



# Nonlinear analysis of factors influencing public health in China based on machine learning

Dehong Sun, Hui Wang

School of Information Engineering, Minnan University of Science and Technology, Quanzhou, China

## Abstract

Public health is a key component of the United Nations Sustainable Development Goals (SDGs) and is central to the all-round development of individuals. Based on panel data from 31 provinces in China between 2010 and 2023, this article adopts the Population Mortality Rate (PMR) as a measure of public health and employs the Generalized Additive Model (GAM) from machine learning to systematically investigate the nonlinear effects of multidimensional factors-including medical resources, environmental pollution, socioeconomic conditions, and technology-on public health. The results of the study show that China's PMR exhibits an overall upward trend, with a spatially uneven distribution and significant regional disparities, as mortality rates in the central and western regions are generally higher than those in the eastern coastal provinces. All influencing factors show significant associations with the PMR. The effects of all influencing factors on mortality exhibit complex nonlinear characteristics, with their impacts varying considerably across different value ranges. Specifically, the influence of medical resources exhibits critical thresholds, such as the Number of Urban Practicing (assistant) Physicians per 10,000 people (NUPP) reaching its maximum health benefit at approximately 40; while the relationship between environmental indicators and mortality reveals potential complex confounding mechanisms, as evidenced by the negative association observed between Sulphur Dioxide Emissions (SDE) and mortality rates within the observed range; whereas the health benefits of basic resources and developmental factors, such as Per Capita Water Resources (PCWR), stabilize after crossing a specific threshold (20,000 cubic meters per person). Finally, some practical policy recommendations are put forward.

**Key words:** public health; Population Mortality Rate (PMR); machine learning; Generalized Additive Model (GAM); nonlinear effects; China.

*Correspondence:* Dehong Sun, School of Information Engineering, Minnan University of Science and Technology, No. 1, Longjing Road, 362000, Quanzhou, China. Tel.: +86.18750516366 - Fax: +86.595.23245687. E-mail: 155065349@qq.com

## Introduction

Public health serves as a core barometer of national development quality and a critical nexus for achieving the United Nations Sustainable Development Goals (SDGs). The Population Mortality Rate (PMR), as an intuitive quantitative indicator reflecting the health status of the entire population, not only constitutes a fundamental basis for assessing public health conditions but also aligns closely with SDG 3, which comprises 13 specific targets, including objectives such as “reducing the global maternal mortality ratio to less than 70 per 100,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births by 2030,” alongside key tasks like “reducing by one-third premature mortality from non-communicable diseases” and “minimizing deaths from environmental health risks.” These efforts ultimately aim to reduce mortality risks across all population groups through multidimensional interventions, thereby achieving health and well-being across the life course.

Currently, with the deepening implementation of the “Healthy China 2030” strategy, China’s public health indicators continue to improve. However, the national overall mortality rate continues to rise. Meanwhile, although premature mortality from non-commu-

nicable diseases has declined compared to 2015 levels, it remains distant from the 2030 target of a one-third reduction. In this context, using PMR as a key metric to systematically investigate the influencing mechanisms of public health in China is imperative for enhancing national health levels, demonstrating dual importance in both theoretical value and practical significance.

Compared with existing research, this paper presents three main innovations. First, innovation in research content perspective when selecting public health indicators: this research adopts a comprehensive analytical framework that integrates multiple core dimensions-including environmental regulation, socioeconomic factors, healthcare conditions, and technological development. Using PMR as the central metric, we systematically examine the synergistic effects of these multifaceted factors on public health outcomes. This integrated approach effectively addresses the limitations of conventional single-factor analyses, thereby providing a more realistic representation of how public health is collectively shaped by multidimensional determinants in real-world scenarios. Second, methodological innovation: while existing research predominantly employs traditional econometric models for analyzing public health determinants-focusing primarily on verifying simple linear relationships between variables-such approaches often fail

to capture the nonlinear interdependencies inherent in complex real-world systems. This study breaks new ground by introducing the Generalized Additive Model (GAM) from machine learning, which transcends the linearity constraints of conventional models and enables in-depth exploration of nonlinear mechanisms linking multidimensional factors with public health outcomes. Compared with traditional regression models, the GAM demonstrates superior capability in accurately capturing dynamic relationships among variables, significantly improves the identification accuracy of health influencing factors, provides more efficient methodological support for quantitative research in public health, and represents an important supplement to the application of Machine Learning (ML) in sustainable health development studies. Third, the validity of the study's conclusions: the application of the GAM framework enables in-depth exploration of dynamic trends in public health impacts, delivering dual analytical advantages. On one hand, it facilitates a more profound understanding of the complex interactions between multidimensional factors and public health, preventing potential misinterpretations in policymaking caused by static conclusions from traditional research. On the other hand, it offers precise, targeted guidance for health policy formulation, effectively mitigating resource misallocation and policy inefficiencies, thereby ensuring that various health interventions maximize their potential to enhance public health outcomes.

## Literature review

### Public health

Public health constitutes a core aspect of the United Nations Sustainable Development Goals and is fundamental to holistic human development and sustainable societal progress. As the most straightforward and comprehensive metric for assessing public health levels, the PMR directly mirrors the effectiveness and challenges of societal health governance through its dynamic fluctuations. Existing studies have identified dual challenges influencing the PMR. On one hand, the transition of disease spectrum poses persistent pressure. Chronic non-communicable diseases—such as cardiovascular conditions (Miao *et al.*, 2024), gastric cancer (Lu *et al.*, 2025), and colorectal cancer (Pan *et al.*, 2025)—constitute major mortality risks, with the phenomenon of multimorbidity becoming increasingly prominent (Zheng *et al.*, 2023). The chronic disease burden among China's elderly population is particularly pronounced, with stroke, ischemic heart disease, and chronic obstructive pulmonary disease representing the top three contributors to Disability-Adjusted Life Years (DALYs) in this demographic (Wu *et al.*, 2024). Meanwhile, threats from traditional infectious diseases persist. During specific periods, influenza has caused excess mortality rates even surpassing those observed during the COVID-19 pandemic (Heudorf & Kowall, 2025), while acute respiratory infections remain a leading cause of death among children under five years old (Merera *et al.*, 2022). On the other hand, the accessibility and quality of healthcare systems directly shape health outcomes. Studies indicate that both financial burdens of healthcare (Rahman *et al.*, 2025) and multidimensional aspects of primary healthcare systems (Adesuyi *et al.*, 2025) are critical constraints on reducing mortality rates. In Indonesia, the implementation of non-communicable disease programs faces multiple challenges including resource constraints, participation issues, and data management difficulties (Prasetyo *et al.*, 2024). Significant regional disparities exist in public health service utilization among China's migrant population, with predisposing characteristics and needs

factors identified as the most explanatory dimensions (Suo *et al.*, 2023). The uneven geographical distribution of these complex challenges likely contributes to the higher mortality rates observed in central and western provinces compared to eastern coastal regions.

