An empirical study on financial forecasting of public health care organization: based on the big data analysis method.

Pei-Yu Hsieh, **Yun-Tien Ma** and **Shiaw–Wen Tien**
Chung Hua University, Hsinchu City, Taiwan

Yuan-Huei Huang
National Taiwan University Hospital, Hsin-Chu Branch, Taiwan.

Keywords
Medical Institutions, Financial Management, Big Data, Taiwan's Diagnostic Related Groups (Tw-DRGs), forecasting.

Abstract

Financial planning and control decisions based primarily on the results of a big data analysis, as managing reference forecast data can provide the most appropriate planning, and decisions.
In this study, use of the past 10 years data within the National Health Insurance database implants decision tree analysis in Taiwan. Taiwan's diagnosis-related group (TW-DRGs) surgery expenses affect the medical profession to analysis and simulation. Grey prediction theoretical simulation analysis and forecasting for future changes in TW-DRGs medical surgical expenses. The results showed that over the past research data in the past few years, most medical project spending trends. This study may provide in the future will produce what kind of cost impact results: 1. Overview the past 10 years, the growth of big data. 2. Existing financial forecast data and verification. 3. Simulation analysis and forecasting of future financial data.

1. Introduction

Taiwan began to implement national health insurance system in 1995. As the financial expenditure imbalance has occurred many times. How to control health care costs is an important topic. Many studies have shown that national health care costs often exceed the growth rate of the country's economic growth, accounting for 20% of GDP. In order to maintain the financial balance of payment systems and sustainable management in four years, the National Health Insurance (NHI) programs in Taiwan Diagnosis Related Groups (TW-DRGs) the integration of health insurance, implanted in 2010 and 2014. Rising health care costs have become a common problem in some developed countries.

The diagnosis-related groups (DRGs) prospective payment system based on the use (PPS) on the fixed costs of the issue option and pay for the hospital introduced in the United States Medicare beneficiaries to provide services. Based DRG-PPS aims to establish financial incentives for hospitals to effectively control the use of resources, does not affect the quality of patient care. Taiwan government to implement the design at Yale University Professor in 1980 diagnosis-related groups (DRGs) system.

However, health insurance information content has reached large areas of data definition. Therefore, this study by computer data statistics, comparison, parsing big data in order to come to objective results, In results show that research may offer what will produce results in the future cost implications:

1. Hospital's operating costs, for reference.
2. The medical insurance budget planning.
3. TW-DRGs policy revision surgery hospital costs.

2. Method

a) Big Data Analysis

Big data set composed by a giant data, the size of these data sets often beyond human collected at an acceptable time, the approach of taking the use, management and processing capabilities. Frequent changes in the size of big data, as of 2012, the size of a single data set to tens of trillion bytes (PB), ranging from a few terabytes (TB). In a 2001 study and related speech, META Group (now Gartner high) Analyst Laney (2011) indicates that the data growth challenges and opportunities in three directions: volume (Volume, data size), speed (Velocity speed data input and output) and multivariate (Variety, diversity), collectively referred to as "3V" or "3Vs." Gartner and now most of the big companies in the information industry, have continued to use 3V to describe big data.

2012 (Gartner) was modified the definition of big data: "Big data is a large, high speed, and / or changing information assets, it needs new approach to facilitate greater decision-making ability, insight and most . good treatment "Further, there is defined a mechanism outside the first four 3V V: authenticity (Veracity) fourth characteristics.

Application examples include a large big data science, RFID, sensing device networks, astronomy, atmospheric science, genomics, biology, large social data analysis [, Internet file handling, making Internet search engines index, Communications record details, military reconnaissance, social networks, commute time forecasting, medical records, photos, images and video archive, a large-scale e-commerce, this study analyzed medical records based.

b) Data Mining

Data mining is to explore ways to resolve big data. Linoff and Berry (1999) defined data mining classification and use of the machine self-learning, association rules, sequence analysis, cluster analysis and other statistical methods, seeking to hide from a lot of different databases, unknown, but very useful information. Therefore, the data mining techniques to extract accurate, previously unknown from big databases, it is important that information and use that information to make important decisions.

