



OPEN

Research on social and economic factors influencing regional mortality patterns in China

Tiantian Li, Shuyin Zhang & Handong Li✉

Regional population mortality correlates with regional socioeconomic development. This study aimed to identify the key socioeconomic factors influencing mortality patterns in Chinese provinces. Using data from the Seventh Population Census, we analyzed mortality patterns by gender and urban–rural division in 31 provinces. Using a functional regression model, we assessed the influence of fourteen indicators on mortality patterns. Main findings: (1) China shows notable gender and urban–rural mortality variations across age groups. Males generally have higher mortality than females, and rural areas experience elevated mortality rates compared to urban areas. Mortality in individuals younger than 40 years is influenced mainly by urban–rural factors, with gender becoming more noticeable in the 40–84 age group. (2) The substantial marginal impact of socioeconomic factors on mortality patterns generally becomes evident after the age of 45, with less pronounced differences in their impact on early-life mortality patterns. (3) Various factors have age-specific impacts on mortality. Education has a negative effect on mortality in individuals aged 0–29, extending to those aged 30–59 and diminishing in older age groups. Urbanization positively influences the probability of death in individuals aged 45–54 years, while the impact of traffic accidents increases with age. Among elderly people, the effect of socioeconomic variables is smaller, highlighting the intricate and heterogeneous nature of these influences and acknowledging certain limitations.

Keywords Mortality patterns, Regional disparities, Determining factors, Functional regression

The distribution of mortality probability by age defines the mortality pattern of a population. Regional population mortality patterns portray the probability distribution of deaths across various age groups in a specific geographic area during a given period. Typically, population censuses categorize individuals into age groups such as 0 years, 1–4 years, ..., 95–99 years, and 100 years and above, for a total of 22 groups. Mortality statistics for each age group are collected, and age-specific mortality rates are converted into probabilities to derive population mortality patterns. Population mortality patterns, influenced by biomedical, demographic, socioeconomic, and environmental factors, are crucial indicators reflecting demographic changes. Among these factors, socioeconomic factors play a significant role in shaping regional population mortality patterns and constitute a key focus of demographic studies.

China, characterized by a vast population and distinct urban–rural dual social structure with extreme regional economic disparities, offers different opportunities and choices for accessing social resources and improving quality of life. This, in turn, impacts mortality patterns in regional populations. This paper aims to answer key questions: What are the primary socioeconomic factors influencing disparities in regional mortality patterns in China? How do these factors affect the distribution of age-specific mortality probabilities in regional populations, and to what extent? Addressing these questions is the primary focus of this research.

The results of this paper provide an important decision-making basis for the government to formulate regional development policies, rationally allocate public medical and health resources, promote sustainable regional social and economic development, and improve regional population health inequality.

The remaining sections of this paper are structured as follows: Section “Literature review” presents a literature review, section “Research method and variable selection” details the research methods and variable selection, section “Empirical analysis” provides the empirical analysis, and the final section contains the conclusion.

School of Systems Science, Beijing Normal University, Beijing, China. ✉email: lhd@bnu.edu.cn

Literature review

The exploration of population mortality and its determinants is a key area in population research, yielding extensive findings. From a socioeconomic perspective, research on regional mortality involves understanding and contrasting patterns, examining socioeconomic influences, and assessing methodological approaches.

Characteristics and comparisons of regional mortality patterns

To explore the temporal and spatial relationships of regional mortality patterns, numerous scholars have conducted comprehensive and extensive research on mortality patterns in various countries and regions, including the United States, Germany, New Zealand, Russia, South Korea, and China, among others, yielding substantial results.

Ivanova & Zhong analyzed regional disparities in male mortality rate changes in Russia, revealing significant stratification¹. Kibele et al. studied German regional mortality rates from 1910 to 2009, noting substantial changes over the century². Van Raalte et al. used Germany as an example and reported that significant regional economic inequality does not necessarily parallel disparities in mortality rates³. Spencer et al. decomposed differences in mortality rates between US urban and rural counties, attributing the widening gap to changes in how characteristics influence mortality⁴.

Many studies have also focused on changes in mortality rates and their distribution in China. Cao et al. reported varying child mortality rates across 81 cities and counties, with higher rates for males overall but a reversal in impoverished areas⁵. Nie and Song reported an uneven regional distribution of infant mortality rates, increasing from coastal to inland and then to remote regions⁶. Li compared mortality levels in North and East China⁷. Zhou, using cluster analysis, explored regional variations, finding influences of both geographical location and economic conditions⁸. Chen et al. used spatial analysis to characterize the spatiotemporal patterns of urban population mortality rates in China, revealing distinct spatial heterogeneity and correlations⁹.

Factors influencing mortality patterns research

Mackenbach et al. conducted a regression-based analysis comparing inequality indices in mortality and self-rated health across 22 European countries, revealing higher mortality rates and poorer self-rated health in socioeconomically disadvantaged groups¹⁰. Kibele et al. found that factors contributing to spatial variations in mortality in Germany changed significantly over time, with a strengthened connection between regional socioeconomic conditions and mortality². Mackenbach's review of theories explaining social inequality suggested that health inequality plays a role in exacerbating social inequality¹¹. Lutz and Kebede's study across 174 countries from 1970 to 2010 found education to be a better predictor, linking improved education to increased income and better health outcomes¹². Gutin and Hummer's review highlighted the association between social status and health, where high socioeconomic status individuals may extend their lives beyond biological expectations, while disadvantages due to social status could lead to premature mortality for lower-status individuals¹³.

Several studies have investigated the intricacies of factors affecting mortality rates in China. Zhao's macrolevel exploration identified socioeconomic factors impacting post-1949 mortality rates¹⁴. Nie & Song empirically analyzed the factors that influenced infant mortality rates between 1996 and 2002⁶. Luo & Xie uncovered contextual determinants for the influence of socioeconomic indicators on elderly mortality, highlighting the nuanced sociopolitical landscape of China¹⁵. Li's work revealed the significant impact of healthcare resource investment on regional disparities in elderly health¹⁶. Chen et al. conducted a comprehensive investigation into various factors shaping population mortality, spanning from wastewater pollution to healthcare improvements¹⁷. Fan's latent socioeconomic status variable stressed the pivotal role of higher socioeconomic status in mitigating mortality risk among Chinese individuals aged 65 and above¹⁸. Chen et al.'s analysis of PM2.5 pollution demonstrated a positive correlation with mortality rates, coupled with spatial spillover effects⁹. Ying & Li highlighted the prominent role of population aging, industrial emissions, climatic conditions, and per capita GDP in shaping mortality rates¹⁹.

Research methods for regional mortality patterns

In the study of mortality patterns and the influencing factors, scholars often use statistical regression models and similar methodologies. For instance, Lutz & Kebede employed multivariate statistical analysis to explore the impacts of education and income on global mortality patterns¹². The Cox proportional hazards model is a common choice, as seen in the work of Luo & Xie, who used it to study the mechanisms through which socioeconomic factors influence mortality rates in China¹⁵. Fan also utilized the Cox proportional hazards regression model to compute hazard ratios and conducted subgroup analyses based on various factors¹⁸. Other techniques, such as multiple regression analysis⁹, global spatial autocorrelation models, local spatial autocorrelation models, and spatial regression analysis methods¹⁷, are employed to investigate various aspects of regional mortality patterns.

In summary, current research on regional mortality disparities in China faces several limitations. First, prior studies often attribute differences in mortality rates to a limited set of socioeconomic factors, and a comprehensive analysis of the overarching socioeconomic forces is lacking. Second, there is a predominant focus on socioeconomic impacts within specific age groups, with insufficient analyses covering the entire mortality pattern. Third, most studies use classical statistical regression analyses, while this paper adopts more recent functional regression models, a method still underutilized in population studies.

Research method and variable selection

In this study, a mortality pattern was employed as a measure of regional population mortality levels. The mortality pattern represents the distribution of age-specific mortality probabilities and can be considered a function of population age. We utilized the function-on-scalar functional regression model for our analysis, drawing

inspiration from the foundational principles outlined in the works of J.O. Ramsay and B.W. Silverman^{20,21} and relevant theoretical literature^{22–29}. The basic principles of the functional regression model are outlined as follows.

