

# Decision Tree Analysis on the Factors Influencing the COI of Infantile Bronchopneumonia Inpatients in a Northern Hospital of China

dongmei pei (✉ [peidm1111@hotmail.com](mailto:peidm1111@hotmail.com))

Shengjing Hospital of China Medical University <https://orcid.org/0000-0002-1947-4447>

**YANGYANG YUE**

Shengjing Hospital of China Medical University

**jiakai guo**

Shengjing Hospital of China Medical University

**qiyong guo**

Shengjing Hospital of China Medical University

---

## Research article

**Keywords:** Infantile bronchopneumonia, cost of illness (COI), decision tree analysis

**Posted Date:** October 28th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.16516/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Background: Bronchopneumonia is a disease with a high death rate for children in developing countries. It not only affects the healthy growth of children, but also puts economic pressure and additional burdens on their families and society. Social development and advancement may change the factors influencing the hospitalization costs. This study aimed to explore reasonable cost control approaches by analyzing the factors related to unreasonable increases in medical costs, so as to standardize the diagnosis and treatment behaviors and determine management methods for scientific medical costs. Methods: Using the decision tree analysis method combined with the characteristic variables of inpatients, data mining and analysis were performed on the assortment of charges for 15,980 infantile bronchopneumonia inpatients in a northern hospital of China during January 2013 to December 2017. Results: The medical costs of infantile bronchopneumonia inpatients tended to decrease year by year. Various factors influencing the hospitalization costs were sequenced in order of decreasing importance: salvage, complications, admission condition, discharge condition, hospital stay, age and medical payment mode. The hospitalization cost of 623 (78.5%) patients with salvage during hospitalization was >RMB 10,000.

Hospitalization cost of

## Background

With the development of economy and science and technology, and an increased percentage of medical and health costs in GDP, governments and scholars in various countries have assigned increased importance to research on the cost of illness (COI) so as to achieve the purpose of reasonably deploying the medical and health resources and scientifically controlling the medical costs. Infantile bronchopneumonia is one of the most common lower respiratory tract infectious diseases. Because children have a weak immune system, the inflammation can develop rapidly and invade other tissues and organs, seriously reducing their quality of life and even threatening life, which not only affects their healthy growth, but also brings economic pressure and an additional burden to their families and society. Many international studies [1–8] have suggested that viruses are the first pathogens inducing bronchopneumonia, with Haemophilus influenzae the major bacterial pathogen of acute pneumonia and Streptococcus pneumoniae also a dominant pathogenic bacterium causing infantile bacterial pneumonia in developed and developing countries. Over 5,000,000 children in developing countries die of S.pneumoniae-induced pneumonia annually, with 81% of death from pneumonia occurring in children  $\leq 2$  years old. As their age increases, children have stronger immunity because of increasingly mature body development and can recover quickly from diseases. The factors influencing the hospitalization costs of children with bronchopneumonia are classified into two types: controllable and uncontrollable. The uncontrollable factors include age, sex and other social factors; controllable factors are treatment charge, diagnosis charge, examination charge, drug expense and other clinical factors. Previous studies pointed out that to reasonably control hospitalization costs of bronchopneumonia in children, we need to enhance the management of controllable factors, reduce occurrence of complications during hospitalization, cut unnecessary diagnosis and treatment charges, and decrease the hospitalization costs of patients. The current COI studies of bronchopneumonia in children mostly involve quantitative analysis on the

composition and influential factors of hospitalization costs, in which the factors associated with the costs in some special period of time or caused by some specific drug and treatment regimens are analyzed dominantly using statistical analysis. However, there have been no comprehensive studies of the composition and influential factors of hospitalization costs. With social development and advancement, there may be new changes in the influential factors for bronchopneumonia-related hospitalization costs. Therefore, large-size data mining and investigational analysis are needed to reflect the current change trend of COI concerning infantile bronchopneumonia, and it is also necessary to explore reasonable cost control approaches by analyzing the factors related to the unreasonable increases in COI, so as to standardize the diagnosis and treatment behaviors and determine methods to manage scientific medical costs.