Therefore, the data mining techniques to extract accurate, previously unknown from big databases, it is important that information and use that information to make important decisions. The common customer portfolio management decision tree analysis. An important feature is the establishment of a branch tree structure classification known examples.

This study used of expert decision tree pruning. The final output is classified information. This study chose the gray prediction method as a predictive tool for research and analysis of surgery, providing a continuation of the direction of development of medical students and hospitals.

3. Formula and Equation

This study is based on data collected Sources NHI study released on CD, we focus on DD and HOSB documents 2001-2011 of two documents, we calculated the 155 research and related internal medicine, surgery, gynecology, obstetrics, orthopedics, urology, ENT and ophthalmology specialists.

Data timetable and project TW-DRG released NHI also be used to simulate and analyze. Integration and process the raw data used to build the tree split in a variety of combinations of related.

Huang, Chang and Hsieh (2013) Take the gray prediction method to build a big data predictive model to analyze the hospital organization's production planning and control

program, won the expenditure projections inference. Decision model application Lin and Yang (2003) selected gray correlation analysis, choice of housing mortgage loans.

Predictive gray system theory, Deng (1989) constructed a gray prediction model of gray system theory. Their results showed that the gray system theory in this case is appropriate. (1,1) of the sequence we created a first-order linear movement of GM. First-order differential equation GM (1,1) model is the grey prediction simulation and adjustment formula based on Huang, Change and Hsieh (2013).

We created a sequence of first-order linear movement GM (1,1). The first-order differential equation for the GM (1,1) model is

$$dx^{(1)} / dt + aX^{(1)} = b$$

Where, t is the independent variable in the system, represents the developed coefficient, b is a gray controlled variables, a and b need to determine the parameters of the model. The variables, including P^((1)),P^((2)),and…P^((n)), are used to construct the Grey forecasting model and accurately predict P(n+1),P(n+2),…and P(n+k). Assumed here that primitive sequences are

$X^{(0)} = \left[P^{(0)}(1), P^{(0)}(2), P^{(0)}(n)\right]$

When building the model, must be applied Grey system original series one-order accumulated generating operation (AGO) to provide architectural model middle messages and weakening trends. X (1) is defined as X (o)' one-order AGO sequence. That is

$$X^{(2)} = \left(P^{(1)}(1), P^{(2)}(2), \ldots, P^{(2)}(n)\right) = \left(\sum_{i=2}^{2} P^{(0)}(t), \sum_{i=2}^{2} P^{(0)}(t), \ldots, \right.$$

From equation. (1) And (3) and the ordinary least-square method, the coefficient a $\hat{a}$ becomes

$a = [a/b] = (B^T B)^{-1} B^T Y_N$.

The accumulated matrix B is

$$B = \begin{bmatrix} -\frac{1}{2}\left[P^{(1)}(1) + P^{(2)}(2)\right] & 1 \\ -\frac{1}{2}\left[P^{(2)}(2) + P^{(2)}(3)\right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[P^{(n)}(n-1) + P^{(2)}(n)\right] & 1 \end{bmatrix}$$

Meanwhile, the constant vector YN is$Y_N = \left[P^{(0)}(2), P^{(0)}(3), \ldots, P^{(0)}(n)\right]^T$.. Substituting $\hat{a}$ to get the differential equation, solving equations, obtain the following approximate relationship.

$$\hat{P}^{(2)}(t+1) = \left(P^{(0)}(1) - \frac{b}{a}\right) e^{-at} + \frac{b}{a}$$

When, $\hat{P}^{(2)}(1) = \hat{P}^{(0)}(1)$when "X", the sequence obtained one-order inverse-accumulated generating operation (IAGO) obtained, the sequence must be reduced as Eq. (6) can be obtained.

$$\hat{P}^{(2)}(t \cdot$$

Given t=1, 2 . . . n, the reduction sequence is obtained as follows (Eq. (7)):

