

The diversity of income effects on mortality across regions in People’s Republic of China: Instrumental variable approach

Artur R. Nagapetyan^{a,*}, Tatiana I. Pavlova^b, Jun Li^c

^a *Far Eastern Federal University, Vladivostok, Russia*

^b *New Economic School, Moscow, Russia*

^c *University of Science and Technology, Liaoning, China*

Abstract

The issue of equality of opportunity is crucial in contemporary society, often examined in relation to income. Previous research has demonstrated the uncertain effect of income on mortality due to the presence of other factors. To empirically assess the difference in the effect of income on mortality rates between more developed and less developed provinces in China, we modeled mortality rates based on the instrumental variables method and the generalized method of moments. The study results indicate that a 1% increase in income reduces mortality by over 2%. The results confirm the hypothesis that income has a stronger negative impact on mortality in more socially developed provinces of the People’s Republic of China compared to less socially developed areas. This difference is statistically significant at the 1% level and is at least 10%. The obtained results and their development can significantly impact the implementation of effective political governance measures aimed at reducing mortality in certain territories and mortality in general.

Keywords: income, mortality, inequalities, instrumental variable, spatial analysis, socio-economic development, China.

JEL classification: C54, I14, I15.

1. Introduction

1.1. Motivation

What is the relationship between income and mortality? Numerous studies are conducted annually to explore the differences in mortality rates among various

* Corresponding author, E-mail address: nagapetyan_ar@dvfu.ru

regions and the impact of income on these rates.¹ In 2019, the all-cause mortality rate in different provinces of China varied significantly, with some regions differing by up to 70%. According to the National Bureau of Statistics of China, Chongqing province had a mortality rate of 7.57 deaths per 1,000 people, while Xinjiang province had a mortality rate of 4.45.² The income of a population can directly impact mortality rates. For instance, it can affect the ability to pay for necessary treatment, medicine, or other actions that can reduce the likelihood of death. Additionally, income can indirectly impact mortality through socio-economic factors, such as education (Clark and Royer, 2013; Lindahl, 2005). As noted in the literature, it is important to consider the effect of accumulated factors on mortality when assessing the impact of income, particularly those resulting from low incomes (Lincoff et al., 2023). Low income in the past can lead to under-education and under-funding of investments in one's own health. This may not only affect future mortality rates but moreover cause no impact of income on mortality in the present for those adversely affected by low income in the early years (Koch et al., 2010a, 2010b).

Such evidence poses serious challenges to the implementation of effective policy interventions to reduce mortality in particular areas and mortality inequalities in general. Answers to questions regarding the impact of income on mortality rates, both generally and in specific areas, can directly influence the need for certain activities. These activities may include the development of health and other types of insurance to assist low-income individuals when their lives are at risk, as well as direct subsidy policies for the poor. The size and necessity of these subsidies should be determined based on the level of social development in the regions. As will be discussed in Subsection 4.3, on the example of the provinces of the People's Republic of China, the level of social development refers to the quality of social infrastructure available to its residents (quality of health care, education—qualifications and competence of persons providing such services, etc.; Xiang et al., 2020; Zhang et al., 2015). These characteristics directly impact people's ability and need to use their income to improve their chances of survival.

1.2. Research problem

This paper explores the challenge of evaluating the effect of income on mortality. Despite the numerous studies conducted on this topic, there is no consensus in the literature. While some researchers have found a negative correlation between income and mortality (Backlund et al., 1996, 1999; Lindahl, 2005; Martikainen et al., 2001; O'Hare et al., 2013), others have reported no statistically significant correlation (Ahammer et al., 2016; Blakely et al., 2004; Koch et al., 2010a). In some articles, the authors suggest the possible presence of a positive influence (Adda et al., 2009; Koch et al., 2010b). Examples of mechanisms that demonstrate a negative impact on mortality include the ability to finance costs that reduce the likelihood of death in the event of adverse conditions. Another mechanism can be described through the effect of population income on other socioeconomic characteristics that reduce the likelihood of mortality, such as education and investment in one's own health. Examples of mechanisms demonstrating the potential

¹ See, for example, <https://pubmed.ncbi.nlm.nih.gov/?term=income+Mortality>

² <https://data.stats.gov.cn/english/easyquery.htm?cn=E0103>

positive effect of income on mortality include the following: individuals with higher incomes are more likely to work in an office, have a sedentary lifestyle, and are less likely to engage in physical labor (Lindsay et al., 2017). Likewise, despite the fact that, on the one hand, high incomes can allow individuals to consume high-quality products, do sports, and relax more in nature, on the other hand, high incomes can directly affect eating behavior, for example leading to the abuse of harmful products (sugar, processed meat, tobacco, alcohol, drugs; Adda et al., 2009; Ettner, 1996; Xu, 2013), which together with a sedentary lifestyle can become the prerequisite for a wide range of diseases, excessive weight gain (Ren et al., 2019; Sahabudhee et al., 2023), lower immunity levels, and increased stress, etc. In recent decades, China has lost its title as one of the slimmest nations and now tops the list of countries with overweight populations (Zhang et al., 2020).³ Being overweight has a positive correlation with population mortality (Flegal et al., 2013; Lincoff et al., 2023). Furthermore, the potential for false negatives and false positives between income and mortality may lead to an underestimation or overestimation of their effect, particularly due to the presence of missing variables. For instance, a false-positive connection may arise if a potential omitted variable is present. This variable could be the level of concentration in the area of industries that are characterized by a high pollution intensity (Arceo et al., 2016). This factor may increase the income of the population but correspondingly lead to higher mortality rates due to the negative impact of pollution on health. The quality of education in a given territory may be a potential prerequisite for a false-negative correlation. Higher quality education may lead to higher income levels and a lower probability of mortality (Backlund et al., 1999).

To a certain extent, the characteristics of territories, such as their level of social development, can cause variations in the influence of the income indicator on mortality. For example, the possibilities available to people, and their need to use income to influence the probability of survival depends, among other things, on the conditions of the environment in which they live. This will be demonstrated below, in the literature review of the cases of countries with different levels of social infrastructure development, and among the relatively more socially developed countries, and among the relatively less socially developed countries (GlobalSurg Collaborative, 2016; Kondo et al., 2009). A statistically significant negative influence was more often found in countries with relatively more developed social infrastructure, such as the U.S. (Backlund et al., 1996, 1999), Finland (Martikainen et al., 2001), and Sweden (Lindahl, 2005), while a statistically insignificant or even positive influence was more often found in the countries with less developed social infrastructure, such as Chile (Koch et al., 2010a), or the countries of Latin America (Koch et al., 2010b). Examples of variation in the effect of income on mortality rates were found (Ahammer et al., 2016; Backlund et al., 1999; Blakely et al., 2004; Fichera and Savage, 2015; Koch et al., 2010a, 2010b; Martikainen et al., 2001). The above evidence allows us to form the following research contradiction.