### Factors influencing public health activities

Current research on public health influencing factors exhibits significant limitations that hinder a comprehensive explanation of the complex patterns in China's PMR. These constraints are primarily manifested in two dimensions: research perspective and methodological approach.

In terms of research perspective, most studies tend to examine isolated factors or remain confined to specific domains. For instance, disease-specific research focuses on independent effects of dietary factors (Lu *et al.*, 2025; Pan *et al.*, 2025), clinical indicators (Zhang *et al.*, 2025a; Shu *et al.*, 2025), or patient behaviours (Zhang *et al.*, 2025b). Other studies have separately examined the roles of policy awareness (Yu *et al.*, 2025), community environment (Sofiyah *et al.*, 2025), and professional capacity (Prasetyo *et al.*, 2024; Wu *et al.*, 2025). Environmental studies reveal that land surface temperature significantly influences mosquito net usage behaviour (Otieno *et al.*, 2024), while willingness to pay for air quality improvement is shaped by multiple factors including education level, income, and health status (Kalhori *et al.*, 2024). This fragmented research approach fails to reveal the interactive effects among different dimensions—such as healthcare resources, environmental pollution, socioeconomic conditions, and technological development—and makes it difficult to assess the relative importance of each factor.

Methodologically, traditional studies predominantly rely on linear relationship assumptions, which significantly deviate from the actual operational mechanisms of influencing factors. In reality, the effects of numerous factors on mortality rates may exhibit complex nonlinear characteristics, such as threshold effects, saturation effects, or inflection point effects. Spatial geographical research demonstrates clear regional differentiation in the cumulative health deficit index among China's elderly population, with the interactive effects of personal income, air quality, and GDP showing nonlinear enhancement characteristics (Xiang *et al.*, 2023). Although ML methods have been applied to health risk prediction (Mohammadi-Pirouz *et al.*, 2024; Zhang *et al.*, 2025a; Shu *et al.*, 2025), these models predominantly focus on individual clinical predictions and suffer from limited interpretability, making them difficult to apply to macro-level policy formulation. Existing research methodologies show particular inadequacy in addressing nonlinear effects of multidimensional factors at the regional level.

In summary, public health constitutes a complex systemic project characterized by diversified influencing factors and sophisticated effect mechanisms. To more accurately understand the dynamic evolution of public health, it is imperative to comprehensively consider various factors and conduct multidimensional, in-depth analyses. Utilizing provincial panel data and employing GAM to systematically investigate the nonlinear effects of healthcare resources, environmental pollution, socioeconomic conditions, and technological factors on PMR maintains the flexibility of ML in handling complex relationships while preserving the strong interpretability of statistical models. This approach aimed to provide scientific evidence for formulating precise public health policies, simultaneously contributing to the advancement of both UN SDG 3 and the Healthy China 2030 initiative, thereby offering crucial support for the global improvement of human health outcomes.

## Research design

The specific conceptual framework diagram used for this study is shown in Figure 1.

## Research methodologies

ML, as a core branch of artificial intelligence, has become an effective tool for quantitative analysis across multiple domains due to its advantages of low cost, high processing efficiency, and low operational complexity (Z. Li *et al.*, 2023). The GAM, as a representative model in the field of ML, offers more flexible analytical capabilities compared to traditional linear models (Ravindra *et al.*, 2019) and is widely regarded as an efficient alternative to Generalized Linear Models (GLM) (Díaz Martínez *et al.*, 2023).

The core logic of the GAM framework lies in extending the standard linear model while preserving overall additivity, by replacing the simple linear relationship between independent and dependent variables in linear models with smooth nonlinear functions corresponding to each independent variable. This approach better captures the complex relationships between variables observed in real-world scenarios. The traditional multiple linear regression model is formulated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \tag{Eq. 1}$$

where  $\beta_0$  is the constant term;  $\beta_i$  ( $i = 1, 2, \dots, p$ ) the regression coefficient; and  $\varepsilon$  the error term. The GAM replaces each term  $\beta_i X_i$  with a smooth nonlinear function  $f_i(X_i)$ . The model formulation is thus adjusted as follows:

$$Y = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) + \varepsilon \tag{Eq. 2}$$

This study implements GAM model fitting using the R programming language, specifically employing the GAM function from the mgcv package. During the fitting process, the model iteratively updates the coefficients of individual variables while keeping other variables fixed, ensuring precise capture of each variable's independent effects within the additive framework. Compared with traditional linear models, GAM demonstrates three significant advantages: i) strong nonlinear fitting capability: without presuming specific variable relationships, GAM flexibly captures nonlinear patterns such as threshold effects and marginal changes through smooth functions, thereby avoiding assumption bias inherent in traditional linear models; ii) balance between interpretability and complexity: by retaining the additive structure of variable effects, GAM allows clear decomposition of individual variable contributions while visualizing impact trends, effectively mitigating the “black-box” problem associated with deep learning; and iii) high tolerance for non-normal data: GAM reduces the influence of outliers and supports direct fitting of both continuous and categorical variables, significantly minimizing data preprocessing efforts.

## Variable definitions

The study area of this research was the administrative divisions of China and the data were obtained from the *China Statistical Yearbook* published by the National Bureau of Statistics of China, covering 31 provinces in China (excluding Hong Kong, Macao and Taiwan) from 2010 to 2023, and the type of data was panel data.