Representation of functional data

The subject of functional data analysis deals with a set of smooth curves, denoted as $\{y_n^0(t), t \in [T_0, T_1], 1 \leq n \leq N\}$, comprising N such curves. For a given sample curve, we can only observe a finite number of points, i.e., $\{y_{n,j}^0 \in R, 1 \leq n \leq N, 1 \leq j \leq J_n\}$, with each curve having J_n observed points. Functional data possess infinite dimensional characteristics, and for the sake of data modeling and statistical inference, dimension reduction techniques are often applied. One common approach is to expand the functional data using basis functions. Given the known discrete points, fitting with basis functions allows the representation of the original functional data as a linear combination of these basis functions. Let $\{\phi_k, 1 \leq k \leq K\}$ be the basis functions, then the fitted function can be expressed as follows:

$$y_n(t) \approx \sum_{k=1}^K c_{n,k} \phi_k(t), 1 \leq n \leq N \tag{1}$$

where $c_{n,k}$ represents the coefficient for each basis function $\phi_k(t)$, and the original data can be expressed as follows:

$$y_{n,j}^0 = y_n(t_{j,n}) + \varepsilon_{n,j}, 1 \leq n \leq N, t_{j,n} \in [T_0, T_1], 1 \leq j \leq J_n \tag{2}$$

where $\varepsilon_{n,j}$ represents the error terms, which are independently and identically distributed with a mean of zero and a variance of σ^2 .

The basis functions can be specified as periodic Fourier bases, nonperiodic B-spline bases, or wavelet bases. Alternatively, data-driven functional principal component bases can be selected, but this necessitates performing principal component analysis (PCA) on the functional data beforehand.

If the original data are affected by a significant level of noise, the expansion of the basis functions becomes increasingly distorted as the number of basis functions increases. Having a relatively small number of basis functions, on the other hand, will also limit the shape of their linear combinations. Therefore, we employ a roughness penalty method to smooth the functional data:

$$PSS_\lambda(c_{n,k}) = \sum_{j=1}^{J_n} [y_{n,j}^0 - y_n(t_{j,n})]^2 + \lambda \int_{T_0}^{T_1} [\mathcal{L}y_n(t)]^2 dt \tag{3}$$

At this point, we need to find the coefficients $c_{n,k}$ to minimize the penalty square sum, where $\lambda \in R_+$ is the smoothing parameter, and \mathcal{L} is a linear differential operator. To strike a balance between overfitting and over-smoothing by adjusting λ , we utilized the generalized cross-validation (GCV) method.

Functional regression model

Functional regression analysis encompasses various models, such as scalar-to-function regression, function-to-scalar regression, function-to-function regression, mixed-variable regression, nonlinear regression of scalars to functions, and functional generalized linear models. Given the specific problem under investigation, this paper employs a function-to-scalar regression analysis model, expressed as follows:

$$y_n(t) = x_{n1}\beta_1(t) + x_{n2}\beta_2(t) + \dots + x_{nq}\beta_q(t) + \varepsilon_n(t), 1 \leq n \leq N \tag{4}$$

The function variable $y_n(t), 1 \leq n \leq N$, observed N times, is an N -dimensional vector. It needs to be generated using the observed values $\{y_{n,j}^0 \in R, 1 \leq n \leq N, 1 \leq j \leq J_n\}$ from the initial data. $x_{ni}, 1 \leq n \leq N, 1 \leq i \leq q$ is a set of q covariate scalars. $\beta_i(t), 1 \leq i \leq q$ represents q functional regression parameters, also known as effect functions, which can be expressed using the basis function $\varphi_1, \varphi_2, \dots, \varphi_K$ as $\beta_i(t) = \sum_{k=1}^K b_{ik}\varphi_k(t), 1 \leq i \leq q$, and $b_{ik}, 1 \leq i \leq q, 1 \leq k \leq K$ represents the coefficient for each basis function $\phi_k(t)$. $\varepsilon_n(t), 1 \leq n \leq N$ represents the error term function.

Assuming that the observation points are $\{t_j : 1 \leq j \leq J\}$, the model can be represented in matrix form as:

$$\mathbf{Y} = \mathbf{XB}\Phi + \boldsymbol{\varepsilon} \tag{5}$$

Here,

$$\mathbf{Y} = (Y_n(t_j))_{\substack{1 \leq j \leq J \\ 1 \leq n \leq N}}, \mathbf{X} = (x_{ni})_{\substack{1 \leq i \leq q \\ 1 \leq n \leq N}}, \mathbf{B} = (b_{ik})_{\substack{1 \leq k \leq K \\ 1 \leq i \leq q}}, \Phi = (\varphi_k(t_j))_{\substack{1 \leq j \leq J \\ 1 \leq k \leq K}}, \boldsymbol{\varepsilon} = (\varepsilon_n(t_j))_{\substack{1 \leq j \leq J \\ 1 \leq n \leq N}}.$$

As a result, the penalized least squares used for the regression can be defined as:

$$P_\lambda(B) = \sum_{n=1}^N \sum_{j=1}^J [Y_n(t_j) - (\mathbf{XB}\Phi)_{nj}]^2 + \sum_{i=1}^q \lambda_i \int [\mathcal{L}\beta_i(t)]^2 dt \tag{6}$$

Similarly, we need to find the coefficients \mathbf{B} to minimize the sum of squares with the penalty mentioned above. In this paper, we assume that a single smoothing parameter $\lambda_i, 1 \leq i \leq q$ imparts a uniform level of smoothness to each component, and we employ the smoothing parameter cross-validation (CV) method for selection.

The advantage of using functional data regression analysis lies in its consideration of the holistic nature of the functional dependent variable. Our study focuses on the overall mortality pattern, representing a complete function curve. In comparison to traditional multivariate regression methods, functional regression analysis

avoids separate regressions for different age groups and instead comprehensively considers the overall variation. Multivariate regression methods often perform local fitting, providing accurate results in specific age ranges but lacking a comprehensive study of the overall mortality pattern and the identification of factors influencing the entire pattern. In supplementary information II, we present a comparative analysis between the results of multivariate regression methods and our approach, demonstrating that functional data regression, in contrast to traditional methods, more comprehensively reveals the characteristics of the mortality pattern. This global analysis contributes to a deeper understanding of the overall patterns in population mortality.

Functional hypothesis testing

In the empirical section, two hypothesis tests were employed: the functional F test and the pointwise t test. The pointwise t test is analogous to the traditional t test, with the initial step involving discretization of functional data as needed, followed by steps identical to those in the traditional t test. Further details are not provided here. The following section introduces the functional F test method derived from functional regression analysis^{27,30}.

Without loss of generality, we assume the original model to be

$$Y(t) = X_1\beta_1(t) + \dots + X_p\beta_p(t) + \epsilon(t) \quad (7)$$

Write the alternative hypothesis (Model 1) as follows:

$$H_1 : Y(t) = X_1\beta_1(t) + \dots + X_p\beta_p(t) + \epsilon(t) \quad (8)$$

Write the null hypothesis (Model 0i) as:

$$H_{0i} : \beta_i(t) = 0, i = 1, \dots, p \quad (9)$$

The functional F test statistic is defined as follows:

$$\mathcal{F}_i = \frac{rss_{0i} - rss_1}{rss_1/(n-p)} \quad (10)$$

Here, $ss = \sum_{n=1}^N \int (y_n(t) - \hat{y}_n(t))^2 dt$, $y_n(t)$ represents the initial functional curve, and $\hat{y}_n(t)$ represents the regression-predicted functional curve.

We reject the null hypothesis if $\mathcal{F}_i > F^{(1-\alpha)}(f_1, f_2)$, where $f_1 = [\text{trace}(E)^2/\text{trace}(E^2)]$, $f_2 = [(n-p)\text{trace}(E)^2/\text{trace}(E^2)]$, and E is the empirical covariance matrix of the error process obtained from the alternative hypothesis.

Data processing and variable selection

The data for this study were derived from the 2020 *China Population Census Yearbook* published by the National Bureau of Statistics of China, the 2021 *China Health Statistics Yearbook*, and annual provincial-level data from the *National Data Website*. These data are given in Online Resource 1. The dataset covers all 31 provinces in mainland China. The dependent variable data consist of mortality probabilities by age, gender, and the urban–rural classification for each of the 31 provincial administrative regions. The independent variables encompass gender, the urban–rural classification, and various socioeconomic variables. Variable selection was carried out through the following steps:

First, we utilized provincial mortality rate data from the Seventh National Population Census as the basis for estimating mortality probabilities for 22 age groups. Mortality probabilities were calculated for both gender and urban–rural subgroups, serving as measures of mortality levels for the 31 provinces.

The original dependent variable data were sex and the urban/rural age-specific mortality probabilities for 31 provincial districts. The age-specific mortality probabilities are the probability that the number of people who have reached a certain age will die before reaching another specific age, usually calculated from the known mortality rate³¹. The age-specific mortality rate data were obtained from the 7th Chinese census, and the formula was

$$m_x = \frac{D_x}{P_x} \quad (11)$$

where m_x represents the mortality rate at the age of x , D_x represents the number of deaths in the age group of x , and P_x represents the population in the age group of x .

If the age group distance is n , the age-specific mortality probabilities can be expressed as

$$q_x = \frac{2 * n * m_x}{2 + n * m_x} (n \text{ years interval}) \quad (12)$$

When $x = 0$:

$$q_0 = \frac{m_0}{1 + (1-r)m_0} \quad (13)$$

In the last age group, the age-specific mortality probabilities are equal to 1.

