*American Journal of Life Sciences*
2024, Vol. 12, No. 2, pp. 33-43
https://doi.org/10.11648/j.ajls.20241202.12

Research Article

# Robust Analysis of the Influencing Factors for Hospitalization Costs of Senile Cataracts Patients in Chengdu Considering Different Types of Insurance

**Haitao Tian[1,*], Tianjun Li[2], Shiqi Lu[3]**

[1]School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China
[2]Department of Medical Insurance, West China Hospital of Sichuan University, Chengdu, China
[3]Department of Statistics, Southwest Jiaotong University, Chengdu, China

## Abstract

Chengdu is one of the earliest pilot cities for urban-rural basic medical insurance integration in China. This study aimed to analyze the influencing factors of hospitalization costs of senile cataract in a tertiary hospital in Chengdu by robust method, especially considering the influence of medical insurance type. A total of 1310 discharged patients from a tertiary hospital from January 2020 to June 2021 who were mainly diagnosed with senile cataracts were selected as the research subjects. Kruskal-Wallis H test and Spearman correlation analysis are used to conduct univariate statistical analysis. The robust multivariate linear regression model and a semi-parametric multivariate regression model are established to obtain the influencing factors for their hospitalization costs. The robust multivariate regression model results show that reimbursement ratio, number of surgeries, type of medical insurance, hospitalization days, number of additional diagnoses and material proportion have significant correlations with the response variable, i.e. total hospitalization costs of the senile cataract patients. In the robust multivariate regression analysis, the type of insurance is significantly associated with the hospitalization costs. Fixing other variables, the hospitalization costs of patients with UEBMI insurance were 7.6% higher than those with URRBMI insurance. Generalized additive model (GAM) can express the nonlinear relationship between explanatory variables and response variable. Because of the nonlinear part of the GAM, the interpretation and description of the model can provide more knowledge than the linear models. In the GAM model, the type of insurance is also significantly related to the total costs. According to the regression effects of reimbursement ratio, number of surgeries, type of medical insurance, hospitalization days, number of additional diagnoses and material proportion on total costs, the paper aims to provide some references for promoting the reform of the local medical system and improving the eye health status and quality of life of middle-aged and elderly groups.

## Keywords

Senile Cataract, Medical Insurance, Hospitalization Costs, Influencing Factors, Robust Regression, Generalized Additive Model

*Corresponding author: tht199304@163.com (Haitao Tian)

**Received:** 12 March 2024; **Accepted:** 3 April 2024; **Published:** 12 April 2024

# 1. Introduction

Over the past few decades, China has embarked on a nationwide effort to reform its healthcare system due to the growing demand for medical services and an increasingly aging population. In the late 1990s, the Chinese government established a medical insurance scheme for urban employees and retirees (the Urban Employee Basic Medical Insurance, UEBMI) [1]. In 2003, the Chinese government introduced the New Rural Cooperative Medical Care System (NRCMS) for the rural population [2]. In 2007, China started the Urban Resident Basic Medical Insurance (URBMI), which will bring urban non-labor population into the coverage of insurance [3]. However, access and equity hinder progress towards universal health care. In order to further narrow the gap between urban and rural areas and improve the treatment of medical services for rural residents, the Chinese government issued the document on integrating the basic medical insurance system for urban and rural residents in 2016 [4]. The Urban and Rural Residents' Basic Medical Insurance (URRBMI) will integrate the NRCMS and URBMI [5].

As a pilot city for the integration of medical insurance for urban and rural residents, Chengdu is in the forefront of medical insurance reform. At present, the medical insurance in Chengdu is mainly divided into the Urban and Rural Residents' Basic Medical Insurance (URRBMI) and the Urban Employee Basic Medical Insurance (UEBMI).

Cataract is one of the main causes of blindness and a major public health problem worldwide [6, 7]. In 2010, cataracts accounted for 33.4% of global blindness and 18.4% of moderate to severe visual impairment [8]. The 2017 Global Burden of Disease Study showed that cataract disability increased by 29.6% globally from 2007 to 2017, reaching 8.01 million [9]. In 2019, the World Vision Report (2020) released by the World Health Organization showed that about 65.2 million patients around the world suffer from moderate and severe far-vision impairment or blindness caused by cataracts, and the incidence of cataracts increases sharply with the acceleration of global population aging [10]. Among them, the incidence of cataracts is high in developing countries, with the majority in Asia [11]. In China, the number of cataract patients in 2020 is estimated to be 5.0625 million, which has caused serious public health problems, greatly affects patients and their families, and also brings a heavy economic burden to the whole society [12-14]. According to the literature, from 2000 to 2020, the total incidence rate of cataracts in the Chinese population over 50 years old was 27.45%, that in rural areas was 28.79%, that in urban areas was 26.66%, and the total coverage rate of cataract surgery was 9.19% [15]. As life expectancy increases and the population ages rapidly, the number of people with visual impairment due to cataracts will increase continuously.

According to the pathological factors of lens opacity, cataracts can be divided into congenital, senile (age-related), metabolic, toxic and radiation cataracts. At present, the corresponding drug treatment for different etiologies fails to delay or reverse lens opacity, the only effective treatment is still surgery [16]. Among them, phacoemulsification combined with intraocular lens (IOL) implantation is one of the most effective surgical methods for cataract treatment, but the costs of the surgery bring an economic burden to patients [17]. In recent years, the relationship between cataract treatment and the social economy has been a hot topic. For example, Den et al. studied that socioeconomic status was negatively correlated with the burden of cataract blindness in 185 countries [18]. They determined that ultraviolet radiation exposure interacts with socioeconomic status and the burden of cataract blindness. Ko et al. found personal familiarity with cataract surgery is important in determining a patient's willingness to pay [19]. Tan et al. believed that cataract diagnosis was associated with a higher utilization rate of medical resources and higher economic burden by studying cataract patients in American medical insurance [20]. At the same time, the study of hospitalization costs is a hot topic, many scholars have conducted related research, such as Bhuiyan et al., Luo et al., Xu et al., Urbich et al., Rodriguez-Martinez et al. [21-25].