On the one hand, it can be assumed that residents of territories with a more developed social infrastructure, given the advantages and opportunities from which they can benefit, make them less likely to face the threat of death, even if they have low income, than residents of territories with less developed social

³ <https://renewbariatrics.com/obesity-rank-by-countries/>

infrastructure (i.e. income has less negative impact on mortality in more socially developed regions). For example, due to more attractive conditions in terms of the social package (competition for workers), existing support measures, the quality of social institutions, employment opportunities, and corporate insurance in more socially developed regions compared to less developed regions.

On the other hand, residents of territories with more developed social infrastructure, given the advantages and opportunities that they can take advantage of, may be more likely to influence the reduction of mortality, especially if they have high income, than residents of territories with less developed social infrastructure (i.e. income has a higher negative impact on mortality in more socially developed regions). For example, if a person has a high income in a region with a low level of social infrastructure development, for example, low quality of health care, etc., he/she will be much less likely to increase his/her chances of survival, compared to a person with a high income living in a territory with a high quality of health care.

1.3. Research question

Is it true that more socially developed provinces in China have a higher negative effect of income on mortality than less socially developed provinces? (i.e., income has a higher negative effect on mortality in more socially developed regions).

The study contributes to the literature in several ways. First, as part of combating endogeneity, potentially caused by omitted variables and other econometric problems. In addition to the empirical consideration of existing approaches to using the instrumental variable method (using local unemployment rate as an instrument to estimate the effect of income on mortality), a new instrument constructed on the basis of the assumed exogenous variation of parameters of neighboring labor markets is proposed. This instrument, bearing in mind the proposed rationale, peculiarities of the labor market in the People's Republic of China (PRC) and approaches to robustness checks can be considered as exogenous to the modeled mortality variable in the region in question (more details in Subsection 4.2). Second, we assessed the presence of variation in the impact of income on mortality in regions with different levels of social infrastructure development, which directly affects the ability and need for the population to use income to improve their chances of survival. In this case, an additional factor that allows us to hope for the accuracy of the obtained estimates is the use of two sources of exogenous variation. The first one is based on the use of an instrumental variable constructed on the basis of the assumed exogenous variation of neighboring labor markets. The other one is based on the fact of formation of groups of PRC provinces differing in the level of social development, largely due to the geographical characteristics of the respective territories (more details in Subsection 4.3).

The results of the study show that a 1% increase in income leads to a reduction of the mortality rate by more than 2%. At the same time, when the instrumental variable method is not used, the size of the estimated impact is less than 1%, which leads to the risk of underestimating it by more than a factor of two. The results furthermore allow us to confirm the hypothesis that the effect of income on mortality in the more socially developed provinces of the PRC has a greater negative effect compared to the less socially developed territories. At

the same time, the detected difference is statistically significant at the 1% level of significance and is not less than 10%.

The rest of the paper is as follows. Section 2 reviews the relevant literature. Section 3 describes the data used in the paper. Section 4 presents the empirical strategy, including a discussion of the choice of the dependent variable, the choice of the instrumental variable used in the literature (4.1) and the proposed instrumental variable in the paper (4.2), and a discussion of the approach to forming groups of PRC provinces according to the level of socio-economic development (4.3). Section 5 presents the main results. Section 6 discusses the approaches to robustness checks. Section 7 presents the main conclusions.

2. Literature review

The literature presents various results characterizing the influence of income on the mortality rate. Among other things, there is variation in the effect of income on mortality in the countries with relatively more developed social infrastructure, as well as in developing countries. Backlund et al. (1999) in a study of U.S. mortality data on more than 400,000 men and women aged 25 to 64 years from the National Longitudinal Mortality Study (NLMS), found a negative relationship between income and mortality rates. The reduction in mortality associated with an increase in income per \$1,000 was much greater for incomes below \$22,500 than for incomes above \$22,500. Another study, similarly based on an analysis of U.S. data, showing a significant negative relationship between income and mortality, noted its variation for different social groups (Arellano and Bond, 1991). Fichera and Savage (2015), using information on precipitation in Tanzania as an exogenous variation, find a positive effect of population income on health outcomes. Martikainen et al. (2001) additionally examine mortality rates in Finland for all men and women over 30 years of age. The researchers conclude that there is a negative relationship between income and mortality indicators. They explain it by the causal effects of poverty, especially by the accumulation of factors increasing mortality. In some studies, to overcome the threat of biased results (due to endogeneity to estimate the effect of income on mortality) using Swedish data, by using as exogenous variation the cash lottery prizes of individuals, the authors conclude that a 10% income increase generates 4–5% of a standard deviation better health and decreases the probability of dying by 2–3 p.p. (Lindahl, 2005).

In contrast, Ahammer et al. (2016), using quasi-experimental methods to analyze the Austrian data, demonstrate the absence of a statistically significant relationship between income and mortality rates. Blakely et al. (2004) on the analysis of New Zealand data show a weak link between income and mortality, especially when considering the influence of other socio-economic factors. An analysis of the Chilean data similarly failed to find an income-mortality connection in the case of individuals adversely affected by low income at an early age, which, among other things, was reflected in their level of education (Koch et al., 2010a). The contradictions are especially exacerbated by the emergence of studies describing the positive effect of income on mortality. For example, in their study by Adda et al. (2009) analyze data on income, expenditures, sociodemographic factors, and other risk factors, using exogenous sources of income variation, and concludes that income has a positive effect on mortality rate. This conclusion is

largely due to more risky health behaviors. Koch et al. (2010b) in their study based on an analysis of Latin American data in countries with economies in transition (adjusting for age, sex, education, and behavioral and biological risk factors for mortality) conclude that there is an adverse effect of income on mortality.

In addition to the fact that people with high incomes are more likely to work in an office, have sedentary lifestyles, and are less likely to do physical labor, there is evidence in the literature demonstrating the effect of income on increasing the preconditions for bad habits, poorer eating habits, etc. For example, Ettner (1996) demonstrates the effect of income growth on increased alcohol consumption which is consistent with the results of Adda et al. (2009) on the increase in the degree of health risk behavior with increasing income. There is likewise evidence demonstrating a positive connection between increased income and increased cigarette consumption in the United States and decreased physical activity (Xu, 2013). Ren et al. (2019) also show in their study based on data from the China Health and Nutrition Survey (CHNS) that body weight and the likelihood of being overweight increase with additional income. They show that low-income people are less likely to be overweight in China (a country in transition) which can be explained by income restrictions on unhealthy foods. In turn, Flegal et al. (2013) based on a meta-analysis of more than 97 studies covering observations of 2.88 million people and more than 270,000 deaths, provide compelling evidence of the impact of being overweight with significantly higher rates of all-cause mortality.