Based on the Seventh National Population Census data, we obtained mortality probabilities for 22 age groups: 0 years, 1–4 years, 5–9 years, 10–14 years, 15–19 years, 20–24 years, 25–29 years, 30–34 years, 35–39 years, 40–44 years, 45–49 years, 50–54 years, 55–59 years, 60–64 years, 65–69 years, 70–74 years, 75–79 years, 80–84 years, 85–89 years, 90–94 years, 95–99 years and 100 years and above. Since the mortality probability for

individuals aged 100 years and above reaches a limiting value of 1 and the data for the 95–99 years age group exhibit substantial variations across provinces, these two age groups were excluded from our analysis.

Second, the impact of socioeconomic factors on population mortality levels is multifaceted. Previous research has established a strong correlation between the level of socioeconomic development, quality of life, medical care, and technological advancements in a region. Additionally, higher levels of education have been shown to effectively reduce mortality probabilities. Drawing from prior studies on the socioeconomic factors influencing mortality patterns and considering the comprehensiveness and availability of data, we selected 22 indicators. These indicators were subjected to functional stepwise regression analysis, and variables with poor regression performance and nonsignificant regression coefficients were removed. Based on previous experience, we initially selected seven indicators, including Per Capita Gross Domestic Product (PGDP), the urbanization rate, education, gender, and the urban–rural status, to construct the foundational functional regression model. The model's coefficient of determination (R^2) was relatively low. To enhance the model's performance, we gradually introduced different factor indicators and conducted variable selection based on p values from pointwise t tests.

During the stepwise introduction and screening of variables, we used pointwise t test p values as the criterion, considering the initial twenty age segments as data points for testing the regression coefficients of each factor. If all values in the p value vector were greater than 0.05, we removed that factor and continued to introduce new factors, repeating the process. Simultaneously, we removed certain independent variables, examined changes in the overall model's pointwise t test significance, and gradually selected statistically significant variables affecting the target variable, resulting in a relatively concise and effective model.

In the final model, although the variables health technicians and insured employees did not pass the pointwise significance test at the 0.05 level, they were important medical human resource indicators and social security variables, and therefore, were retained. Additionally, considering the overall p values of all independent variables, their inclusion strengthened the pointwise significance of all variables. The model also included two resource-related medical indicators (hospital beds) and social security variables (insured residents) corresponding to them. Possible collinearity issues may have resulted in lower significance for the former two variables. The retention of these variables also facilitated a comparative analysis of their marginal effects with the corresponding variables.

Ultimately, we selected the gender indicator and thirteen socioeconomic indicators that influence regional mortality patterns, as shown in Table 1, and categorized them into three groups: demographic attributes, economic attributes, and social attributes.

Demographic attribute indicators include age-specific mortality probabilities and gender variables. Economic attributes include PGDP and tertiary sector. Social attributes encompass a variety of indicators. The urban–rural indicator is a binary control variable with a value of 0 or 1. The urbanization rate serves as an indicator of the urban–rural development level, while income and consumption represent indicators of residents' living standards. Hospital beds are indicators of physical resource capacity in healthcare development, and health technicians serve as indicators of human resource capacity in healthcare development. The fatality rate is also used as an indicator of the level of healthcare development. Healthcare security level indicators include insured residents and insured employees. These two insurance indicators were calculated based on the number of people covered

Category	Attribute Name	Attribute Definition
Demographic Attributes	Mortality Probabilities	Age-specific mortality probabilities. An important indicator for assessing regional population mortality patterns
	Gender	A control variable: equal to 1 for males and 0 for females
Economic Attributes	PGDP	Per capita GDP. Measures the wealth or ability to withstand social risks possessed by the population
	Tertiary Sector	Proportion of the tertiary sector (%). Reflects the level of transformation and development in the regional economic structure
Social attributes	Urban–Rural	A control variable for urban and rural classification, with a value of 1 indicating urban and 0 indicating rural
	Urbanization	Urbanization rate (%). A measure of urbanization level and the urbanization process
	Income	Per capita disposable income of urban and rural residents (CNY). Measuring the income level of the population, thereby reflecting residents' living standards and well-being ³²
	Consumption	Per capita consumption expenditure of urban and rural residents (CNY). An important indicator reflecting the standard of living and quality of life of residents
	Hospital Beds	Number of hospital beds per 10,000 people. An important physical resource indicator for evaluating the supply capacity of medical resources
	Health Technicians	Number of health technicians per 10,000 people. An important human resource indicator for evaluating the supply capacity of medical resources
	Fatality Rate	Mortality rate of medical and health institutions (%). An indicator reflecting the state of the regional healthcare system and its level of development
	Insured Residents	Proportion of urban and rural residents covered by basic medical insurance (%). An indicator reflecting the coverage of social security for non-working populations
	Insured Employees	Proportion of employees covered by basic medical insurance (%). An indicator reflecting the coverage of social security for working populations
	Education	Average years of education for the population aged 15 and above (Years). A crucial indicator for measuring the level of education and human capital development
	Traffic Accidents	Total number of traffic accidents. An indicator measuring public safety levels and the risk of population mortality

Table 1. Attributes and definitions. Source: China Population Census Yearbook, China Health Statistics Yearbook and National Data Website, 2020–2021.

by basic medical insurance and the total number of permanent residents, as reported in the *National Statistical Yearbook*. Education served as an indicator of educational attainment. Traffic accidents are used as indicators of public safety levels. To standardize the different scales of the data for comparison, we normalized the covariate indicators using Z score transformation based on the mean and standard deviation of the original indicator data.

Empirical analysis

Using the outlined methods and weighted least squares fits, mortality pattern curves for China's provincial regions in 2020 were derived (Fig. 1) and categorized by gender and urban–rural classification.

Key observations from Fig. 1 include the following: (1) Mortality patterns exhibit a "J"-shaped distribution, with higher probabilities in infants and a gradual increase after age 40, increasing significantly after 70. (2) Females in both urban and rural populations show lower mortality probabilities across age groups than males. (3) Urban populations generally have lower mortality probabilities than do their rural counterparts of the same gender. (4) In urban areas, female mortality increases at approximately age 50, while mortality increases for urban males and rural females at approximately age 45, and rural males show a substantial increase in mortality at approximately age 40. (5) Substantial provincial variation exists, with divergence in mortality patterns for rural males starting at approximately age 35, those for urban males and females starting at approximately age 50, and rural females exhibiting increasing divergence at approximately age 45. Notably, considerable variations exist in mortality probabilities for the elderly population (those over age 60) among provinces.

These findings not only lay the groundwork for identifying the factors influencing mortality patterns regionally but also offer important insights into the mechanisms through which these driving factors affect mortality probabilities across different age groups.

Overall regression results of the model

We established the functional regression model as follows:

$$y_i(t) = \beta_0(t) + GD_i\beta_1(t) + GDP_i\beta_2(t) + TS_i\beta_3(t) + UR_i\beta_4(t) + URR_i\beta_5(t) + PCI_i\beta_6(t) + PCC_i\beta_7(t) + HB_i\beta_8(t) + HT_i\beta_9(t) + FR_i\beta_{10}(t) + IR_i\beta_{11}(t) + IE_i\beta_{12}(t) + EDU_i\beta_{13}(t) + TA_i\beta_{14}(t) + \varepsilon_i(t), i = 1, \dots, 124 \quad (14)$$

Here, $y_i(t)$ represents the smooth curves of mortality probabilities for the four mortality patterns across the 31 provincial regions, totaling 124 curves. $\beta_0(t)$ is the functional constant term, $\varepsilon_i(t)$ is the functional random error term, and $\beta_1(t) \sim \beta_{14}(t)$ are the functional regression coefficients for various variables. The meanings of the other symbols are detailed in Table 2. Figure 2 shows the regression coefficient curves for 14 explanatory variables.

Examining Fig. 2's 14 regression coefficient curves reveals that the influence of covariates on regional mortality patterns intensifies notably after age 45. Both demographic and socioeconomic factors show a significant increase in their impact on mortality probabilities, as evidenced by the substantial increase in the absolute values of the regression coefficients. Among the 14 coefficients, gender and urban–rural status had the greatest impact on overall mortality probabilities. The gender coefficient consistently holds a positive value, indicating higher male mortality probabilities across different age groups, all else being equal. Conversely, the urban–rural coefficient remains consistently negative, signifying lower mortality probabilities in urban areas than in rural areas across all age groups, attributable to various factors.

Figure 3 shows the goodness of fit of the functional regression model by age group.

