Zhang et al. Archives of Public Health (2024) 82:195
https://doi.org/10.1186/s13690-024-01386-2

Archives of Public Health

RESEARCH

Open Access

A measurement study of the environmental quality and medical expenditures of elderly individuals: causal inference based on machine learning

Yu Zhang[1], Sheng Chen[2*] and Dewen Liu[1]

Abstract

Background The global surge of environmental pollution exacerbates health issues, disease incidence, and economic strain. In China, the increasing healthcare costs of the elderly population necessitate addressing this challenge as part of the "Healthy China" strategy. We explore the impact of environmental quality on elderly healthcare expenses.

Methods This study devised a comprehensive environmental quality index for 30 Chinese provinces, excluding Tibet, which was correlated with medical expenses for individuals older than 60 years, using China Family Panel Studies (CFPS) data. Because the traditional econometric model cannot solve the endogeneity problem and the selection of instrumental variables is subjective, a new machine learning algorithm is adopted based on the traditional ordinary least squares (OLS) model and the fixed effect model to conduct causal analysis to ensure the reliability of the results. Finally, heterogeneity analysis was conducted based on the generalized random forest algorithm.

Results Southern provinces such as Jiangxi and Guangxi exhibited superior environmental qualities. A regional analysis revealed a gradient where environmental quality decreased from west to east and from south to north. Both conventional and machine learning methodologies underscored a pivotal finding: enhanced environmental qualities significantly curtail elderly healthcare expenses. A heterogeneity assessment revealed that such improvements predominantly benefit elderly people in the eastern and central regions, with marginal impacts in the west. For different groups, the improvement of environmental quality can significantly reduce the medical expenditure of people aged 60 to 75, with bedtime hours between 9 and 11 PM and a lower household income.

Conclusions This study, employing machine learning and traditional models, demonstrates that enhancements in environmental quality significantly reduce medical costs for the elderly in China, especially in the eastern and central regions, and among demographics such as individuals aged 60–75 and low-income households. These findings underscore the potential of environmental policies to lower medical costs within the "Healthy China" initiative framework. However, the study's scope is limited by the environmental quality index and the extent of data coverage, indicating a need for further research expansion.

Keywords Environmental quality, Medical expenditure, Entropy weights-TOPSIS method, Generalized random forests

*Correspondence:
Sheng Chen
rainy0139@163.com
Full list of author information is available at the end of the article

Text box 1. Contributions to the literature

- The study advances public health research by integrating machine learning with traditional econometric methods, offering deeper insights into the causal effects of environmental quality on health outcomes.
- It highlights regional disparities in environmental health impacts, emphasizing the need for localized public health strategies tailored to specific geographic areas.
- The research underscores the importance of health policy interventions in mitigating the negative health effects of environmental degradation, providing valuable guidance for developing effective public health policies.

Introduction

In recent years, environmental pollution has significantly impacted people's health and lives worldwide. On January 9, 2021, China's Ministry of Ecology and Environment released the "Guidance on Integrating and Strengthening Efforts to Address Climate Change and Ecological and Environmental Protection", which indicates that during the 14th Five-Year Plan period, environmental and ecological prevention and control will be considered in an integrated manner, combining the prevention and control of pollution with the region's climate and economy. Given the growing concerns about environmental issues, numerous studies have focused on this area, such as Zhou et al. [1] from the perspective of cardiovascular diseases, Giovanis and Ozdamar [2] from the perspective of mental health and Wei and Zhou [3] from the perspective of the willingness of talent mobility to study the impact of environmental quality, especially air pollution, and the impact of environmental quality on health care spending-related literature on the impact on medical expenditure is relatively rare. Individual medical cost expenditure is one of the main costs caused by environmental quality, and for elderly people without a stable source of income, an increase in medical cost expenditure will decrease individuals' willingness to spend on other consumption expenditures; in particular, the cost of major illnesses due to environmental factors may even affect individuals' standard of living.

China's National Health Plan for the Fourteenth Five-Year Plan states that preventive health care for elderly people should be strengthened, and cooperation mechanisms between medical and health care and elderly care organizations should be improved to promote the health of elderly people. A full understanding of the extent to which environmental quality, as well as the physical quality and living habits of elderly people, affects their medical expenditures will help us understand the economic costs of environmental quality for elderly people and will also provide constructive suggestions for policy formulation and implementation to improve environmental quality. The data used come from the China Family Panel Studies (CFPS), the Economy Prediction System (EPS) database, and the China Environmental Statistics Yearbook (CESY), from which data on environmental quality indicators are filtered out, and data on medical expenditure of the elderly and related population health indicators are filtered out from the CFPS. The difficulty of the study lies in two aspects: first, choosing appropriate indicators to measure environmental quality. The indicators of environmental quality are not only air quality but also urban noise pollution and water pollution. There are two main sources of data on environmental quality indicators collected in this paper: one is from databases such as China Stock Market & Accounting Research Database (CSMAR) and the EPS. The other is the data from China's provincial statistical bureaus, which are relatively objective and can be matched with the data from the CFPS. The second is endogeneity. Most of the previous studies by scholars on the effects of changes in environmental quality on the medical cost expenditure of elderly people have used the instrumental variable approach, but the selection of instrumental variables often involves a large number of subjective factors; therefore, we used the generalized random forest algorithm in machine learning for causal identification to ensure that the results were reliable, and the algorithm was used to conduct further analysis of heterogeneity.

We match the data on gender, age, number of exercise sessions, number of pension insurance sessions, etc., of elderly respondents in the CFPS database with the air quality, green area, and urban environmental noise pollution of the living environment of elderly respondents in 2014, 2016, 2018, and 2020 and construct an empirical framework to explore the impact of environmental quality indicators on the medical expenditure of elderly respondents. In the long run, improvements in environmental quality can significantly reduce the cost of medical expenditures for elderly people, and younger people who have better living habits and lower incomes benefit more from improvements in environmental quality; in other words, improvements in environmental quality can significantly reduce the cost of medical expenditures for this group of elderly people.

Environmental quality, population health and medical expenditures

Previous literature has shown that poor environmental quality has a significant negative impact on human health. Specifically, poor environmental quality, especially when air pollution is severe, can lead to an increase in neonatal mortality. Arceo et al. [4], in his study of the effects of air pollution on human health, found that an increase in the concentration of carbon monoxide and PM10 can increase the neonatal mortality rate. Manisalidis et al. [5] studied the extent of the negative effects of air pollution on human systems and found that the nervous system and respiratory system in the human body are affected by air pollution. In a study on the health of older people in the U.S.A., Deryugina et al. [6] reported that 25% of elderly people exposed to severe air pollution were in relatively poor health. Yang and Liu [7] used data from approximately 16,000 households distributed in 25 provinces in the CFPS to analyse the impact of air quality on the health of the population, and the study showed that an increase in industrial emissions can threaten the health of the population; for example, the incidence of respiratory illnesses and the rate of visits to doctors increase with increasing industrial emissions. Di et al. [8] examined the effects of short-term exposure to ambient PM2.5 and warm-season ozone on mortality in the U.S. Medicare population aged 65 years and older and showed that exposure to PM2.5 and warm-season ozone was significantly associated with increased mortality, even at levels below the current National Air Quality Standards.

Environmental quality problems not only pose a threat to human health but also increase residents' medical expenditures, which increases the burden on residents' lives. Preker et al. [9] provide a preliminary estimate of overall quantifiable medical expenditures due to anthropogenic pollution and show that the share of medical expenditures for pollution-related diseases is substantial in very low-income countries. Akpalu and Normanyo [10] used the utility assessment method to assess the impact of carbon dioxide emissions on local medical expenditure in Ghana, Africa, and the results of the study showed that a decrease in the quality of the environment would increase the medical cost expenditure of the population. Yang and Zhang [11] examined the impact of air pollution exposure on household medical expenditures in China using the Urban Household Survey (UHS) database. The results showed that every 1% increase in annual exposure to fine particulate matter (PM2.5) correspondingly led to a 2.942% increase in household medical expenditure. Aydin and Bozatli [12] explored the interactions between environmental quality and medical expenditures in five North African countries over the period of 1990–2019, considering key socioeconomic variables as determinants, with results showing an increasing trend in medical expenditures, highlighting the potential threat of carbon dioxide emissions to healthcare costs, and highlighting synergistic relationships between healthcare, political stability and economic development.

