Reporting Payment in Each Level of Hospital

Department / Level	All Patients	Medical	Medical Digestive	Family
Academic Hospital	※	1391.3	1110.2	1187.6
Regional Hospital	957.0	1063.5	1042.8	880.2
District Hospital	※	903.3	839.1	1238.5
Clinic	634.2	940.0	1770.3	637.7

Table 9 Different Rate

BOGs / Level	All Patients	Medical	Medical Digestive		Average
	P3	P12	P24	P27	
Academic Hospital	※	40.11%	0.65%	49.42%	30.06%
Regional Hospital	5.98%	7.10%	-5.46%	40.35%	16.00%
District Hospital	※	-8.94%	※	※	-8.94%
Clinic	※	※	※	※	※
Average	5.98%	12.76%	-2.41%	44.89%	
Average of Department	5.98%	5.18%		44.89%	

Table 10 Result of t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Different Rate	4	15.3050	20.6763	10.3381

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Different Rate	1.480	3	.235	15.3050	-17.5956	48.2056

5. Discussion

Generally, numerous combinations of order items exist, which comprise two to ten order items. Prescribing such a large number of order items is counterproductive for physicians and patients. Meanwhile, a Fixed Amount Payment System involving order items will not accord with the costs/benefits.

aRule$_D$(T) and *TOP* computing are used to adjust the support, confidence and number of order items. Thirty-three PBOGs were obtained through calculating *aRule$_D$(T)* and *TOP*. Subsequently, Neural Clustering SOM computing was applied to obtain five SBOGs.

In computing, the PBOGs of each department are done by SBOGs. Although some PBOGs of one department may be subsets of other PBOGs of other departments, PBOG ownership is unimportant. One reason for this phenomenon is that the department to which the PBOGs belongs is random, and another reason is the character of the physicians prescribing an order for the code of 571.4 patients. That is, physicians have the same ideas regarding cure and diagnosis. From this case, PBOG P28 is a subset of some PBOGs, for example P1, P2, P3…P25. Meanwhile, some PBOGs are the same between departments, for example P 4 and P32.

Four draft BOGs were obtained by calculating the Value Rate Algorithm of the system, namely P3, P12, P24 and P27. The 'Different Rates' for these four draft BOGs were 5.98%, 12.76%, -2.41% and 44.89%, respectively. The result was accepted based on statistical hypotheses testing. Therefore, these four draft BOGs are termed BOGs.

To summarize, the followings sub-conclusions can be obtained.

1. BOG P3 and BOG P27 are the best BOGs.
2. BOG P12 works best in Academic Hospitals and Regional Hospitals, but care is required in District Hospitals.
3. Care is required in employing BOG P24 and the problem of costs/benefits must be carefully considered.

Meanwhile, these BOGs are highly elastic, and can be adjusted depending on patient disease status. For example, when patients require half yearly '*SONAGRAPHY FOR UPPER ABDOMEN*' for tracing and observation purposes, the BOGs can be added to this order item. Moreover, its purpose is fulfilled. Meanwhile, the subset of BOG can be considered as the order when patient illness status is stable.

Finally, this study decodes these BOGs to order name as follows：

P3：*[TOTAL PROTEIN]+[ALBUMIN]+[ALKP]+[GOT]+[GPT]+[ALPHA-FETOPROTEIN]+ [LDH]*

P12：*[ALBUMIN]+[ALKP]+ [TOTAL BILIRUBIN]+ [DIRECT BILIRUBIN]+ [ALPHA-FETOPROTEIN] +[LGH]*

P24：*[TOTAL BILIRUBIN]+[DIRECT BILIRUBIN]+[GOT]+[GPT]+[HbeAg]+ [ALPHA-FETOPROTEIN]*

P27：*[ALBUMIN]+[TOTAL BILIRUBIN]+[DIRECT BILIRUBIN]+[GOT]+[GPT]+[HbeAg]*

6. Conclusion

Maintaining a financial balance given limited medical payments is essential for health insurance payment units and hospitals(McIlvance, 1983)(Hurzeler & Leary, 1983). The Fixed Amount Payment System can not only avoid unnecessary medication examinations, but also reduces financial expenses within the global budget system. In this study, these Basic Order Groups, which were created using the model of the ART_N.V. scheme, assist the Payment Unit in producing the BOGs for every disease.