Our contribution is mainly reflected in three aspects. First, the sample we analyzed is representative. The sample is from the best local hospital, which is located in relatively developed city in southwest China, with a population of more than 20 million. Therefore, the analysis results have reference value. Second, we used two different regression models to make up for the disadvantages of the single use of classical multivariate linear regression analysis. We analyzed the influencing factors of hospitalization costs through a robust multivariate regression model. Moreover, the GAM was established, and this semi-parametric regression model could reveal more useful information. Third, we focus on the medical insurance type variable and find that it is significantly related to total costs under current health policies. The patients with UEBMI spend more and get better medical care. There is still a gap in medical resources. Therefore, it is suggested that the government continue to reform the content of medical insurance for urban and rural residents, and strive to make low-income groups enjoy more quality medical services without increasing the medical burden.

# 2. Materials and Methods

## 2.1. The Data Source

The sample in this paper is from the medical insurance office of a tertiary hospital in Chengdu. From January 2020 to June 2021, all 1310 patients with senile cataracts underwent phacoemulsification combined with IOL implantation. The sample included many potential influencing factors and covariates such as age, gender, medical insurance types (URRBMI, UEBMI), hospitalization time, discharge time, hospitalization days, medical total costs, reimbursement of costs, number of surgeries, and personal out-of-pocket costs, and all kinds of hospitalization costs (treatment fee, examination fee, material fee, ward fee, radiology examination fee, nursing fee, laboratory fee, anesthesia fee, blood dispensing examination fee, surgery fee, western medicine fee and meal fee), medical record catalog name, medical record catalog surgery name, surgical name, surgical grade, surgical anesthesia, type of incision and healing, etc.

## 2.2. Descriptive Analysis of Data

To gain a more comprehensive understanding of the information provided by the sample, a descriptive statistical analysis is performed first. Among the 1310 senile cataract patients who underwent phacoemulsification combined with IOL implantation, the average reimbursement rate was 40.3%. Among all kinds of hospitalization costs, the cost of materials averagely accounted for 55.0% of the total hospitalization costs (much higher than others), followed by the average cost of surgeries accounted for 23.5%, examination fees accounted for 6.8%, and other costs accounted for less than 5%. Materials cost accounts for the largest proportion of total hospitalization costs in our study, and some other studies have shown that the hospitalization cost of using the domestic low-priced IOL is much lower than that of using an imported high-priced lens [26].

Non-parametric Kruskal-Wallis H test and Spearman correlation test are adopted. The corresponding results are shown in Table 1 and Table 2. In the statistical hypothesis test, P value less than 0.05 is considered to have a significant difference at the significance level of 5%; otherwise, there is no significant difference. Less than 0.1 is considered to have a significant difference at the significance level of 10%, whereas otherwise, there is no significant difference.

***Table 1.*** *Test results of each variable between total costs.*

| **Variable** | | **Num** | **%** | **total costs (RMB)** | | | |
|---|---|---|---|---|---|---|---|
| | | | | **Mean** | **Sd** | **Median** | **P value** |
| Age | 40~60 | 141 | 0.11 | 10650.40 | 3912.67 | 8419.57 | 0.504 |
| | 60~80 | 998 | 0.76 | 10047.68 | 4922.88 | 8416.95 | |
| | >80 | 168 | 0.13 | 10406.69 | 4981.44 | 8563.96 | |
| Gender | M | 489 | 0.37 | 9931.99 | 4884.46 | 8401.26 | 0.522 |
| | F | 818 | 0.63 | 10294.46 | 5089.54 | 8435.11 | |
| *Insurance* | *R* | *199* | *0.15* | *9429.74* | *3765.95* | *8134.07* | *0.005* |
| | *E* | *1108* | *0.85* | *10289.80* | *5198.25* | *8461.16* | |
| Reimbursement ratio | <=0.3 | 162 | 0.12 | 17428.88 | 8518.23 | 16795.10 | <0.001 |
| | 0.3~0.4 | 299 | 0.23 | 8168.05 | 2625.60 | 7245.45 | |
| | 0.4~0.5 | 806 | 0.62 | 9478.75 | 3211.50 | 8491.58 | |
| | >0.5 | 40 | 0.03 | 9300.46 | 4105.41 | 8156.40 | |
| Material proportion | <=0.4 | 25 | 0.02 | 12019.54 | 8026.26 | 8792.84 | <0.001 |
| | 0.4~0.5 | 373 | 0.29 | 9687.91 | 3103.14 | 8722.71 | |
| | 0.5~0.6 | 615 | 0.47 | 9071.33 | 3387.89 | 8315.22 | |
| | >0.6 | 294 | 0.22 | 12873.01 | 7752.11 | 9612.35 | |
| Number of surgeries | 1 | 1145 | 0.88 | 8872.86 | 2795.91 | 8257.85 | <0.001 |
| | 2 | 162 | 0.12 | 19248.10 | 7319.70 | 16820.50 | |
| Additional diagnoses | 0 | 189 | 0.15 | 11847.10 | 7221.35 | 8678.72 | <0.001 |
| | 1-2 | 722 | 0.55 | 10157.41 | 4597.19 | 8451.97 | |
| | 3 | 239 | 0.18 | 9410.62 | 3967.44 | 8232.85 | |
| | >=4 | 157 | 0.12 | 9272.14 | 4572.97 | 7865.32 | |

Note: The number of surgeries refers to single eye surgery or double eye surgery; The number of additional diagnoses refers to the number of diagnoses other than the main diagnoses on the first page of medical records.