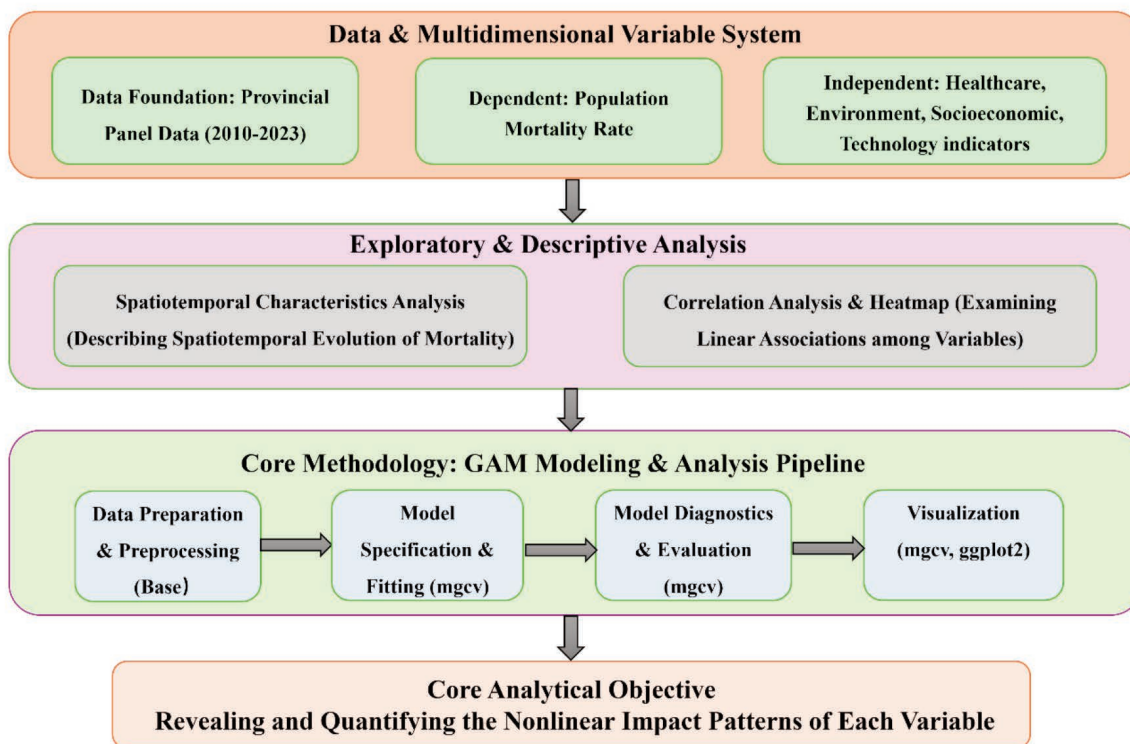


Figure 1. Conceptual framework diagram.

The dataset is balanced with no missing values for the selected variables across the study period.

Public health status can be reflected through multiple dimensions, with PMR serving as a key metric. Considering the conceptual definitions of various variables, this study analyzed influencing factors of public health from multiple perspectives. For environmental regulation, Sulphur Dioxide Emissions (SDE) and Chemical Oxygen Demand Emissions (CODE) were selected. In terms of socioeconomic factors, the analysis covered Urbanization Rate (UR), Per Capita Water Resources (PCWR), Harmless Treatment rate of Household Waste (HTHW) and Local Government Expenditure on Environmental Protection (LFEEP). Regarding medical conditions, the study includes Number of Beds per 10,000 people in Medical Institutions (NBMI), Number of Urban Practicing (assistant) Physicians per 10000 people (NUPP) and Number of Medical and Health Institutions (NMHI). For technological development, the Number of Domestic Patent Applications Accepted (NDPAA) was used as an indicator. PMR was the dependent variable. Definitions of all variables are presented in Tables 1 and 2 presents descriptive statistics of the variables. As shown in the table, the PMR ranges from 4.210‰ to 9.840‰, with a mean value of 6.383‰.

## Results and Discussion

### *Spatiotemporal distribution of public health influencing factors*

#### *Temporal changes in PMR*

Data on PMR from 2010 to 2023 were collected from the China Statistical Yearbook, with 2010 as the base period to calculate the year-on-year growth rate of population mortality. Figure 2 illustrates the temporal evolution of both the PMR and its growth rate. As shown in Figure 2, the PMR generally exhibits an upward trend. The temporal evolution of the PMR can be divided into three main phases: from 2010 to 2019, the PMR showed a relatively moderate upward trend; from 2019 to 2020, the rate of increase accelerated significantly; after 2020, the upward trend continued, reaching a relatively high level (approximately 7.8‰) in 2023. The marked acceleration in mortality growth around 2019–2020 coincides with the onset of the COVID-19 pandemic, which likely contributed to excess mortality directly and indirectly through health-care system disruption. Subsequent fluctuations may reflect the evolving pandemic impact and recovery patterns. When considering the growth rate trends of the PMR, it becomes evident that the changes exhibit significant fluctuations. From 2010 to 2023, the growth rate demonstrated considerable volatility overall. The growth rate peaked in 2020, subsequently experienced a decline,

**Table 1.** Variables and definitions.

Variable	Unit	Definition
PMR	‰	Population Mortality Rate
NUPP	per 10,000 People	Number of Urban Practicing (assistant) Physicians per 10000 people
NMHI	number	Number of Medical and Health Institutions
NBMI	number	Number of Beds per 10000 people in Medical Institutions
SDE	10000 tons	Sulphur Dioxide Emissions
CODE	10000 tons	Chemical Oxygen Demand Emissions
PCWR	cubic meters/person	Per Capita Water Resources
HTHW	%	Harmless Treatment rate of Household Waste
LFEEP	100 million CNY	Local Fiscal Expenditure on Environmental Protection
NDPAA	item	Number of Domestic Patent Applications Accepted
UR	%	Urbanization Rate

**Table 2.** Variable description statistics.

Variable	Minimum	Q1	Median	Mean	Q3	Maximum
PMR	4.210	5.630	6.240	6.383	7.010	9.840
NUPP	16.00	32.00	38.00	39.18	44.00	75.00
NMHI	4129	18425	28050	32059	37571	92825
NBMI	25.10	44.91	54.66	55.18	65.64	89.26
SDE	0.110	8.512	19.595	36.663	55.578	182.740
CODE	1.76	18.56	43.52	60.09	90.21	198.25
PCWR	51.9	429.8	1697.2	6421.4	3154.2	153681.9
HTHW	38.00	91.60	99.20	93.32	100.00	100.00
LFEEP	11.77	77.80	121.47	146.34	190.88	747.44
NDPAA	162	17597	49112	108568	127987	993480
UR	22.65	50.72	58.37	59.06	65.70	89.58

but then rose again. Moreover, the growth rates varied markedly across different years, revealing an unstable and fluctuating growth pattern.

### Spatial distribution of PMR

Significant disparities in population characteristics exist across Chinese provinces, with PMR demonstrating notable spatial heterogeneity. Figure 3 presents the top 15 provinces ranked by average PMR from 2010 to 2023. As shown in Figure 3, Sichuan Province exhibits the highest average mortality rate, followed by Hunan Province and Chongqing Municipality. The differences among Shandong, Guizhou, and Heilongjiang provinces are relatively minor.

