Study on the influencing factors and prediction of the medical cost of chronic renal failure in China based on a decision tree algorithm

pingping dai (276160176@qq.com)

Central South University Third Xiangya Hospital https://orcid.org/0000-0003-1024-7114

Aijing Luo

Central South University Xiangya School of Medicine

Weifu Chang

Key laboratory of medical information research,Third Xiangya Hospital, Central South University

Zirui Xin

Key laboratory of medical information research, Third Xiangya Hospital,Central South University

Haiwei Cheng

Department sociology, Central South University

Wei Ouyang

Key laboratory of medical information,Third Xiangya Hospital, Central South University

---

Research article

Keywords: decision tree, CHAID algorithm, CART algorithm, chronic renal failure, hospitalization costs, influencing factors, prediction

Posted Date: January 27th, 2020

DOI: https://doi.org/10.21203/rs.2.21909/v1

Study on the influencing factors and prediction of the medical cost of chronic renal failure in China based on a decision tree algorithm

Aijing Luo[12], Pingping Dai[1], Weifu Chang[13], Zirui Xin[13], Haiwei Cheng[4],Yangwei Ou[13]

ABSTRACT

Purpose: The objective of this study is to explore the factors that affect the hospitalization costs for patients with chronic renal failure (CRF) and to extract relevant characteristics for modeling, so as to make predictions about hospitalization costs for CRF patients.

Methods: This study collected the data on the first page of the 2014 medical records of three first-class tertiary hospitals in Beijing. Using IBM SPSS Modeler software, we used the chi-squared automatic interaction detector (CHAID) and classification and regression tree (CART) algorithms to construct a prediction model of the hospitalization costs for CRF patients and conducted a comparative analysis. The data of the 1819 cases in this study included the index variables on the first page of the medical records, which covered social economics, clinical characteristics, and medical consumption. The input variables included medical payment method, sex, age, marital status, length of hospital stay, main diagnosis, number of other diagnoses, major surgery, and number of surgeries. The target variable was the total medical expense.

Results: Our results showed that medical payment method, sex, age, marital status, length of hospital stay, main diagnosis, number of other diagnoses, major surgery, and number of surgeries all had an effect on hospitalization cost. There was no significant difference in the prediction models of the total hospitalization cost constructed using the CHAID and CART algorithms.

Conclusions: Major surgery and length of hospital stay were important predictor variables for modeling with both CHAID and CART algorithms. The length of hospital stay should be included in the grouping variables when disease-related grouping and prediction of medical costs are done for CRF patients.

Keywords: decision tree; CHAID algorithm; CART algorithm; chronic renal failure; hospitalization costs; influencing factors; prediction.

Background

Chronic renal failure (CRF) refers to chronic progressive renal parenchymal damage of various causes, resulting in significant atrophy of the kidney and its inability to maintain basic function. It is a clinical syndrome characterized by the main clinical manifestations of retention of metabolites and water, electrolyte and acid-base disorders, and the involvement of other organ systems. CRF has become a major public health problem worldwide. CRF can occur in all age groups, and there is much diversity in the affected population [1]. With the improvement of people's living standards, the incidence of CRF is increasing year by year, and CRF has become one of the major chronic diseases affecting the health of Chinese people. In 2010, a systematic report of kidney data of the United States showed that medical insurance expenditures in the United States were $29 billion in 2009, accounting for approximately 6% of the annual medical budget. A study by the Korean Society of Nephrology also showed that the incidence of end-stage renal disease in Korea is 70% of that in the United States, and the number of patients is increasing year by year, with the cost of treatment also increasing accordingly. Thus, the increase in the number of CRF patients has significantly added to the pressure on the medical budget [2-6].

The treatment of CRF includes conservative medical treatment and surgical treatment, such as continuous peritoneal dialysis, hemodialysis, and kidney transplantation. The cost of treatment varies greatly between treatments. The cost of hospitalization for CRF patients with different comorbidities and concomitant diseases is also different [7-10]. The study by Khan S of Tufts University School of Medicine in Boston found that secondary hyperparathyroidism was associated with high costs for CRF patients with cardiovascular complications [11]. The findings by Zhao Y et al. suggested that medical insurance payment coverage would affect the medical cost of hospitalization. [12] Different medical insurance reimbursement payment systems can affect the choice of treatments for end-stage renal disease, thereby affecting the allocation of relevant resources and ultimately the national medical budget [13]. Thus, many factors affect the hospitalization costs of CRF patients, and the patient grouping and medical insurance payment standards are more complex, necessitating further investigation [14].

Data mining is a process of extracting implicit and potentially useful information and patterns for decision-making from vast amounts of noisy and messy data by various methods. According to the discipline to which the algorithm belongs, data mining algorithms can be divided into machine learning algorithms and statistical learning algorithms. Data mining algorithms have been extensively applied, and machine learning algorithms and traditional statistical learning algorithms each have their own advantages. For example, the use of electronic medical records has generated a large number of clinical medical datasets that do not meet the assumptions of parametric methods in traditional statistics and have diverse types of data, but these clinical datasets can be mined and analyzed using machine learning algorithms.