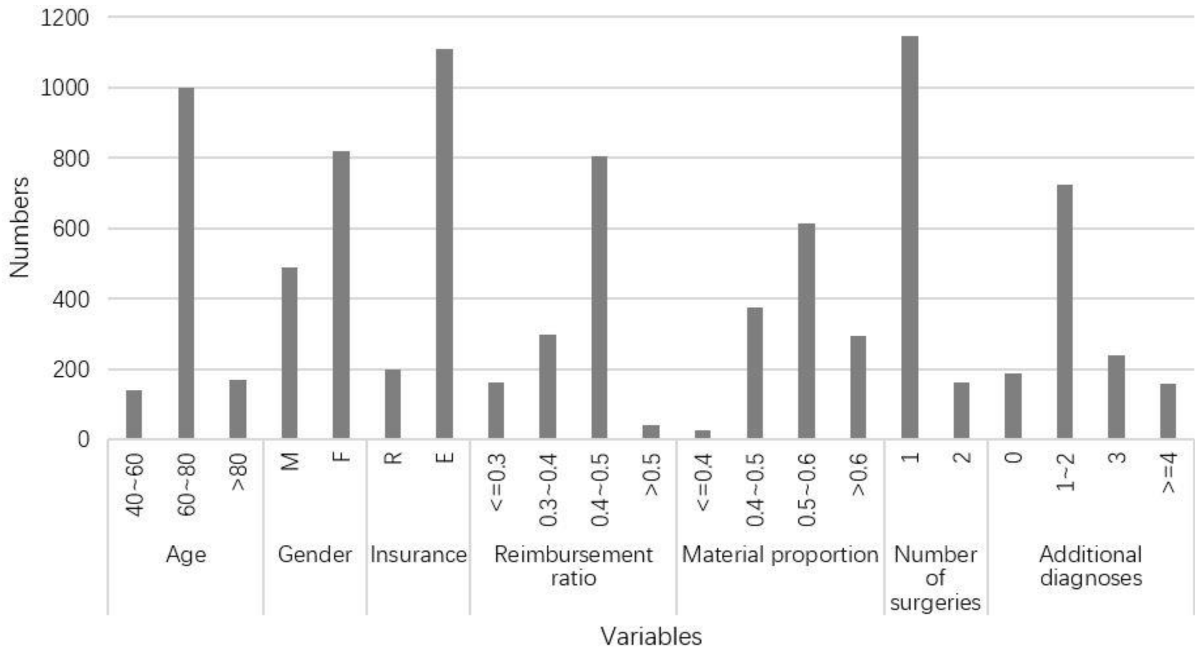


***Figure 1.*** *Histogram of variables.*

As shown by Table 1 and Figure 1, among the 1310 patients, the majority of senile cataract patients undergoing phacoemulsification combined with IOL implantation were aged 60-80 years, accounting for 76%. At the same time, the P value obtained by the Kruskal-Wallis H test indicates that age has no significant influence on total hospitalization costs. There is no significant difference between male and female in hospitalization days and total costs.

There are two types of medical insurance in this sample, namely URRBMI and UEBMI, among which UEBMI account for 85%. It is found that the average total hospitalization cost of employee insured patients is higher than that of resident insured patients, and the P value of the medical insurance type for total hospitalization cost is 0.005, less than 0.05, indicating that a significant difference between the two types of medical insurance can be inferred. The median total medical coats of patients with UEBMI insurance are 327.09 RMB higher than those with URRBMI insurance. We believe this is due to the relatively higher incomes of patients with UEBMI insurance, who typically choose more expensive treatment options and therefore incur higher total costs.

Similarly, from the perspective of reimbursement ratio, there are significant differences in hospitalization costs of patients with different reimbursement ratio. The total hospitalization costs of patients with a reimbursement ratio less than or equal to 0.3 is nearly twice than that of other patients. For the Material proportion factor, the P value of the corresponding test for the total hospitalization costs is less than 0.001, which can be regarded as a significant correlation. As for the number of surgeries, the average length of hospitalization days for patients undergoing monocular surgery was 1.23, while the average length of hospitalization days for patients undergoing binocular surgery was 5.62. The average total hospitalization cost of patients undergoing binocular surgery was twice than that of patients undergoing monocular surgery, with a P value less than 0.001, showing a significant relationship between the number of surgeries and total cost. As the number of additional diagnoses increased, the total hospitalization costs decreased, and the P value was less than 0.001. However, the P value of the corresponding test between the number of additional diagnoses and the hospitalization days was 0.432. As the cost of materials and the cost of surgery accounted for 55.0% and 23.5% of the total costs on average, the correlation between the materials fee and other variables, the correlation between the surgery fee and other variables can be explored, as shown in Table 2 for further analysis.