Thus, in the reviewed literature there is no consensus about the influence of income on mortality rates. At the same time, there is evidence of negative, zero, and positive impacts, including evidence based on quasi-experimental methods, which further exacerbates the problem under consideration.

The variable characterizing the unemployment rate is often used as an instrumental variable in studies assessing the impact of income on health outcomes (Ettner, 1996; Kuehnle, 2014; Wei and Feeny, 2019; Xu, 2013). Subsection 4.1 will consider the use of unemployment rate as an instrumental variable in the reviewed studies.

Potential variation in the influence of income indicator on mortality can be explained by the presence of heterogeneity associated with different characteristics of the territories where the data were collected for the study. For example, as analysis of the literature shows (Ahammer et al., 2016; Backlund et al., 1999; Blakely et al., 2004; Fichera and Savage, 2015; Koch et al., 2010a, 2010b; Martikainen et al., 2001), the level of social development of the territory (measures of social support, quality of health care), determines the possibility and the need to use income to affect the probability of survival in the event of certain risks to life. At the same time, the present analysis of the literature revealed the variation in the influence of income on mortality both in countries with relatively more developed social infrastructure and in developing countries. Still, a statistically significant negative influence was more often found in the countries with relatively more developed social infrastructure, such as the U.S. (Backlund et al., 1996, 1999), Finland (Martikainen et al., 2001), Sweden (Lindahl, 2005), while a statistically insignificant or even positive influence was more often found in the countries with less developed social infrastructure, such as Chile (Koch et al., 2010a), and countries in Latin America (Koch et al., 2010b).

3. Data and variables

The study is based on regional data from the 31 provinces of Mainland China, covering the years 2010 to 2019, which were obtained from the National Bureau of Statistics of China.

Mortality rate was chosen as the dependent variable. Some variables were used unchanged in the paper, while others were modified to be included in the study (see Table 1 for details). For example, the Per capita disposable income variable was modified to obtain Per capita real disposable income. For this purpose, the Consumer Price

Table 1

Variables and their designations.

Variable name	Description	Variable
<i>Demography & sociology</i>		
1. Mortality rate	Crude death rate, deaths per 1,000 person	<i>death_rate_</i>
2. Birth rate	Crude birth rate, births per 1,000 person	<i>birth_rate_</i>
3. Sex ratio (female = 100)	Number of men per 100 women (sample survey)	<i>sex_</i>
4. Older people rate	Old-age dependency ratio (sample survey, 65 and older), %	<i>old_</i>
5. Urban population rate	Number of urban residents per 1 person	<i>city_p_</i>
6. Never married rate	Population aged 15 and over, never married (sample survey), per 1 person	<i>marriage_no_p_</i>
7. Divorced population rate	Population aged 15 and over, divorced (sample survey), per 1 person	<i>divorce_p_</i>
8. Higher education rate	Population aged 6 and over, college and higher level (sample survey), per 1 person	<i>educ_high_p_</i>
<i>Economy</i>		
9. Real income	Per capita disposable income adjusted by Consumer Price Index, yuan	<i>income_real_</i>
10. Unemployment rate	Unemployment rate in urban area, %	<i>unempl_</i>
11. Real gross regional product per capita	Gross regional product per capita adjusted by Consumer Price Index, 1,000,000 yuan per person	<i>gdp_p_r_</i>
12. Average unemployment rate in neighboring regions	Average unemployment rate in urban area in neighboring regions. Regions with a common border (group A) and regions with a common border with regions from group A, % (see subsection 4.2)	<i>unempl_i7_</i>
13. Level of social development of region (3 sectors)	The provinces of China were divided into three groups depending on the level of socio-economic development. Categorical variable. 1—low level of social development, 2—average level of social development, 3—high of social development (see subsection 4.3)	<i>sector3</i>
<i>Health</i>		
14. Number of doctors per population	Number of licensed doctors per 1 person	<i>doctors_lic_p_</i>
15. Number of medical beds per population	Number of beds of medical institutions per 1 person	<i>beds_p_</i>
16. Visits in health institutions per population	Number of visits in health institutions per population, 10 000 person-times per 1 person	<i>visits_p_</i>
<i>Ecology</i>		
17. Air pollution per territory	Nitrogen Oxides Emission in Waste Gas, 10,000 tonnes nitrogen oxides emission in waste gas per 1 square km	<i>air_pollut_</i>

Source: Compiled by the authors.

Index (2010 = 100) was used, which in turn was calculated based on the Consumer Price Index variable (preceding year = 100) in the database we used.

Part of the data in the National Bureau of Statistics of China is based on the sample survey. We are talking about such variables as Sex ratio, Proportion of older people, Population that never married, Divorced population, Higher education.

In addition to the instrumental variable used in the literature for the real income, variable in the form of unemployment rate within the region under consideration, was constructed a variable—Average unemployment rate in neighboring regions (*unempl_i7_*) based on the assumed exogenous variation of neighboring labor markets.

The Level of social development of region variable (No. 13 in Table 1) divides China's regions according to their level of social and economic development (more details in Subsection 4.3). Fig. 1 shows a heat map showing the level of development of territories in the context of the Real gross regional product per capita. It can be observed that the richer as well as the more socially developed territories are located near the coast (this will be justified in Subsection 4.3). In the west—the least developed and in the center—regions with an average level of social development.

Detailed information on the descriptive statistics of the variables under consideration is given in Table 2 (data for all years are considered). Data on key variables for PRC provinces is presented in Supplementary material (Table A1).

Detailed information on the descriptive statistics of the variables under consideration, taking into account the division into groups of provinces according to the level of social development is given in Table 3 (data for all years are considered).

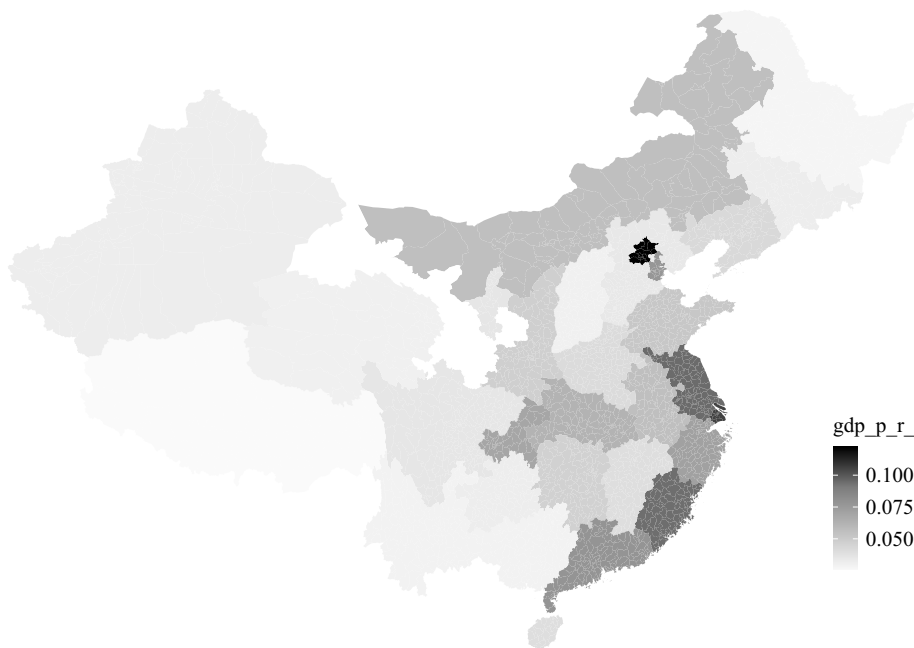


Fig. 1. Heat map showing the level of development of territories in the context of the variable Real gross regional product per capita, 2019.