## Methods

### 2.1 Data source

All cases of bronchopneumonia in children of 1–14 years old who were admitted to a northern hospital of China from January 2013 to December 2017 were included in the study. The following data were collected: demographic information (age, sex and medical insurance mode) and diagnosis and treatment information (hospitalization date, hospital stay, hospitalization costs, admission condition, discharge condition, complications and salvage). The patients with the same identity information were regarded as repeat inpatients, and data for their first hospitalization were collected.

The age stages of children follow: infant stage,  $\leq 1$  year (12 month); toddler stage, 1–3 years ( $>1$  year and  $\leq 3$  years); preschool stage, 3–6 years ( $>3$  years and  $\leq 6$  years); and school stage,  $>6$  years.

The medical payment mode was classified into the following types: medical insurance for urban residents, new-type rural cooperative medical insurance, business insurance and out-of-pocket payment.

The hospital stay was divided into two types:  $\leq 7$  and  $>7$  days.

Based on the medical situations, the admission condition was divided into three levels: critical, emergent and general.

Based on the turnover, discharge condition was classified into five types: improved, cured, uncured, death and other.

### 2.2 Analysis methods

The decision tree in machine learning is a learning-supervising method that is used to judge the possible event probability by calculating the probability of known events and then constructing a classification tree, and it is widely applied in classification and regression tasks and has such advantages as intuitiveness, high efficiency and measurability. The datasets are classified using the decision tree

classifier J48 in WEKA3.8.1 (Waikato Environment for Knowledge Analysis) tool, and then a classification tree based on the optimal information attributes is constructed with the same method [10]. The main purpose of the decision tree is to make a prediction model of target variables according to the predictable factors, and the decision tree is a supervision method to classify the samples into the categories of interest using the “if–then” rules[11].

This algorithm is used to find the most important independent variable and is first set at the root node and then introduces the next optimal fitting variable (known as bifurcation). The tree flows in a top-to-bottom manner, from the root node to internal nodes (independent variables) and then to terminal leaf nodes (class prediction)[12–17]. In the decision tree, the first variable (root) is the most important factor, and the variable furthest from the root is the next most important factor for the data classification[18]. All variables in a path are regarded as predictors (“if” part) and the class labels of leaf nodes are the expected results (“then” part). To avoid over-fitting and maintain parsimony, the unnecessary terminal branches can be removed according to the defined algorithm so as to prune the tree-generated model, without influencing classification accuracy[19].

## Results

### 3.1 General data

Table 1 shows the overall change trend of hospitalization costs of 15,980 infantile bronchopneumonia inpatients in a northern hospital of China during 2013–2017. The medical costs of infantile bronchopneumonia inpatients tended to decrease year by year. The consumer price index was 100 in 2013 and 132.1 in 2017[9], suggesting that the medical costs of infantile bronchopneumonia inpatients had a yearly decreasing trend, although there was an improvement in standard of living, an increase in commodity prices and changes in some social and economic factors.

We created a database of 15,980 infantile bronchopneumonia inpatients who received treatment in a northern hospital of China from January 2013 to December 2017. The database consisted of 10 variables: eight input and two target variables. The input variables were age, sex, medical payment mode, hospital stay, salvage, complications, admission condition and discharge condition. The two target variables were hospitalization costs of >RMB 10,000 and <RMB 10,000. The classification characteristics were compared between the two target groups using  $\chi^2$  test in SPSS 19.0 statistical software, and the analysis showed significant differences (Table 2).

A decision tree with 13 nodes and 15 leaf nodes was constructed from the datasets of infantile bronchopneumonia inpatients using the 10-fold computation of J48 algorithm (Fig. 1). The decision tree showed various factors influencing the hospitalization costs in order of decreasing importance: salvage, complications, admission condition, discharge condition, hospital stay, age and medical payment mode.

Table 3 lists all 13 “if–then” rules used for modeling. The hospitalization cost of 623(78.5%)bronchopneumonia patients with salvage during hospitalization was >RMB 10,000. Hospitalization cost of <RMB 10,000 was seen in 14,688(96.2%) patients without salvage, 13,921(87.1%) patients without complications (regardless of salvage),13,860(86.7%)patients without salvage and complications, and 948(81.6%) patients without salvage and with complications and a general admission condition.