Geographically, among the regions with higher average mortality rates, except for Liaoning, Shandong, and Hebei provinces, the remaining twelve are predominantly located in central and western China. Potential explanations for this pattern include: in mountainous provinces such as Sichuan, Hunan, Chongqing, and Guizhou, complex topography accelerates outmigration of working-age adults, resulting in elevated aging rates. Concurrently, medical resources remain concentrated in urban centres, creating barriers to healthcare access in rural mountainous areas. This combination leads to delayed management of chronic conditions among the elderly, thereby increasing mortality. In north-eastern industrial bases like Liaoning and Heilongjiang, population decline exacerbates aging, while harsh winter conditions amplify cardiovascular and cerebrovascular disease risks, creating a dual burden of aging and elevated mortality. Although Shandong and Henan are situated in the Central Plains, their large populations contribute to relatively low per capita healthcare resources and insufficient primary care capacity, indirectly affecting mortality outcomes. North-western provinces such as Gansu face heightened respiratory disease risks due to arid climates and winter air pollution. These provinces, whether due to topography, regional positioning, or climatic factors, experience overlapping challenges in demographic structure, medical resource distribution, and health risks, collectively driving their mortality rates to rank among the highest nationally.

## GAM evaluation

### Correlation analysis

To further investigate the associations between various variables and public health, this study first conducted correlation analysis on relevant variables. Figure 4 presents a heatmap of correlation coefficients between each variable and the PMR, where colour intensity visually reflects the strength of correlations. The analysis reveals distinct patterns across variable categories: Among medical resource variables, NBMI shows a moderate positive correlation with PMR ( $r = 0.59$ ), whereas NUPP exhibits a weak negative correlation ( $r = -0.14$ ). For environmental indicators, both SDE ( $r = -0.12$ ) and CODE ( $r = 0.37$ ) display weak linear correlations with PMR. Most socioeconomic and technological development indicators, including PCWR ( $r = -0.20$ ), UR ( $r = -0.03$ ), and NDPAA ( $r = -0.02$ ), show very weak or negligible linear associations with PMR. However, it is important to note that correlation analysis can only characterize linear co-variation trends between variables, neither establishing causal relationships nor discerning complex nonlinear dependency structures. To overcome these limitations and further reveal the mechanisms through which various variables influence PMR, this study adopts a regression analysis framework and specifically selects the GAM from the field of machine learning, which is renowned for its statistical theoretical foundation. By incorporating flexible smoothing terms, the model effectively captures potential nonlinear relationships among variables, thereby revealing their association patterns more authentically. Compared to linear models with fixed parametric forms, GAM achieves an optimal balance between functionality and interpretability, providing a more scientific and refined basis for public health decision-making.

### GAM results

Table 3 presents the fitting results of the GAM model. From a machine learning perspective, the p-values reflect the significance of each feature’s explanatory power for the prediction target. With the exception of LFEEP, the smooth terms of all other variables have p-values less than 0.05, indicating that the selected variables have significant effects on PMR. The Effective Degree of Freedom (EDF) serves as an important metric for evaluating variables’ con-

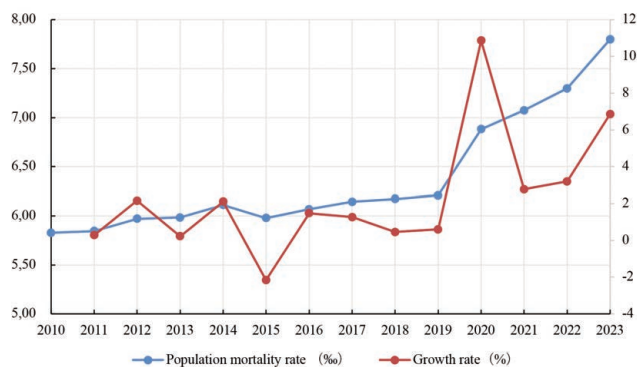


Figure 2. Time-Series evolution of PMR and growth rate from 2010 to 2023.

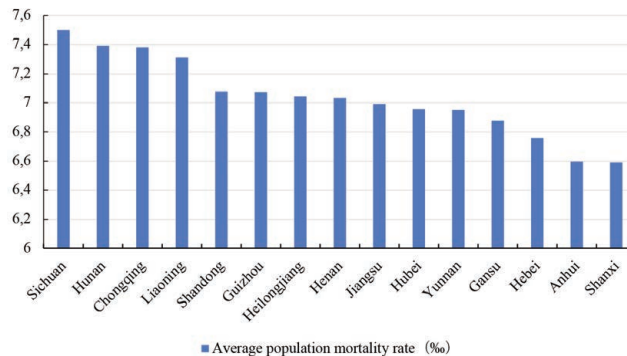


Figure 3. Top 15 provinces in average PMR from 2010 to 2023.

tribution to the model. If a smooth term has an EDF greater than 1 and is statistically significant, it demonstrates the existence of a nonlinear relationship between that variable and the response variable, aligning with machine learning’s requirement for modelling complex relationships. In this model, as shown in Table 3, with the exception of HTHW, all other smooth terms exhibit EDF values greater than 1. The ranking of variables by their contribution to the model is as follows: SDE (8.809), NMHI (8.318), PCWR (7.923), NDPAA (7.523), UR (7.340), NUPP (6.996), NBMI (6.076), CODE (4.691), LFEEP (2.679), and HTHW (1.000). The GAM-adjusted R<sup>2</sup> value fitted in this paper is 0.8, with deviance explained of 82.6%. These metrics indicate that the model possesses strong interpretability and goodness-of-fit, effectively capturing the complex nonlinear relationships between the variables and PMR.