Source: Compiled by the authors.

Table 2

Descriptive statistics of the variables in question.

Variable	Mean	St. dev.	Count	Min	Max
<i>Demography & sociology</i>					
Mortality rate	6.03	0.78	310	4.21	7.57
Birth rate	11.34	2.72	310	5.36	17.89
Sex ratio (female = 100)	104.98	4.61	310	86.96	123.17
Older people rate	13.55	3.48	310	5.60	23.80
Urban population rate	0.57	0.13	310	0.23	0.90
Never married rate	0.17	0.03	310	0.10	0.27
Divorced population rate	0.02	0.01	310	0.00	0.05
Higher education rate	0.12	0.07	310	0.02	0.48
<i>Economy</i>					
Real income	15,697.38	7431.40	310	5055.70	51,544.10
Unemployment rate	3.28	0.65	310	1.20	4.50
Real gross regional product per capita	0.04	0.02	310	0.01	0.12
Average unemployment rate in neighboring regions	3.31	0.22	310	2.69	3.94
Level of social development of regions (3 sectors)	1.97	0.86	310	1.00	3.00
<i>Health</i>					
Number of doctors per population	0.00	0.00	310	0.00	0.00
Number of medical beds per population	0.00	0.00	310	0.00	0.01
Visits in health institutions per population	0.0005	0.0002	310	0.0003	0.0011
<i>Ecology</i>					
Air pollution per territory	0.00	0.00	310	0.00	0.01

Source: Authors' calculations.

The analysis of the data in Table 3 makes it possible to assess the main characteristics of the groups of PRC provinces highlighted in Subsection 4.3. For example, let us consider the Real income indicator. On average in Chinese provinces it is 16,697 yuan, while on average in regions with a low level of social development, an average level of social development and high level of social development, it is 11,602.8, 13,509.8 and 21,755.2 yuan, respectively. Columns 6, 7, 8 provide information on testing different hypotheses about the relationship of these indicators. We check whether the difference between low level of social development and high level of social development is statistically significant. We look specifically at the most developed and least developed to determine the difference between the extreme cases. For example, according to the results obtained, the difference between the average values of the group of regions with low level of social development and high level of social development is 10,152.4 yuan (column 5) and is statistically significant. The data in Table 5 clearly shows the difference between the development indicators of the selected groups of provinces. According to the results, the average values of Gross regional product, Real gross regional product per capita, Income, Real income, Number of doctors, Number of doctors per population, Number of medical beds, Visits in health institutions, Visits in health institutions per population, Higher education of regions with high level of social development are statistically significantly higher than the corresponding average values in regions with low level of social development.

Consider the histograms of the distributions of the values of key variables, including in order to consider the level of variation and take the first steps to identify outliers(tales).

Table 3
Descriptive statistics of the variables under consideration, taking into account the division into groups of provinces according to the level of social development.

Variables	Mean (all)	Mean (1) (sector3 = 1)	Mean (2) (sector3 = 2)	Mean (3) (sector3 = 3)	Mean (3)	diff = mean(1) – mean(3)	Ha: diff < 0 Pr(T < t)	Ha: diff != 0 Pr(T > t)	Ha: diff > 0 Pr(T > t)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
<i>Demography & Sociology</i>									
Death rate	6.03	6.02	6.27	5.86	0.16	0.929	0.143	0.071	
Birth rate	11.34	12.52	10.74	10.50	2.02	1	0.000	0.000	
Sex ratio (female = 100)	104.98	104.55	104.51	105.81	-1.26	0.029	0.059	0.971	
Older people rate	13.55	12.83	14.16	13.90	-1.07	0.014	0.029	0.986	
Urban population rate	0.57	0.48	0.53	0.68	-0.20	0.000	0.000	1	
Never married rate	0.17	0.17	0.15	0.17	0.00	0.274	0.548	0.726	
Divorced population rate	0.02	0.02	0.02	0.01	0.00	1.000	0.000	0.000	
Higher education rate	0.12	0.10	0.10	0.16	-0.06	0.000	0.000	1	
<i>Economy</i>									
Real income	15,697.40	11,602.80	13,509.80	21,755.20	-10,152.40	0.000	0.000	1	
Unemployment rate	3.28	3.34	3.46	3.07	0.28	0.999	0.002	0.001	
Real gross regional product per capita	0.036	0.027	0.030	0.052	-0.025	0.000	0.000	1	
Average unemployment rate in neighboring regions	3.31	3.33	3.35	3.28	0.05	0.955	0.090	0.045	
Level of social development of regions (3 sectors)	1.97	1.00	2.00	3.00	-2.00				
<i>Health</i>									
Number of doctors per population	0.00	0.00	0.00	0.00	0.00	0.000	0.000	1	
Number of medical beds per population	0.0050	0.0051	0.0052	0.0047	0.0005	1.000	0.0005	0.0003	
Visits in health institutions per population	0.00	0.00	0.00	0.00	0.00	0.000	0.000	1	
<i>Ecology</i>									
Air pollution per territory	0.00053	0.00045	0.00043	0.00068	-0.00023	0.000	0.000	1	

Source: Authors' calculations.