The performance of the decision tree for the hospitalization costs of children with bronchopneumonia was determined using a confusion matrix. The analysis accuracy of the decision tree was further evaluated by confusion matrix analysis of datasets (Table 4). The correct classification rate, wrong classification rate, precision and recall of the prediction model were 78.6%(12,559/1598), 21.4% (3421/15,980), 78.1% and 78.6%, respectively; the precision and recall both had significant equilibrium.

## Discussion

The hospitalization costs can be used as an important evaluation indicator for standard medical behaviors, and excessive hospitalization costs will bring a heavy economic burden to patient families and cause resource waste of medical insurance payments. It is necessary to analyze the structure of component charges in the hospitalization costs or the hospitalization cost, which can provide a foundation for the health economic research of diseases and a reference basis for the health management departments formulating the relevant policies of COI. The demonstration of cost study results and data allows hospital managers to understand the features of COI, and offers a data support to standardizing reasonable clinical treatment, managing the diagnosis and treatment behaviors of medical staff, and achieving the access of patients to the best medical services.

In the present study, we described the change trend and influencing factors of hospitalization costs by analyzing the COI and the influencing factors for infantile bronchopneumonia inpatients from a northern hospital of China during 2013–2017. In a view of the overall COI trend, the hospitalization costs of infantile bronchopneumonia inpatients tended to decrease yearly during 2013–2017. There was no increase in either hospitalization costs of disease burden due to the regulation of policies, although the living standards of people improved gradually and material life was enriched.

Salvage and complications had a greater impact on hospitalization costs and were the top two influential factors. Thus disease of patients having developed to an emergent state at admission and so requiring salvage, or having complications, should be avoided, because these can increase the expense burden.

Analysis showed that the following measures can be used to reasonably reduce the hospitalization costs of infantile bronchopneumonia: (i) actively carry out health education of patients and their families and popularize disease prevention and treatment knowledge to avoid serious disease states before admission; (ii) hospital managers should urge medical staff to avoid unnecessary examinations and charges and shorten the average hospital stay by finishing various necessary diagnostic examinations of patients in outpatient departments and by not requiring repeated examinations after hospitalization; (iii)

provide timely examination results of patients by perfecting the information system so that clinicians can initially access these results by computer, without waiting for paper reports; and (iv) improve the business ability and technical level of doctors and determine management methods for scientific medical costs.

However, this study only involved analysis of a 5-year inpatient database and analysis of the factors affecting hospitalization costs was limited. Greater time range and coverage of the study would allow more rigorous and all-round analysis results.

## Conclusions

We used a decision tree to analysis the factors influencing the cost of illness of infantile bronchopneumonia from a database of inpatients in a large northern Chinese hospital. We used seven features of patients as predictor variables. The classifier, trained by J48, indicates that factors influencing the hospitalization costs were sequenced in order of decreasing importance: salvage, complications, admission condition, discharge condition, hospital stay, age and medical payment mode. The demonstration of study results allows hospital managers to understand the factors influencing the COI of infantile bronchopneumonia inpatients. Such study can provide a foundation for the health economic research.

## Abbreviations

COI	Cost of illness
SPSS	Statistical Package for Social Sciences
WEKA	Waikato Environment for Knowledge Analysis

## Declarations

### Ethics Approval and consent to participate

Ethics approval was obtained by Shengjing Hospital of China Medical University Ethics Committee (ref. Ethics 2019PS089J). We obtained written informed consent from all parents for participants under 16 years old.

### Consent for publication

Not applicable

### Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

## Competing interests

All authors declare that they have no conflicts of interests.

## Funding

No funding

## Authors' contributions

QG and DP had the initial conception of the idea and background for this study. DP and YY made contributions in the acquisition and analysis of the data. JG performed the literature search. All authors contributed to the writing, reviewing and final approval of the manuscript.

## Acknowledgements

We are grateful to Hospital Administration Office in Shengjing Hospital of China Medical University for supporting this research.