Disease order items can be determined by BOG using the Fixed Amount Payment System. This study of the ART_N.V. scheme employs the case of ICD-9-CM 571.4, which does not belong to the Fixed Amount Payment System, to provide its BOGs to reduce the current impact. Moreover, this study obtains BOGs to compute the cost/benefit of the payment system. Employing the BOG in the Fixed Amount of Payment System thus can achieve a suitable payment standard for reference.

To summarize, more benefits are as follows:

1. Patients receive high quality medical care.
2. Hospital materials, medicine and examination costs are reduced, and system management is improved.
3. This Basic Order Group formulates the basic required order items for every disease, adjusts and pays every disease, and establishes a suitable standard system in the Fixed Amount Payment and AP-DRG systems.
4. Hospitals establish the Basic Order Group system to control medical cost and model medical care quality.
5. The Basic Order Group is very feasible, because it can be adjusted based on patient characteristics and illness severity.
6. Basic Order Group can adjust how hospitals face the current and future challenges of the Global Budget System and Fixed Amount Payment System.

The Basic Order Group not only efficiently provides valid disease order items, but also reduces costs and effectively manages the Fixed Amount Payment System and the AP-DRG of the Global Budget System.

Reference

1. Agrawal, Rajesh and Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th VLDB Conference,* Santiago, Chile. 1994.
2. Birkhoff, G., "lattice Theory." Providence, R.I.: Am Math. Soc., Colloquium Publication, 1967, vol. 25.
3. Chiou, Peterson, "Image Compression by Self-Organized Kohonen Map", *DaYeh University Electrical Engineering*, Master Theses (no-published), 1998.
4. Chaudhuri, Surajit and Umeshwar Dayal, "An Overview of Data Warehousing and OLAP Technology," ACM SIGMOD Record, 1997, 26:65-74.
5. Chiang, Jaw-Ching and Shi-Jen Lin, "A Study of Data Mining Medical Informatics", *The Journal of China Association for Medical Informatics*, Vol 9, 1999, pp. 71-82.
6. Chiang, Jaw-Ching and Shi-Jen Lin, "The Model of Payment System under Case Payment – APORES Model", *The Journal of Taiwan Association for Medical Informatics*, Vol 14, February 2002, pp. 1-16
7. DOH, Executive Yuan, "Taiwan Public Health Report 2003", *DOH, Executive Yuan*, Taipei, Taiwan, 2003.
8. Han, Jiawei, Yandong Cai and Nick Cercone, "Data-Driven Discovery of Quantitative Rules in Relational Database," *IEEE Transaction on Knowledge and Data Engineering*, Vol. 5, Feb. 1993, pp29-40.
9. Hurzeler, R. Johnson and Robert J. Leary, "Diagnosis Related Group as Casemix/Management Tool for Hospice Patient", *IEEE,* 1983.
10. IBM, "Using the Association Visualizer", Version 6 Release 1, 1999.
11. IBM, "Using Intelligent Miner for Data", Version 6 Release 1, 1999.
12. IBM, DB2 Intelligent miner for Data, V6.1, 1996.
13. IBM, DB2 UDB Enterprise for Windows, V7, 1995.
14. Kohonen, T, "The Self-Organizing Map," Proceedings of the IEEE, vol. 78, no. 9, pp. 1464-1480, Sep. 1991.
15. Lee, C.-C. and J. P. de Gyvez, "Color Image Processing in a Cellular Neural-Network Environment," *IEEE Trans. on Neural Networks*, vol. 7, no. 5, pp. 1086-1098, Sep. 1996.
16. McIlvance, Mary Elizabeth, "How a Data Base Management System Integrates Patient casr and Financial Data to Manage Diagnosis Related Group", 1983, *IEEE*
17. Microsoft, Office Premium, Version 2000.
18. Ng, Raymond and Jiawei Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994, pp144-155.

19.NHI, 2000, http://www.nhi.gov.tw

20.Rothermel, Kurt, "Extended Concepts for Association Rule Discovery", 1997.

21.SPSS, SPSS for Windows, Release 8.0.0, 1997.