Although existing literature generally focuses on the direct connection between environmental quality and health status, these studies often utilize cross-sectional data, which makes it difficult to address the long-term effects of environmental changes on health. This research introduces machine learning methods to analyze the data, allowing for a more accurate capture of the long-term impacts of environmental quality changes on the medical expenditures of the elderly. This provides a new perspective for analyzing how improvements in environmental quality can potentially reduce medical expenditures and enhance the health of the elderly, enhancing the timeliness of the research and the reliability of causal inference.

Mechanism of the impact of environmental quality on medical expenditures of the elderly

Environmental quality has a direct impact on the population's medical expenditures by influencing health status. When the quality of the environment is poor, the health of the population is jeopardized, making them more susceptible to a wide range of illnesses and, in severe cases, reducing life expectancy. Huang et al. [13] used a log-linear exposure response equation to analyse the relationship between environmental quality and health in the Pearl River Delta region, and the final results showed that particulate pollutants in the Pearl River Delta in 2006 had a serious impact on the health of local residents. Liao et al. [14] quantified the impact of air pollution on the healthcare utilization and healthcare costs of Chinese residents and showed that PM2.5 significantly increased people's healthcare costs and out-of-pocket expenses. Zhang and Zhang [15] constructed a PMG-ARDL model to analyse the relationship between toxic air pollution and healthcare expenditures within the EU region, showing that air pollution has a significant negative impact on affecting the health of the population and on healthcare expenditures.

With economic development, the more rapid the economic growth is, the more serious the environmental pollution will be; at the same time, the residents' income will increase steadily, which will improve the residents' lifestyle, and then the residents will increase their expenditure on medical expenditures even more. Lu et al. [16] investigated the integrated dynamic relationships among environmental quality, economic development and public health in China, and the findings confirmed the negative impact of environmental pollution on public health and revealed that economic and social factors also affect public health levels. Among European countries, Ireland is the country with the highest use of diesel engines, with 43.57%; therefore, Dey et al. [17] studied the impact of diesel transportation in Dublin on environmental quality and residents' health, and the results showed that the ban on diesel engines resulted in significant reductions in nitrogen oxides, sulfides and suspended particulate emissions, as well as savings of €43.8 million during the ban, and a significant boost to the local economy and consequently to public health. Li et al. [18] studied the impact of pollution-intensive industrial agglomeration on residents' health expenditures and found that pollution-intensive industrial agglomeration both increases and decreases health expenditures and that the influencing factors include the environmental effect, the economic effect, and the public service effect. Yacour et al. [19] focused their research on North African countries, using panel cointegration analysis to explore the interactions between environmental quality, economic growth, and medical expenditures.

In summary, environmental quality further affects residents' physical health and the regional economy by influencing their medical expenditures. As shown in Fig. 1, environmental quality affects medical expenditure in direct and indirect ways. On the one hand, elderly people have poorer immunity and physical conditions, and poor environmental conditions threaten the health of elderly people and increase medical expenditures. On the other hand, environmental quality affects the local economic situation and increases the income of residents, which in turn indirectly affects medical expenditures. On the other hand, environmental quality indirectly affects the medical expenditure of elderly people by influencing the local economic situation and increasing the income of residents.

To summarise previous studies, the main focus has been on the impact of environmental quality on overall public health, with fewer detailed analyses of populations such as the elderly and their healthcare expenditure. We explore in detail how environmental quality affects the health of the elderly and thereby influences their medical expenditures, offering a more nuanced analysis approach. Moreover, this research also considers regional and individual differences, making the results more specific and targeted, which aids in the formulation of more effective regional and demographic-specific health policies. Additionally, most studies have employed econometric models, which are not well-suited for handling high-dimensional data and are better at predicting outcomes for low-dimensional data. Furthermore, linear regression methods may face challenges in handling complex data, especially in accounting for individual differences in estimation results. Therefore, this study adopts a relatively novel approach—machine learning (specifically, Generalized Random Forests, GRF) for causal inference research, which naturally controls for unobserved confounding variables by using a large number of data features for non-parametric and automatic interaction selection. This method considers individual heterogeneity more fully than traditional linear regression methods, thus ensuring the reliability of the estimation results.

Machine learning and causal inference

With the development of big data and artificial intelligence, new machine learning methods have gradually been used in various studies and are considered to be able to effectively compensate for the shortcomings of traditional measurement methods [20]. Machine learning has powerful generalizability and explanatory ability, and now, other domestic and foreign scholars use machine learning to solve causal problems.

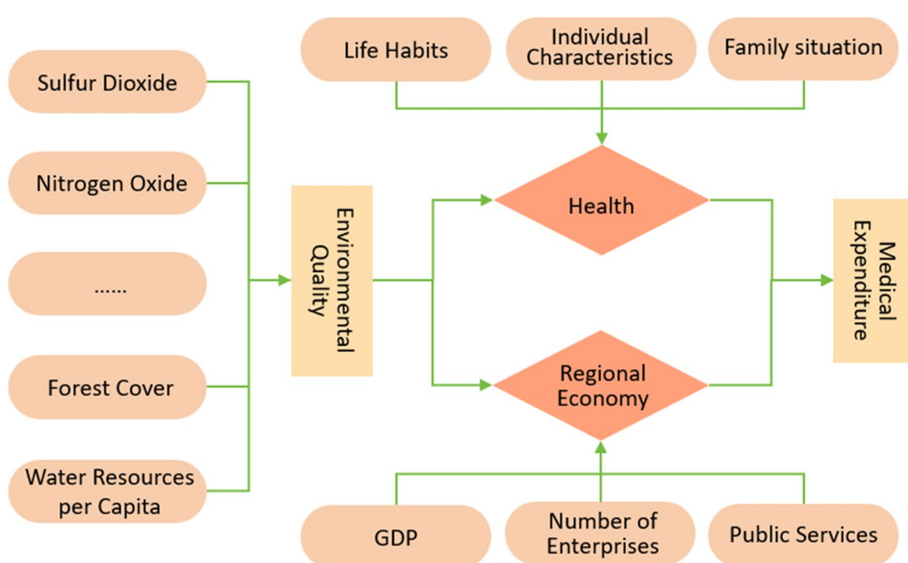

Fig. 1 Mechanism of the impact of environmental quality on medical expenditures in old age

Mullainathan [21] noted that machine learning algorithms can be used in the first stage of the two-stage regression of instrumental variables to solve the problem of weak instrumental variables. Kamińska [22] modelled the regression relationship between air pollutant concentrations and nine variables, such as meteorological conditions, temporal conditions, and traffic flow, using the random forest approach, with the fit of the model and the significance of each predictor variable varying by season. Hu et al. [23] used causal forests to analyse the relationship between "skewed" policies and regional development. Baiardi and Naghi [24] discussed the application of machine learning, particularly causal machine learning methods such as double machine learning and causal random forests, in economic research. They emphasized the nonparametric advantages of these methods in handling large numbers of covariates and high-dimensional data. The literature stresses that these methods are more capable of identifying and dealing with heterogeneity treatment effects in data compared to traditional methods [24].

Although machine learning techniques have been used to handle big data and perform causal inference in existing research, they have rarely been applied at the intersection of environmental science and health economics. By introducing the generalized random forest model, this paper not only improves the accuracy of model predictions but also enhances the explanatory power of causal relationships, making the research results more practically valuable. Additionally, although the generalized random forest model performs well in this study, its external validity in different contexts and populations may be questionable. Therefore, further validation is needed in other settings to ensure the generalizability of the model.

The main contributions of this paper are as follows: First, we have expanded the perspective of existing research, not only analyzing the impact of air quality on the medical expenses of the elderly but also considering multiple environmental factors such as land pollution and urban greening. By constructing a comprehensive environmental quality assessment system, we thoroughly analyzed these factors' impact on the medical expenditures of the elderly. Secondly, this study employed machine learning methods, specifically the generalized random forest algorithm, for causal inference. The introduction of this method not only improved the accuracy of the model's predictions but also enhanced our ability to handle complex data structures. Lastly, we took into account individual behaviors and regional differences, incorporating individual health habits and socioeconomic statuses as control variables in the model, making the research results more targeted and practically applicable.