$\hat{X}^{(0)} = \left(\hat{P}^{(0)}(1), \hat{P}^{(0)}(2), \ldots, \hat{P}^{(0)}(n+1)\right)$

Where $\hat{P}^{(0)}(n+1)$ is a basic predictive value Grey P (n +1).. After the model generation and development, further testing is necessary to understand the predicted value and the actual value

of the error. To demonstrate the efficiency of the proposed prediction model, we use the residual error testing methods to compare actual and predicted values. Equations (8) and (9) are used to compute the Grey forecasting residual and average residual errors.

$$Error = \left|\frac{P(t) - P(t)}{P(t)}\right| \qquad j = 1,2, \ldots, n$$

$$Average\ error = \frac{1}{n}\sum_{i=2}^{n}\left|\frac{P(t) - P(t)}{P(t)}\right| \qquad j = 1,2, \ldots, n$$

4. Result

The big data analysis shows, the cost of surgery over the 10 years by the Health Insurance Bureau of Taiwan's 10 spending big data image conversion as operating costs of hospital surgical procedures, refer to budget planning shown in Figure 1

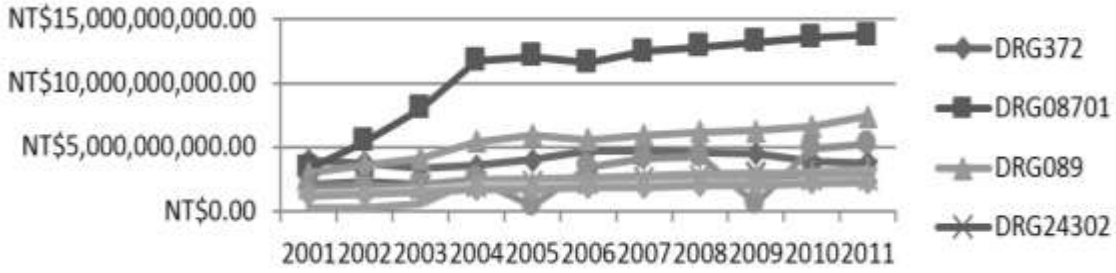


Figure 1：Taiwan's 10 spending big data image conversion as operating costs.
(From this Study)

Huang, Chang and Hsieh (2013) gray prediction model and the data are presented in Table 1. Grey prediction model is used to predict surgical medical costs, in 2012, 2013 and 2014. 2012, 2013 and 2014 used to predict health care costs 4:00 forecasts.

Unit: NTD.

DRG	2010 health care cost	2011 health care cost	2011 health care cost forcast	2012 health care cost forcast	2013 health care cost forcast	2014 health care cost forcast
DRG372	3896256136	3744638422	3969710866	3689167257	3173910621	3089088296
DRG08701	13536068039	13556670550	13908753844	14277993442	14665216937	15056524257
DRG089	6626552284	7396331395	7377842110	7564638009	7738485593	7914823676
DRG24302	2754037418	3096214015	3019822927	3137932392	3273102477	3414053616
DRG245	2581326594	2451900614	2602385518	2771700977	2857825177	2982564962
DRG41601	4867218387	5227552863	4165320277	4594521743	5067796611	5589634217
DRG12101	3045162290	3290369412	3258347969	3507898926	3758182598	4026138237
DRG132	2867738149	2849731361	3069780957	3045296217	3120395482	3197340292
DRG01401	2268590472	2275692156	2298252207	2355432849	2423808129	2494161054
DRG320	2179055029	2183145828	2260146827	2319390377	2414920672	2573907261

Table 1: In the 2006 value of 97% prediction accuracy; 2007:92%; 2008:97%; 2009:99%; 2010：95%; 2011:93%. Gray forecasting models accurately predict the value of an average of 95%. For this study, 2001-2011 surgery information, projections 2012, 2013, 2014, surgical inpatient medical expenses..(Huang, Chang and Hsieh, 2013)

Patients with a positive predictive 2010-2011 budget planning in health insurance are shown in Figure 2 and 3.