In this study, inpatients with CRF in three tertiary first-class hospitals in Beijing were selected as the study subjects. IBM SPSS Modeler software was used to analyze the influencing factors of the hospitalization cost of CRF patients and conduct a grouping study to analyze the reference value of medical costs in each group, so as to provide a reference basis that relevant medical decision-making departments could use to control the unreasonable growth of medical expenses and medical insurance payments.

Methods

Data collection

The data were derived from the first page of medical records in three first-class tertiary hospitals, and the inpatients whose main diagnosis code (ICD-10) for the discharge diagnosis was N18 (CRF) in 2014 were selected to generate a dataset of 1819 patients.

Guided by related reports [15-19], we included three index variables found on the first page of the medical records: social economics, clinical characteristics, and medical consumption. The input variables included medical payment method, gender, age, marital status, the length of hospital stay, main diagnosis, the number of other diagnoses, major surgery and operation and the number of surgeries. The target variable was the total medical cost (Table 1).

Table 1 Variables in the research

Variables	Variable assignment or variable description	Type of variable
Socioeconomic variables		
Medical payment method	1 = Medical insurance; 2 = poverty relief; 3 = full public expense;	Nominal Variables

	4 = full self-pay; 5 = Other.	
Sex	1 = male; 2 = female.	Marker variable
Age	Age values of each patient.	Continuous variables
Marital status	1 = unmarried; 2 = married.	Nominal Variables
Variables of Clinical Characteristics		
Main diagnosis	1 = CRF; 2 = the uremia stage of CRF; 3 = the azotemia stage of CRF; 4 = chronic kidney disease stage 1; 5 = chronic kidney disease stage 2; 6 = chronic kidney disease stage 3.	Nominal Variables
Number of other diagnoses	Number of other diagnoses for each patient during hospitalization.	Continuous variables
Major surgery	0 = None; 1 = Kidney transplantation; 2 = Hemodialysis; 3 = Electrocardiogram (ECG); 4 = Medical preparation for renal dialysis; 5 = Arteriovenous fistula repair for renal dialysis; 6; Peritoneal dialysis; 7 = Venipuncture; 8 = Ultrasound-guided puncture renal biopsy; 9 = computed tomography (CT) examination; 10 = Ultrasonography; 11 = Nuclear magnetic resonance examination; 12 = Oxygen inhalation; 13 = Other surgical procedures; 14 = 24-hour blood pressure monitoring.	Nominal Variables
Number of surgeries	Number of surgeries and operations performed for each patient during hospitalization.	Continuous variables
Medical consumption variables		
Length of hospital stay	Length of hospital stay during hospitalization for each patient.	Continuous variables
Total medical expenses	The total medical cost for each patient during hospitalization.	Continuous variables

Analytic tool selection

In this study, IBM SPSS Modeler software was used to perform pre-preparation and preprocessing of the collected data. The chi-squared automatic interaction detector (CHAID) algorithm and the classification and regression tree (CART) algorithm were used for modeling and analysis.

The CHAID algorithm was proposed by Kass et al. in 1980. It is a self-stratification method based on the target variable. It takes the dependent variable as the root node, categorizes each independent variable, and calculates the categorical chi-squared value. If the categorizations of several variables are all significant, the significances of these categorizations are compared, and then the independent variables in the significant categorizations are selected as the child nodes. Finally, the corresponding decision tree is obtained by repeatedly categorizing the data.

The CART algorithm was proposed in 1984 by Breiman et al. of Stanford University and the University of California, Berkeley. During the growth of CART, corresponding calculation is needed for each input variable to determine the optimal grouping variable. However, the calculation strategies of the classification tree and regression tree are different, and there are also differences in the calculation strategies between the numerical input variables and the categorical input variables [20].

The reason we chose the CHAID algorithm and the CART algorithm was that the input variables and output variables of these two algorithms can be both categorical

and numerical, a feature that met the requirement for data types in this study. The output variable of this study, the total medical cost is a numerical variable. Therefore, in this study, we mainly used the CHAID algorithm and CART algorithm to perform regression prediction analysis.

Results

In this study, IBM SPSS Modeler software was used to process and analyze the dataset, and the analysis results were as follows.

Assessment and adjustment of data quality

The missingness, outliers, and extremes of the dataset of the 1819 cases collected in this study were assessed. The completeness of the data of these 1819 cases was good, with no missingness, but there were some outliers and extremes. For example, age, length of hospital stay, and total medical cost all had individual outliers or extremes. To prevent outliers and extremes from affecting our data analysis results, the outliers and extremes were corrected. The correction method was to replace the outliers with the normal values closest to them, while the extreme samples were eliminated. In the end, 1802 cases were included.

Feature selection

If all the input variables were involved in the modeling without feature selection, not only might the computational efficiency of the model be affected, but more importantly, due to the possible correlations between the input variables, the obtained model might not be usable for prediction. In this study, Feature Selection node in the Modeler program was used to perform feature selection from the perspectives of the variable itself and of the correlations between input variables and output variables, and the analysis results are shown in Table 2. It can be seen that medical payment method, sex, age, marital status, length of hospital stay, main diagnosis, number of other diagnoses, major surgery, and number of surgeries were all important for predicting the total medical cost. Moreover, there were no variables in the （screened field） box, i.e., from the point of view of the variable itself, there were no unimportant variables. Therefore, all nine selected variables were included in the modeling in this study.