***Table 2.*** *Spearman correlation analysis results of various variables with material fee and surgery fee.*

| **Variable** | **materials fee** | | **surgery fee** | |
|---|---|---|---|---|
| | **correlation coefficient** | **P value** | **correlation coefficient** | **P value** |
| Age | -0.010 | 0.726 | 0.072 | 0.009 |
| Gender | -0.060 | 0.031 | -0.066 | 0.018 |
| Hospitalization days | 0.437 | <0.001 | 0.637 | <0.001 |
| Reimbursement ratio | -0.353 | <0.001 | 0.026 | 0.346 |
| *Insurance* | *0.087* | *0.002* | *0.034* | *0.217* |
| Number of surgeries | 0.467 | <0.001 | 0.830 | <0.001 |
| Material proportion | 0.308 | <0.001 | -0.028 | 0.317 |
| Additional diagnoses | -0.140 | <0.001 | -0.043 | 0.117 |

According to Table 2, in the 1310 senile cataract patients, the P values of Spearman correlation tests of all variables except for age and material cost were all less than 0.05, indicating that these variables were significantly correlated with materials fee by univariate analysis. The P values of the corresponding tests for reimbursement ratio, type of medical insurance, the material proportion, number of additional diagnoses on surgery fee were greater than 0.05. It can be found that the medical insurance type is significantly correlated with the materials fee, but not with the surgery fee.

## 2.3. Introduction to Statistical Models

To obtain more robust analysis, multivariate robust multivariate regression model and generalized additive model (semi-parametric regression model) are used to analyze the sample.

A brief introduction to multivariate linear regression models is conducted as following. $Y$ denotes response variable and $X = (X_1, X_2, \cdots, X_p)$ of $p$ dimension stands for explanatory variables, where $\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)^T$ is the coefficient vector to be estimated, $\varepsilon$ is a random error. The multivariate linear regression model is

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \quad (1)$$

The assumptions of multivariate linear regression using ordinary least squares (OLS) estimation are: $\varepsilon$ follows a normal distribution, homoscedasticity and uncorrelated. Due to the large sample size in this paper, according to the large sample theory, we only need to check whether the model meets the Gauss-Markov assumption, that is, whether the residuals are homoscedasticity and uncorrelated.

Robust regression is one of the robust methods in statistics, which can be used in medical data research [27, 28]. Since the OLS is very sensitive to outliers, the classical least-squares loss function of OLS is modified in robust regression, which can be used to fit the structures in many data sets. In addition, robust regression can identify potential strong influence points, outliers and the corresponding structures when model assumption unsatisfied. When the error follows normal distribution, the estimation of robust regression is almost as good as OLS. But, the results of robust regression are better than the OLS when the Gauss-Markov assumption is not satisfied. When sample data contain outliers, the residuals estimated by OLS are not normally distributed and tend to be heavy-tailed. Considering the sum of squares of residuals, the dominating role of outliers is greatly increased, resulting in poor accuracy of estimation. Therefore, the robust regression model is used in this paper. Total hospitalization costs are highly right-skewed, so they are log transformed.

M-estimation was proposed by Huber, widely used for robust regression, it is also called generalized maximum likelihood estimation [29]. When applied to the multivariate regression model, the object function to be minimized is as follows,

$$\hat{\beta}_M = argmin \sum_{i=1}^{n} \rho(Y - X\beta) = argmin \sum_{i=1}^{n} \rho(\varepsilon).$$

where $\rho(t)$ is the loss function, which satisfies the following assumptions,

$\rho(t)$ is a smooth positive real function;

when $t \geq 0$, $\rho(t)$ does not decrease;

$$\rho(0) = 0, max\, \rho(t) = 1.$$

Obviously, when $\rho(\varepsilon) = \varepsilon^2$, the M-estimation is just OLS, so the OLS estimation is a special case of M-estimation.

OLS estimation and M-estimation were respectively used to establish the model (1). Combined with the results presented below, the model established by M-estimation has better stability than OLS and is generally not affected by outliers.

However, model (1) only describes the linear relationship between $X$ and $Y$, and cannot explain whether there is a nonlinear relationship with unknown function. Therefore, in order to better establish the model between $X$ and $Y$, this paper will build a generalized additive model (GAM) by semi-parametric regression estimation. The model can contain both parametric components and non-parametric components. It can express the linear and nonlinear relationship between explanatory variables and response variables in details, reducing the estimation bias caused by linear settings [30]. Therefore, the GAM has a good effect when applied to medical data [31, 32]. The mathematical formula of GAM is shown as follows,

$$Y = \beta_0 + \sum_i \beta_i X_i + \sum_j f_j(X_j) + \varepsilon \quad (2)$$

Where, $X_i$ is the explanatory variable assumed to be linear, $\beta_i$ is the corresponding parameter, $f_j(\cdot)$ is the smoothing function corresponding to the explanatory variable $X_j$ with unknown form, and $\varepsilon$ is the random error term.

In summary, R language was used to establish multivariate regression models including classical regression model, robust regression model (M-estimation) and GAM, to analyze the influencing factors and their influencing degree on hospitalization costs of senile cataract patients. The following is the construction of dummies for qualitative variables to be adopted in the process of establishing multivariate regression models (1) and (2) in this paper.

***Table 3.*** *Dummy variable Settings.*

| | |
|---|---|
| Gender | F→0; M→ 1 |
| Insurance | R→0; E→ 1 |

As gender and medical insurance type are qualitative variables in this paper, these variables need to be transformed as dummies, as shown in Table 3.

After the transformation of dummy variables, the multivariate regression models can be fitted. Explanatory variables and response variables included in models (1) and (2) are described in Table 4.