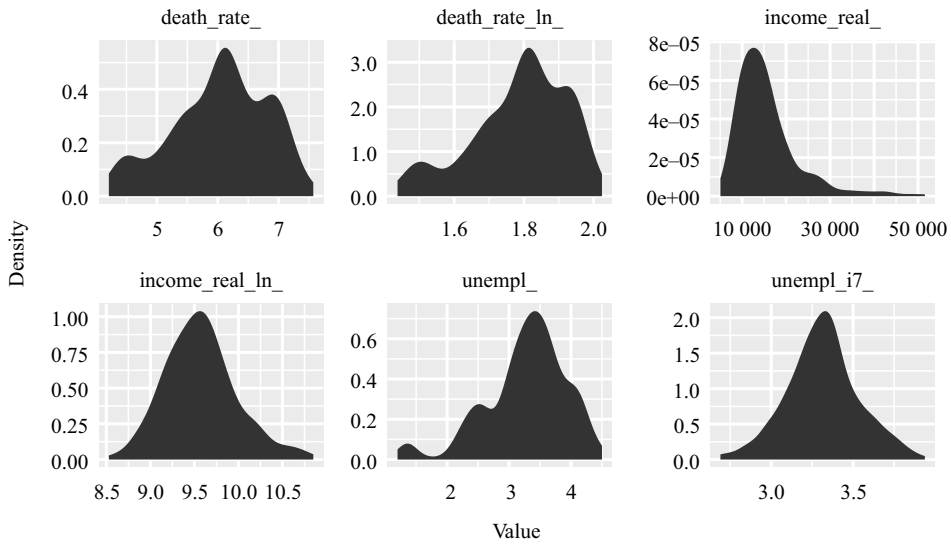


Fig. 2. Distributions of key variables (data for all years are considered).

Note: Unless otherwise stated, the presence of “ln” in the designation of a variable means that the logarithm of the corresponding value is considered.

Source: Compiled by the authors.

Fig. 2 shows histograms of distributions (data for all years are considered) of the dependent variable, the key independent variable whose influence we want to evaluate, as well as the distribution of indicators (candidates for the role of instrumental variables) characterizing the unemployment rate in the region under consideration, as well as the variable Average unemployment rate in neighboring regions (more details in Subsection 4.2).

The data in Fig. 2 show a positive skewness for mortality (*death_rate_*) and unemployment (*unempl_*), as well as a positive kurtosis. Like other variables, the real income variable (*income_real_*) has outliers (heavy tails). Since the increase in absolute terms for the same number of income units can have a different value depending on the current amount of income, it was decided to use the logarithm of this variable, which helps to combat this problem. Taking the logarithm of such variables as the death rate (*death_rate_*) and unemployment (*unempl_*) does not help fight outliers and bring the data to a more “normal” form. However, it was decided to take the logarithm of the death rate variable (*death_rate_*) for easier interpretation of the regression results. The variable Average unemployment rate in neighboring regions (*unempl_i7_*) has a more normal distribution. Supplementary material (Fig. A1) shows histograms of distributions of key variables for 2019, which show similar results.

Fig. 3 shows histograms of distributions (considering data for all years) of other variables used, including control variables. The variables considered in Fig. 3, including such control variables as the birth rate (*birth_rate_*), real GRP per capita (*gdp_p_r_*), sex (*sex_*) are characterized by kurtosis and asymmetry, which differ from the normal distribution. This may be due to the differentiation of socio-economic indicators of the development of the PRC regions. For a more accurate analysis of data related to key variables, we will build graphs showing outliers using the Box Plots package (boxplot function) in R (see Supplementary material, Figs. A2–A4)

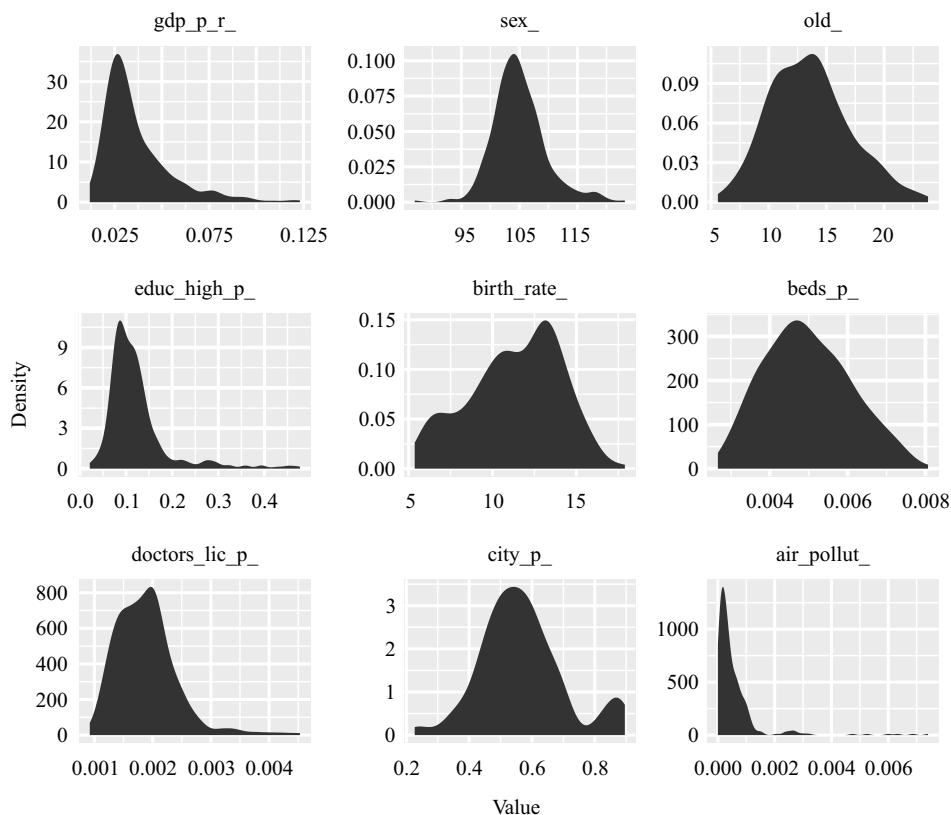


Fig. 3. Distributions of other variables used, including control variables (data for all years are considered).

Source: Compiled by the authors.

Fig. A2 in Supplementary material highlights two regions that can be categorized as outliers when considering the real income variable (*income_real_*). We are talking about such provinces as Beijing, Shanghai. This result is quite intuitive, because these regions are among the most developed in China. Outliers were not found when considering the death rate variable (*death_rate_ln_*). Outliers were similarly not found for variables characterizing the average unemployment rate in neighboring regions (*unempl_i7*). For unemployment (*unempl_*), Beijing with the lowest unemployment rate can be considered as an outlier. For the variable that characterizes the value of real GRP per capita, two regions can also be distinguished, namely Beijing, Shanghai, which confirms the results obtained earlier for the real income variable. In order to neutralize the potential negative impact on the quality of modeling of detected outliers, it was decided to evaluate the corresponding models both with and without these regions and subsequently draw conclusions about the degree and direction of their influence. The results of the main regressions where these two regions are excluded are shown in Table 6 (more details in Subsection 6.2—robustness check 2). Almost all of the main results associated with the application of the instrumental variable proposed in the study remain unchanged. Thus, the results associated with estimating the effect of income on mortality for the PRC provinces as a whole remain unchanged. In regressions where the difference in

the effect of income on mortality in different groups of provinces is searched for, the difference between low level of social development regions and high level of social development regions becomes statistically less significant (10% significance level) for the model with all controls. This can be explained by the fact that by removing the two most developed regions from the group of high level of social development regions, this group becomes more similar to the group of low level of social development regions.