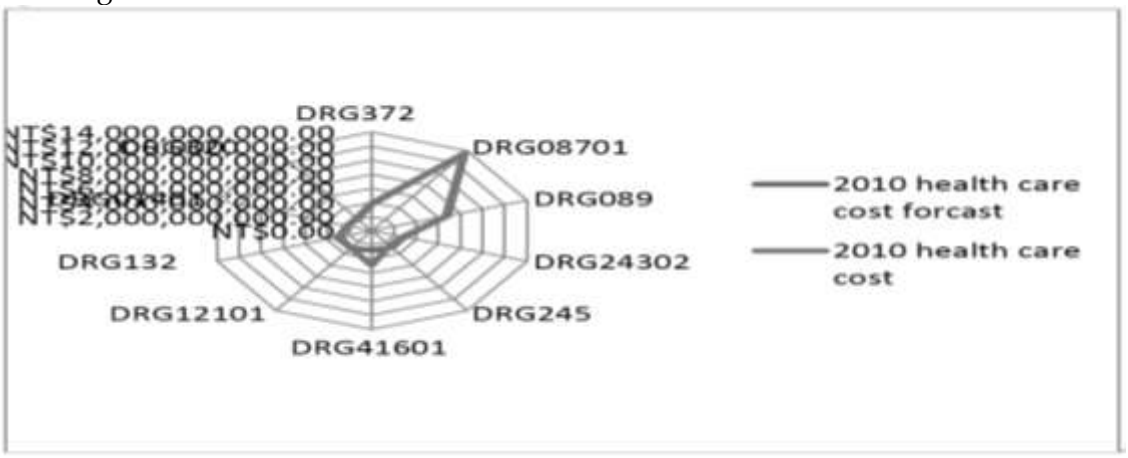


Figure 2: Patients with a positive predictive 2010. (From this Study)