## References

- [1] Michelow C, Olsen K, Lozano JE, et al. Epidemiology of clinical characteristics of community acquired pneumonia in hospitalized children[J]. *Pediatrics*, 2004; 113(4):701-707.
- [2] Jain S, Williams DJ, Arnold SR, et al. Community-acquired pneumonia requiring hospitalization among U.S. children[J]. *New England Journal of Medicine*, 2015, 372(9):835-845.
- [3] Rudan I, O'Brien KL, Nair H, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. [J]. *Journal of Global Health*, 2013, 3(1):010401.
- [4] Pennington JE. Treating respiratory infections in the era of cost control[J]. *American Family Physician*, 1986, 33(2):153-160.
- [5] Greenwood B. The epidemiology of pneumococcal infection in children in the developing world[J]. *Philosophical Transactions of The Royal Society of London Series B-Biological Sciences*, 1999, 354(1384):777-785.
- [6] Oliwa J, Karumbi J, Marais B, et al. Tuberculosis as a cause or co-morbidity of childhood pneumonia in tuberculosis-endemic areas: a systematic review[J]. *Lancet Respiratory Medicine*. 2015, 3(3):235-243.
- [7] Zhang S, Sammon PM, King I, et al. Cost of management of severe pneumonia in young children: systematic analysis[J]. *Journal of Global Health*. 2016, 6(1): 010408.

- [8]Juven T, Mertsola J, Waris M. Etiology of community-acquired pneumonia in 254 hospitalized children[J]. *Pediatric Infectious Disease Journal*, 2000, 19(4):293.
- [9]National Bureau of Statistics. 2017 statistical bulletin of the People's Republic of China on national economic and social development[2018-2-28]
- [http://www.stats.gov.cn/tjsj/zxfb/201802/t20180228\\_1585631.html](http://www.stats.gov.cn/tjsj/zxfb/201802/t20180228_1585631.html).
- [10]Worachartcheewan, A., et al., Identification of metabolic syndrome using decision tree analysis. *Diabetes Res Clin Pract*, 2010. 90(1): p. e15-8.
- [11]Al-Turaiki, I., M. Alshahrani, and T. Almutairi, Building predictive models for MERS-CoV infections using data mining techniques. *J Infect Public Health*, 2016. 9(6): p. 744-748.
- [12]Maniruzzaman, M., et al., Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Comput Methods Programs Biomed*, 2017. 152: p. 23-34.
- [13]Yoo, I., et al., Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst*, 2012. 36(4): p. 2431-48.
- [14]Tayefi, M., et al., The application of a decision tree to establish the parameters associated with hypertension. *Comput Methods Programs Biomed*, 2017. 139: p. 83-91.
- [15]Samanta, B., et al., Prediction of periventricular leukomalacia. Part I: Selection of hemodynamic features using logistic regression and decision tree algorithms. *Artif Intell Med*, 2009. 46(3): p. 201-15.
- [16]Ravikumar, P., et al., Prevalence and risk factors of diabetes in a community-based study in North India: the Chandigarh Urban Diabetes Study (CUDS). *Diabetes Metab*, 2011. 37(3): p. 216-21.
- [17]Reis, J.P., et al., Lifestyle factors and risk for new-onset diabetes: a population-based cohort study. *Ann Intern Med*, 2011. 155(5): p. 292-9.
- [18]Ramezankhani, A., et al., Applying decision tree for identification of a low risk population for type 2 diabetes. *Tehran Lipid and Glucose Study. Diabetes Res Clin Pract*, 2014. 105(3): p. 391-8.
- [19]Zhang, Q, et al, Development and validation study of a non-alcoholic fatty liver disease risk scoring model among adults in China. *Fam Pract*, 2017. 34(6): p. 667-672.

## Tables

Table 1 Changes in hospitalization costs of children with bronchopneumonia during 2013-2017

Year	N	Per capita cost (yuan/person)
2013	2346	10237.04±10417.32
2014	2409	9139.85±9200.02
2015	3107	9260.93±8670.43
2016	4168	9529.16±8893.55
2017	3950	8835.5±6614.00

F=10.556, P<0.001

Table 2 Variable characteristics of influential factors related to hospitalization costs of infantile bronchopneumonia inpatients.