Methods

Data

In this paper, we use three evaluation indices—the ecological environment, economic environment and social environment—as the explanatory variables of the model. The data are obtained from the CSMAR and EPS databases, and some missing data are filled in according to the data from the statistical bureaus of each province in China. According to the availability of data, comprehensive evaluation indices of the environmental quality of 30 provinces excluding Tibet in 2014, 2016, 2018 and 2020 were selected as the research objects.

The explanatory variables used medical expenditures from the CFPS data, and the covariates included 24 variables in three categories: living habits, household situation, and individual characteristics. The data come from the China Family Panel Studies (CFPS), which is a tracking survey of representative samples of villages, households, and family members across the country to reflect China's economic development and social changes. The microdata are selected from the 2014, 2016, 2018, and 2020 survey data of people older than 60 years. For handling partially missing data and outliers, we employed methods such as moving averages, ARIMA forecasting, and interpolation. These methods ensured the reliability and accuracy of our research results. Consequently, we confirmed a usable sample size of 18,862. Table 1 shows the descriptions of the microvariables and descriptive statistics.

Variable

Dependent variable

Medical Expenditure: The primary dependent variable of the study, the medical expenditure of the elderly, is measured through patients' medical insurance records and direct medical cost records, including the total of outpatient and hospitalization expenses.

Independent variable

Environmental Quality Index: The main explanatory variable in this study is the composite environmental quality index, assessed using various environmental monitoring parameters such as PM2.5, water quality levels, and noise levels. This index is developed through a comprehensive evaluation system of environmental quality, utilizing the entropy weight-TOPSIS method to calculate a composite index for each province as a proxy indicator.

Table 1 Description of microvariables and descriptive statistics (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Variable	Average value	Standard deviation	Maximum value	Minimum value
Medical Expenditures	4793.924	15,588.730	0.000	450,000.000
Environment	71.000	21.041	25.561	119.253
Household	0.369	0.483	0.000	1.000
Age	68.084	6.350	60.000	102.000
Sex	0.518	0.500	0.000	1.000
Receiving pension insurance or not	0.722	0.448	0.000	1.000
Per capita household income	21,805.460	56,992.560	0.000	1,858,333.000
Happiness	7.692	2.184	0.000	10.000
Trust in doctors	6.989	2.379	0.000	10.000
Degree of health	2.513	1.225	1.000	50.000
Whether have a chronic disease within 6 months	0.307	0.461	0.000	1.000
Satisfaction with the point of care	3.731	0.770	1.000	5.000
Level of medical care at the point of care	3.598	0.861	1.000	5.000
Frequency of exercise	4.138	3.683	0.000	50.000
Smoked cigarettes in the past month	0.287	0.452	0.000	1.000
Drinking alcohol 3 times a week in the past month	0.165	0.371	0.000	1.000
Lunch break	0.636	0.481	0.000	1.000
Length of lunch break	45.369	45.999	0.000	240.000
Sleeping time at night	21.046	2.533	0.000	24.000
Satisfaction with life	4.129	0.951	1.000	5.000
BMI	22.844	3.640	8.964	35.500
Hospitalization in the past year	0.210	0.408	0.000	1.000
Number of meals with family	6.095	2.231	0.000	7.000
Number of family members	3.812	2.131	1.000	21.000
Marital status	0.812	0.391	0.000	1.000
Intelligence level	4.788	1.447	1.000	7.000

Control variables

To ensure the accuracy of our study results, we have included the following control variables, covering 24 variables listed in Table 1 aside from medical expenses and environmental factors:

> Individual Characteristics: This includes demographic factors such as age, gender, education level, and residential status, as well as health-related variables like chronic disease status, past hospitalizations, and overall health condition. Psychological factors such as satisfaction with life, medical services, and trust in doctors are also included, as these have been shown to significantly correlate with health status and medical needs.
>
> Family Circumstances: Economic status indicators such as per capita family income and pension receipt, along with family dynamics such as the number of family members and frequency of meals shared with family, affect an individual's access to and affordability of medical services.
>
> Lifestyle Habits: Behaviors including smoking, drinking, frequency of physical activities, and sleep duration, which can impact an individual's health status.

Measurement of the comprehensive environmental quality assessment index

Comprehensive evaluation index system for environmental quality

Environmental problems urgently need to be solved. Domestic and foreign research organizations mostly use index evaluation methods to measure environmental quality, and the determination of an evaluation index is the basis of this method. In the selection of an evaluation index, it is necessary to consider whether the index can include all the core contents of environmental quality. Therefore, this paper reviews relevant literature and documents such as the National Environmental Protection Standards of the People's Republic of China and refers to the studies of Bernardo et al. [25], Mourhir et al. [26] and Li and Wang [27] and combines the constraints of benefit

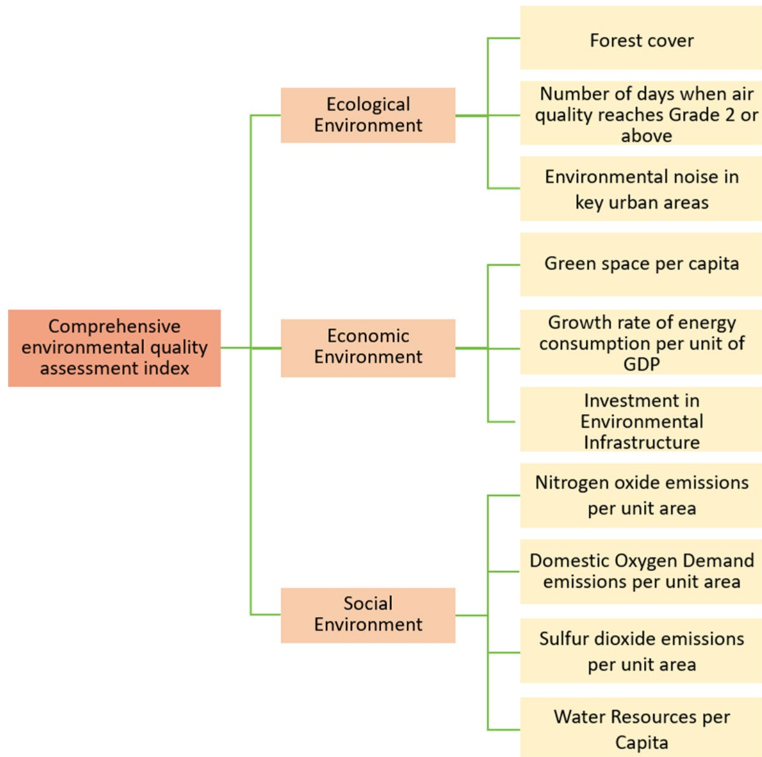

Fig. 2 Comprehensive environmental quality evaluation system

analysis, scientificity, practicability, and data availability to construct a comprehensive evaluation index system of environmental quality, which contains three criteria layers for ecological, economic, and social environments and 10 indicator layers, as shown in Fig. 2.

Next, the indicators are further explained. For the ecological environment, forest coverage is an important indicator of ecological balance and a reflection of abundant resources, representing the condition of the original ecological environment, while an air quality of Grade 2 or above indicates a low pollution index, with less PM2.5, PM10, etc., and the environmental noise in key urban areas reflects the liveability of the area; for the economic environment, the green space per capita reflects the green space in a certain period of time; for the economic environment, the green space per capita reflects the green space in a certain period of time. In terms of the economic environment, the green space per capita reflects the development of the urban economy in a certain period of time, and the greening difference between regions is gradually decreasing, while the growth rate of energy consumption per unit of GDP and the investment in environmental infrastructure reflect the level of energy consumption and the economic cost invested in environmental pollution. In terms of the social environment, the amount of water resources per capita, the amount of nitrogen oxides per unit of nitrogen oxides, the amount of domestic oxygen demand and the amount of sulfur dioxide emissions are chosen to measure the social governance capacity.

Entropy weight-TOPSIS method

Entropy weight method is an interpretation based on the basic principles of information theory, and it is also an objective assignment method to determine the weights according to the degree of change of each indicator, which fully demonstrates the information value of the data itself and avoids the possibility of human judgment and manipulation. The basic idea of entropy weight method is that the smaller the entropy of the evaluation index, the greater the degree of its discrete, the more information it contains, the greater the role in the comprehensive judgment system, so the corresponding weight is greater, and the TOPSIS method is a sorting method close to the ideal solution to evaluate the advantages and disadvantages of each index according to the size of the order. The entropy weight-TOPSIS method combines the two, weighting by entropy weighting method and sorting by TOPSIS method, so that the evaluation results have stronger objectivity.