### Impacts of various variables on PMR

#### The impact of NUPP on PMR

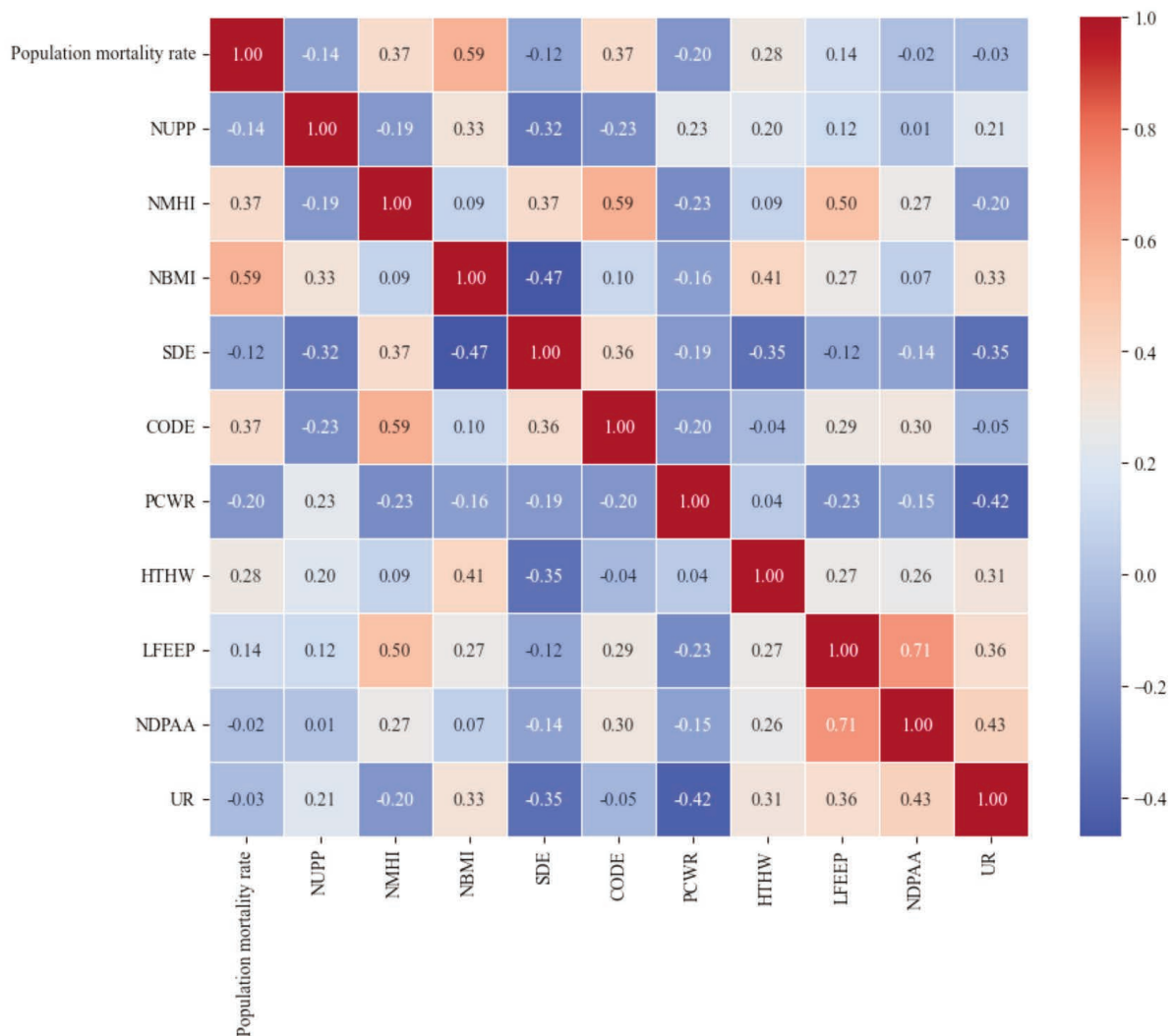
As shown in Figure 5(a), the impact of NUPP on population mortality can be divided into two phases based on the slope changes of the curve: the 20-40 range and the 40-70 range. In the first phase (20-40), the curve shows a remarkably significant

downward trend, indicating that increasing the number of physicians in this interval most effectively reduces population mortality. It is noteworthy that in the second phase (40-70), the downward trend of the curve noticeably slows and levels off, demonstrating

**Table 3.** GAM results.

Variable	EDF	F	p
S(NUPP)	6.996	5.717	8.66e-07
S(NMHI)	8.318	8.337	<2e-16
S(NBMI)	6.076	26.517	<2e-16
S(SDE)	8.809	8.486	<2e-16
S(CODE)	4.691	5.743	2.00e-05
S(PCWR)	7.923	9.405	<2e-16
S(HTHW)	1.000	8.646	0.00349
S(LFEEP)	2.679	2.413	0.05697
S(NDPAA)	7.523	4.839	7.00e-06
S(UR)	7.340	5.634	1.00e-06

EDF, Effective degree of freedom.



**Figure 4.** Heat map of variable correlation.

the marginal diminishing returns in health benefits. NUPP serves as a core indicator for measuring the allocation of regional medical human resources, and its relationship with public health is not simply linear. Figure 5(a) clearly demonstrates that when NUPP exceeds 40, the health benefits gained from further increasing the number of physicians become highly limited. Therefore, in regions where resources have not yet reached the standard, the focus should be on overcoming resource bottlenecks and striving to achieve the critical level of around 40 for NUPP. For regions with already sufficient resources ( $NUPP > 40$ ), the strategic emphasis should shift from expanding scale to improving quality, optimizing service structures, and enhancing service efficiency to unlock health potential, thereby maximizing the benefits of resource allocation.

### *The impact of NMHI on PMR*

As shown in Figure 5(b), NMHI exhibits a significant nonlinear relationship with population mortality. When NMHI is at relatively low levels, although the curve shows some fluctuations, mortality generally decreases with increasing resources. However, when NMHI exceeds the threshold of approximately 40,000, mortality begins to show an upward trend as resources continue to increase. This indicates that expanding healthcare resource coverage during the resource-scarce phase is the most efficient strategy for reducing mortality, with the most notable marginal health benefits. Conversely, once the number of institutions surpasses the saturation threshold, continued indiscriminate expansion may lead to diseconomies of scale due to resource dilution, insufficient service coordination, or redundant construction, ultimately adversely affecting health outcomes.

As a key indicator for measuring regional medical resource density, the nonlinear relationship between NMHI and mortality rates profoundly reveals the complexity of resource allocation efficiency. Given its non-monotonic linear influence and clear characteristics of phased heterogeneity and diminishing marginal returns, regional health policies must adopt differentiated strategies based on response patterns at different development stages. For low-resource regions, priority should be given to expanding effective supply and rapidly improving basic medical accessibility. For regions where the number of healthcare institutions approaches or exceeds the threshold, the policy focus should shift from scale expansion to system integration and efficiency enhancement. By optimizing structures and strengthening coordination, health benefits can be maximized, ultimately achieving systematic optimization of public health resources and sustained improvement in population health levels.

### *The impact of NBMI on PMR*

As shown in Figure 5(c), the effect of NBMI on population mortality exhibits significant nonlinear characteristics, with the overall curve shape revealing a pattern of diminishing marginal returns from resource allocation. Notably, the curve demonstrates its steepest upward trajectory in the NBMI range of 50-60, indicating that this interval represents a critical zone where resource pressure and health risks interact synergistically to amplify mortality. As a crucial reflection of healthcare resource supply capacity, NBMI plays a significant role in safeguarding public health. However, in the process of increasing bed resources to improve health outcomes, resource increments do not uniformly enhance impact effects across all ranges. Despite continuous growth in the number of beds, constraints such as regional aging trends, shifts in disease spectrum, or inefficiencies in the medical service system may limit their effectiveness. Nevertheless, as illustrated in Figure

5(c), the NBMI range of 50-60 exhibits the most pronounced increase in risk, underscoring the need to prioritize this interval in resource planning and risk prevention strategies.