The model's goodness of fit, which surpassed 60%, notably exceeded 80% for the 10–14 years, 15–19 years, and 65–69 years age groups. In the nine other age groups (5–9 years, 20–24 years, 25–29 years, 30–34 years, 35–39 years, 40–44 years, 60–64 years, 70–74 years, and 75–79 years), the goodness of fit exceeded 70%. This indicates that the 14 selected demographic and socioeconomic indicators effectively explain the mortality probabilities for these age groups. However, the model's fit decreased notably for the older age groups, especially those older than 85 years. This implies the necessity of additional factors to elucidate mortality probabilities for the last two age groups. The lower coefficient of determination R^2 for the last two age groups, particularly for the 90–94 years age group, suggests the influence of unconsidered factors, such as biological and environmental variables, which will be incorporated in future research.

In Table 3, we use the functional F test to test the significance of each coefficient. The variables gender and urban–rural attributes are significant at the 0.1% level, which is consistent with the analysis above. The remaining

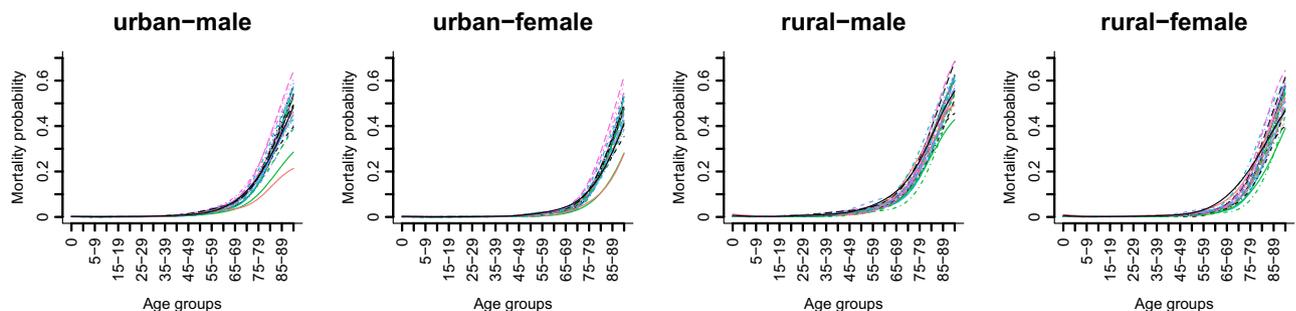


Figure 1. The mortality pattern curves for the 31 provincial regions.

Attribute Name	Symbol	Mean (Standard Deviation)
Gender	GD	
PGDP	GDP	70,786.58(31,325.74)
Tertiary Sector	TS	53.89 (7.57)
Urban–Rural	UR	
Urbanization	URR	62.46(13.47)
Urban Income	PCI	42,252.67(11,129.24)
Rural Income	PCI	17,814.39 (5833.54)
Urban Consumption	PCC	26,080.76(6043.74)
Rural Consumption	PCC	13,848.59(3299.88)
Hospital Beds	HB	64.68(8.72)
Health Technicians	HT	77.61(11.43)
Fatality Rate	FR	0.49 (0.39)
Insured Residents	IR	71.00 (20.55)
Insured Employees	IE	24.32 (14.77)
Education	EDU	9.88(0.99)
Traffic Accidents	TA	7892.71(6096.92)

Table 2. Descriptive statistics of attributes. Urban income and rural income are collectively referred to as ‘income’ in the analysis, and urban consumption and rural consumption are collectively referred to as ‘consumption’.

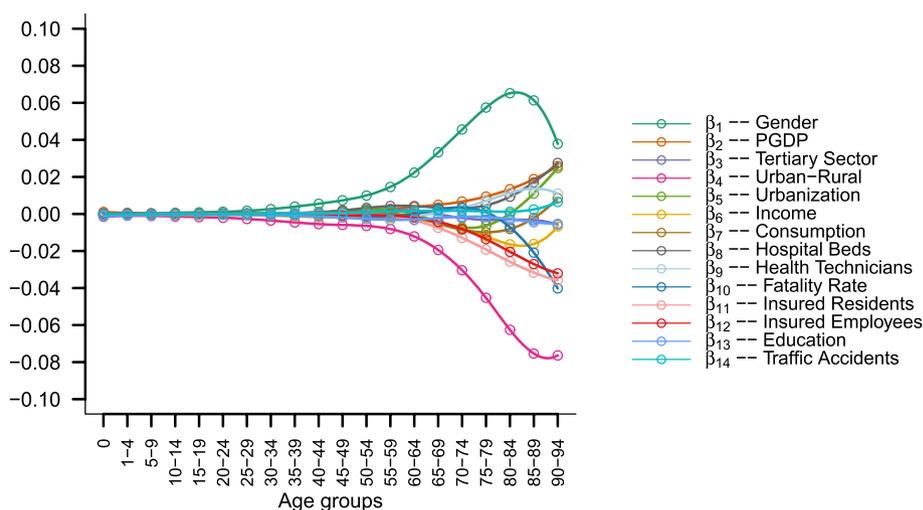


Figure 2. Functional regression coefficient results.

two significant functions are hospital beds and fatality rate. In this way, we selected four indicators as global influencing variables. The purpose of testing the functional regression coefficients is to determine whether each regression coefficient function is significantly different from zero. The mortality probabilities are close to zero in the initial age groups (1–39 years), leading to very large and nonsignificant *p* values for the regression coefficients, as these coefficients were already close to zero. However, the lack of significance of the overall functional regression coefficients does not imply that they are not significant in every age group. Subsequently, we conducted pointwise *t* tests to observe the significance of each functional coefficient in different age groups.

Analysis of the age-specific effects of the variables gender, urban–rural, hospital beds, and fatality rate

Table 4 shows the regression coefficient values and significance levels of the variables gender, urban–rural, hospital beds and fatality rate.

The regression coefficients for the gender variable are significant for all age groups except the 1–4-year-old age group. For neonates (0 years old), the gender variable is significant at the 0.1 level. For the 90–94 age group, the significance level for the gender variable is 0.05. For all the other age groups, the *p* values of the *t* tests for the gender variable are less than 0.01. This finding aligns with previous research confirming that male mortality probabilities are generally greater than female mortality probabilities. Our study further revealed that in each

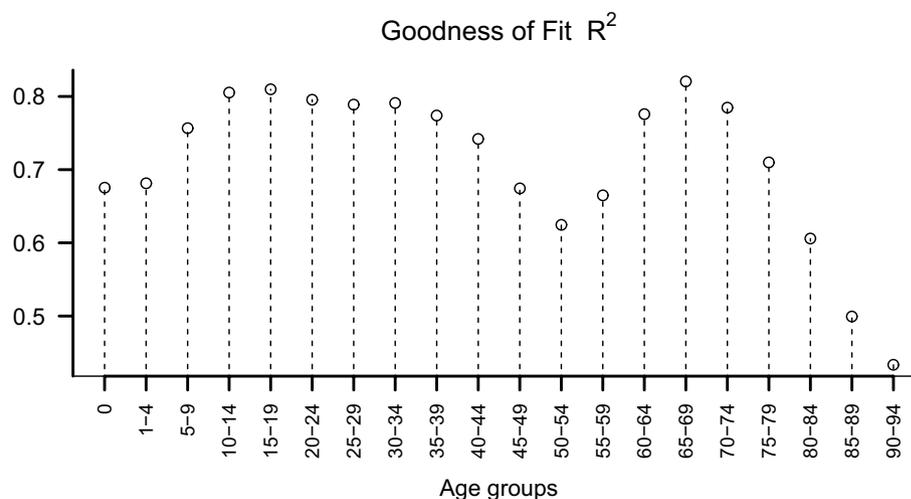


Figure 3. Goodness-of-fit values for each age group.

Attribute category	Functional regression coefficients	The p -value from the functional F-test
Demographic attributes	$\beta_1(t)$ - Gender	9.041112e-29***
Economic attributes	$\beta_2(t)$ - PGDP	0.2427897
	$\beta_3(t)$ - Tertiary Sector	0.958046
Social attributes	$\beta_4(t)$ - Urban-Rural	4.689741e-32***
	$\beta_5(t)$ - Urbanization	0.1561508
	$\beta_6(t)$ - Income	0.6637036
	$\beta_7(t)$ - Consumption	0.8381877
	$\beta_8(t)$ - Hospital Beds	0.0004727471***
	$\beta_9(t)$ - Health Technicians	0.1635833
	$\beta_{10}(t)$ - Fatality Rate	0.0002020485***
	$\beta_{11}(t)$ - Insured Residents	0.08622965
	$\beta_{12}(t)$ - Insured Employees	0.6487598
	$\beta_{13}(t)$ - Education	0.9240515
	$\beta_{14}(t)$ - Traffic Accidents	0.7131577

Table 3. The p values from the functional F tests for the functional regression coefficients. One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

age group, under other equal conditions, the male mortality probabilities were consistently greater than the female mortality probabilities. This indicates that gender-specific mortality patterns in China have transitioned into a "modern pattern"³³.