Compared to other environmental quality assessment methods such as Principal Component Analysis (PCA), Analytic Hierarchy Process (AHP), and Data Envelopment Analysis (DEA), the entropy weight-TOPSIS method determines weights based on the amount of information inherent in the data itself, thus minimizing the influence of subjective factors. It comprehensively considers the impact of various indicators to produce relatively reasonable comprehensive evaluation results. Additionally, its calculation process is relatively simple, making it suitable for processing and analyzing large-scale data. Zou et al. [28] utilized the entropy weight-TOPSIS method to assess the progress of China's green energy consumption revolution. The study indicates that the entropy method can objectively determine the weights of various indicators, while the TOPSIS method is used to comprehensively evaluate the progress [28]. The specific steps are as follows:

Step 1: Construct the original decision matrix. Assuming that there are m provinces, and each province has n evaluation indicators, the value of evaluation indicator j for province i is x_{ij}.

$$A=(x_{ij})_{m\times n}, (i = 1, 2, ..., m, j = 1, 2, ..., n) \tag{1}$$

Step 2: Normalization of decision matrix. $x_{\min}$ is the minimum value of evaluation indicator j for province i, $x_{\max}$ is the maximum value of evaluation indicator j for province i.

The positive indicator is calculated by the formula:

$$Q_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

The negative indicator is calculated as:

$$Q_i = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \tag{3}$$

Then through the normalization process, all the indicators are converted to dimensionless, and the standardized decision matrix is obtained as $B(h_{ij})$. In the formula:

$$h_{ij} = x_{ij} / \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \tag{4}$$

h_{ij} denotes the quantitative value of evaluation indicator j for province i.

Step 3: Calculate the entropy value of evaluation indicators. m is the number of provinces in China, in this paper, m=30. n is the number of evaluation indicators, in this paper, n=10.

$$E_j = -\frac{1}{\ln(m \times n)} \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} \ln h_{ij} \tag{5}$$

When $h_{ij} = 0$, $\ln h_{ij}$ is meaningless, then $h_{ij} \ln h_{ij} = 0$ can be considered.

Step 4: Calculate the informativeness weights of evaluation indicators

$$H_j = 1 - E_j \tag{6}$$

Step 5: Calculate the entropy weight of evaluation indicators

$$W_j = H_j / \sum_{i=1}^{m} H_j \tag{7}$$

Step 6: Construct the weighting matrix

$$I = B \times W \tag{8}$$

Step 7: Determine the ideal solution x^+, negative ideal solution x^- and virtual worst solution x^*.

$$x^+ = \{(\max_i) I_{ij} | i = 1, 2, ..., n\} = \{x_1^+, x_2^+, ..., x_m^+\} \tag{9}$$

$$x^- = \{(\min_i) I_{ij} | i = 1, 2, ..., n\} = \{x_1^-, x_2^-, ..., x_m^-\} \tag{10}$$

$$x^* = \{3x_1^- - 2x_1^+, 3x_2^- - 2x_2^+, ..., 3x_m^- - 2x_m^+\} = \{x_1^*, x_2^*, ..., x_m^*\} \tag{11}$$

Step 8: Calculate the distance.

The distance from the infuence indicator of each province to the ideal solution x^+ is:

$$S_{ij}^+ = \sqrt{\sum_{i=1}^{m} (I_{ij} - x_i^+)^2} \tag{12}$$

The distance from the infuence indicator of each province to the virtual worst solution x^* is:

$$S_{ij}^* = \sqrt{\sum_{i=1}^{m} (I_{ij} - x_i^*)^2} \tag{13}$$

Step 9: Calculate the relative proximity index.

$$c_{ij} = S_{ij}^* / (S_{ij}^+ + S_{ij}^*) \tag{14}$$

where $c_{ij} \in [0, 1]$, denotes the comprehensive evaluation index of environmental quality. According to the calculated value of c_{ij}, the evaluation objects are ranked in order of superiority or inferiority from the largest to the smallest.

Weighting analysis

After completing the construction of the comprehensive environmental quality evaluation index system, the entropy weight-TOPSIS method was used to calculate the weights of each index, as shown in Table 2.

As shown in Table 2, at the indicator level, there are five positive and five negative indicators. Among the positive indicators, water resources per capita have the greatest weight, followed by investment in environmental infrastructure, forest coverage, green space per capita in parks, and the number of days when air quality reaches level 2 or higher. This indicates that when more water is available per capita, there is less pollution and better ecological quality. The weighting of environmental noise in key urban areas in the negative indicators is too large compared to that in the other indicators, and the lower the environmental noise is, the better the sense of residents' living experience.

For the criterion layer, the weights of the ecological environment, economic environment and social environment are 26.84%, 30.55% and 42.61%, respectively; i.e., the entropy weight of the social environment is the highest, and that of the ecological environment is the lowest. According to the principle of the entropy weight method combined with the distribution results, the social environment data contain the least information entropy, and the degree of dispersion and the difference between provinces are the greatest; therefore, environmental quality problems are difficult to solve by focusing on the social environment. The social environment is a large group, from the state down to enterprises and individuals. The rapid development of industry is accompanied by extremely serious pollution. In 2021, the National Development and Reform Commission issued a policy system to improve the dual-control policy of energy consumption, and some regions have begun to limit electricity to reduce the emissions of nitrogen oxides (NOx), sulfur dioxide (SO_2), and other pollutants from enterprises and factories and to achieve the target tasks of carbon peaking and carbon neutrality. For the ecological environment, which carries the least weight, there is less variation among provinces, and combined with the influence of some common factors such as laws, policies and culture, people have a stronger sense of protecting the ecological environment.

Environmental quality analysis

The data of 30 provinces in 2014, 2016, 2018 and 2020 are then put into the entropy weight-TOPSIS model to obtain the comprehensive evaluation index of environmental quality, which is shown in Table 3. The distribution map of environmental quality is drawn according to the relevant data, as shown in Fig. 3.

As shown in Table 3, the top three provinces are Jiangxi, Guangxi and Sichuan. With the continuous increase in energy savings and emission reduction, environmental quality improved annually in 2018 compared with 2014, which is attributed to the cost invested in China's current environmental governance. Among them, Shanghai has the lowest overall environmental quality evaluation index compared with the other provinces. As an economically developed region with a population of nearly 30 million people, Shanghai has a large amount of automobile exhaust emissions that cause serious air pollution, and at the same time, the surrounding areas are densely

Table 2 Weights of the evaluation indicators (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

System Layer	Criteria Layer	Indicator Layer	Attributes	Weights
Comprehensive Environmental Quality Assessment Index	Ecological Environment	Forest cover	Positive	0.155
		Number of days when air quality reaches Grade 2 or above	Positive	0.057
		Environmental noise in key urban areas	Negative	0.057
	Economic Environment	Green space per capita	Positive	0.067
		Growth rate of energy consumption per unit of GDP	Negative	0.020
		Investment in Environmental Infrastructure	Positive	0.215
	Social Environment	Nitrogen oxide emissions per unit area	Negative	0.010
		Domestic Oxygen Demand (DOD) emissions per unit area	Negative	0.008
		Sulfur dioxide emissions per unit area	Negative	0.010
		Water resources per capita	Positive	0.398