Variable	Definition	Hospitalization cost		<i>P</i> -value
		<RMB 10,000 N=11,924	>RMB 10,000 N=4056	
Medical payment mode	Medical insurance for urban residents	6818 (57.2%)	2102 (51.8%)	0.063
	New-type rural cooperative medical insurance	2395 (20.1%)	1110 (27.4%)	
	Business insurance	767 (6.4%)	169 (4.2%)	
	Out-of-pocket payment	1944 (16.3%)	675 (16.6%)	
Age	≤1 year	2709 (22.7%)	1276 (31.5%)	0.091
	1-3 years	4609 (38.7%)	1171 (28.9%)	
	3-6 years	3496 (29.3%)	1037 (25.6%)	
	>6 years	1110 (9.3%)	572 (14.1%)	
Sex	Male	6812 (57.1%)	2415 (59.5%)	0.007
	Female	5112 (42.9%)	1641 (40.5%)	
Hospital stay	≤7 days	8415 (70.6%)	1845 (45.5%)	<0.001
	>7 days	3509 (29.4%)	2211 (54.5%)	
Salvage	No	11798 (98.9%)	3468 (85.5%)	<0.001
	Yes	126 (1.1%)	588 (14.5%)	
Complications	No	11217 (94.1%)	3037 (74.9%)	<0.001
	Yes	707 (5.9%)	1019 (25.1%)	
Admission condition	Critical	90 (0.8%)	292 (7.2%)	<0.001
	Emergent	1755 (14.7%)	661 (16.3%)	
	General	10079 (84.5%)	3103 (76.5%)	
	Turnover			
Turnover	Cured	1068 (9.0%)	299 (7.4%)	<0.001
	Improved	10614 (89.0%)	3653 (90.1%)	
	Uncured	107 (0.8%)	39 (1.0%)	
	Death	2 (0.01%)	4 (0.01%)	

Other

133 (1.1%)

61 (1.5%)

---

Table 3 Thirteen rules for hospitalization costs of infantile bronchopneumonia inpatients extracted from the decision tree

---

Rule 1: If the inpatient had no salvage and no complications, then hospitalization cost was <RMB 10,000 (13,860 patients).

---

Rule2: If the inpatient had no salvage but had complications and a critical or emergent admission condition, then hospitalization cost was >RMB 10,000(245).

---

Rule3: If the inpatient had no salvage but had complications, a general admission condition and age >6 years, then hospitalization cost was >RMB 10,000(119).

---

Rule4: If the inpatient had no salvage but had complications, a general admission condition, age <6 years and hospital stay of  $\leq 7$  days, then hospitalization cost was <RMB 10,000 (593).

---

Rule5: If the inpatient had no salvage but had complications, a general admission condition, age <6 years, hospital stay of >7 days and medical payment mode was medical insurance for urban residents, then hospitalization cost was <RMB 10,000(235).

---

Rule6: If the inpatient had no salvage but had complications, a general admission condition, age <6 years, hospital stay of >7 days and medical payment mode was new-type rural cooperative medical insurance, business insurance or out-of-pocket payment, then hospitalization cost was >RMB 10,000(214).

---

Rule7: If the inpatient had salvage and complications and the turnover was uncured, death or other, then hospitalization cost was <RMB 10,000(10).

---

Rule8: If the inpatient had salvage and complications and the turnover was cured or improved, then hospitalization cost was >RMB 10,000(290).

---

Rule9: If the inpatient had salvage and a hospital stay of >7 days but no complications, then hospitalization cost was >RMB 10,000(157).

---

Rule10: If the inpatient had salvage and a hospital stay of  $\leq 7$  days but no complications, and the turnover was uncured, death or other, then hospitalization cost was <RMB 10,000(30).

---

Rule11: If the inpatient had salvage, a hospital stay of  $\leq 7$  days, age >6 years, no complications and the turnover was cured or improved, then hospitalization cost was >RMB 10,000(93).

---

Rule12: If the inpatient had salvage, a hospital stay of  $\leq 7$  days, age  $\leq 6$  years, no complications, the turnover was cured or improved and medical payment mode was out-of-pocket payment or business insurance, then