Figs. 4–7 illustrate the association between the mortality rate and real income in 2019 across all provinces of Mainland China, encompassing less developed regions, regions with medium social development, and more developed regions. The data in Figs. 4–7 within the framework of the primary analysis demonstrate the following associations between the considered variables: negative for all

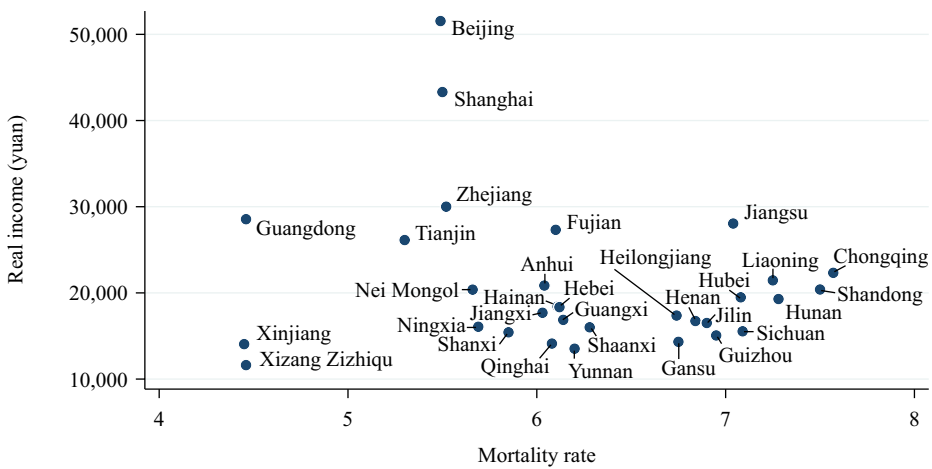


Fig. 4. The association between mortality rate and real income for regions of Mainland China: all regions.

Source: Compiled by the authors.

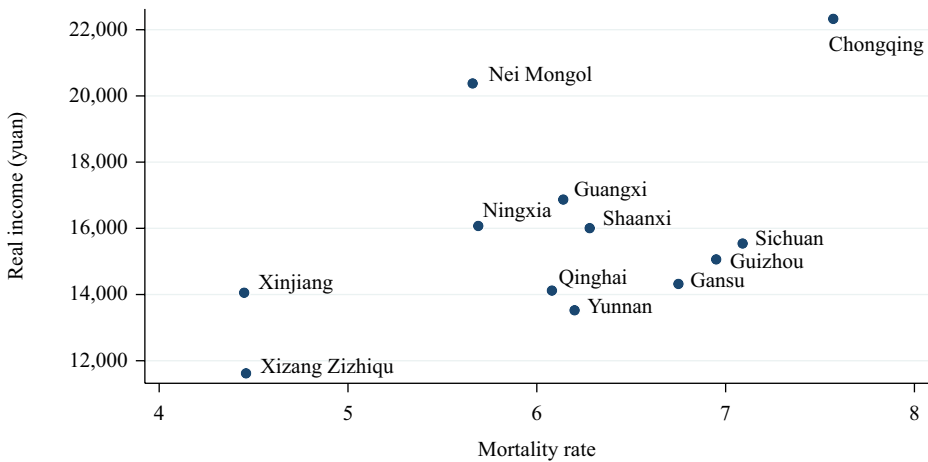


Fig. 5. The association between mortality and real income for regions of Mainland China: regions with low level of social development.

Source: Compiled by the authors.

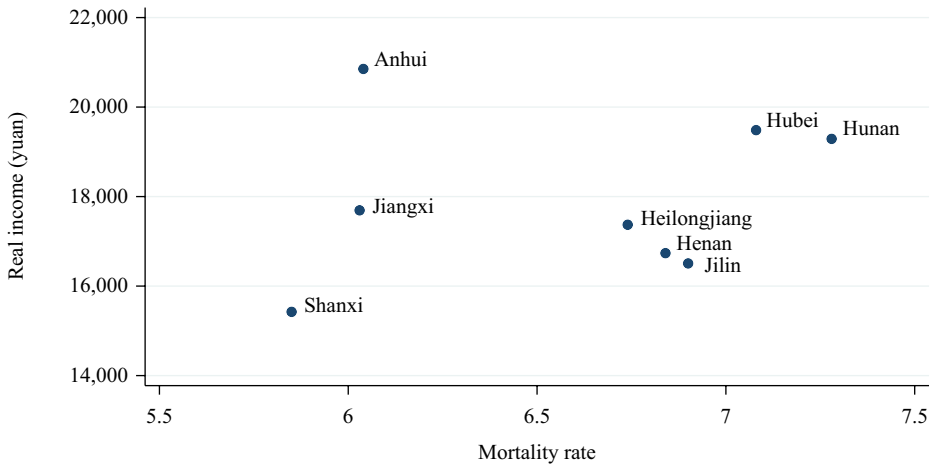


Fig. 6. The association between mortality rate and real income for regions of Mainland China: regions with average level of social development.

Source: Compiled by the authors.

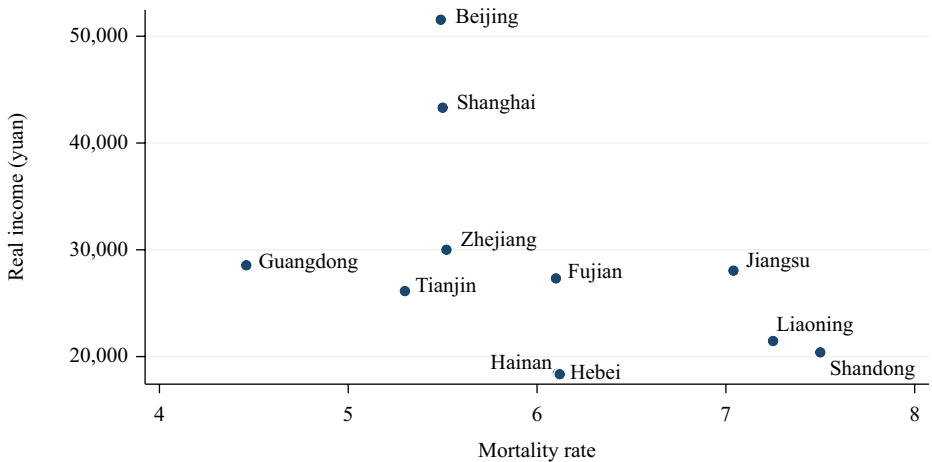


Fig. 7. The association between mortality and real income for regions of Mainland China: high developed regions.