Table 3 Measurement results of the comprehensive evaluation indices of the environment quality in each province (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Province	2014	2016	2018	2020	Average	Ranking
Anhui	0.008	0.009	0.009	0.008	0.008	15
Beijing	0.009	0.010	0.009	0.007	0.009	9
Fujian	0.009	0.012	0.008	0.008	0.009	7
Gansu	0.005	0.006	0.006	0.005	0.006	26
Guangdong	0.007	0.008	0.008	0.009	0.008	18
Guangxi	0.011	0.011	0.010	0.012	0.011	2
Guizhou	0.009	0.008	0.007	0.008	0.008	19
Hainan	0.009	0.010	0.009	0.007	0.009	11
Hebei	0.008	0.007	0.007	0.008	0.008	21
Henan	0.007	0.006	0.009	0.009	0.007	22
Heilongjiang	0.008	0.007	0.007	0.009	0.008	20
Hubei	0.008	0.010	0.007	0.009	0.009	12
Hunan	0.009	0.009	0.007	0.009	0.009	10
Jilin	0.006	0.006	0.006	0.007	0.006	24
JiangSu	0.009	0.008	0.008	0.008	0.008	16
Jiangxi	0.011	0.013	0.010	0.012	0.012	1
Liaoning	0.006	0.006	0.005	0.005	0.005	27
Inner Mongolia	0.010	0.009	0.007	0.007	0.008	17
Ningxia	0.005	0.006	0.006	0.005	0.005	28
Qinghai	0.010	0.010	0.009	0.009	0.009	6
Shandong	0.009	0.008	0.009	0.008	0.009	13
Shanxi	0.006	0.006	0.006	0.006	0.006	25
Shanxi	0.008	0.008	0.007	0.007	0.007	23
Shanghai	0.003	0.003	0.004	0.003	0.003	30
Sichuan	0.009	0.009	0.011	0.011	0.010	3
Tianjin	0.006	0.004	0.004	0.004	0.005	29
Xinjiang	0.010	0.011	0.009	0.010	0.010	5
Yunnan	0.009	0.010	0.011	0.009	0.010	4
Zhejiang	0.009	0.010	0.009	0.009	0.009	7
Chongqing	0.009	0.008	0.008	0.010	0.009	14

populated with industries, which pollutes industries and results in poor environmental quality. Figure 3 shows that the environmental quality in China is generally greater in the western region than in the eastern region and greater in the southern region than in the northern region. Due to the low population density in the western region, the environmental pressure is low, and the emission of pollutants is low. The northern region is colder than the southern region, uses more coal, and has a dry climate with low vegetation cover and low precipitation, leading to an increase in airborne pollutants and unfavourable precipitation.

Model construction

Traditional linear regressions are first created to analyse the effects of environmental quality and medical expenditures, including least squares estimation and fixed effects models. Compared with least squares regression, fixed effects are able to control for the effects of unobservable factors caused by individuals and time. The fixed effects model was developed as follows:

$$Y_{i,t} = \beta_0 + \beta_1 W_{i,t} + \beta_2 X_{i,t} + \gamma_i + \sigma_t + \varepsilon_{i,t} \quad (15)$$

where $Y_{i,t}$ denotes the medical expenditure of individual i in year t, W denotes the composite air quality index of the time and area where individual i is located, $X_{i,t}$ denotes the control variables, γ_i denotes the fixed effect of area i, σ_t is the fixed effect of year t, and $\varepsilon_{i,t}$ is the random disturbance term.

Results

As shown in Table 4, the results of both the least squares regression and the fixed effects model indicate that environmental quality can significantly reduce the medical expenditure of elderly people.

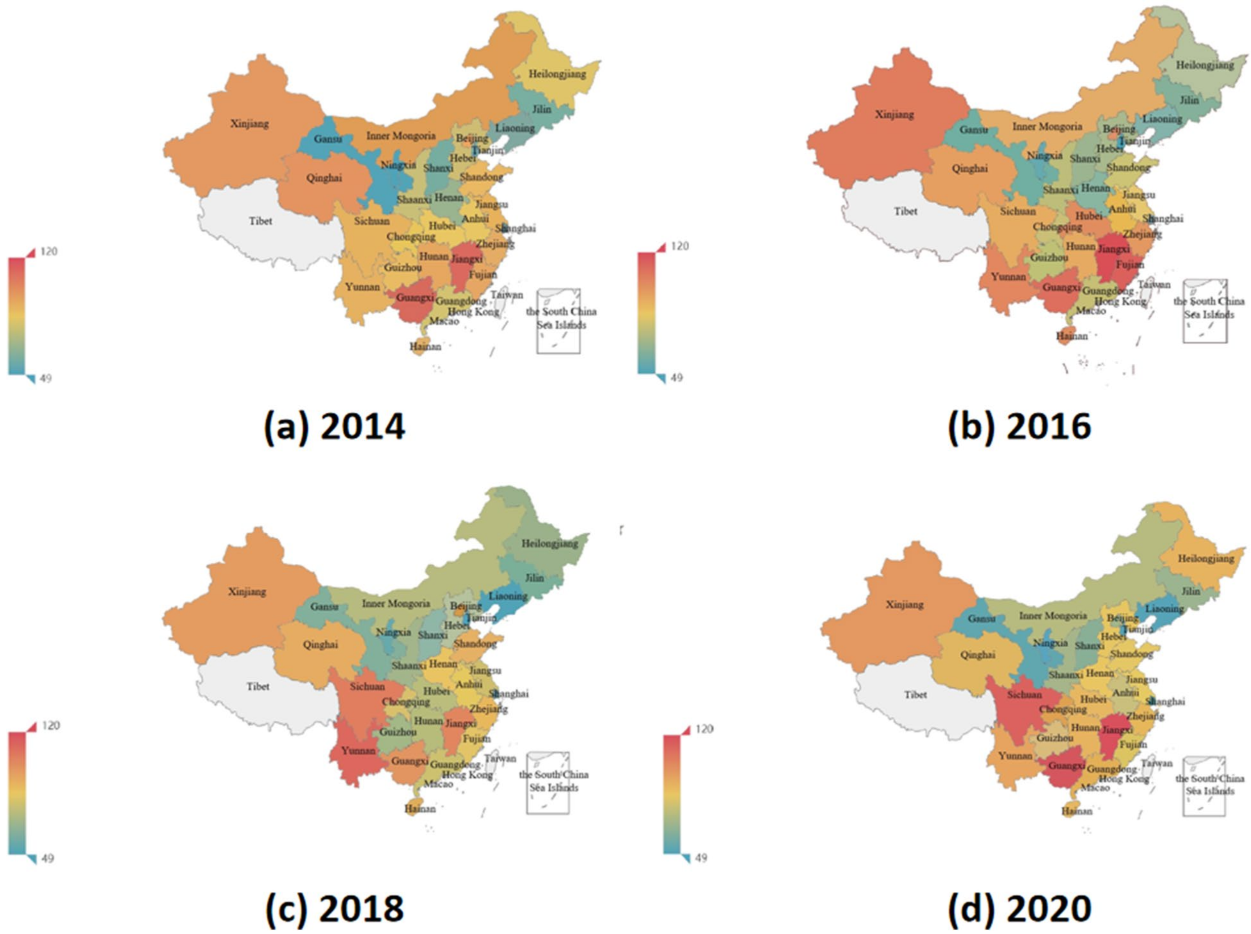

Fig. 3 Distribution of environmental quality in each province (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Table 4 Basic regression results (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Variable	(1) Medical Expenditures	(2) Medical Expenditures	(3) Medical Expenditures
Environmental Quality	-24.780***	-22.600***	-31.711*
	(5.392)	(5.108)	(16.244)
Control Variable	No	Yes	Yes
Year Fixed Effects	No	No	Yes
Individual Fixed Effects	No	No	Yes
Observed Value	18,862.000	18,862.000	18,862.000
Method	OLS	OLS	FE
R	0.001	0.164	0.178

(1), (2), (3) distinguish the results of models using different fixed effects

*is significant at the 10% level, ***is significant at the 1% level

Causal analysis based on generalized random forests

The results of both the least squares estimation and the fixed effects estimation used above indicate that the improvement of environmental quality can significantly reduce the medical expenditures of elderly people, but these two methods cannot solve the endogeneity problem that exists in the model. Therefore, this paper adopts the generalized random forest

Table 5 Importance of sample characteristics (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Variable	Importance
Hospitalization in the past year	0.400
BMI	0.095
Age	0.084
Intelligence level	0.053
Sleeping time at night	0.047
Level of medical care at point of care	0.043
Health level	0.035
Per capita household income	0.033
Happiness	0.032
Household	0.027
Gender	0.027
Length of lunch break	0.026
Satisfaction with life	0.015
Trust in doctors	0.014
Satisfaction with point of care	0.014
Frequency of exercise	0.013
Number of family members	0.013
Whether there is chronic disease within 6 months	0.012
Marital status	0.010
Number of meals with family	0.004
Receive pension insurance or not	0.004
Smoked cigarettes in the past month	0.001
Drinking more than 3 times a week for the past month	0.000
Lunch break or not	0.000

model to conduct the analysis to ensure the reliability of the results.

The sample characteristic variables Xi used for classification, including gender, age, popularity, happiness, trust in doctors, health, BMI, lunch break length, sleeping time, intelligence level, number of times of exercise, etc., were put into the GRF model for identification to derive the degree of importance of each variable as a criterion for classification. Variables with a weighting of less than 0.01 were excluded, and then the remaining variables were entered into the model for analysis. The remaining variables were put into the model for analysis. This method can improve the accuracy of model estimation by excluding variables with poor explanatory power [29].