Table 2 Feature selection results

Field	Measurement	Importance	Value
Major surgery	Nominal	Important	1.0
Length of hospital stay	Continuous	Important	1.0
Number of surgeries	Continuous	Important	1.0
Age	Continuous	Important	1.0
Main diagnosis	Nominal	Important	1.0
Other diagnoses	Continuous	Important	1.0
Medical payment method	Nominal	Important	1.0
Marital status	Nominal	Important	0.999
Sex	Marker	Important	0.994

Note: Important > 0.95, 0.9 < moderately important ≤ 0.95, not important < 0.9.

Modeling

Seventy percent of the data were drawn as a training set for building the model, and the remaining 30% were used as the test set for the prediction model. The decision tree obtained using the CHAID model was a four-level decision tree. The

classification rules of the decision tree are shown in Figure 1 . Because the decision tree had many bifurcations, Figure 1 only shows the classification rules of the two-level decision tree. It can be seen from Table 3 that the middle node of the first level of this decision tree used the major surgery as the optimal grouping variable to generate three groups, and the middle node of the next layer of these three groups used the length of the hospital stay as the optimal grouping variable to further categorize them.

The first group was the patients who received medical preparation for renal dialysis, arteriovenous fistula repair for renal dialysis, ultrasound-guided puncture biopsy, ultrasonography, and 24-hour blood pressure monitoring and those who did not undergo surgery. The average of their total medical expenses was 8211.4 yuan. The second group was the patients who received renal allograft transplantation. The average of their total medical expenses was 68,494.2 yuan. The third group was the patients who received hemodialysis, electrocardiogram, peritoneal dialysis, venipuncture, CT examination, nuclear magnetic resonance examination, oxygen inhalation, and other surgical operations. The average of their total medical expenses was 13,451.5 yuan.

When looking at the middle node of the lower level of each group, each group was further divided using the length of hospital stay as the optimal grouping variable. For example, the first group was divided into seven subgroups using the length of hospital stay as the optimal grouping variable, and the specific grouping criteria were length of hospital stay ≤ 3, 3 < length of hospital stay ≤ 6, 6 < length of hospital stay ≤ 7, 7 < length of hospital stay ≤ 8, 8 < length of hospital stay ≤ 12, 12 < length of hospital stay ≤ 18, and length of hospital stay > 18 days.