***Table 4.*** *Description of variables used in multivariate regression models.*

| ***Y*** | **Log (Hospitalization costs)** |
|---|---|
| $X_1$ | Age |
| $X_2$ | Gender |
| $X_3$ | Hospitalization days |
| $X_4$ | Reimbursement ratio |
| $X_5$ | Insurance |
| $X_6$ | Number of surgeries |
| $X_7$ | Material proportion |
| $X_8$ | Additional diagnoses |

## 3. Results

To detect outliers in the sample, we first establish a general multivariate linear regression model by OLS and then use it to calculate the Cook's distance of all data points. As shown by Figure 2, sample points NO. 220, 662, 689, 790 and 1297 can be regarded as outliers. After eliminating these five outliers, we established a multivariate regression model based on OLS. In contrast, when modeling with M-estimation, there is no need to eliminate outliers because of its robustness.

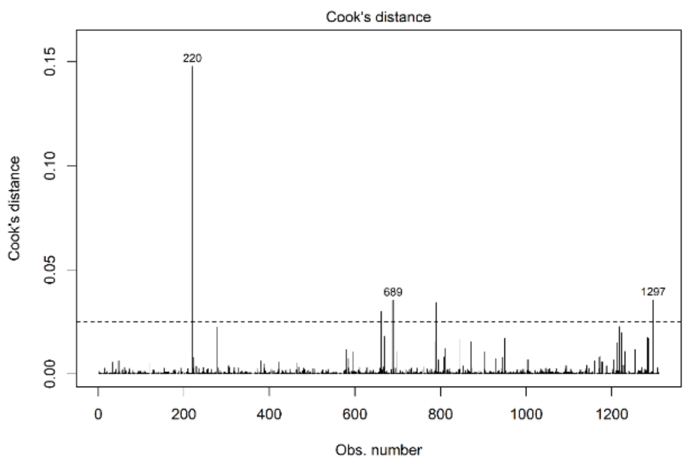


***Figure 2.*** *Cook's distance of all observation points.*

The results of two multivariate linear regression models are shown in Table 5.

***Table 5.*** *Results of classical and robust multivariate linear regression models.*

| | **Classical multivariate regression (OLS)** | | **Robust multivariate regression (M-estimation)** | |
|---|---|---|---|---|
| | **Estimator** | **P value** | **Estimator** | **P value** |
| Intercept item | 9.440 | <0.001 | 9.675 | <0.001 |
| Age | -0.0002 | 0.674 | -0.0004 | 0.548 |
| Gender | 0.011 | 0.250 | 0.011 | 0.258 |
| Hospitalization days | 0.068 | <0.001 | 0.062 | <0.001 |
| Reimbursement ratio | -2.298 | <0.001 | -2.510 | <0.001 |
| *Insurance* | *0.080* | *<0.001* | *0.076* | *<0.001* |
| Number of surgeries | 0.514 | <0.001 | 0.526 | <0.001 |
| Material proportion | -0.132 | 0.277 | -0.388 | 0.002 |
| Additional diagnoses | -0.022 | <0.001 | -0.022 | <0.001 |
| AIC | -970.366 | | -921.324 | |
| Sd(res) | 0.167 | | 0.169 | |
| MSE | 0.027 | | 0.029 | |
| MAE | 0.138 | | 0.137 | |

Remark: AIC is short for Akaike information criterion; Sd(res) means the standard deviation of residuals; MSE represents mean square error; MAE stands for mean absolute error.

In modeling, the classical multivariate linear regression uses the data stripped of the outliers, while the robust multivariate linear regression uses all data. We will further examine the significance of the material proportion in GAM.

For the classical model obtained by OLS method, Spearman correlation test is used for homoscedasticity assumption. Considering the Spearman correlation tests for the absolute values of residuals and $X_i(i = 1, 2, \cdots, 8)$, we find that the hospitalization days, reimbursement ratio, number of surgeries, and material proportion are significantly correlated with the absolute values of residuals with P values far less than 0.001, so the assumption of homoscedasticity is much likely to be unsatisfied. In contrast, the robust multivariate regression does not need to meet the assumption, and its application range is usually much wider.

Next, we discuss explanatory variables who reject the null hypothesis. The hospitalization days is positively correlated with the total costs. When other variables are fixed, the hospitalization costs increase by 6.2% for each additional day of stay. There is a negative correlation between the reimbursement ratio and the hospitalization costs. The number of surgeries is positively correlated with the total costs. There is a negative correlation relationship between the material proportion and the total costs of senile cataract patients. As shown in Table 5, the regression coefficient for the number of additional diagnoses is -0.022, showing a negative correlation relationship between the total costs.

The results above mainly show the relationships of factors and covariates to hospitalization costs by linear mode. Next, in the process of GAM, we selected significant variables $X_i(i = 1, 2, \cdots, 8)$ as the nonlinear part of GAM. To check the existence of nonlinear relationship, the Spearman correlation test is used for the residuals and $X_i$, the reimbursement ratio and the material proportion show significant results, with P values far less than 0.001. Therefore, in the establishment of GAM, the reimbursement ratio and material proportion are taken as nonlinear parts, and the nonlinear part is fitted with regression spline basis. Other variables are still set as linear. In short, the results of GAM include not only the coefficient estimation of linear regression and the significance test results of the estimators, but also the hypothesis test results of whether some nonlinear relationships exist or not. The results of GAM are shown in Table 6. Figure 3 is the smooth curve output after the spline smoothing method is used to fit the nonlinear variables.