As shown in Table 5, according to the ranking of importance, six variables, namely, marital status, number of meals with family members, whether or not they received pension insurance, whether or not they smoked in the past month, whether or not they drank alcohol in the past month, and whether or not they had taken a lunch break, are excluded; then, the results of the model are calculated under different decision trees. Since random forest is an improvised learning method constructed based on decision trees, the number of decision trees will have an important impact on the results of random forest, and when the number of decision trees is small, the estimation results will produce a large error or even an overfitting phenomenon.

The first three columns of Table 6 show that environmental quality has a negative effect on the medical expenditure of elderly people, the standard deviation of the model stabilizes at approximately 5.0 with the change in the number of decision trees, and the average treatment effect stays at -21. This means that the average medical expenditure of elderly people will be reduced by 21.08 dollars for every improvement in environmental quality.

Heterogeneity analysis

Heterogeneity test

Compared with the traditional linear regression model, the random forest can obtain the average treatment effect of each individual and then calculate the overall average treatment effect. Figure 4 shows the distribution of the treatment effect of environmental quality on the medical expenditure of elderly people. The average treatment effect for most individuals is between -50 and 10, the impact of environmental quality on medical expenditure

Table 6 Impact of environmental quality on the medical expenditures of elderly people (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Variable	(1) Medical Expenditures	(2) Medical Expenditures	(3) Medical Expenditures
Environmental Quality	-21.103***	-21.408***	-21.083***
	(5.175)	(5.214)	(5.149)
Number of Decision Trees	500.000	1000.000	2000.000
Model	Causal Forest	Causal Forest	Causal Forest
Observed Value	18,862.000	18,862.000	18,862.000

(1), (2), (3) distinguish the results of using different numbers of decision trees in the model

***is significant at the 1% level

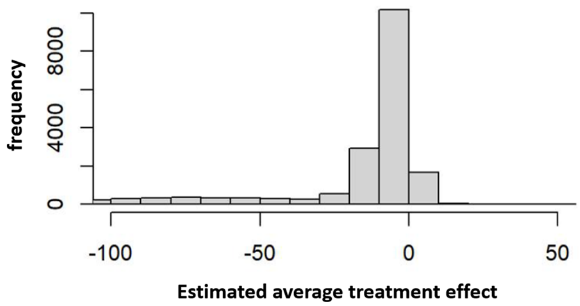

Fig. 4 Distribution of the treatment effects of environmental quality on medical expenditures (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Table 7 Heterogeneity test (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Variable	(1) Medical Expenditures	(2) Medical Expenditures
Mean	0.994***	0.998***
	(0.244)	(0.233)
Differential	1.226***	1.201***
	(0.300)	(0.277)
Number of Decision Trees	1000.000	2000.000
Model	Causal Forest	Causal Forest
Observed Value	18,862.000	18,862.000

(1) and (2) distinguish the results of using different numbers of decision trees in the model

***is significant at the 1% level

is mostly negative, and the impact of environmental quality on different groups of elderly people is also different.

Chernozhukov et al. [30] proposed a method called the "best linear predictor" to test the heterogeneity of the average treatment effect, split the average treatment effect estimates into two C_i and D_i parts, and constructed the following regression model:

$$(Y_i - \widehat{m}^{(-i)}(X_i)) = \beta_1 C_i + \beta_2 D_i \tag{16}$$

where $C_i = \overline{\tau}(W_i - \widehat{e}^{(-i)}(X_i)), D_i = (\widehat{\tau}^{(-i)}(X_i) - \overline{\tau})(W_i - \widehat{e}^{(-i)}(X_i))$,$\overline{\tau}$ is the mean of the mean treatment effect estimates for the samples estimated outside the package. Heterogeneity is considered to exist when β_2 is significantly positive, and the regression results are shown in Table 7.

As shown in Table 7, the coefficient of β_2 is significantly positive under different decision trees, indicating that there is significant heterogeneity in the impact of environmental quality on the medical expenditures of elderly people, and the heterogeneity of environmental quality on different groups of elderly people will be further analysed later.

Analysis of regional heterogeneity

Since the eastern, central, and western parts of the country are quite different in terms of economy, geographic location, and environmental factors, the samples were divided into eastern, central, and western parts of the country according to the provinces to analyse whether there is any difference in the impacts of environmental quality in the three regions. The results are shown in Table 8 and Fig. 5.

Table 8 shows that improving environmental quality can significantly reduce the medical expenditures of elderly people in the eastern and central regions, but the negative effect on medical expenditures in the western region is not significant. Figure 5 shows that the treatment effect in the western region is approximately the same when the position is greater than zero and less than zero, so it has a nonsignificant impact on the western region, and the treatment effect in the eastern and central regions is mainly concentrated when the position is less than zero. The potential reasons for the insignificant impact of environmental quality improvements on healthcare expenditures in the western region may include: the remoteness of the area, leading to significant disparities in infrastructure, medical resource distribution, and the pace and intensity of environmental recovery compared to the eastern and central regions. Additionally, the complex terrain

Table 8 Average treatment effects in the eastern, central and western regions (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

Explanatory Variable	(1) Eastern Region	(2) Central Region	(3) Western Region
Environmental Quality	-20.504***	-16.061***	6.719**
	(8.084)	(10.022)	(10.011)
Number of Decision Trees	2000.000	2000.000	2000.000
Model	Causal Forest	Causal Forest	Causal Forest
Observed Value	9110.000	5485.000	4266.000

(1), (2), (3) distinguish the results from different regions

**is significant at the 5% level, ***is significant at the 1% level

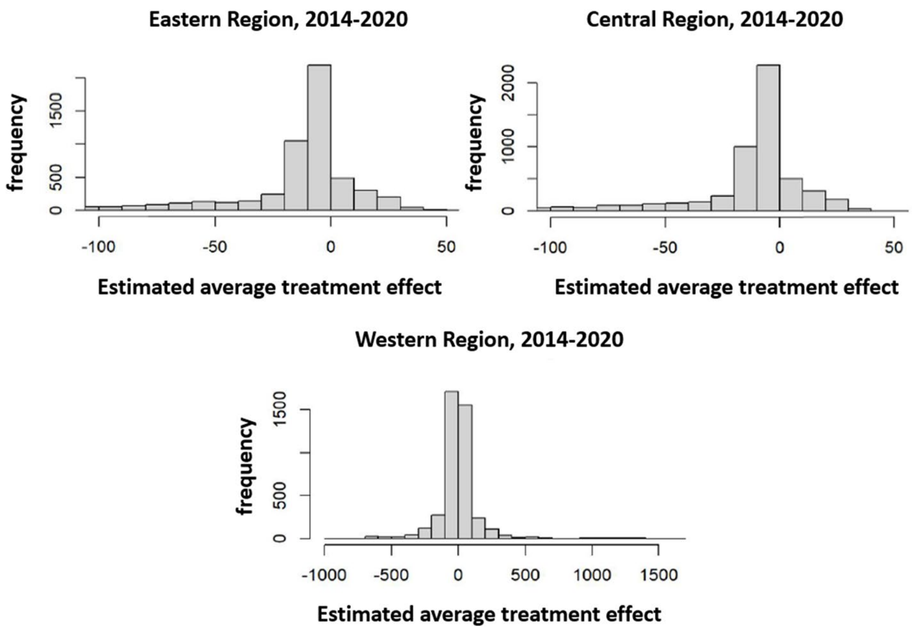

Fig. 5 Distribution of treatment effects in the eastern, central and western regions (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

and geomorphology of the western region complicate environmental issues, making improvements less effective. Furthermore, the lower population density and sparse industrial distribution in the west also reduce the health benefits of environmental enhancements.

Heterogeneity analysis of differences in individual characteristics

The impact of the environment on medical expenditures may vary with age in the elderly population. This paper analyses the heterogeneity of the impact of environmental quality on older adults of different ages, as shown in Fig. 6.