The urban-rural variable has a negative influence on mortality probabilities across all age groups, and its impact is significant at the 0.001 level for all 20 age groups. This implies that, under other equal conditions, urban mortality probabilities are generally lower than rural mortality probabilities. Unlike urban areas, rural regions face challenges with limited medical facilities, scarce healthcare resources, and poor health conditions, which can impact medical accessibility. Rural areas are more vulnerable to infectious diseases due to poor infrastructure. Urban areas, with better education and health awareness, exhibit improved lifestyles and overall enhanced health. Economic advantages in urban regions enable access to advanced medical care. The hospital beds variable significantly positively impacts the mortality probability for individuals aged 20-29, 40-69, and 85-94 years. This effect may be linked to variations in healthcare facility quality. Simply having more medical beds does not guarantee lower mortality if healthcare quality is inadequate. For the fatality rate, a significant impact is noted in the 85-94 age group, and it is important to highlight that this effect is negative.

Analysis of the effects of other indicators by age group

In Table 5, further regression analysis indicates that the variables health technicians and insured employees lack significance across all age groups, leading to their exclusion from further examination. Consequently, our focus narrows to 12 variables—4 global and 8 local. We will proceed to assess the impact of these eight remaining

Age groups	$\beta_1(t)$	$\beta_4(t)$	$\beta_8(t)$	$\beta_{10}(t)$
0	4.68E-04*	-1.54E-03***	-2.57E-05	-4.82E-05
1-4	1.43E-04	-9.83E-04***	7.93E-05	-1.26E-04
5-9	2.77E-04***	-1.02E-03***	8.99E-05	-1.15E-04
10-14	5.97E-04***	-1.33E-03***	8.07E-05	-9.32E-05
15-19	8.75E-04***	-1.68E-03***	1.15E-04	-1.21E-04
20-24	1.18E-03***	-2.08E-03***	1.83E-04*	-1.67E-04
25-29	1.69E-03***	-2.66E-03***	2.45E-04*	-1.57E-04
30-34	2.59E-03***	-3.55E-03***	2.73E-04	-3.45E-05
35-39	3.89E-03***	-4.59E-03***	4.02E-04	8.31E-05
40-44	5.50E-03***	-5.49E-03***	8.70E-04*	-3.35E-05
45-49	7.33E-03***	-5.96E-03***	1.89E-03***	-5.94E-04
50-54	9.96E-03***	-6.45E-03***	3.25E-03***	-1.31E-03
55-59	1.46E-02***	-8.12E-03***	4.26E-03***	-1.38E-03
60-64	2.23E-02***	-1.22E-02***	4.24E-03***	-1.64E-05
65-69	3.32E-02***	-1.94E-02***	3.33E-03*	2.37E-03
70-74	4.56E-02***	-3.04E-02***	2.94E-03	3.49E-03
75-79	5.74E-02***	-4.53E-02***	4.58E-03	9.11E-04
80-84	6.52E-02***	-6.25E-02***	9.31E-03	-7.16E-03
85-89	6.13E-02***	-7.54E-02***	1.70E-02*	-2.09E-02*
90-94	3.78E-02**	-7.64E-02***	2.75E-02**	-4.02E-02**

Table 4. Coefficients and significance of the variables gender, urban–rural, hospital beds, and fatality rate. The number represents the coefficient, and the symbol in the upper right corner represents the significance. Table 5 shares the same format. One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

Age groups	$\beta_2(t)$	$\beta_3(t)$	$\beta_5(t)$	$\beta_6(t)$	$\beta_7(t)$	$\beta_9(t)$	$\beta_{11}(t)$	$\beta_{12}(t)$	$\beta_{13}(t)$	$\beta_{14}(t)$
0	1.02E-03**	2.36E-04	-5.87E-04*	-1.60E-04	-1.06E-03**	-5.67E-05	-8.70E-04*	-3.21E-05	-1.10E-03***	-1.14E-04
1-4	5.85E-04***	2.11E-04	-2.16E-04*	-1.70E-04	-5.88E-04***	-3.64E-05	-4.89E-04*	2.98E-05	-7.31E-04***	-5.90E-05
5-9	3.35E-04**	1.59E-04	-7.16E-05	-3.58E-04*	-1.39E-04	3.44E-05	-2.86E-04	1.50E-05	-5.22E-04***	-2.65E-05
10-14	2.30E-04	1.07E-04	-4.34E-05	-5.40E-04**	1.47E-04	8.90E-05	-1.92E-04	-5.52E-06	-4.32E-04***	-9.24E-06
15-19	2.27E-04	7.76E-05	-3.13E-05	-5.65E-04**	1.60E-04	7.48E-05	-1.43E-04	2.24E-05	-4.17E-04**	8.73E-07
20-24	2.77E-04	8.25E-05	1.73E-07	-5.13E-04*	-1.90E-06	3.43E-05	-1.29E-04	4.26E-05	-4.59E-04***	1.39E-05
25-29	3.24E-04	1.29E-04	6.00E-05	-5.61E-04*	-1.59E-04	4.90E-05	-1.61E-04	-4.64E-05	-5.47E-04***	4.05E-05
30-34	3.25E-04	2.29E-04	1.63E-04	-8.65E-04	-1.44E-04	1.85E-04	-2.44E-04	-3.30E-04	-6.77E-04*	9.26E-05
35-39	3.43E-04	4.39E-04	3.96E-04	-1.35E-03	8.07E-05	3.38E-04	-3.32E-04	-7.09E-04	-8.97E-04*	1.92E-04
40-44	5.14E-04	8.47E-04	8.88E-04	-1.77E-03	4.63E-04	2.91E-04	-3.45E-04	-9.66E-04	-1.30E-03*	3.68E-04
45-49	9.68E-04	1.53E-03	1.75E-03*	-1.92E-03	9.37E-04	-1.59E-04	-2.16E-04	-8.98E-04	-1.95E-03*	6.47E-04
50-54	1.72E-03	2.25E-03*	2.53E-03*	-1.89E-03	1.16E-03	-8.58E-04	-2.44E-04	-6.68E-04	-2.71E-03**	1.00E-03*
55-59	2.67E-03	2.46E-03	2.22E-03	-2.10E-03	4.80E-04	-1.29E-03	-1.09E-03	-8.11E-04	-3.19E-03*	1.36E-03*
60-64	3.71E-03	1.63E-03	-1.66E-04	-2.99E-03	-1.72E-03	-9.48E-04	-3.41E-03	-1.87E-03	-3.01E-03	1.64E-03*
65-69	4.92E-03	-2.24E-04	-4.20E-03	-4.87E-03	-5.23E-03	5.23E-04	-7.47E-03	-4.25E-03	-2.21E-03	1.77E-03
70-74	6.68E-03	-2.17E-03	-7.21E-03*	-7.87E-03	-8.52E-03	3.18E-03	-1.29E-02	-8.11E-03	-1.50E-03	1.70E-03
75-79	9.37E-03	-3.22E-03	-6.33E-03	-1.21E-02	-9.94E-03	7.05E-03	-1.92E-02	-1.36E-02	-1.61E-03	1.40E-03
80-84	1.33E-02	-3.09E-03	1.38E-04	-1.64E-02	-8.10E-03	1.14E-02	-2.58E-02	-2.04E-02	-2.87E-03	1.21E-03
85-89	1.88E-02	-3.33E-03	1.10E-02	-1.62E-02	-2.20E-03	1.37E-02	-3.17E-02	-2.71E-02	-4.45E-03	2.42E-03
90-94	2.59E-02	-5.72E-03	2.48E-02	-6.66E-03	8.47E-03	1.10E-02	-3.59E-02	-3.20E-02	-5.37E-03	6.48E-03

Table 5. Coefficient values and significance of t Tests for regression coefficients of other indicators in various age groups (two-tailed). One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

factors in age-specific groups, ranking them based on the magnitude of influence (sorted by the absolute values of coefficients). The results will be presented in three stages: low age group (0–29 years), mid age group (30–64 years), and high age group (65–94 years). We will discuss the noteworthy socioeconomic factors influencing each of these stages.