Source: Compiled by the authors.

regions (correlation coefficient is -0.12), positive for group of low level of social development regions (correlation coefficient is 0.19), for group of average level of social development regions (correlation coefficient is 0.35), negative for group of high level of social development regions (correlation coefficient correlation is -0.41). As part of the overall course of the study, it is important to note that the relationships found for the considered groups of regions are different. More rigorous modeling in the future will better explain this behavior in the data. It is also better to evaluate the value of the coefficients more strictly and establish how even in such a simple way, such correlations see this relationship. To a certain extent, the results are consistent with our hypothesis that the negative effect of income on mortality is higher in the more socially developed provinces of China than in the less developed provinces.

4. Analytical framework and empirical strategy

The dependent variable chosen was the mortality rate. And not the all-cause mortality rate for the following reasons. First, income can affect mortality not only through diseases, but also through other channels: frostbite; death from hunger; infant, child, maternal mortality due to insufficient nutrition of mother and child (poor families may give birth at home, which moreover increases the risk of mother or child death even in the absence of diseases); extreme sports and entertainment (in high income groups). Second, and more importantly, a large number of deaths from various diseases are not actually counted in statistics, because of the low detection rate (in the provinces of the PRC). For example, because poor people may simply not go to the doctor for a variety of reasons: “bad” social package (no sick insurance), lack of time (need to feed family), especially for poor people; very long queues to doctors, which is especially relevant for the PRC because of the large number of inhabitants. It is important to consider that regions with higher incomes have significantly higher rates of disease detection than poor regions (relevant for PRC) (in fact, considering the variable mortality from all diseases can lead to false positives: higher incomes, higher rates of disease detection, higher rates of disease mortality, compared to regions with lower detection rates). Thus, the merits of the mortality indicator include the fact that it is much less dependent on the level of disease detection, and also takes into account a wider range of channels of influence of income on mortality.

Because of the endogeneity problems in the model described in the Introduction section (e.g., because of missing variables), the results of standard regressions may be biased, and the estimated effects underestimated or overestimated.

Our main identification strategy is to apply the instrumental variable method. The analysis of the literature has shown that studies of similar cases often use as an exogenous variation the characteristics of the labor market of the region in question, in particular the unemployment rate for instrumental variable in assessing the impact of income on health outcomes (Ettner, 1996; Kuehnle, 2014; Wei and Feeny, 2019; Xu, 2013). Despite this, as will be shown in Subsection 4.1, there is a debate about the extent to which the exogeneity property is satisfied for this instrument.

In view of this, we propose to consider another instrumental variable constructed on the basis of the assumed exogenous variation of neighboring labor market parameters—Average unemployment rate in neighboring regions (*unempl_i7_*) (regions with a common border (group A) and regions with a common border with regions from group A). The details of the rationale for the use of this variable as an instrument and its calculation will be discussed in Subsection 4.2.

Considering that the research question involves checking that in more socially developed provinces of China the influence of income on mortality is higher than in less developed provinces, Subsection 4.3 provides a rationale for forming groups of provinces of China with low level of social development, average level of social development and high level of social development based on literature, their historical, geographical and economic characteristics (Xiang et al., 2020; Zhang et al., 2015).

The design of the research project likewise involves the implementation of several robustness checks, including for testing the stability of the results by adding control variables (Average Death Rate in neighboring regions (*death_rate_i7_*) and unemployment in the region itself) (Subsection 6.1), removing outliers (Subsection 6.2)

and testing the hypothesis of a relationship between the death rates of different regions (Subsection 6.3)

The following basic models will be evaluated in this study (the model numbering see in Table 4).

- Pooled regression (1).

$$\begin{aligned} \log(\text{death_rate}_i) = & \alpha_0 + \beta_1 \times \log(\text{income_real}_i) + \beta_2 \times \text{sex}_i + \beta_3 \times \text{old}_i + \\ & + \beta_4 \times \text{beds_p}_i + \beta_5 \times \text{city_p}_i + \beta_6 \times \text{educ_high_p}_i + \beta_7 \times \text{doctors_lic_p}_i + \\ & + \beta_8 \times \text{birth_rate}_i + \beta_9 \times \text{marriage_no_p}_i + \beta_{10} \times \text{divorce_p}_i + \\ & + \beta_{11} \times \text{gdp_p_r}_i + \beta_{12} \times \text{air_pollut}_i + \beta_{13} \times \text{visits_p}_i + \varepsilon_i, \end{aligned} \quad (1)$$

where $\log(\text{death_rate}_i)$ is the natural logarithm of the variable death_rate_i ; the designations of other variables are given in Table 1.

- Panel data model with fixed effects (2). The advantages of this model include the possibility to solve omitted variable bias in terms of variables that are different for different territories but do not change over time (region fixed effects) and variables that change over time but are the same for all territories (time fixed effects).

$$\begin{aligned} \log(\text{death_rate}_{it}) = & \alpha_i + \beta_1 \times \log(\text{income_real}_{it}) + \beta_2 \times \text{sex}_{it} + \beta_3 \times \text{old}_{it} + \\ & + \beta_4 \times \text{beds_p}_{it} + \beta_5 \times \text{city_p}_{it} + \beta_6 \times \text{educ_high_p}_{it} + \beta_7 \times \text{doctors_lic_p}_{it} + \\ & + \beta_8 \times \text{birth_rate}_{it} + \beta_9 \times \text{marriage_no_p}_{it} + \beta_{10} \times \text{divorce_p}_{it} + \\ & + \beta_{11} \times \text{gdp_p_r}_{it} + \beta_{12} \times \text{air_pollut}_{it} + \beta_{13} \times \text{visits_p}_{it} + \\ & + \beta_{14} \times \text{factor}(\text{year}_t) + \varepsilon_{it}, \end{aligned} \quad (2)$$

where $\text{factor}(\text{year}_t)$ is the dummy variable per year.

- A fixed-effects panel data model using the instrumental variable method. Among other things, the instrumental variable method will provide more accurate estimates by using exogenous variation and partially solving the endogeneity problem. The unemployment rate (unempl_i) (3) and the Average unemployment rate in neighboring regions (unempl_i7_i) proposed in this paper (4) will be considered as instrumental variables.
- Model of panel data with fixed effects including interactions between the variable Real Income and the categorical variable—Level of social development of regions (3 sectors) to estimate the difference in the effect of income on mortality for groups of regions with different levels of social development (5).

$$\begin{aligned} \log(\text{death_rate}_{it}) = & \alpha_i + \beta_1 \times \log(\text{income_real}_{it}) + \beta_2 \times \text{sex}_{it} + \beta_3 \times \text{old}_{it} + \\ & + \beta_4 \times \text{beds_p}_{it} + \beta_5 \times \text{city_p}_{it} + \beta_6 \times \text{educ_high_p}_{it} + \beta_7 \times \text{doctors_lic_p}_{it} + \\ & + \beta_8 \times \text{birth_rate}_{it} + \beta_9 \times \text{marriage_no_p}_{it} + \beta_{10} \times \text{divorce_p}_{it} + \\ & + \beta_{11} \times \text{gdp_p_r}_{it} + \beta_{12} \times \text{air_pollut}_{it} + \beta_{13} \times \text{visits_p}_{it} + \beta_{14} \times \text{factor}(\text{year}_t) + \\ & + \beta_{15} \times \log(\text{income_real}_{it}) \text{sector3}(= 2) + \\ & + \beta_{16} \times \log(\text{income_real}_{it}) \text{sector3}(= 3) + \varepsilon_{it}, \end{aligned} \quad (5)$$

Table 4
Regressions results (basic models).