The results show that the negative impact of environmental quality on medical expenditure decreases as age increases, and the negative impact of environmental quality on medical expenditure is most significant for those aged between 60 and 75, with confidence intervals less than zero. The negative effect of the environment on medical expenditures decreases among elderly

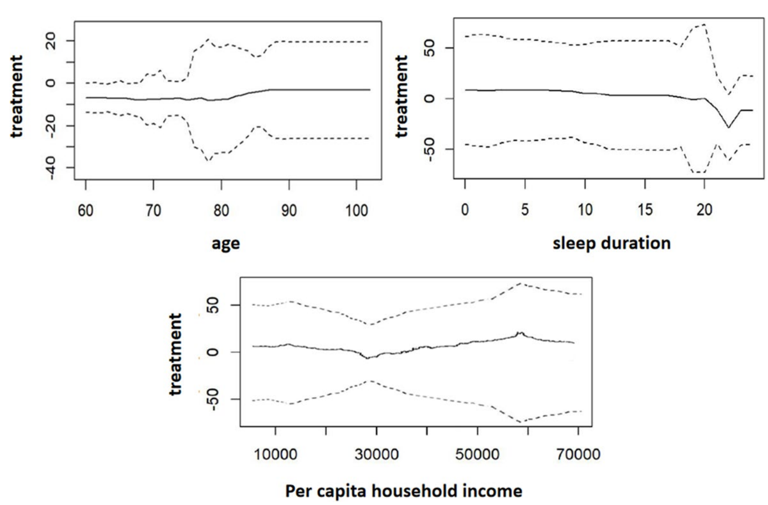

Fig. 6 Impact of environmental quality on people of different ages, sleep durations and economic conditions (years: 2014, 2016, 2018, 2020; place: 30 provinces in China)

people aged 75 and above due to their own physical conditions, and the distribution of confidence intervals shows that the negative effect of the environment is also not significant.

In addition to age, the different living habits of elderly people may also lead to a weakening of the inhibitory effect of environmental quality on medical expenditure, and different sleeping hours may lead to different effects of environmental quality on medical expenditure. Environmental quality significantly reduces the effect of health care spending when older adults sleep between 21:00 and 23:00. This may be because those who go to sleep between 21:00 and 23:00 follow the rhythm of the biological clock and maintain a good routine, so the negative effect of environmental quality on medical expenditure is more significant for this group.

At the same time, the impact of environmental quality on elderly people is different for different household economic conditions. For families with a per capita income of less than 50,000, good environmental quality can reduce medical expenditures, but for people with incomes higher than 50,000, the effect of environmental quality is not significant. The reason for this may be that high-income families pay more attention to their health and increase their personal medical expenditures, including medical expenditures, self-purchased medicines, health care products, health care appliances and medical appliances, which leads to a nonsignificant negative effect on environmental quality. For lower-income households, there is less need for such expenditures, so most of the expenditures are concentrated on medical fees, medicines, and other expenditures for therapeutic purposes, and environmental quality mainly affects this aspect of medical expenditures.

Discussion

This research offers a granular examination of the relationship between environmental quality and healthcare expenditures among elderly people in China, a demographic increasingly burdened by health-related costs amid environmental challenges. By blending machine learning algorithms with conventional econometric models, this study not only reaffirms the critical role of environmental quality in public health but also pioneers methodological advances in environmental health economics. Our findings, which indicate a significant reduction in healthcare costs with improved environmental conditions, particularly in China's eastern and central regions, resonate with and extend the narratives of prior studies.

The observed regional disparities in the impact of environmental quality on healthcare expenditures highlight the intricate interplay between geographical, environmental, and socioeconomic factors. This nuanced understanding aligns with global environmental health research, which consistently underscores the heightened vulnerability of specific regions and populations to environmental hazards. Furthermore, our study's emphasis on the differential impact across various demographic cohorts, notably the significant benefits observed among individuals aged 60 to 75 and lower-income households, underscores the potential of environmental interventions to address health inequities—a cornerstone of the World Health Organization's agenda.

However, the study has methodological and conceptual limitations. Relying on a comprehensive but single environmental quality index may not capture all environmental factors affecting health, such as local air pollutants, water quality, and green space usage. Future research should aim to include a wider range of environmental indicators for a more comprehensive analysis. For example, Zhao and Irfan [31] emphasized the necessity of comprehensive environmental indices, and Cheng et al. [32] highlighted the multifaceted and region-specific nature of environmental factors, suggesting that a single index might oversimplify the impact of the environment on healthcare expenditures. This highlights the need for more detailed and localized environmental data to better understand the subtle impacts on healthcare costs.

Moreover, although using generalized random forest algorithms helps address endogeneity issues and improve causal inference, it also adds complexity to the model's interpretability. The black-box nature of machine learning models, such as generalized random forests, can make it challenging to understand the underlying mechanisms driving observed relationships. Additionally, this study faces the risk of ecological fallacy, where conclusions about individual outcomes may be incorrectly drawn from group-level data. Studies by Robinson (2009) and Akande and Slain (2023) have shown that observational study results are particularly susceptible to ecological fallacy [33, 34]. While provinces with better environmental quality often show lower average healthcare costs, this does not necessarily mean that all individuals benefit equally. Differences in income, healthcare access, and personal behaviors can lead to significant individual variations that group data cannot capture. Although advanced techniques like Generalized Random Forests were used to control for heterogeneity and confounding factors, the risk of ecological fallacy remains. Future research should incorporate more granular data to better reflect individual experiences and outcomes, providing a more comprehensive understanding of the relationship between environmental quality, health, and medical expenditures. Developing more interpretable models or combining machine learning methods with traditional econometric approaches could enhance the transparency and reliability of the results.

Another limitation is potential omitted variable bias. While we included a range of covariates to account for individual behaviors, such as lifestyle habits and exercise frequency, other unmeasured factors, such as genetic predispositions, healthcare accessibility, and socio-economic status beyond income, may also play a significant role in determining healthcare expenditures. Including these additional variables in future research could provide a more comprehensive understanding of the determinants of healthcare costs.

Additionally, integrating pilot health policies is crucial for understanding the actual impact of environmental quality improvements on healthcare costs. Breed et al. [35] showed that policy interventions are important for improving human health conditions when dealing with issues like climate change and environmental degradation. Future research should explore the specific impact of health policies on environmental health outcomes to provide more targeted and effective policy recommendations.

By addressing these limitations, future research can build on the current study and contribute to a more detailed and actionable understanding of the complex relationship between environmental quality and health economics. This, in turn, can support the formulation of more effective and equitable environmental health policies.

Conclusions

In this paper, we first measured the environmental quality of each region in China from 2014 to 2020, and the results show that the distribution of environmental quality is high in the east, low in the west, high in the north and low in the south. Then, the measured environmental quality of different years and provinces is matched with the CFPS data from 2014, 2016, 2018 and 2020. The study revealed that the traditional least squares method, fixed effects model, newly emerging machine learning method, and generalized random forest significantly reduced the medical expenditures of elderly people through improvements in environmental quality. Heterogeneity analysis based on generalized random forests reveals that improving environmental quality significantly reduces medical expenditures for elderly people in the Eastern and Central Regions, but the negative effect on medical expenditures for elderly people in the West is not significant. In addition, as environmental quality improves, the medical expenditures of elderly people between 60 and 75 years old are relatively lower; elderly people who go to bed between 21:00 and 23:00 have good work routines; and the medical expenditures of elderly people with higher income families are not significantly affected by environmental quality, while better environmental quality has a significant effect on the medical expenditures of elderly people with low-income families. This study significantly advances our understanding of the intricate relationship between environmental quality and healthcare expenditures for elderly people in China.

The evidence presented in this study underscores the critical need for comprehensive environmental policies that are attuned to the socioeconomic and demographic realities of China's diverse regions. To this end, we recommend the following policy actions:

Health-oriented urban planning

Implement strict emission controls in industrially dense areas and promote clean production technologies to reduce industrial pollution. Additionally, enhance public transportation systems, encourage the use of electric and hybrid vehicles, and gradually phase out high-emission vehicles to control traffic pollution. Expanding urban green spaces and parks can also improve air quality and provide healthier living environments for residents.

Policy incentives and economic measures

Introduce environmental incentive policies, such as tax breaks and subsidies, to encourage businesses and individuals to adopt green technologies and products. Establish green funds to support environmental projects and enterprises, and promote green bonds and loans to provide financial backing for sustainable initiatives. These economic measures can help accelerate the transition to environmentally friendly practices.