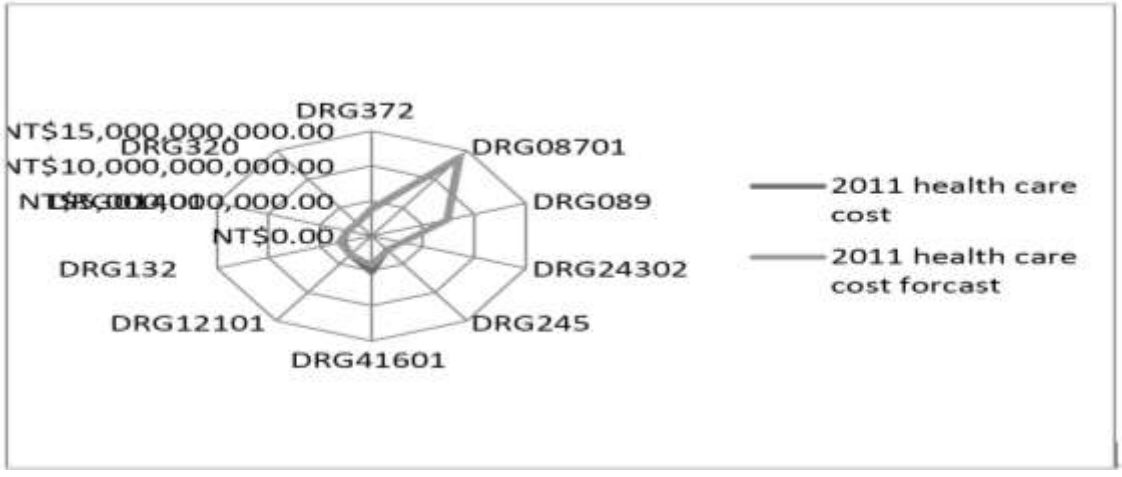
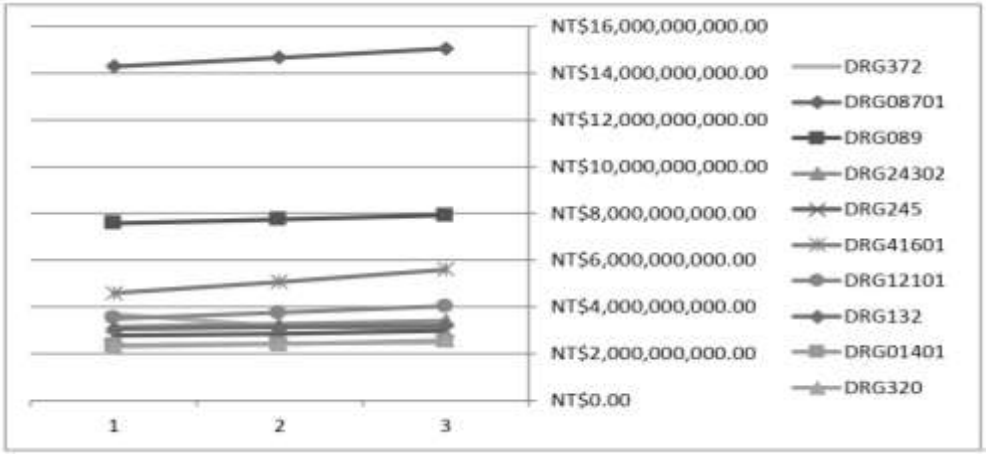
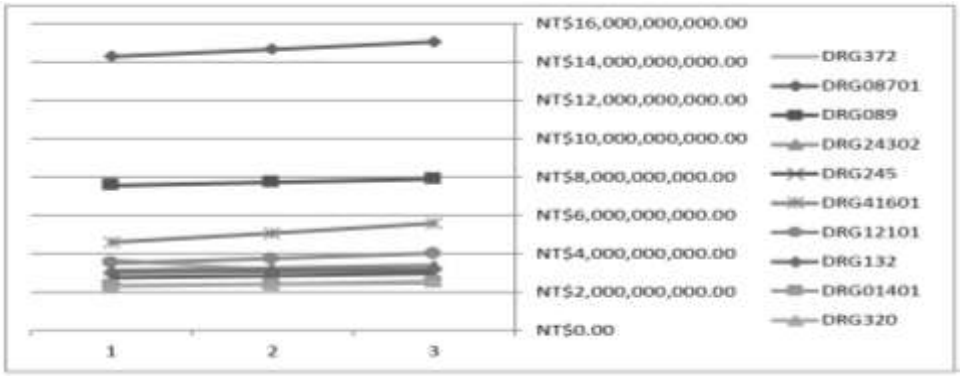


Figure 2: Patients with a positive predictive 2011. (From this Study)

The last three years, 10 major operations expenditure projections, presented below in Figures 3.





5. Conclusion

a) Discussion

The results of this study should indeed should permit Huang, Chang and Hsieh (2013) National Health Insurance in Taiwan surgical medical spending is the growing; however, this study through big data updates from the National Health Insurance database, and use images way to display more detailed information, as a hypothetical question to answer: way to display more detailed information, as a hypothetical question to answer:

1. Hospital's operating costs, for reference. In this study, the records of the past 10 years found that because the National Health Insurance gradually lowering the annual growth in the medical surgical expenses , reflecting the need to focus on investment in medical surgical unit cost increases , medicines , supplies , equipment and human resources ... and so on. In this study, the records of the past 10 years found that because the National Health Insurance gradually lowering the annual growth in the medical surgical expenses , reflecting the need to focus on investment in medical surgical unit cost increases , medicines , supplies , equipment and human resources etc..

For example : DRG08701 (pulmonary edema and respiratory failure) the National Health Insurance surgical expenses of the project, after the 2003-2004 growing twice annually sustained increase of 7% -10 % , which means that rising health care costs , hospital investment in this operating budget of the project also needs to be adjusted annually to meet the changing needs of the medical and disease . Operating budget of the project also needs to be adjusted annually to meet the changing needs of the medical and disease.