***Table 6**. GAM fitting results.*

| **Parameter estimation of GAM parameter section** | | |
|---|---|---|
| **Variable** | **Estimator** | **P value** |
| Intercept item | 8.361 | <0.001 |
| Age | 0.001 | 0.028 |
| Gender | -0.002 | 0.699 |
| Hospitalization days | 0.036 | <0.001 |
| *Insurance* | *0.037* | *<0.001* |
| Number of surgeries | 0.580 | <0.001 |
| Additional diagnoses | -0.007 | 0.002 |
| Results of the non-parametric part of GAM | | |
| Smooth component | Df | P value |
| Reimbursement ratio | 8.311 | <0.001 |
| Material proportion | 8.572 | <0.001 |
| AIC: -2330.992 | GCV: 0.010 | Sd(res): 0.097 |
| MSE: 0.009 | MAE: 0.065 | |

Remark: AIC is short for Akaike information criterion; GCV stands for generalized cross validation; Sd(res) means the standard deviation of residuals; MSE represents mean square error; MAE stands for mean absolute error.

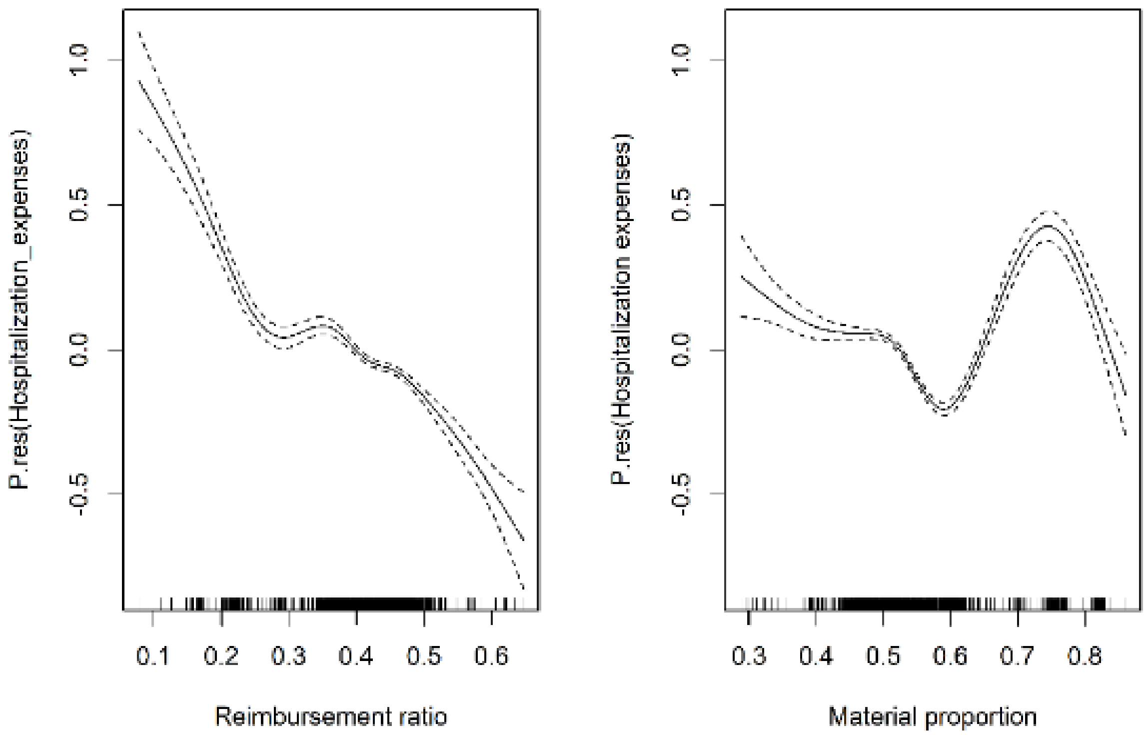


***Figure 3**. Smooth curves.*

The dashed lines indicate the 95% confidence interval of the fitted values, and the estimated curves of reimbursement ratio and material proportion are shown in the Figure 3.

## 4. Discussion

According to the results of classical and robust multivariate linear regression models, classical multivariate linear regression and robust multivariate linear regression have the close standard deviation of residuals, mean square error and mean absolute error. In the case that the outliers are not deleted, the model evaluation indexes of the model built by the robust method are similar to those built by the classical method, indicating that the robust method is resistant to outliers. There is a small difference between the model results of the two methods. The main difference lies in the significance of the material proportion, which is considered insignificant by OLS and significant by M-estimation.

According to the results of robust multivariate linear regression in Table 5, the P value for $\widehat{\beta_1}$ (estimator of $\beta_1$ for variable age) is 0.548, much higher than 0.05 and 0.1, indicating that the null hypothesis has not been rejected, thus variable age has no significant correlation with the total costs. The P value for $\widehat{\beta_2}$ (estimator of $\beta_2$ for variable gender) is 0.258, indicating that the null hypothesis has not been rejected and there is no significant correlation between gender and the total costs.

According to the robust multivariate regression model, with other variables unchanged, the hospitalization costs of patients with UEBMI insurance are 7.6% higher than those with URRBMI insurance. It indicates that the current two types of medical insurance in Chengdu have significant correlation with the total costs of senile cataract.