Public health and preventive measures

Develop community health programs that include environmental health monitoring networks to track air quality, water quality, and noise levels in real-time. Offer regular health check-ups for the elderly, focusing on detecting and preventing environment-related diseases such as respiratory and cardiovascular conditions. Issue timely health warnings to guide the public in taking preventive actions during high pollution periods.

Education and awareness enhancement

Integrate environmental protection and health education into school curriculums at all levels to foster environmental awareness from a young age. Conduct public awareness campaigns using media and community events to educate the broader population about the importance of environmental quality and the specific steps they can take to reduce pollution and protect their health.

Interdepartmental and international cooperation

Establish collaboration mechanisms between environmental, health, and economic departments to address environmental health challenges comprehensively. Create joint action plans that align environmental improvements with public health and economic goals. Additionally, learn from the successful experiences of developed countries in environmental protection and health management, and engage in regional cooperation with neighboring countries to tackle transboundary environmental issues.

To build on the findings of this study, future research should explore the longitudinal effects of public health policies on healthcare expenditures, incorporate more detailed environmental and health data, and employ interpretable machine learning models that facilitate clearer insights into the causal mechanisms at play. Expanding the scope of research to include comparative analyses across different countries could also provide valuable global perspectives on the interplay between environmental quality and healthcare economics.

In sum, this research highlights the vital role of environmental quality in shaping healthcare expenditures for elderly people, offering actionable insights for policymakers committed to fostering sustainable, healthy, and equitable societies.

Abbreviations

CFPS	China Family Panel Studies
OLS	Ordinary least squares
CSMAR	China Stock Market & Accounting Research Database
EPS	Economy Prediction System database
CESY	China Environmental Statistics Yearbook
RF	Random Forest
GRF	Generalized Random Forest
GDP	Gross Domestic Product

Acknowledgements
Not applicable.

Authors' contributions
YZ (Yu Zhang) conceptualized the study, led the data collection and analysis, conducted the experiments, drafted the manuscript. SC (Sheng Chen) provided technical support, conducted the experiments, and assisted with the data processing. Dewen Liu (DL) offered insights into the research design and helped with the literature review and discussion. All authors reviewed and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
The datasets used and analysed during the current study are available from the corresponding author upon reasonable request.

Data availability
The datasets generated by the survey research during and analyzed during the current study are available in the CFPS repository, https://opendata.pku.edu.cn/dataverse/CFPS.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Advanced Interdisciplinary Studies, Hunan University of Technology and Business, 569 Yuelu Ave., Yuelu District, Changsha 410205, China. [2]School of Science, Hunan University of Technology and Business, 569 Yuelu Ave., Yuelu District, Changsha 410205, China.

Received: 28 April 2024 Accepted: 29 August 2024
Published online: 30 October 2024

References

1. Zhou X, Song Y, Chen W, Zhang X. The progress of epidemiological study on the effects of traffic related air pollution on the cardiovascular system. J Public Health Prev Med. 2020;31(4):106–10.
2. Giovanis E, Ozdamar O. Health status, mental health and air quality: evidence from pensioners in Europe. Environ Sci Pollut Res. 2018;25:14206–25.
3. Wei H, Zhou Y. The impact of international talent on environmental pollution: firm-level evidence from China. Energy Econ. 2023;125:106800.
4. Arceo E, Hanna R, Oliva P. Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. Econ J. 2016;126:257–80.
5. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: a review. Front Public Health. 2020;8:505570.
6. Deryugina T, Heutel G, Miller NH, Molitor D, Reif J. The mortality and medical costs of air pollution: evidence from changes in wind direction. Am Econ Rev. 2019;109:4178–219.
7. Yang T, Liu W. Does air pollution affect public health and health inequality? Empirical evidence from China. J Clean Prod. 2018;203:43–52.
8. Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, Dominici F. Association of short-term exposure to air pollution with mortality in older adults. JAMA. 2017;318:2446–56.
9. Preker AS, Adeyi OO, Lapetra MG, Simon DC, Keuffel E. Health care expenditures associated with pollution: exploratory methods and findings. Ann Glob Health. 2016;82:711–21.
10. Akpalu W, Normanyo AK. Gold mining pollution and the cost of private healthcare: the case of Ghana. Ecol Econ. 2017;142:104–12.
11. Yang J, Zhang B. Air pollution and healthcare expenditure: Implication for the benefit of air pollution control in China. Environ Int. 2018;120:443–55.
12. Aydin M, Bozatli O. The impacts of the refugee population, renewable energy consumption, carbon emissions, and economic growth on health expenditure in Turkey: new evidence from Fourier-based analyses. Environ Sci Pollut Res. 2023;30:41286–98.
13. Huang D, Xu J, Zhang S. Valuing the health risks of particulate air pollution in the Pearl River Delta, China. Environ Sci Policy. 2012;15:38–47.
14. Liao L, Du M, Chen Z. Air pollution, health care use and medical costs: evidence from China. Energy Econ. 2021;95:105132.
15. Zhang C, Zhang L. The relationship between toxic air pollution, health expenditure, and economic growth in the European Union: fresh evidence from the PMG-ARDL model. J Environ Sci Pollut Res. 2024;31:21107–23.
16. Lu ZN, Chen H, Hao Y, Wang J, Song X, Mok TM. The dynamic relationship between environmental pollution, economic development and public health: evidence from China. J Clean Prod. 2017;166:134–47.
17. Dey S, Caulfield B, Ghosh B. Potential health and economic benefits of banning diesel traffic in Dublin, Ireland. J Transp Health. 2018;10:156–66.

18. Li H, Lu J, Li B. Does pollution-intensive industrial agglomeration increase residents' health expenditure? Sustain Cities Soc. 2020;56:102092.
19. Yacour S, Soumbara S, El Ghini A. Environmental quality, economic growth, and healthcare expenditure nexus for North Africa: a panel cointegration analysis. J Environ Model Assess. 2024;29:307–21.
20. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: opportunities and challenges. Neurocomputing. 2017;237:350–61.
21. Mullainathan S, Spiess J. Machine learning: an applied econometric approach. J Econ Perspect. 2017;31:87–106.
22. Kamińska JA. The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław. J Environ Manage. 2018;217:164–74.
23. Hu Z, Gu J, Chen Y. "Biased" Policy, Changes of production sectors and development gap between the north and the south in China: causal inference from machine learning. J Financ Econ. 2022;48:93–107.
24. Baiardi A, Naghi AA. The value added of machine learning to causal inference: evidence from revisited studies. Economet J. 2024;27(2):213–34.
25. Bernardo M, Casadesus M, Karapetrovic S, Heras I. How integrated are environmental, quality and other standardized management systems? An empirical study. J Clean Prod. 2009;17:742–50.
26. Mourhir A, Rachidi T, Papageorgiou EI, Karim M, Alaoui FS. A cognitive map framework to support integrated environmental assessment. Environ Model Softw. 2016;77:81–94.
27. Li N, Wang J. Comprehensive eco-environment quality index model with spatiotemporal characteristics. Sensors. 2022;22(24):9635.
28. Zou T, Guo P, Wu Q. Applying an entropy-weighted TOPSIS method to evaluate energy green consumption revolution progressing of China. Environ Sci Pollut Res. 2023;30(14):42267–81.
29. Basu S, Kumbier K, Brown JB, Yu B. Iterative random forests to discover predictive and stable high-order interactions. Proc Natl Acad Sci. 2018;115:1943–8.
30. Chernozhukov V, Demirer M, Duflo E, Fernandez-Val I. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India (No. w24678). National Bureau of Economic Research; 2018. https://doi.org/10.3386/w24678.
31. Zhao W, Irfan M. Does healthy city construction facilitate green growth in China? Evidence from 279 cities. Environ Sci Pollut Res. 2023;30:102772–89.
32. Cheng C, Ren X, Zhang M, Wang Z. The nexus among CO2 emission, health expenditure and economic development in the OECD countries: new insights from a cross-sectional ARDL model. Environ Sci Pollut Res. 2024;31:16746–69.
33. Robinson WS. Ecological correlations and the behavior of individuals. Int J Epidemiol. 2009;38(2):337–41.
34. Akande MY, Slain KN. Neighborhood condition and PICU admission: facts and fallacies. Chest. 2023;164(6):1341–2.
35. Breed MF, Cross AT, Wallace K, et al. Ecosystem restoration: a public health intervention. EcoHealth. 2021;18:269–71.

Publisher's Note