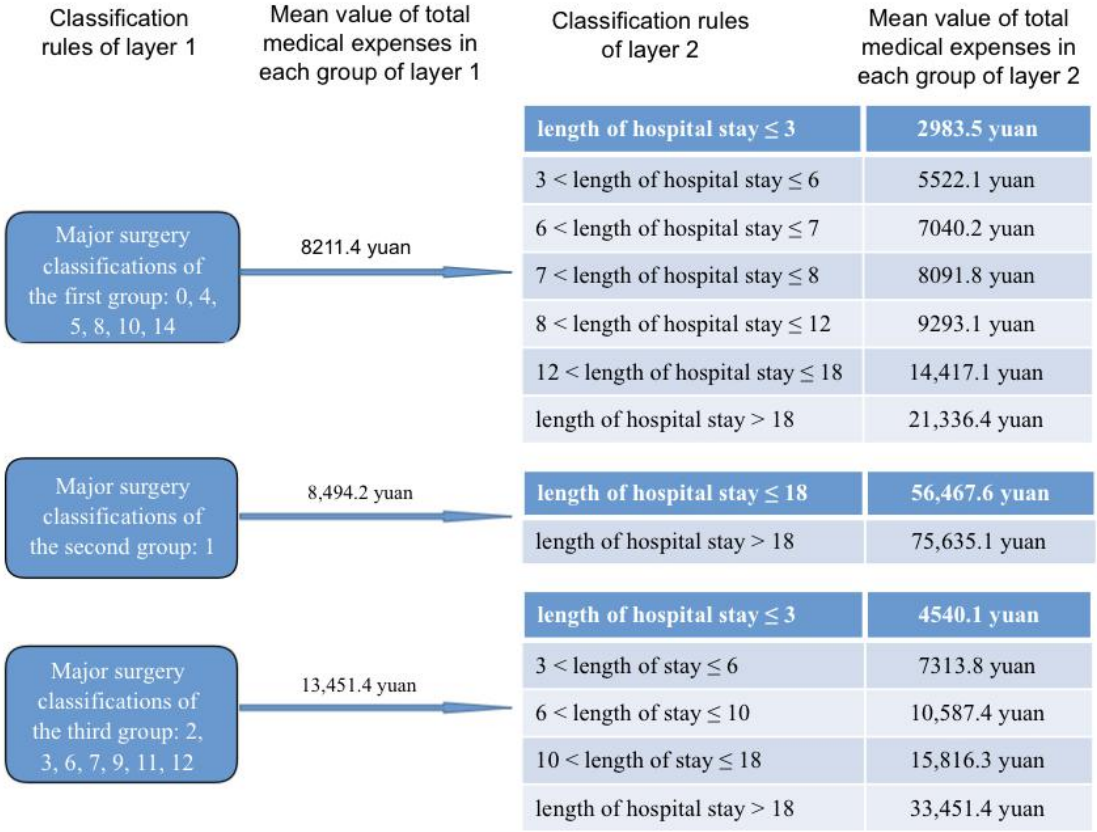

Figure 1 Calculation results of the CHAID model

When modeling with the CART algorithm, a 5-level binary decision tree was

obtained, and its classification rules are shown in Figure 2. This decision tree was first divided into two groups using major surgery as the optimal grouping variable. The first group was the CRF patients who underwent renal allograft transplantation, and the average of their total medical expense was 66,634.5 yuan, while patients who underwent other surgeries or operations were in another group, whose average total medical expense was 9816.1 Yuan. The difference in total medical expenses varied greatly with the length of hospital stay in each group, so the decision tree used the length of hospital stay as the optimal grouping variable for deep growth. Among them, patients who underwent renal allograft transplantation and had a hospital stay of ≤ 21.5 days were included as one group, whose average of total medical expense was 57,982 yuan, and patients who underwent renal allograft transplantation and had a hospital stay of > 21.5 days were in another group, whose average of total medical expense was 77,843 yuan. The CRF patients who received other operations were further divided into 7 subgroups with the length of hospital stay as the optimal grouping variable, and the specific grouping criteria were length of hospital stay ≤ 2.5 as a group, average total medical expense 27,63.6 yuan; 2.5 < length of hospital stay ≤ 4.5, average total medical expense 4638.4 yuan; 4.5 < length of hospital stay ≤ 6.5, average total medical expense 6371.3 yuan; 6.5 < length of hospital stay ≤ 11.5, average total medical expense 9158.3 yuan; 11.5 < length of hospital stay ≤ 17.5, average total medical expense as 14,651.6 yuan; 17.5 < length of hospital stay ≤ 21.5, average total medical expense 22,300.8 yuan; and length of hospital stay 21.5, average total medical expense 33,262.5 yuan.

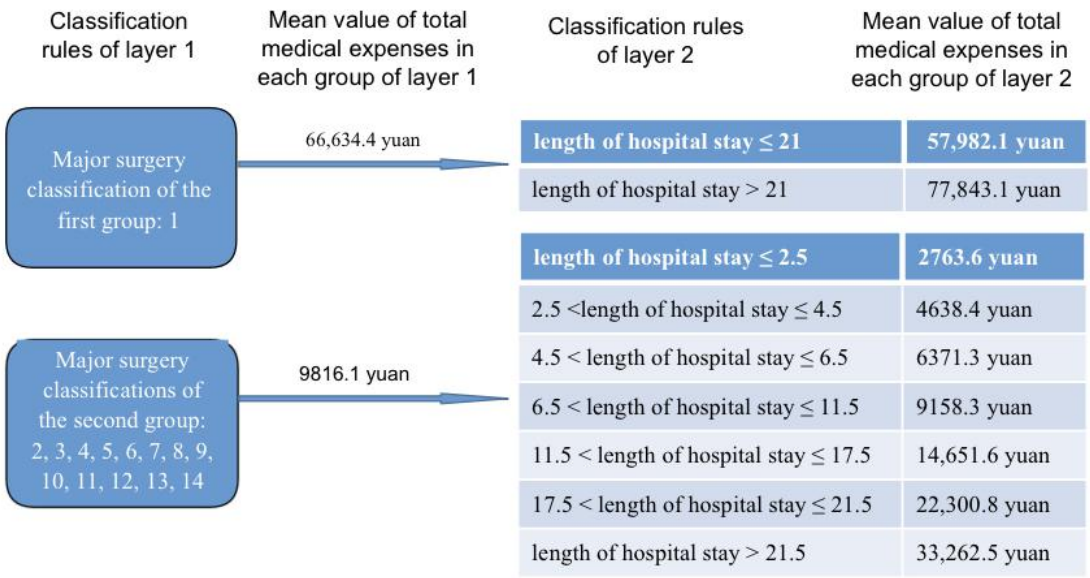

Figure 2 Calculation results of the CART model

According to the calculated importance of each predictor variable in the CHAID model in Table 3, seven influencing factors were identified, and the influencing factors involved in inference rules only included the major surgery, length of hospital stay, main diagnosis, sex, number of surgeries, number of other diagnoses, and age, not marital status or medical payment method. It is worth noting that this does not indicate that marital status and medical payment method had no effect on the total medical expense. According to the obtained importance of each predictor variable in Table 3, major surgery had the strongest effect on the grouping results, and the importance of length of hospital stay, main diagnosis, sex, number of surgeries,

number of other diagnoses, and age for grouping results decreased in that order. According to the obtained importance of each predictor variable by the CART model in Table 3, eight influencing factors were identified, including the major surgery, length of hospital stay, number of surgeries, marital status, medical payment method, main diagnosis, age, and other diagnoses. Sex was not included, but this does not indicate that sex had no effect on total medical cost. According to the calculated importance of each predictor variable in Table 3, major surgery had the most important effect on the grouping results, followed by length of hospital stay, number of surgeries, marital status, medical payment method, main diagnosis, age, and number of other diagnoses.