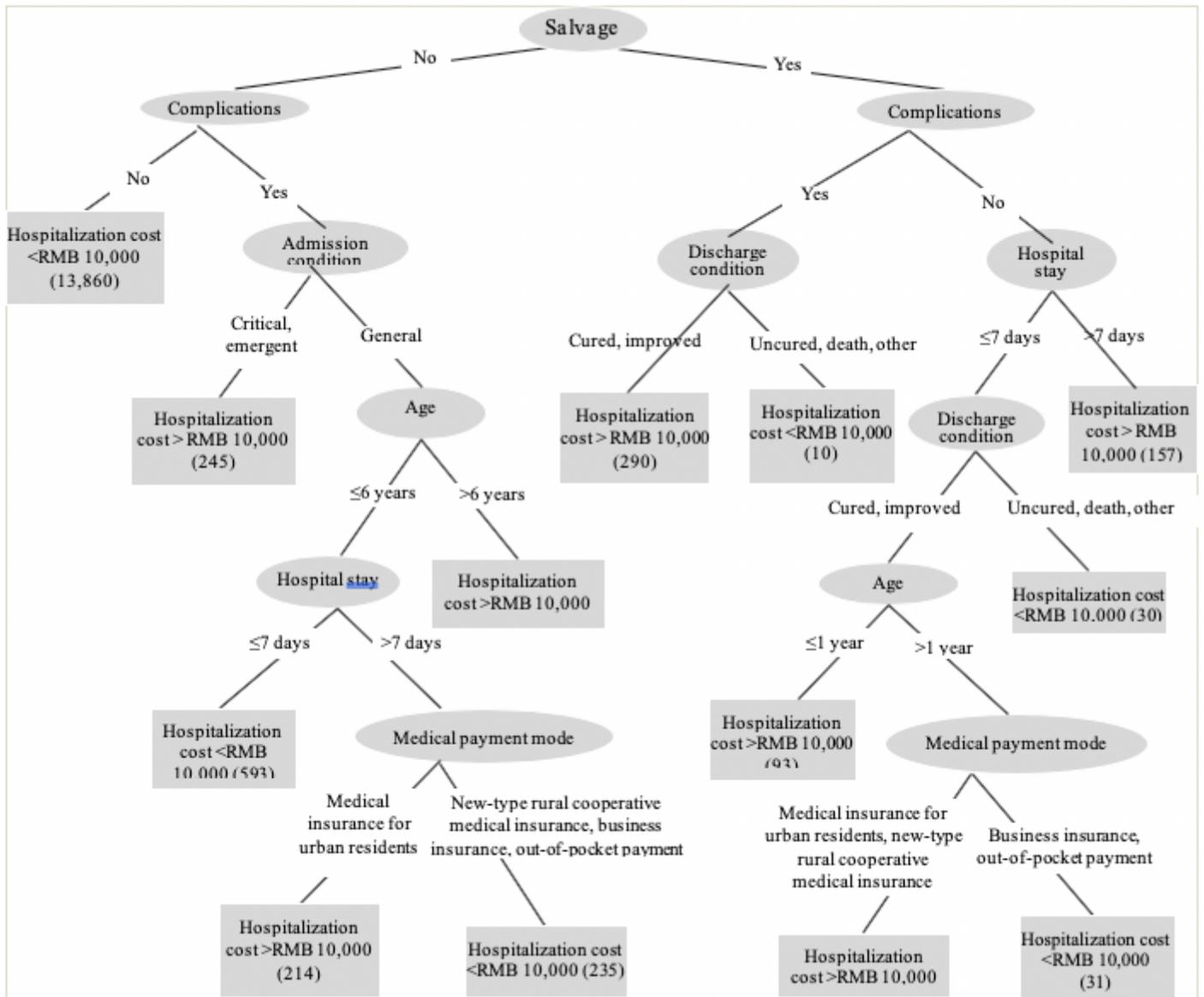
hospitalization cost was <RMB 10,000(31).

Rule13: If the inpatient had salvage, a hospital stay of  $\leq 7$  days, age  $\leq 6$  years, no complications, the turnover was cured or improved and medical payment mode was medical insurance for urban residents or new-type rural cooperative medical insurance, then hospitalization cost was >RMB 10,000(83).

Table 4 Dataset matrix chart of infantile bronchopneumonia inpatients

Actual results	Prediction results	
	Hospitalization cost < RMB 10,000	Hospitalization cost > RMB 10,000
Hospitalization cost < RMB 10,000	11,631	293
Hospitalization cost > RMB 10,000	3128	928

## Figures



**Figure 1**

Dataset decision tree constructed with J48 algorithm (the digit in each node bracket represents the sample size)