Ranking	0	1–4	5–9	10–14	15–19	20–24	25–29
1	β_4 -UR 1.54E-03 4.45E-11***	β_4 -UR 9.83E-04 5.74E-20***	β_4 -UR 1.02E-03 1.76E-26***	β_4 -UR 1.33E-03 6.01E-26***	β_4 -UR 1.68E-03 3.26E-31***	β_4 -UR 2.08E-03 1.79E-39***	β_4 -UR 2.66E-03 2.71E-40***
2	β_{13} -EDU 1.10E-03 3.92E-05***	β_{13} -EDU 7.31E-04 1.22E-09***	β_{13} -EDU 5.22E-04 1.23E-07***	β_1 -GD 5.97E-04 8.83E-09***	β_1 -GD 8.75E-04 9.46E-14***	β_1 -GD 1.18E-03 6.31E-21***	β_1 -GD 1.69E-03 9.03E-25***
3	β_7 -PCC 1.06E-03 3.32E-03**	β_7 -PCC 5.88E-04 2.70E-04***	β_6 -PCI 3.58E-04 2.48E-02*	β_6 -PCI 5.40E-04 9.31E-03**	β_6 -PCI 5.65E-04 8.60E-03**	β_6 -PCI 5.13E-04 1.24E-02*	β_6 -PCI 5.61E-04 2.97E-02*
4	β_2 -GDP 1.02E-03 3.10E-03**	β_2 -GDP 5.85E-04 1.35E-04***	β_2 -GDP 3.35E-04 9.28E-03**	β_{13} -EDU 4.32E-04 5.83E-04***	β_{13} -EDU 4.17E-04 1.58E-03**	β_{13} -EDU 4.59E-04 3.80E-04***	β_{13} -EDU 5.47E-04 7.19E-04***
5	β_{11} -IR 8.70E-04 4.00E-02*	β_{11} -IR 4.89E-04 1.16E-02*	β_1 -GD 2.77E-04 1.89E-04***			β_8 -HB 1.83E-04 2.16E-02*	β_8 -HB 2.45E-04 1.47E-02*
6	β_5 -URR 5.87E-04 1.86E-02*	β_5 -URR 2.16E-04 4.16E-02*					
7	β_1 -GD 4.68E-04 2.80E-02*						

Table 6. Absolute values, significance, and ranking of the regression coefficients for the low age group. In each cell, the first row represents the indicator name, the second row denotes the absolute coefficient values, and the third row indicates the p value along with the significance symbols. Tables 7 and 8 share the same format. One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

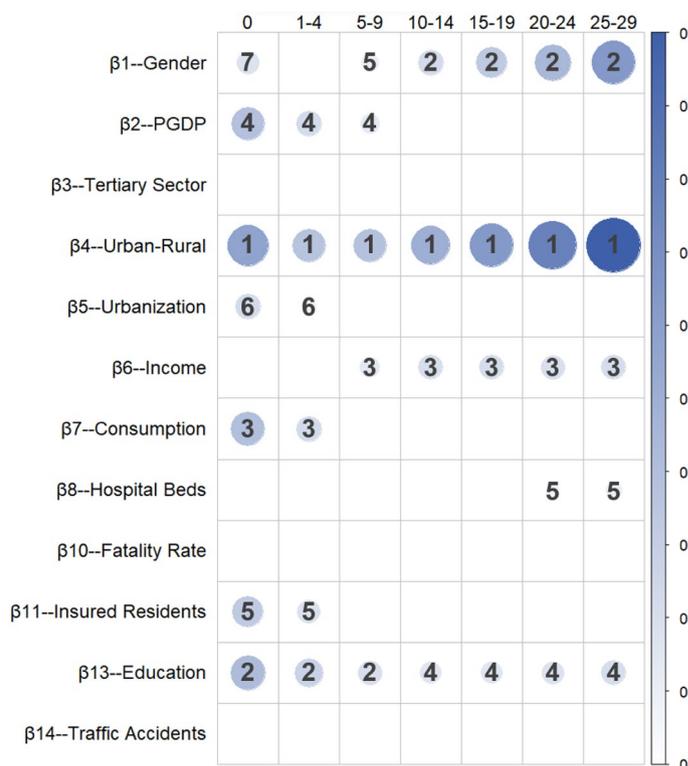


Figure 4. Marginal impact ranking of factors in the low age group. Note: The size and color intensity of the circles represent the magnitude of the absolute values, where larger circles with darker colors indicate larger absolute values. (The same applies throughout).

Ranking	30–34	35–39	40–44	45–49	50–54	55–59	60–64
1	β_1 -UR 3.55E-03 6.09E-32***	β_1 -UR 4.59E-03 3.94E-24***	β_1 -GD 5.50E-03 7.18E-19***	β_1 -GD 7.33E-03 1.23E-19***	β_1 -GD 9.96E-03 3.22E-22***	β_1 -GD 1.46E-02 1.57E-26***	β_1 -GD 2.23E-02 2.58E-31***
2	β_1 -GD 2.59E-03 3.28E-22***	β_1 -GD 3.89E-03 1.24E-19***	β_4 -UR 5.49E-03 7.86E-19***	β_4 -UR 5.96E-03 6.99E-15***	β_4 -UR 6.45E-03 2.03E-12***	β_4 -UR 8.12E-03 2.17E-12***	β_4 -UR 1.22E-02 9.69E-15***
3	β_{13} -EDU 6.77E-04 1.26E-02*	β_{13} -EDU 8.97E-04 4.56E-02*	β_{13} -EDU 1.30E-03 4.71E-02*	β_{13} -EDU 1.95E-03 2.08E-02*	β_8 -HB 3.25E-03 1.45E-06***	β_8 -HB 4.26E-03 5.87E-07***	β_8 -HB 4.24E-03 1.22E-04***
4			β_8 -HB 8.70E-04 3.35E-02*	β_8 -HB 1.89E-03 4.15E-04***	β_{13} -EDU 2.71E-03 9.02E-03**	β_{13} -EDU 3.19E-03 1.45E-02*	β_{14} -TA 1.64E-03 4.93E-02*
5				β_5 -URR 1.75E-03 2.94E-02*	β_5 -URR 2.53E-03 1.05E-02*	β_{14} -TA 1.36E-03 3.13E-02*	
6					β_3 -TS 2.25E-03 4.73E-02*		
7					β_{14} -TA 1.00E-03 4.51E-02*		

Table 7. Absolute values, significance, and ranking of the regression coefficients for the mid age group. One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

Analysis of the impact of socioeconomic factors on mortality patterns—low age group (0–29 years)

Table 6 and Fig. 4 present the marginal effects and ranking of various factors in the mortality patterns for the low age group.

From Table 6 and Fig. 4, it is clear from the results that the urban–rural factor ranks first in terms of its impact on the seven population groups within the low age range. After urban–rural, the variables education and gender follow, with gender having a more significant impact on newborns (0 years) and young children (5–9 years), whereas gender has no significant influence on the 1- to 4-year-old age group. However, the influence of gender begins to increase in the subsequent 10–29-year age group, ranking second in terms of its effect on mortality probabilities. Newborns are affected by seven significant factors, while the 10–14-year-old and 15–19-year-old groups are influenced by only four factors.

Education significantly influences the 0–29 years age group, with p values mostly below 0.01 (except for the 15–19 age group). Higher education has a notable negative impact on mortality, particularly in the 0–9 age range, ranking second in influencing infant and toddler mortality. Overall, women with higher education levels exhibit increased self-care and prevention behaviors, contributing to reduced infant mortality probabilities⁹. In the 10–29 age group, education ranks fourth in terms of the degree of impact. Individuals with higher education levels tend to adopt healthier lifestyles and possess more health awareness.

Income reflects people's living standards and purchasing power. The mortality probability for the age group of 5–29 years is significantly influenced by income, ranking it as the third most influential factor. The data show that income has the most significant impact on the 15- to 19-year-old population, reducing mortality probability by 5.65E-4.

The proportion of urban and rural residents covered by basic medical insurance has a significant impact on the 0–4-year-old age group, with the most pronounced effect occurring for the 1–4-year-old age group, where the influence value peaks at -4.89E-4.

The urbanization rate variable significantly affects the mortality probabilities of infants and young children aged 0–4 years, ranking as the sixth most influential factor. For each percentage point increase in the urbanization rate, the mortality probability for newborns decreases by 5.87E-4, and for 1–4-year-olds, the mortality probability decreases by 2.16E-4. Urban areas have superior medical facilities, more healthcare institutions, and more advanced medical services, ensuring that residents access higher-quality medical resources⁵.

The variable per capita GDP significantly affects infants and young children aged 0–9 years, consistently ranking fourth. However, the findings of this study reveal a positive influence of per capita GDP on infant mortality probabilities, which warrants further discussion.

Analysis of the impact of socioeconomic factors on mortality patterns—mid age group (30–64 years)

Table 7 and Fig. 5 present the marginal effects and ranking of various factors on the mortality patterns for the mid age group.