From the results of the classification rules of the two models, it was seen that both models used the major surgery as the optimal grouping variable for the first layer of grouping and used the length of hospital stay as the optimal grouping variable for the second layer of grouping. From the calculated importance of each predictor variable in the two models, the major surgery and length of hospital stay were significantly more important in predicting the total medical expense than other variables (Table 3).

Table 3 Importance of predictor variables for CHAID and CART models

Variables	**Measure values of the importance of each variable in the CHAID model**	**Measure values of the importance of predictor variables in the CART model**
Major surgery	0.73	0.67
Length of hospital stay	0.26	0.32
Main diagnosis	0.01	0
Sex	0	-
Number of surgeries	0	0
Number of other diagnoses	0	0
Age	0	0
Marital status	-	0
Medical Payment Method	-	0

Note: The importance of the ith input variable is defined as: $\text{Evaluation}_i = \frac{1-P_i}{\sum(1-P_i)}$, and the sum of the importance of all input variables included in the model is 1.

Assessment of the model

The analysis node in the output tab in Modeler was used to compare and analyze the total medical expenses predicted by the two models and the actual medical expenses, and the obtained results are shown in Table 4. The correlation between the total medical cost predicted by the CHAID model and the actual total medical cost was 0.912, and the absolute mean error was 4140.9. The correlation between the total medical cost predicted by the CART model and the actual total medical cost was 0.912, and the absolute mean error was 4305. Thus, the correlations of the two models in predicting total medical cost were good, but there were still errors, and the accuracy should be further improved.

Table 4 Results of the comparison and analysis of the total medical expense predicted by CHAID model and the actual total medical expenses.

	Minimum error	Maximum error	Mean error	Absolute Mean error	Standard deviation	Linear correlation	Incidence
CHAID	49,699.7	71,248.7	0.0	4140.9	7663.1	0.912	1802
CART	50,335.8	73,956.3	123.1	4305.0	7670.6	0.912	1802

Discussion

Our study showed that medical payment method, sex, age, marital status, length of hospital stay, main diagnosis, number of other diagnoses, major surgery, and the number of surgeries all had an effect on the total medical expense of a single hospitalization for CRF patients. The importance of different variables for the establishment of the CHAID and CART models varied because of differences in the methods used to select the optimal grouping variables between the two models. However, from the evaluation results of CHAID and CART models, the results of the two were not significantly different.

The variable that had the greatest impact on the total cost of a single hospitalization for CRF patients was the major surgery, which is also the reason, at present, why major surgery is used as the criterion for the division into medical group, surgical group, and non–operating room operation when countries or regions, such as the United States, Australia, and Beijing, design diagnosis-related groupings (DRGs) [21-22]. In the CHAID model of this study, major surgery was used as the optimal grouping variable for CRF patients, and the CRF patients could also be divided into a medical group, surgical group, and non–operating room operation. The second group was patients who received renal allograft transplantation, and their total medical expense was significantly higher than those in the other two groups. The first group was patients who underwent medical preparation for renal dialysis and arteriovenous fistula repair for renal dialysis, and their total medial expense was higher than that of the patients in the third group, who underwent hemodialysis and peritoneal dialysis. The above grouping results were consistent with the actual clinical situation. Hemodialysis and peritoneal dialysis have different costs. However, these studies focused on comparing the overall cost of each treatment regimen rather than the cost of a single hospitalization, which is different from the focus of this study [23]. In addition, when CRF patients are admitted to large medical institutions for hemodialysis or peritoneal dialysis, they generally have comorbidities requiring examinations and treatments, and the treatment cost is also affected by other factors, so the difference is not very significant in the grouping results.

Our results also showed that next to major surgery, the length of hospital stay was another important variable affecting the total cost of a single medical treatment for CRF patients. For example, for CRF patients undergoing renal allograft transplantation, the difference in the total cost of a single medical treatment was caused by the length of hospital stay, while the difference in the length of hospital stay was mainly caused by the postoperative complications. For example, pulmonary infection, pleural effusion, and pulmonary edema are relatively common pulmonary complications after renal transplantation. The occurrence of complications could cause an increase in the consumption of medical resources, and the hospitalization cost could also increase [24]. For CRF patients who received medical treatment and non–operating room surgical treatment, the influence of the length of hospital stay on the total medical cost should be given even more attention, because the change in the length of hospital stay was the most important variable driving the change in the total medical cost of hospitalization in the absence of the influence of the major surgery variable. It is regrettable that the BJ-DRGs in Beijing do not include the length of hospital stay of CRF patients as a grouping node. Including the length of hospital stay as a case grouping variable and giving different expense reimbursements to CRF patients who have different lengths of hospital stay is beneficial in mitigating the problem of inadequate medical care that arises in the currently implemented DRG medical insurance payments.

Our results suggest that the factors affecting the medical costs of CRF patients are complex, and future studies need to explore in depth the objective factors affecting their length of hospital stay, such as age, complications, and comorbidities, so as to develop a disease-appropriate length of hospitalization for individual patients and to avoid affecting the clinical efficacy and prognosis due to insufficient time of hospitalization and treatment, and at the same time, the length of hospitalization can be effectively controlled to improve medical efficiency and effectively allocate medical resources.