The results of GAM show that gender is not significant, while age, hospitalization days, type of medical insurance, number of surgeries, number of additional diagnoses, reimbursement ratio and material proportion are all significantly correlated with the hospitalization costs, among which reimbursement ratio and material proportion show a curving relationship with hospitalization costs. The reimbursement ratio can mainly be divided into three stages, with a rapid decreasing trend when it is less than 0.25. Between 0.25 and 0.4, it is a fluctuating state. When it is above 0.4, there is also a rapid decreasing trend. The material proportion can also be divided into three sections. When it is less than 0.6, there is a downward trend. When it is within (0.6, 0.76), there is an upward trend. When it is greater than 0.76, it shows a rapid downward trend. Meanwhile, it is found that the AIC value (-2330.992), standard deviation of residual (0.097), mean square error (0.009) and mean absolute error (0.065) of the model are far better than the results of the above two multivariate linear regression models. Because GAM can express the nonlinear relationship of explanatory variables, factors and covariates to response variable. It is worth noting that the GAM results consider the material proportion as a significant non-parametric variable. Therefore, the results of the above two types of multivariate linear regression on the material proportion further demonstrate the robustness of the robust regression method.

## 5. Conclusions

Based on robust analysis, we studied the influencing factors of hospitalization costs of senile cataract patients with different insurance types in Chengdu. The following conclusions and suggestions are presented: The cost of materials, such as intraocular lenses and surgical instruments, are the key factor influencing total hospitalization costs. It is suggested that doctors should choose surgical products in a more comprehensive and reasonable way, and expect the new technology to reduce the cost of intraocular lens. There is still an imbalance in medical resources. Therefore, it is suggested that the government continue to reform the content of medical insurance for urban and rural residents, and make efforts to allow low-income groups to enjoy more high-quality medical services without increasing the medical burden, so as to improve the welfare level of these groups and alleviate the imbalance of medical resources. In addition, the reimbursement ratio, the number of surgeries, hospitalization days, number of additional diagnoses and material proportion are key influencing factors to the hospitalization costs of senile cataract patients. In order to reduce the economic burden of senile cataract treatment, hospitals can use more reasonable surgical supplies and reduce the hospitalization days of patients while ensuring the surgical effect. The relevant local health insurance institutions may consider adding more specific types of health insurance for the elderly. In summary, this paper conducted a robust analysis to study the influencing factors of hospitalization costs for senile cataract patients, with emphasis on the impact of medical insurance type. The sample used was from Chengdu, one of the earliest pilot cities for urban-rural health insurance integration. The conclusions can provide reference for the management and payment of hospitalization expenses for senile cataract in the future.

## Abbreviations

UEBMI: Urban Employee Basic Medical Insurance
NRCMS: New Rural Cooperative Medical Care System
URBMI: Urban Resident Basic Medical Insurance
URRBMI: Urban and Rural Residents' Basic Medical Insurance
WHO: World Health Organization
GAM: generalized additive Model
OLS: ordinary least squares
IOL: intraocular lens
AIC: Akaike information criterion
Sd: standard deviation
MSE: mean square error
MAE: mean absolute error

## Conflicts of Interest

The authors declare no conflicts of interest.

## References


[1] Yip WCM, Hsiao WC, Chen W, et al. Early appraisal of China's huge and complex health-care reforms. Lancet. 2012; 379(9818): 833-42. https://doi.org/10.1016/S0140-6736(11)61880-1

[2] You XD, Kobayashi Y. The new cooperative medical scheme in China. Health Policy. 2008; 91: 1-9. https://doi.org/10.1016/j.healthpol.2008.11.012

[3] Lin W, Liu GG, Chen G. The Urban Resident Basic Medical Insurance: A landmark reform towards universal coverage in China. Health Economics. 2009; 18: S83-S96. https://doi.org/10.1002/hec.1500

[4] Fan X, Su M, Si Y, et al. The benefits of an integrated social medical insurance for health services utilization in rural China: evidence from the China health and retirement longitudinal study. International Journal for Equity in Health. 2021; 20: 126. https://doi.org/10.1186/s12939-021-01457-8

[5] Xue ZX, Li BW. The Impact of the Integration of Urban and Rural Medical Insurance on Migrant Workers' Overwork: Evidence From China. Frontiers in Public Health. 2022; 10: 934524. https://doi.org/10.3389/fpubh.2022.934524

[6] Reis T, Lansingh V, Ramke J, et al. Cataract as a Cause of Blindness and Vision Impairment in Latin America: Progress Made and Challenges Beyond 2020. American Journal of Ophthalmology. 2021; 225: 1-10. https://doi.org/10.1016/j.ajo.2020.12.022

[7] To KG, Meuleners LB, Fraser ML, et al. The impact of cataract surgery on vision-related quality of life for bilateral cataract patients in Ho Chi Minh City, Vietnam: a prospective study. Health and Quality of Life Outcomes. 2014; 12: 16. https://doi.org/10.1186/1477-7525-12-16

[8] Khairallah M, Kahloun R, Bourne R, et al. Number of People Blind or Visually Impaired by Cataract Worldwide and in World Regions, 1990 to 2010. Investigative Ophthalmology & Visual Science. 2015; 56(11): 6762-6769. https://doi.org/10.1167/iovs.15-17201

[9] GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392(10159): 1859-1922. https://doi.org/10.1016/S0140-6736(16)31678-6

[10] World Health Organization (2019). World Report on Vision. Available at: https://www.who.int/publications/i/item/9789241516570

[11] Nangia V, Jonas JB, George R, et al. Prevalence and causes of blindness and vision impairment: magnitude, temporal trends and projections in South and Central Asia. The British Journal of Ophthalmology. 2019; 103(7): 871-877. https://doi.org/10.1136/bjophthalmol-2018-312292

[12] Zhu M, Yu J, Zhang J, et al. Evaluating vision-related quality of life in preoperative age-related cataract patients and analyzing its influencing factors in China: a cross-sectional study. BMC Ophthalmology. 2015; 15: 160. https://doi.org/10.1186/s12886-015-0150-8

[13] Xiao BX, Yi JL, Limburg H, et al. The prevalence of blindness, visual impairment and cataract surgery in Tuoketuo and Shangdu Counties, Inner Mongolia, China. Open Journal of Ophthalmology. 2015; 5(1): 23-30. https://doi.org/10.4236/ojoph.2015.51005

[14] Zheng YF, Huang WY, Xiao BX, et al. Preferences for participation in shared decision-making among cataract patients in urban southern China: a cross-sectional study. Lancet. 2016; 388(1): S56.