### *The impact of SDE on PMR*

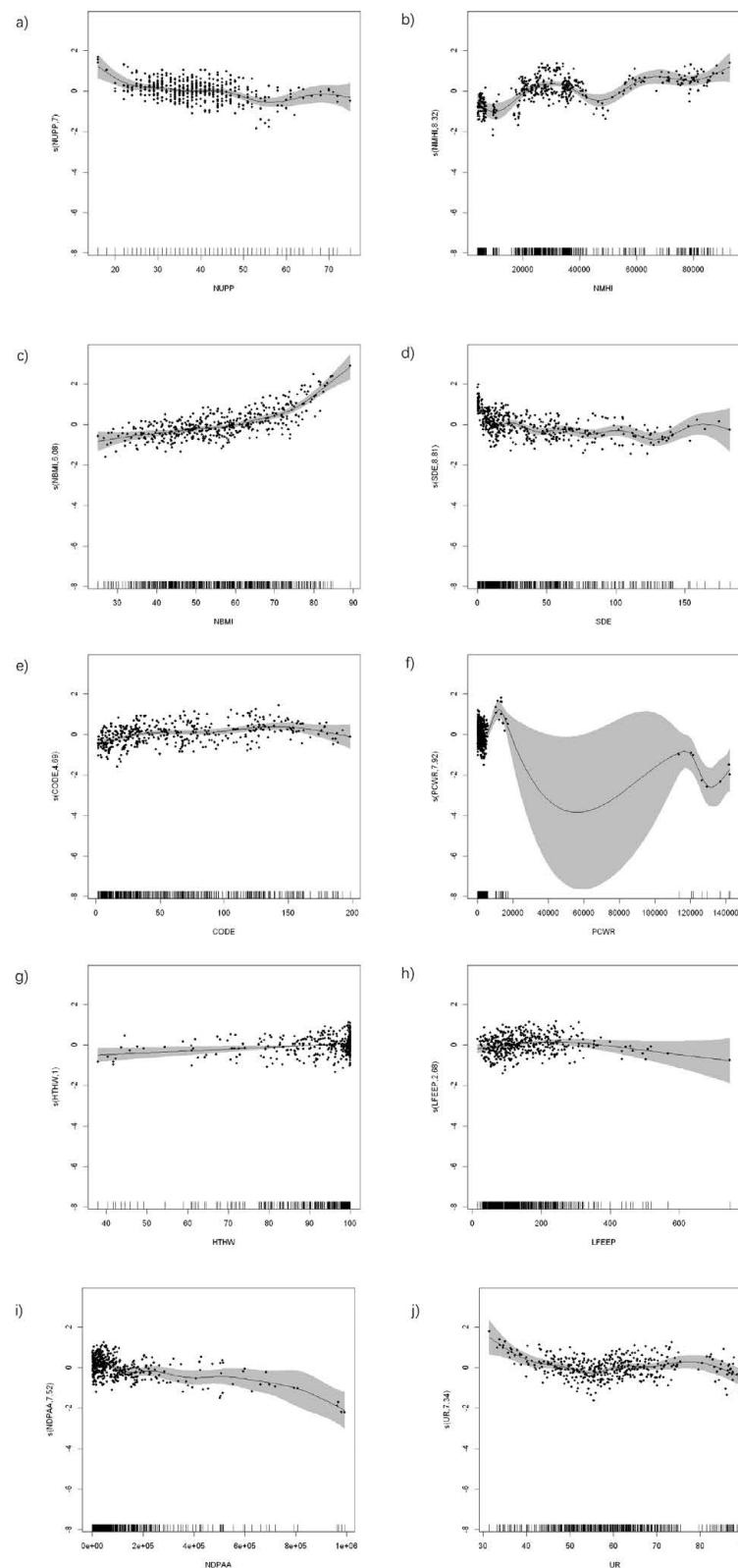
As shown in Figure 5(d), SDE exhibits a complex nonlinear relationship with population mortality. Specifically, within the range where SDE levels are below 50, a significant negative correlation with mortality is observed. However, when SDE exceeds this threshold, the strength of this association substantially diminishes. As a key indicator of air pollution, SDE has significant implications for public health. Figure 5(d) suggests the potential presence of insufficiently controlled confounding factors. For instance, SDE may be highly correlated with variables such as regional industrialization progress and economic development levels, which themselves exert potentially positive effects on health. The observed negative correlation at lower emission levels may reflect concurrent improvements in socioeconomic conditions and healthcare infrastructure that accompany early industrial development, which could offset the direct adverse health effects of emissions. This highlights the complex interplay between pollution and development factors. During the process of industrial development, increases in sulphur dioxide emissions may coincide with improvements in regional healthcare conditions, household income, and social security levels. The beneficial health effects of these factors could partially mask the adverse health impacts of sulphur dioxide emissions. This finding underscores the critical importance of controlling for socioeconomic and other confounding variables in environmental health research. It also highlights the need for a more comprehensive consideration of the interactions among various socioeconomic factors in environmental policy formulation, so as to avoid oversimplified conclusions regarding the health effects of pollutants.

### *The impact of CODE on PMR*

As shown in Figure 5(e), the impact of CODE on population mortality exhibits nonlinear fluctuations. Notably, when the CODE value is below 50, the curve demonstrates a distinct upward trend. As a key indicator of organic water pollution, the observed association pattern of CODE presents complexity. The upward trend in the low concentration range may be linked to characteristics of the early stages of regional industrialization—a phase typically marked by the initial emergence of environmental pollution, while corresponding environmental protection infrastructure and health protection measures remain underdeveloped. Figure 5(e) suggests that for regions with low CODE levels but undergoing rapid industrialization, it is crucial to prioritize the balance between economic development and environmental protection. Strengthening environmental regulation and health protection measures at an early stage of industrialization may represent the most cost-effective strategy for health risk prevention and control.

### *The impact of PCWR on PMR*

As shown in Figure 5(f), PCWR demonstrates a complex nonlinear relationship with population mortality. Analysis of the PCWR data distribution reveals that the per capita water resource allocation primarily clusters within the 20,000 to 140,000 range. Particularly noteworthy is the persistence of higher PMR when per capita water resources fall below approximately 20,000. PCWR serve as a fundamental indicator for assessing regional water resource adequacy and exert a foundational impact on public health. Sufficient water availability is a prerequisite for ensuring drinking water safety, maintaining adequate sanitary conditions,



**Figure 5.** **a)** The impact of Number of Urban Practicing (assistant) Physicians per 10000 people on Population Mortality Rate; **b)** The impact of Number of Medical and Health Institutions on Population Mortality Rate; **c)** The impact of Number of Beds per 10000 people in Medical Institutions on Population Mortality Rate; **d)** The impact of Sulphur Dioxide Emissions on Population Mortality Rate; **e)** The impact of Chemical Oxygen Demand Emissions on Population Mortality Rate; **f)** The impact of Per Capita Water Resources on Population Mortality Rate; **g)** The impact of Harmless Treatment rate of Household Waste on Population Mortality Rate; **h)** The impact of Local Fiscal Expenditure on Environmental Protection on Population Mortality Rate; **i)** The impact of Number of Domestic Patent Applications Accepted on Population Mortality Rate; **j)** The impact of Urbanization Rate on Population Mortality Rate.

and supporting agricultural production. The findings demonstrate that maintaining PCWR above 20,000 cubic meters per person represents a basic requirement for mitigating health risks associated with water scarcity.