Among the socioeconomic factors significantly affecting the mortality probabilities of mid age group, the factor gender gradually surpasses the urban–rural factor to become the primary factor. In addition, the most pronounced effect is observed for the education indicator, which ranks third in terms of its influence on the 30–49-year-old population and then falls to fifth place among the 50–59-year-old population, with no significant impact on the 60–64-year-old group. Educational attainment is often associated with occupation and economic

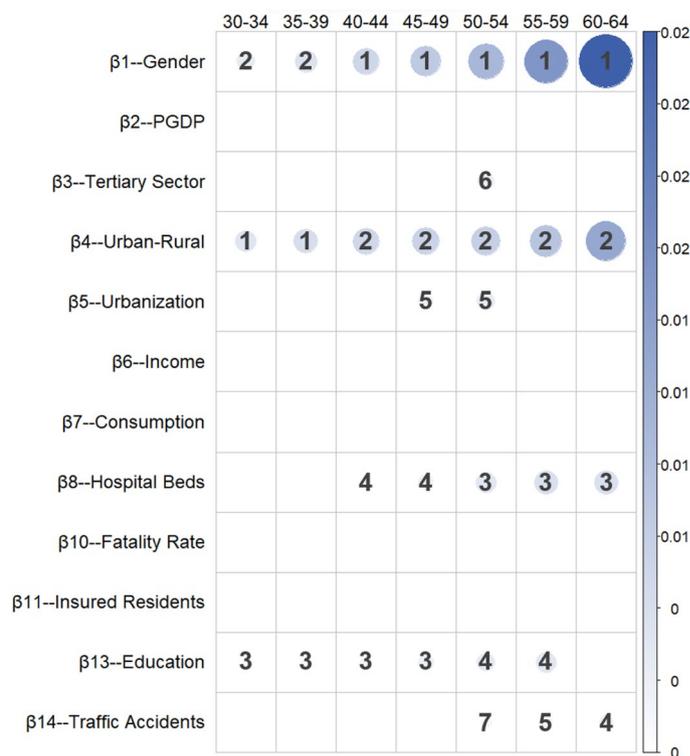


Figure 5. Marginal impact ranking of factors in the mid age group.

Ranking	65–69	70–74	75–79	80–84	85–89	90–94
1	β ₁ -GD 3.32E-02 4.93E-33***	β ₁ -GD 4.56E-02 1.94E-29***	β ₁ -GD 5.74E-02 9.64E-23***	β ₁ -GD 6.52E-02 1.23E-15***	β ₄ -UR 7.54E-02 3.27E-12***	β ₄ -UR 7.64E-02 5.85E-09***
2	β ₄ -UR 1.94E-02 2.33E-17***	β ₄ -UR 3.04E-02 6.44E-18***	β ₁ -UR 4.53E-02 9.57E-17***	β ₄ -UR 6.25E-02 9.01E-15***	β ₁ -GD 6.13E-02 4.43E-09***	β ₁₀ -FR 4.02E-02 1.05E-03**
3	β ₈ -HB 3.33E-03 2.99E-02*	β ₅ -URR 7.21E-03 4.34E-02*			β ₁₀ -FR 2.09E-02 2.94E-02*	β ₁ -GD 3.78E-02 2.25E-03**
4					β ₈ -HB 1.70E-02 2.59E-02*	β ₈ -HB 2.75E-02 4.55E-03**

Table 8. Absolute values, significance, and ranking of the regression coefficients for the high age group. One asterisk (*) indicates significance at the 5% level, i.e., $p < 0.05$; two asterisks (**) indicate significance at the 1% level, i.e., $p < 0.01$; three asterisks (***) indicate significance at the 0.1% level, i.e., $p < 0.001$.

status. Higher educational levels are typically accompanied by better employment opportunities and higher income levels, thus reducing health risks.

The urbanization rate significantly impacts the mortality probabilities of mid age group, particularly individuals aged 45–54, ranking fifth in terms of its influence. The effect is positive, implying that urban living might be associated with greater social and economic pressures, which can negatively impact the health of middle-aged individuals. Similarly, the proportion of the tertiary sector significantly affects the mortality probability of the 50- to 54-year-old group. With increasing age, traffic accidents have a progressively greater impact on the mortality probability, with coefficient values showing an upward trend from 9.26E-5 to 1.64E-3 and significant effects among the 50- to 64-year-old population.

Analysis of the impact of socioeconomic factors on mortality patterns—high age group (65–94 years)

Table 8 and Fig. 6 present the marginal effects and ranking of various factors in the mortality patterns for the high age group.

In the elderly population, the main influencing factors are the four global variables analyzed earlier. In the 65–84-year-old group, the primary influencing factor is the gender, followed by the urban–rural factor. However, among the 85–94-year-old population, the influence of gender is lower than that of urban–rural.

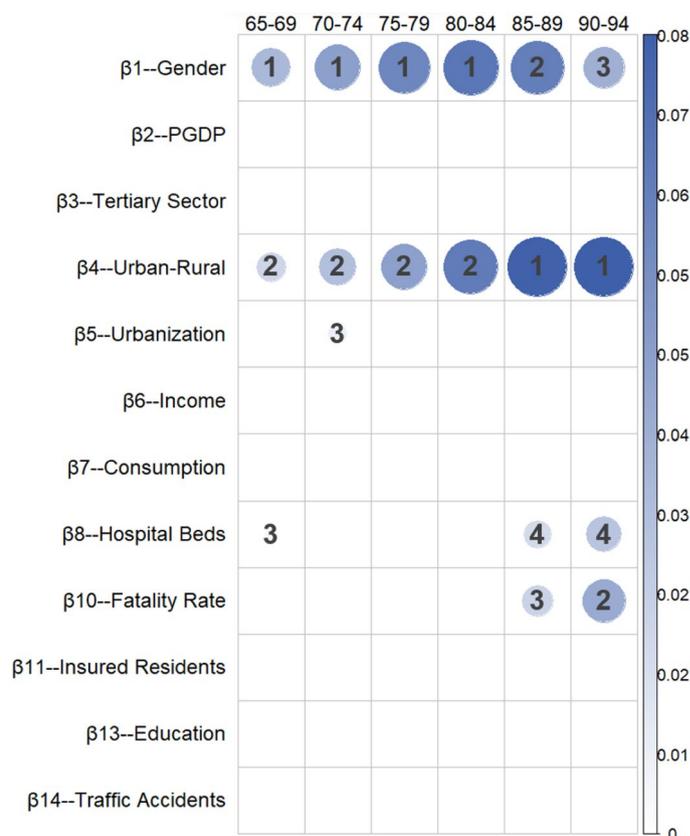


Figure 6. Marginal impact ranking of factors in the high age group.

The urbanization rate has a more significant impact on the mortality probability of individuals aged 70–74 years. Education has no significant effect on the mortality probability of the high age group compared with the low and mid age groups. Some research has suggested that education might not be a strong indicator of the socioeconomic status of China's elderly population¹⁵.

For the elderly population, the chosen variables inadequately explain the mortality probabilities, suggesting that additional factors play a crucial role in this age group. Age-related factors are prominent as the body undergoes natural aging, increasing the likelihood of mortality. Socioeconomic factors are not the primary determinants for this elderly segment.

In the high age group, some factors exhibit large coefficient values but remain nonsignificant. For example, both insured residents and insured employees have substantial negative effects on the mortality probability, and their impact becomes more pronounced with increasing age. Furthermore, income and consumption also have a certain effect on reducing the mortality probability of the elderly, but these factors were not significant.

Conclusion

Leveraging data from China's 7th Population Census, our research delved into the socioeconomic impact factors and mechanisms of mortality patterns, accounting for interprovincial disparities, urban–rural divides, and gender differences. This study also scrutinized their marginal effects and variations across age groups. The key findings include:

- (1) Regional mortality patterns in China show significant gender and urban–rural disparities across most age groups. In the same age group, the mortality probabilities are generally greater for males than for females, and they generally are greater in rural areas than in urban areas. The factor urban–rural dominates the mortality probabilities of individuals aged 39 and younger, while the factor gender becomes most significant in the 40–84 age group.
- (2) Socioeconomic factors, apart from gender and urban–rural distinctions, on mortality patterns typically become evident after the age of 45, with less noticeable variations in their impact on early-life mortality patterns.
- (3) Factors such as PGDP, tertiary sector, urban–rural, urbanization, income, consumption, hospital beds, fatality rate, insured residents, education, and traffic accidents have different influence directions on mortality patterns across age groups, and the marginal effects of each factor are different. For example, in the 0–29 age group, education exerts a significant negative effect on mortality patterns, and the marginal effect is larger than that of most other factors. The education index still has a negative effect on the mortality

- probabilities of the population aged 30–59, and there is no significant effect on the high age group. The rate of urbanization exerts a significant positive influence on the mortality probability of the 45–54 age group. With increasing age, the impact of traffic accidents on the mortality probability gradually increases. The mortality pattern of the elderly group is mainly affected by the factors hospital beds and fatality rate, but other socioeconomic variables have no significant affected.
- (4) The socioeconomic indicator mechanism displays an age structure effect, with lower explanatory power for mortality patterns in older age groups. Further analysis in older age groups should consider additional explanatory variables.