Our study employed the CHAID algorithm and CART algorithm of Modeler to construct the decision tree prediction mode. They have the following advantages: ① their input variables and output variables can be both categorical and numerical, which are suitable for datasets with complex data types; ② they have relatively diverse output types, which can be both a classification decision tree and inference rules; ③ their output results are easily interpretable, which is suitable for researchers to explore the inherent pattern of the data and make classification predictions for subjects. At the same time, there is also an insurmountable shortcoming in using Modeler's CHAID algorithm and CART algorithm to construct decision tree grouping model: both algorithms can only select the optimal grouping variable among the input variables but cannot integrate the effect of multiple input variables on the target variable. Therefore, subsequent studies can use factor analysis and other methods to effectively integrate multiple input variables before using them in modeling [20].

Our study showed that it is necessary for health administrative departments and medical insurance administrative departments to revise the current DRGs for CRF patients and develop an individualized medical insurance payment system that is suitable for China's national conditions, so that CRF patients can receive feasible and fair medical services. Some other countries have already worked in this direction and put it into practice [14]. Our study can provide a reference and suggestions for health administrative departments and health insurance management departments, as well as provide methodological guidance for case grouping and cost prediction for each specific disease.

Competing interests

We declare that we have no conflicts of interest.

Acknowledgements

Thanks to the members of the Hunan Provincial Key Laboratory of Medical Information for their suggestions on the data processing and interpretation of the research results, and to some medical staff in the Department of Nephrology and Transplantation of the Third Xiangya Hospital of Central South University for their communication and guidance on clinical issues of this study.

Author details

1. Key Laboratory of Medical Information Research, Third Xiangya Hospital, Central South University, Changsha, Hunan Province, People’s Republic of China;
2. The Second Xiangya Hospital, Central South University, Changsha, Hunan Province, People’s Republic of China;
3. Department of Medical Information，School of life science, Central South University, Changsha, Hunan Province, People’s Republic of China;

4. Department of Sociology, Central South University, Changsha,Hunan Province, People’s Republic of China.

References

1. Harambat J, van Stralen KJ, Kim JJ, Tizard EJ:Epidemiology of chronic kidney disease in children .Pediatr Nephrol. 2012 Mar;27(3):363=73.
2. Dong-ChanJin, Sung RoYun, Seoung WooLee, Sang WoongHan, WonKim, JonghaPark,Yong KyunKim:Lessons from 30 years' data of Korean end-stage renal disease registry, 1985-2015 .Kidney Res Clin Pract. 2015 Sep;34(3):132-9.
3. Amanda A, Honeycutt ; Joel E, Segel ; Xiaohui, Zhuo ; Thomas J, Hoerger ; Kumiko, Imai ; Desmond, Williams:Medical costs of CKD in the Medicare population .J Am Soc Nephrol. 2013 Sep;24(9):1478=83.
4. Hwang SJ, Tsai JC, Chen HC:Epidemiology, impact and preventive care of chronic kidney disease in Taiwan .Nephrology (Carlton). 2010 Jun;15 Suppl 2:3=9.
5. Josef, Coresh ; Elizabeth, Selvin ; Lesley A, Stevens ; Jane, Manzi ; John W, Kusek ; Paul, Eggers ; Frederick, Van Lente ; Andrew S, Levey: Prevalence of chronic kidney disease in the United States . JAMA 298: 2038–2047, 2007 .
6. Ploth DW, Shepp PH, Counts C, Hutchison F:Prospective analysis of global costs for maintenance of patients with ESRD.Am J Kidney Dis. 2003 Jul;42(1):12-21.
7. Al=Shdaifat EA1, Manaf MR.The economic burden of hemodialysis in Jordan .Indian J Med Sci. 2013 May=Jun;67(5=6):103=16.
8. Hap K, Madziarska K, Hap W, Mazanowska O:Phenotypic risk factors for new=onset diabetes mellitus (NODAT) in renal transplant recipients .Postepy Hig Med Dosw (Online). 2014 Nov 18;68:1347=51.
9. Tsao NW, Lo C, Leung M, Marin JG, Martinusen D:A budget impact analysis of the introduction of erythropoiesis stimulating agent subsequent entry biologics for the treatment of anemia of chronic kidney disease in Canada .Can J Kidney Health Dis. 2014 Nov 11;1:28.
10. Imran S, Sheikh A, Saeed Z, Khan SA, Malik AO, Patel J, Kashif W, Hussain A:Burden of chronic kidney disease in an urban city of Pakistan, a cross=sectional study .J Pak Med Assoc. 2015 Apr;65(4):366=9.
11. Khan S:Secondary hyperparathyroidism is associated with higher cost of care among chronic kidney disease patients with cardiovascular comorbidities . Nephron Clin Pract. 2007;105(4):c159-64. Epub 2007 Jan 25.
12. Zhao Y, Kang B, Liu Y, Li Y, Shi G, Shen T, Jiang Y, Zhang M, Zhou M, Wang L:Health insurance coverage and its impact on medical cost: observations from the floating population in China.PLoS One. 2014 Nov 11;9(11):e111555.
13. Naito H:The Japanese health-care system and reimbursement for dialysis.Perit Dial Int. 2006 Mar-Apr;26(2):155-61.
14. Hornberger J, Hirth RA:Financial implications of choice of dialysis type of the revised Medicare payment system: an economic analysis .Am J Kidney Dis. 2012 Aug;60(2):280-7. doi: 10.1053/j.ajkd.2012.03.010. Epub 2012 May 5.
15. Muñoz E, Thies H, Maesaka JK, Angus G,Goldstein J, Wise L:Diagnosis related groups, resource utilization, age, and outcome for hospitalized nephrology patients .Am J Kidney Dis. 1988 Jun;11(6):481-8.