### *The impact of HTHW on PMR*

As shown in Figure 5(g), the impact of HTHW on population mortality generally exhibits a relatively flat, approximately linear relationship. Specifically, the smoothed curve remains largely stable overall, displaying an extremely gradual upward trend.

As a key indicator of environmental infrastructure, the marginal diminishing trend in the health benefits of HTHW warrants attention. Figure 5(g) indicates a stable association between increased HTHW and reduced mortality risk, yet the protective benefits show a slight pattern of marginal diminishment as the treatment rate rises. When the treatment rate has already reached a high level, this factor may primarily function as a foundational safeguard. To achieve further breakthroughs in health outcomes, policy focus may need to shift from solely increasing the treatment rate to building a more comprehensive health promotion system. This includes multidimensional measures such as optimizing the allocation of medical resources and strengthening the prevention and control of chronic diseases.

### *The impact of LFEEP on PMR*

As shown in Figure 5(h), LFEEP demonstrates a certain degree of nonlinear fluctuation in its impact on population mortality. Analysis of the LFEEP data distribution reveals that when LFEEP exceeds 200, the curve begins to show a declining trend. LFEEP serves as a key metric for evaluating local government investment in environmental governance, demonstrating a clear correlation with the improvement of public health levels. Figure 5(h) reveals that increasing and maintaining LFEEP above the threshold of 200 effectively enhances its health-protective role. Therefore, for regions with lower environmental protection expenditure levels, raising investment to this critical threshold should be prioritized as a strategy to mitigate residents' health risks.

### *The impact of NDPAA on PMR*

As shown in Figure 5(i), NDPAA exhibits a complex nonlinear relationship with population mortality. Analysis of the NDPAA data distribution shows that when the number of patent applications remains at very low levels (below approximately 200,000), the curve displays a pronounced decline, suggesting that enhancing innovation capacity within this range serves as an effective strategy for reducing PMR. As a key indicator measuring regional technological innovation capability and vitality, NDPAA exerts a profound impact on public health. Figure 5(i) clearly demonstrates that during phases of limited innovation resources, knowledge accumulation and technology spillover exert a strong driving force in improving public health outcomes. Therefore, for regions still in the early stages of innovation capacity development, increasing investment in science and technology can yield the greatest health returns.

### *The impact of UR on PMR*

As shown in Figure 5(j), UR demonstrates a significant nonlinear relationship with population mortality. Analysis of the UR data distribution reveals a declining trend in mortality as urbanization rate approaches the 50%-60% range, indicating an association between urbanization progress and reduced mortality risk during this phase. As a key indicator of socioeconomic development and structural transformation, the Urbanization Rate (UR) exhibits

complex phase-specific characteristics in its relationship with public health. During the early and middle stages of urbanization, resource concentration and infrastructure development likely drive health gains, whereas at higher urbanization levels, "urban disease" effects—such as increased life stress, environmental pollution, and rising chronic disease risks—may become more pronounced. Therefore, public policy should account for the varying health impact mechanisms across urbanization stages. In the middle to late phases, efforts should focus on mitigating health pressures associated with high-density urban living through optimized urban governance, improved environmental quality, and enhanced chronic disease prevention and control, thereby achieving the goal of healthy urbanization.

## Conclusions

This study, applying the GAM model utilizing data from 31 Chinese provinces (excluding Hong Kong, Macao, and Taiwan) from 2010 to 2023 reveals pervasive complex nonlinear association mechanisms between multidimensional factors and public health in China. The main conclusions are as follows: First, healthcare resources exhibit distinct threshold effects and diminishing marginal returns on public health outcomes. When NUPP reaches approximately 40 physicians per 10,000 people, the growth rate of health benefits slows significantly. Furthermore, when NMHI exceeds around 40,000 institutions, any additional increase is associated with elevated mortality risk. These findings highlight the necessity for healthcare resource allocation to shift from quantitative expansion toward structural optimization and efficiency enhancement. Second, the relationship between environmental pollution indicators and mortality is substantially confounded by socioeconomic factors, resulting in statistical associations that contradict theoretical expectations. SDE demonstrates a negative correlation with mortality across the observed range, while CODE exhibits multi-stage fluctuations in its effects. This strongly suggests the presence of inadequately controlled confounding factors in the model—such as simultaneously occurring economic development and improvements in healthcare standards. Subsequent research should employ more rigorous causal inference frameworks to elucidate the true health effects of these environmental factors. Third, basic resources and development factors exhibit phased transition characteristics in their health impacts. When PCWR falls below 20,000, mortality risk remains elevated, while exceeding this threshold reveals discernible health benefits. Meanwhile, UR demonstrates a directional transition in its effects within the 50-60 range, reflecting a fundamental shift in the relationship between urbanization and population health outcomes.

Based on the findings, the following actionable recommendations are proposed: First, in terms of healthcare resource allocation, differentiated and targeted strategies should be adopted: For regions with NUPP below 40, efforts should focus on strengthening the primary healthcare workforce and rapidly improving the accessibility of basic medical services. For regions where NMHI exceeds 40,000, the policy focus should shift from scale expansion to resource integration and service quality enhancement. This can be achieved through the establishment of medical alliances and implementation of tiered healthcare systems to improve resource allocation efficiency. Second, in the field of environmental governance, a more scientific assessment framework needs to be established. Given the complex relationship between environmental pollution indicators and mortality rates, it is recommended to adopt

health impact assessment methods in environmental policy formulation, fully considering the interactive effects of socioeconomic and other confounding factors. Special emphasis should be placed on strengthening environmental supervision in regions experiencing rapid industrialization but lagging in environmental infrastructure, so as to achieve synergistic progress in economic development and health protection. Third, in urban-rural development and basic resource assurance, key threshold effects must be prioritized. For regions with PCWR below 20,000, ensuring drinking water safety and basic sanitary water supply should be a primary focus. In areas where UR exceeds 50%, targeted efforts are needed to address health pressures stemming from “urban ailments,” such as by optimizing urban governance and enhancing health service facilities to consolidate the health benefits of urbanization. Fourth, in guiding technological innovation, emphasis should be placed on innovation quality and health orientation. In regions where NDPAA remains below 200,000, continued expansion of innovation scale can be encouraged. For areas with already vibrant innovation activities, science and technology innovation resources should be directed toward fields that directly impact public health—such as medical health and environmental governance—to promote the effective translation of technological achievements into tangible health benefits. Through the implementation of these multidimensional and differentiated systematic governance strategies, a health-oriented public policy system can be established, providing a scientific pathway to achieve the strategic goals of “Healthy China 2030.”