These conclusions contribute significantly to our understanding of the impact of socioeconomic factors on mortality patterns according to demographic data. They underscore the complexity of socioeconomic factors' influence on mortality patterns and deepen our comprehension of age-specific differences in mortality patterns.

This paper has limitations, primarily focusing on the influence of socioeconomic factors while overlooking biological and environmental factors. This limitation results in a relatively low explanatory power of the model, especially for the mortality probabilities of elderly people. Future research will explore the impact of geographical and environmental resource factors on mortality patterns and their mechanisms.

Data availability

The data that support the findings of this study are available upon request from the corresponding author.

Received: 13 November 2023; Accepted: 3 May 2024

Published online: 09 May 2024

References

- Ivanova, E. И & Zhong, J. P. Analysis of the causes and regional differences of changes in the mortality rate of men in Russia. *Acad. J. Rus. Stud.* **05**, 89–94 (2011).
- Kibele, E. U., Klüsener, S. & Scholz, R. D. Regional mortality disparities in Germany: long-term dynamics and possible determinants. *Kolner Zeitschrift für Soziologie und Sozialpsychologie.* **67**(Suppl 1), 241–270. <https://doi.org/10.1007/s11577-015-0329-2> (2015).
- van Raalte, A. A., Klüsener, S., Oksuzyan, A. & Grigoriev, P. Declining regional disparities in mortality in the context of persisting large inequalities in economic conditions: the case of Germany. *Int. J. Epidemiol.* **49**(2), 486–496. <https://doi.org/10.1093/ije/dyz265> (2020).
- Spencer, J. C., Wheeler, S. B., Rotter, J. S. & Holmes, G. M. Decomposing mortality disparities in urban and rural US counties. *Health Serv. Res.* **53**(6), 4310–4331. <https://doi.org/10.1111/1475-6773.12982> (2018).
- Cao, L. H., Lin, L. M., Liu, Y. L., Liu, B. Q. & Zheng, M. G. Mortality surveillance of Chinese children under 5 years & causes analysis of different mortality rates by sex in poverty—strichen area. *Chin. Primary Health Care.* **11**(10), 37–39 (1997).
- Nie, F. Q., Song, G. J. The control of infant mortality in China: an empirical analysis on regional difference and government intervention. *Chin. J. Popul. Sci.* (04), 75–80+96 (2005).
- Li, M. A Comparative study of death level in north and east China. (Master's Thesis, Anhui University). Retrieved from https://kns.cnki.net/kcms2/article/abstract?v=1TIORdBTwpai2uTsNp2jwIEl7WjRI3sMRTdRVBs5AgOT4JRJJWdF5_8xnGHG3lO6K2eNL46yD_Ph3AJfE2Lnk01a24e7WR9V4Gp8dG9_l3iyqfSHMd01Wvgxhr71YB-EcLtyfMQF_MdK8QyW3eSusA==uniplatform=NZKPTLanguage=CHS (2019).
- Zhou, Y. G. A study of regional mortality rate disparities in China using cluster analysis. (Master's Thesis, Shanghai University of Finance and Economics). Retrieved from <https://link.cnki.net/doi/https://doi.org/10.27296/d.cnki.gshcu.2021.001295> (2021).
- Chen, M., Huang, B. S. & Liu, Y. Study on demographic spatial distribution of health risks caused by PM_{2.5}: a case from Shenzhen. *Progr. Geogr.* **41**(6), 1028–1040 (2022).
- Mackenbach, J. P. *et al.* Socioeconomic inequalities in health in 22 European countries. *N. Engl. J. Med.* **358**(23), 2468–2481. <https://doi.org/10.1056/NEJMsa0707519> (2008).
- Mackenbach, J. P. Persistence of social inequalities in modern welfare states: Explanation of a paradox. *Scandinavian J. Public Health.* **45**(2), 113–120. <https://doi.org/10.1177/1403494816683878> (2017).
- Lutz, W. & Kebede, E. Education and health: Redrawing the preston curve. *Popul. Dev. Rev.* **44**(2), 343–361. <https://doi.org/10.1111/padr.12141> (2018).
- Gutin, I. & Hummer, R. A. Social inequality and the future of US life expectancy. *Ann. Rev. Sociol.* **47**(1), 501–520. <https://doi.org/10.1146/annurev-soc-072320-100249> (2021).
- Zhao, J. H. An analysis of the socioeconomic factors of changes in mortality trends in China. *Northwest Popul. J.* **4**, 49–53 (1994).
- Luo, W. & Xie, Y. Socio-economic disparities in mortality among the elderly in China. *Popul. Studi.* **68**(3), 305–320. <https://doi.org/10.1080/00324728.2014.934908> (2014).
- Li, Z. The effects of health resources inputs on population health in China. (Doctoral dissertation, Xiamen University). Retrieved from <https://link.cnki.net/doi/https://doi.org/10.27424/d.cnki.gxmd.2020.003110doi:10.27424/d.cnki.gxmd.2020.003110> (2020).
- Chen, B. Z. An empirical study on the influencing factors of population mortality at provincial level in China from the environmental perspective. (Master's thesis, Hainan Normal University). Retrieved from <https://link.cnki.net/doi/https://doi.org/10.27719/d.cnki.ghnsf.2021.000244doi:10.27719/d.cnki.ghnsf.2021.000244> (2022).
- Fan, L. W. Association of socioeconomic and all-cause mortality among older people in China based on latent class model. (Master's thesis, Zhengzhou University). Retrieved from <https://link.cnki.net/doi/https://doi.org/10.27466/d.cnki.gzzdu.2022.004968doi:10.27466/d.cnki.gzzdu.2022.004968> (2022).
- Ying, K. & Li, X. D. The evolution of spatial pattern and influencing factors of population mortality in China. *World Reg. Stud.* **2**, 440–452. <https://doi.org/10.3969/j.issn.1004-9479.2022.02.2020402> (2022).
- Ramsay, J. O. & Silverman, B. W. *Functional Data Analysis* 2nd edn. (Springer, 2005).
- Ramsay, J. O., Hooker, G. & Graves, S. B. *Functional Data Analysis with R and MATLAB* (Springer, 2009).
- Cardot, H., Ferraty, F., Mas, A. & Sarda, P. Testing hypotheses in the functional linear model. *Scandinavian J. Stat.* **30**, 241–255 (2003).
- Goia, A. Selection model in functional linear regression models for scalar response. *Developments in Applied Statistics*. Eds. Feligoy, A and Mrvar, A., Ljubljana, Slovenia: FDV, 97–110 (2003).
- Reiss, P. T., Huang, L., and Mennes, M. Fast function-on-scalar regression with penalized basis expansions. *Int. J. Biostat.* **6**(1) (2010).

25. Zhang, J. T. & Chen, J. Statistical inferences for functional data. *Ann. Stat.* **35**(3), 1052–1079 (2007).
26. Yang, X. W., Shen, Q., Xu, H. Q. & Shoptaw, S. Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures. *Stat. Med.* **26**, 1552–1566 (2007).
27. Shen, Q. & Faraway, J. J. An F test for linear models with functional responses. *Statistica Sinica.* **14**, 1239–1257 (2004).
28. Gertheiss, J., Rugamer, D., Liew, B.X., & Greven, S. Functional data analysis: an introduction and recent developments (2023).
29. Chen, X., Li, H., Liang, H. & Lin, H. Functional response regression analysis. *J. Multivar. Anal.* **169**, 218–233 (2019).
30. Yang, X., Shen, Q., Xu, H. & Shoptaw, S. Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures. *Stat. Med.* **26**(7), 1552–1566. <https://doi.org/10.1002/sim.2609> (2007).
31. Wen, Y. & Yin, Q. *Demography* (Southeast University Press, 2006).
32. National Bureau of Statistics. Questions and Answers on Statistics for Leading Cadres: What is Per Capita Disposable Income of Residents. Retrieved September 25, 2023, from http://www.stats.gov.cn/zs/tjwjs/tjzb/202301/t20230101_1903384.html (2023, January 1).
33. Ren, Q., You, Y. Z., Zheng, X. Y., Song, X. Y., & Chen, G. The level, pattern and regional differences in Chinese deaths since the 80s of the 20th century. *Chin. J. Popul. Sci.* (03), 21–31+81 (2004).

Acknowledgements

We would like to thank the Ministry of Education of China's Major Research Project (Grant number: 19JZD049) for funding this article.

Author contributions

T.L. (First Author) was in charge of conceptualization, data curation, formal analysis, methodology, resource management, software utilization, validation, visualization, and drafting the original manuscript. S.Z. contributed to data curation and manuscript proofreading. H.L. (Corresponding Author) was responsible for conceptualization, research supervision, validation, and review and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61262-5>.

Correspondence and requests for materials should be addressed to H.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024