16. LIU Xia,QIU Jie,SI Qin.Verification of DRGs packet flow in chronic renal failure with path analysis .Chinese Journal of Health Policy,August 2014， Vol. 7 No. 8.
17. AiJing Luo WeiFu Chang ZiRui Xin Hao Ling Jun-Jie Li PingPing Dai Xuan-Tong Deng; Lei Zhang ShaoGang Li .Diagnosis related group grouping study of senile cataract patients based on E-CHAID algorithm .Int J Ophthalmol, Vol. 11, No. 2, Feb.18, 2018.
18. Yuan Suwei, Li Na, Zhang Yunting, Gu Qian, Liu Yan, Ma Jin. Impact of different medical security levels on hospitalization costs of patients with acute myocardial infarction . Journal of Shanghai Jiaotong University: Medical Edition, 2013, 33 (2): 214-219.
19. Wang Zhenyu, Leng Mingxiang, Wan Bin. Impact of different payment methods on the average hospital stay and medical expenses of patients . Medicine and Society, 2012, 25 (7): 63-6474.
20. Xue Wei: Data Mining Based on SPSS Modeler (Second Edition) ., Renmin University of China Press, 2014.10.
21. Common wealth of Australian. Australian refined diagnosis related groups,Version 5.0,Definitions Manual.2002,Canberra.
22. Grimaldi PL,Michelrtti JA.DRGs update:Medicare’s prospective payment plan.Chicago:Pluribus Press,1984.
23. Atapour A, Eshaghian A, Taheri D, Dolatkhah S:Hemodialysis versus peritoneal dialysis, which is cost-effective ?Saudi J Kidney Dis Transpl. 2015 Sep;26(5):962-5.
24. Xu Xia, Guo Shuliang:Clinical Analysis of 125 Cases of Pulmonary Complications after Renal Transplantation. Chongqing Medical Journal, Volume 44, Issue 7: 913-918, March 2015.