[15] Du YF, Liu HR, Zhang Y, et al. Prevalence of cataract and cataract surgery in urban and rural Chinese populations over 50 years old: a systematic review and Meta-analysis. International Journal of Ophthalmology. 2022; 15(1): 141-149. https://doi.org/10.18240/ijo.2022.01.21

[16] Yao K and Wang W. A summary of 70 years of cataract diagnosis and treatment in China. Chinese Journal of Ophthalmology. 2020; 56(5): 321-324. https://doi.org/10.3760/cma.j.cn112142-20200116-00026

[17] Khan A, Amitava AK, Rizvi SA, et al. Cost-effectiveness analysis should continually assess competing health care options especially in high volume environments like cataract surgery. Indian Journal of Ophthalmology. 2015; 63(6): 496-500. https://doi.org/10.4103/0301-4738.162600

[18] Deng Y, Yang D, Yu JM, et al. The Association of Socioeconomic Status with the Burden of Cataract-related Blindness and the Effect of Ultraviolet Radiation Exposure: An Ecological Study. Biomedical and Environmental Sciences. 2021; 34(2): 101-109. https://doi.org/10.3967/bes2021.015

[19] Ko F, Frick KD, Tzu J, et al. Willingness to pay for potential enhancements to a low-cost cataract surgical package in rural southern China. Acta Ophthalmologica. 2011; 90: e54-e60. https://doi.org/10.1111/j.1755-3768.2011.02207.x

[20] Tan H, Xie L, Wang Y, et al. Examine the Burden of Illness of US Medicare Patients Diagnosed with Cataract. Value in Health. 2016; 19: A123-A124. https://doi.org/10.1016/j.jval.2016.03.501

[21] Bhuiyan MU, Luby SP, Alamgir NI, et al. Costs of hospitalization with respiratory syncytial virus illness among children aged <5 years and the financial impact on households in Bangladesh, 2010. Journal of Global Health. 2017; 7(1): 010412. https://doi.org/10.7189/jogh.07.010412

[22] Luo AJ, Chang WF, Xin ZR, et al. Diagnosis related group grouping study of senile cataract patients based on E-CHAID algorithm. International Journal of Ophthalmology. 2018; 11(2): 308-313. https://doi.org/10.18240/ijo.2018.02.21

[23] Xu F, Liu Y, Wheaton AG, et al. Trends and Factors Associated with Hospitalization Costs for Inflammatory Bowel Disease in the United States. Applied Health Economics and Health Policy. 2019; 17(1): 77-91. https://doi.org/10.1007/s40258-018-0432-4

[24] Urbich M, Globe G, Pantiri K, et al. A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014–2020). PharmacoEconomics. 2020; 38(11): 1219-1236. https://doi.org/10.1007/s40273-020-00952-0

[25] Rodriguez-Martinez CE, Sossa-Briceño MP, Castro-Rodriguez JA. Direct medical costs of RSV-related bronchiolitis hospitalizations in a middle-income tropical country. Allergologia et Immunopathologia. 2020; 48(1): 56-61. https://doi.org/10.1016/j.aller.2019.04.004

[26] Gogate PM, Deshpande M and Wormald RP. Is manual small incision cataract surgery affordable in the developing countries? A cost comparison with extracapsular cataract extraction. The British Journal of Ophthalmology. 2003; 87(7): 843-846. https://doi.org/10.1136/bjo.87.7.843

[27] Fu L, Wang Y and Cai F. A working likelihood approach for robust regression. Statistical Methods in Medical Research. 2020; 29(12): 3641-3652. https://doi.org/10.1177/0962280220936310

[28] Yang T, Gallagher C and McMahan CS. A robust regression methodology via M-estimation. Communications in Statistics - Theory and Methods. 2018; 48(5): 1-16. https://doi.org/10.1080/03610926.2018.1423698

[29] Huber PJ. Robust Regression: Asymptotics, Conjectures and Monte Carlo. The Annals of Statistics. 1973; 1(5): 799-821. https://doi.org/10.1214/aos/1176342503

[30] Hastie T and Tibshirani R. Generalized Additive Models. Statistical Science. 1986; 1(3): 297-318. https://doi.org/10.1214/ss/1177013604

[31] Yang J, Zhang M, Chen Y, et al. A study on the relationship between air pollution and pulmonary tuberculosis based on the general additive model in Wulumuqi, China. International Journal of Infectious Diseases. 2020; 96: 42-47. https://doi.org/10.1016/j.ijid.2020.03.032

[32] Dastoorpoor M, Khanjani N, Moradgholi A, et al. Prenatal exposure to ambient air pollution and adverse pregnancy outcomes in Ahvaz, Iran: a generalized additive model. International Archives of Occupational and Environmental Health. 2021; 94(2): 309-324. https://doi.org/10.1007/s00420-022-01899-9