2. Medicare budget planning. In this study, forecasting and empirical research findings, taken Huang, Chang and Hsieh (2013) gray prediction equation is accurate simulations, the average error in 2011 is less than 3.5%, in 2012 0.8% of error is smaller , of course, as amended in 2011 requires empirical data, the research community on behalf of the results provided in the Taiwan National Health Insurance plan can provide effective budget forecasting and simulation. For example : DRG41601 (septicemia)

3. TW-DRGs policy revision surgery costs. The project of this study confirmed the predicted future data, provides almost all of the rising budget speculated that surgery can refer to adjust the distribution of the budget forecasts expenditure on renovation policy, more importantly, how to make surgical medical expenses no longer continue to rise because that is the National Health Insurance, the economic burden of patients and the government.

For example: In addition DRG372 (vaginal delivery), surgical expense of other projects which are increasing every year, according to the 2001-2010 study combined data 2010-2014, he was the main reason for the population decline in fertility problems.

b) Acknowledgments and Legal Responsibility

This study deeply grateful to National Taiwan University Hospital, Hsinchu Branch Research Unit research funding provided by (NTUHHCB research projects, Document Number: HCH102-05).

The study is based on data from the National Health Insurance Research database data portion. Interpretations and conclusions contained herein do not represent the Department of National Health Insurance Bureau, Health or the National Institutes of Health.

References

Beyer, Mark., 2011. Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data. Velocity and Variety, Gartner.

C.T. Lin, S.Y. Yang., 1999, Selection of home mortgage loans using Grey relational analysis. *J. Grey Syst*, 11 (4), 359-368.

Chang Chin-wen., 1984, The United States the reasons for using the DRG pre-price system? *Hospitals*, 17 (6), 24-27.

Chin-Tsai Lin, Shih-Yu Yang., 2003, Forecast of the output value of ai an's opto-electronics industry using the Grey forecasting model. *Technological Forecasting & Social Change*, 70, 177-186.

Department of Health, Executive Yuan., 2010, 2008 National Health Expenditure Statistical Report, 74-46,43-48.

Department of Health, Executive Yuan., 2004, Overview of the National Health Insurance Reform. 2004b,46-47,12-16.

Department of Health, Executive Yuan., 2004, Towards the powers and responsibilities consistent with national health insurance system. 2004a, 118-146, 14-43.

Douglas, Laney., 2001, 3D Data Management: Controlling Data Volume. *Velocity and Variety*, Gartner.

Douglas, Laney., 2012, The Importance of 'Big Data': A Definition. *Velocity and Variety*, Gartner.

Erik Cambria, Dheeraj Rajagopal., 2013, Daniel Olsher, and Dipankar Das. *Big social data analysis//Big Data Computing*, Taylor & Francis.

Halper Fern., 2013, IBM What Is Big Data. *Bringing Big Data to the Enterprise*, Available at website: http://www-01.ibm.com/software/data/bigdata/. [Accessed April 2014]

J.L. Deng.,1982, Control problems of Grey system. *Syst. Lett*, 5,288–294.

J.L. Deng., 1989, Introduction Grey system theory. *J. Grey Syst*, 1,1–24.

Lu JR, Hsieh. C. R., Editor., 2000, Economic incentives of the payment system. *Health Economics*, Pro-Ed Publishing Company, 297-323.

OECD health data 2010. Available at website: http://www.oecd.org/health/healthdata, [Accessed April 2014]

Plested, William G., 2002, DRG system. *Journal of the American College of Surgeons*, 194(1), 101-102.

Shiaw-Wen Tien, Pei-Yu Hsieh, Yuan-Huei Huang, Che-Wei Chang., 2013, Construction predictive models to analyze Tw-DRGs surgical implants affect hospital revenues. *International Conference on Medical Innovation and Computing Service*.

Yuan Huei Huang, Che Wei Chang, Pei Yu Hsieh.,2013, Construction a Big Data Predictive Model to Analysis the Production Planning and Control Plan of Hospital Organization, Applied Mechanics and Materials, 437(3),977-980.