Figures

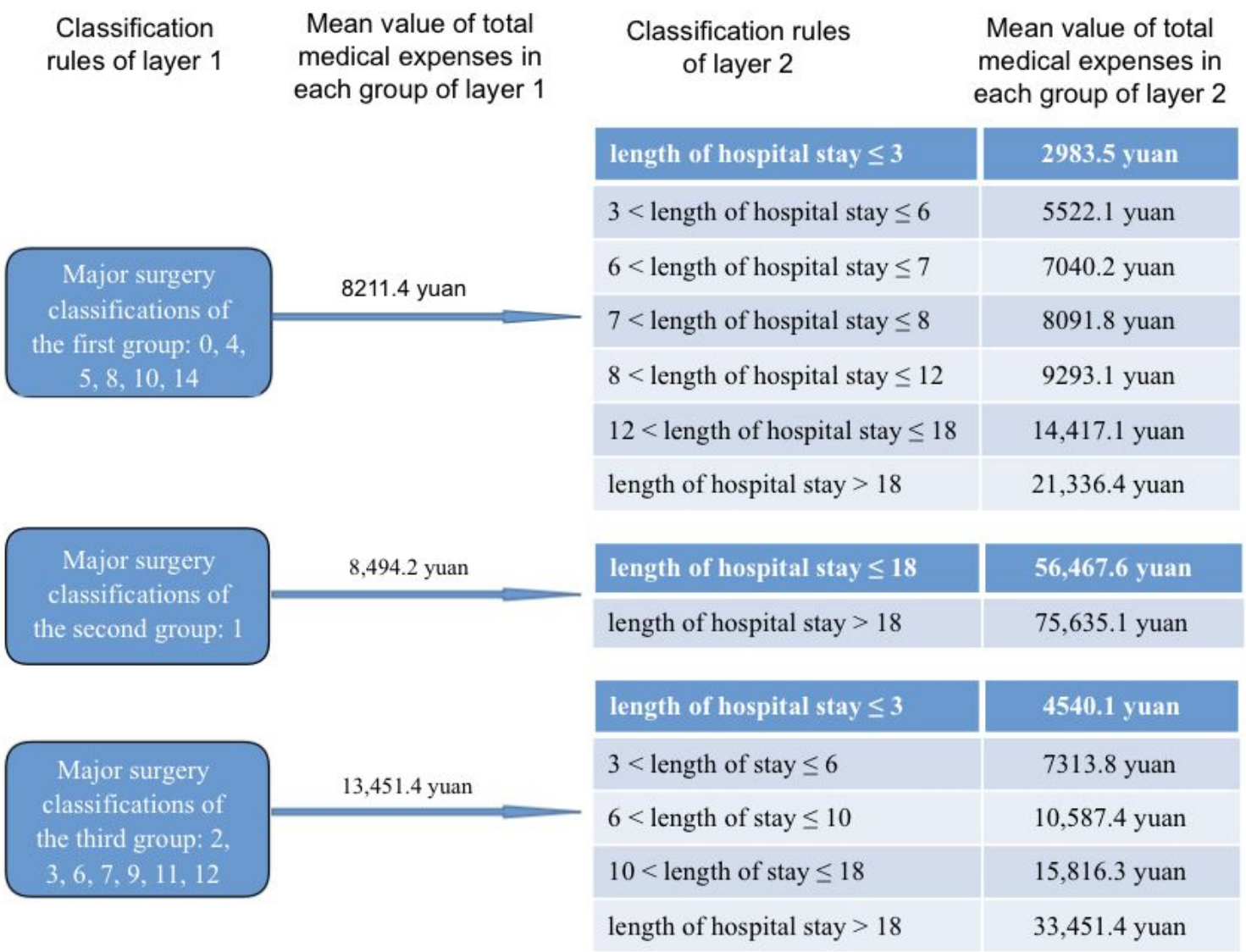

Figure 1

Calculation results of the CHAID model

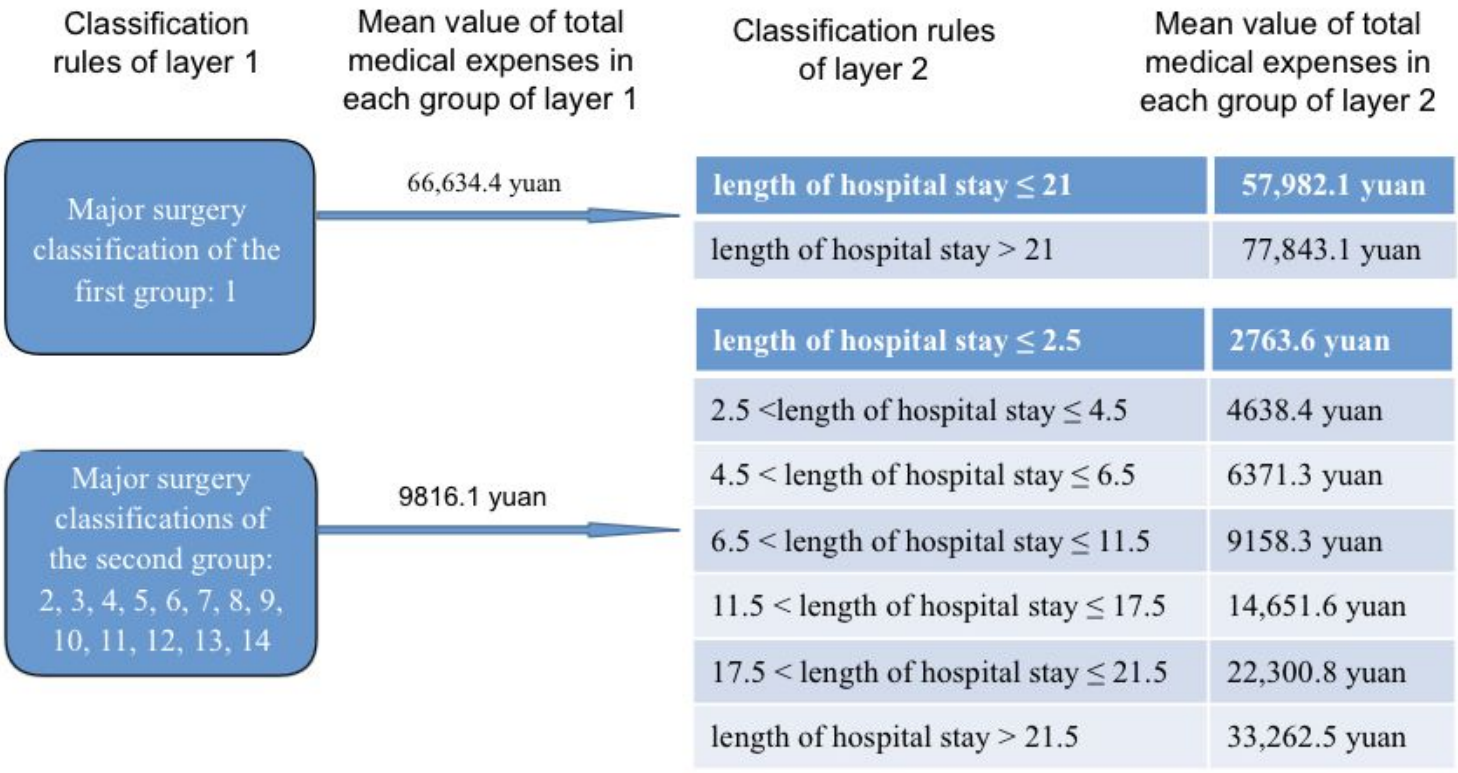

Figure 2

Calculation results of the CART model