## References

- Adesuyi EO, Olabode OO, Olarinde OC, Oyama BO, Aderemi IO, Alao GO, Ogunlowo BC, 2025. Factors Influencing Primary Healthcare System in the Achievement of Universal Health Coverage in the WHO-AFRO Region: An Integrative Review. *Public Health Chall* 4:e70081.
- Díaz Martínez Z, Fernández Menéndez J, García Villalba LJ, 2023. Tariff analysis in automobile insurance: is it time to switch from generalized linear models to generalized additive models? *Mathematics* 11:3906.
- Heudorf U, Kowall B, 2025. Mortality in Frankfurt am Main, Germany, 2020–2023: Higher excess mortality during an influenza wave in 2022 than during all COVID-19 waves altogether. *GMS Hyg Infect Control* 20:Doc04.
- Kalhorji S, Abolhasani L, Sabouhi M, Sarkhosh M, 2024. The Estimation of Willingness to Pay and the Influential Factors for Air Quality Improvement in the City of Mashhad. *J Environ Stud* 38:243–60.
- Li Z, Wang W, Ji X, Wu P, Zhuo L, 2023. Machine learning modeling of water footprint in crop production distinguishing water supply and irrigation method scenarios. *J Hydrol* 625:130171.
- Lu K, Jiang X, Lv G, Yuan J, 2025. Trends, burden, and risk factors of stomach cancer in China, 1990–2021. *Lancet Reg Health West Pac* 55:101428.
- Merera A, Asena T, Senbeta M, 2022. Bayesian multilevel analysis of determinants of acute respiratory infection in children under the age of five years in Ethiopia. *BMC Pediatr* 22:123.
- Miao LP, Ren KH, Li MD, Lyv JC, 2024. Trend analysis and prediction of cardiovascular disease mortality in China from 2009 to 2021. *Chin Gen Pract* 27:2260–4.
- Mohammadi-Pirouz Z, Hajian-Tilaki K, Sadeghi Haddat-Zavareh M, Amoozadeh A, Bahrami S, 2024. Development of decision tree classification algorithms in predicting mortality of COVID-19 patients. *Int J Emerg Med* 17:126.
- Otieno WP, Ankaopong EAK, Nomura K, Matsumoto-Takahashi ELA, 2024. Higher temperatures and being an ethnic minority reduce mosquito net use in Lao PDR: an analysis of Lao PDR’s Multiple Indicator Cluster Survey and Earth observation satellite data. *Trop Med Health* 52:99.
- Pan HF, Tang YQ, Zhu HY, Sun YW, Chi P, Huang Y, 2025. Global burden, trends, and risk factors of early-onset and late-onset colorectal cancer from 1990 to 2021, with projections to 2040: a population-based study. *BMC Gastroenterol* 25:486.
- Prasetyo YB, Ramadan SW, Wardoyo SSI, Dewi YS, Huriah T, 2024. Understanding enablers and barriers in Indonesia’s non-communicable disease program: A qualitative study from nurses’ perspectives. *Int J Noncommun Dis* 9:166–175.
- Rahman MM, Hossain MS, Nakamura R, Alam MA, Azmaf SK, Sato M, 2025. Unmet healthcare needs, out-of-pocket payments and catastrophic health expenditures among hypertensive adults in Bangladesh. *BMC Health Serv Res* 25:1088.
- Ravindra K, Rattan P, Mor S, Aggarwal AN, 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. *Environ Int* 132:104987.
- Shu P, Wang X, Wen Z, Chen J, Xu F, 2025. SHAP combined with machine learning to predict mortality risk in maintenance hemodialysis patients: a retrospective study. *Front Med* 12:1615950.
- Sofiyah ES, Sianipar IMJ, Rahman A, Rafida HA, Suhardono S, Lee C-H, Suryawan IWK, 2025. Advancing the 2030 agenda with community importance-performance perspective and public relations strategies for community-based sanitation. *Sustain Futures* 9:100633.
- Suo Z, Shao L, Lang Y, 2023. A study on the factors influencing the utilization of public health services by China’s migrant population based on the Shapley value method. *BMC Public Health* 23:2328.
- Wu H, Wang Y, Zhang H, Yin X, Wang L, Wang L, Wu J, 2024. An investigation into the health status of the elderly population in China and the obstacles to achieving healthy aging. *Sci Rep* 14:31123.
- Wu S, Yi D, 2025. Improving nurses’ emergency response competence through the clinical rotation training program. *J Integr Nurs* 7:137–46.
- Xiang LC, Yamada M, Feng W, Li D, Nie HS, 2023. Spatial variations and influencing factors of Cumulative Health Deficit Index of elderly in China. *J Health Popul Nutr* 42:66.
- Yu Q, He Y, Li L, Xu D, Li Y, Yin Q, Lu Y, He M, Ma S, 2025. Study on the willingness to enroll in the urban and rural residents basic medical insurance: from the perspectives of policy awareness and institutional trust. *Front Public Health* 13:1535558.
- Zhang J, Xiong W, Zhang C, Huang C, Li W, Liu L, Wang W, Sang Y, Zhen H, Tan C, Yang J, Yang J, 2025a. Enhanced machine learning models for predicting three-year mortality in Non-STEMI patients aged 75 and above. *BMC Geriatr* 25:1–12.
- Zhang X, Xu Q, Li C, Zhang Y, Jin Y, Li P, Guo H, 2025b. Prevalence of low-level viremia and related influencing factors among people living with HIV in China: a systematic review and meta-analysis. *Front Public Health* 13:1661253.
- Zheng X, Tian F, Chen Y, Xue B, Shi L, Zhang C, 2023. Research Hotspots and Evolutionary Trends on Multimorbidity in China from 2002 to 2022. *Chin Gen Pract* 26:2567–73.

---

Received: 7 October 2025; Accepted: 17 January 2026.

Contributions: Dehong Sun designs the experimental process, collects data, analyzes data, writes articles, and Hui Wang creates charts and graphs.

Conflict of interest: the authors declare that they have no competing interests.

Ethics approval: not applicable.

Availability of data and materials: the data used in this study were obtained from the China Statistical Yearbook (<http://www.stats.gov.cn/tjsj/ndsj/>). The dataset is publicly accessible.

Funding: this study was supported by the Fujian Provincial Science and Technology Department (Grant No. 2024H0038) and Technology Innovation Team of Minnan University of Technology (Grant No. 23XTD114).

*Publisher's note: all claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.*