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	FE	FE (iv: Unemployment rate)	FE (iv: Average unemployment rate in neighboring regions)	FE for 3 sectors	FE (iv: Interactions between Unemployment Rate and Level of social development of regions) for 3 sectors	FE (iv: Interactions between Average unemployment rate in neighboring regions and Level of social development of regions) for 3 sectors)
Real income	-0.089* (0.052)	-0.841*** (0.166)	-4.603 (7.101)	-1.987*** (0.424)	-0.910*** (0.176)	-4.069 (3.404)	-2.302*** (0.555)
Real income when level of social development is average (additionally)					-0.006 (0.029)	-0.021 (0.093)	-0.071 (0.043)
Real income when level of social development is high (additionally)					-0.058 (0.040)	0.004 (0.130)	-0.223*** (0.074)
Sex ratio	-0.001 (0.001)	-0.001** (0.001)	-0.003 (0.003)	-0.002** (0.001)	-0.001* (0.001)	-0.003 (0.002)	-0.001 (0.001)
Older people rate	0.028*** (0.003)	0.009*** (0.003)	-0.006 (0.029)	0.005 (0.003)	0.010*** (0.003)	-0.004 (0.014)	0.009*** (0.003)
Number of medical beds per population	16.455* (9.187)	36.317*** (11.974)	55.787 (41.773)	42.246*** (12.673)	29.902** (12.785)	55.481* (32.519)	22.110 (13.485)
Urban population rate	-0.078 (0.136)	-0.284 (0.342)	4.301 (8.666)	1.112* (0.591)	-0.302 (0.342)	3.645 (4.108)	1.076* (0.630)
Higher education rate	0.051 (0.209)	-0.038 (0.181)	-0.140 (0.358)	-0.069 (0.189)	-0.042 (0.181)	-0.117 (0.268)	-0.067 (0.184)
Number of doctors per population	-52.246** (25.948)	-2.177 (41.324)	89.583 (186.203)	25.765 (44.208)	7.321 (41.823)	72.604 (96.954)	56.137 (45.954)
Birth rate	-0.002 (0.003)	0.016*** (0.003)	0.012 (0.009)	0.015*** (0.003)	0.017*** (0.003)	0.012** (0.006)	0.017*** (0.003)
Never married rate	-0.367* (0.222)	0.169 (0.210)	0.152 (0.351)	0.164 (0.220)	0.170 (0.211)	0.168 (0.305)	0.197 (0.215)
Divorced population rate	-2.599** (1.021)	-1.193 (1.721)	-7.045 (11.401)	-2.975 (1.897)	-1.771 (1.767)	-6.168 (6.004)	-5.252** (2.213)
Real gross regional product per capita	1.322 (0.911)	4.048*** (0.979)	12.117 (15.305)	6.505*** (1.318)	4.341*** (1.007)	10.979 (7.525)	7.780*** (1.650)
Air pollution per territory	12.845 (8.209)	47.057*** (9.167)	33.841 (29.220)	43.033*** (9.677)	48.590*** (9.240)	35.918** (17.194)	49.478*** (9.427)
Visits in health institutions per population	-195.760*** (62.330)	-114.965 (117.025)	-232.704 (295.292)	-150.818 (122.905)	-104.919 (117.557)	-225.140 (215.140)	-131.788 (120.510)
Observations	310	310	310	310	310	310	310
R-squared	0.629	0.470		0.371	0.475	0.817	0.908
Time fixed effects	-	+	+	+	+	+	+
Number of region		31	31	31	31	31	31

Note: Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: Authors' calculations.

Model of panel data with fixed effects using the method of instrumental variable including interactions between the variable Real Income and the categorical variable—Level of social development of regions (3 sectors) to estimate the difference in the effect of income on mortality for groups of regions with different levels of social development. Interactions between Unemployment rate (*unempl_*) and the categorical variable—Level of social development of regions (6) and Average unemployment rate in neighboring regions (*unempl_i7_*) and the categorical variable—Level of social development of regions (7) proposed in this work will be considered as instrumental variables.

Base Regression results model number in the Table 4 header matches the above model numbering, adjusted for the use of additional control variables or deleted outliers.

4.1. Discussion of potential threats when using the unemployment variable as an instrumental variable

The use of the unemployment variable as an instrumental variable is a common practice in assessing the impact of income on health outcomes. For example, Ettner (1996) uses the local unemployment variable as an instrumental variable to estimate the impact of income on a large number of different health outcomes. A similar strategy identification is used in their works in assessing the impact of income on various health outcomes by Xu (2013), Kuehnle (2014), Wei and Feeny (2019), and Chen (2019).

All the above authors consider various arguments for and against the possibility of using the unemployment variable as an instrumental variable. Most often the discussion arises because of the exogeneity property of this variable. There is evidence in the literature that this instrument has its limitations, at least when considering the mortality of certain categories of the population. At the same time, in some cases, for example, when considering child health outcomes, the use of this instrument is practically uncriticized. In this study, we test the feasibility of this instrument in terms of its satisfying the two main properties that an instrumental variable must meet. The relevance property is met in this case, e.g., the corresponding first-stage *F*-statistics for this variable is greater than 40, which is less than the second proposed instrument, but greater than 10 (Supplementary material Table A2: Regression results for *F*-test). Relevance can be substantiated, among other things, on the basis of economic theory, according to which an increase in unemployment, other things being equal, leads to an increase in labor supply in the factor market and, thus, to a decrease in labor costs. A potential channel for breaking the exogeneity property can be the intervention of the relevant provincial authorities, for example, more effective provincial governance can, on the one hand, affect the unemployment rate in the region, reducing it, and, on the other hand, lead to a decrease in the mortality rate due to more efficient operation of the health care system. The fact that in China the private sector accounts for a substantial share of enterprises and about 80% of the jobs in cities and towns can be an argument in favor of defending the exogeneity of this instrument. In a sense, this allows a greater degree of faith in the greater degree of autonomy of labor markets. Subsection 4.2 will consider an approach to the design of another instrument, which, to some extent due

to the specifics of labor markets in the PRC, can partially solve the problems associated with variable unemployment as an instrument.

4.2. Construction of a new instrumental variable

The paper proposes an approach to the construction of another instrument—Average unemployment rate in neighboring regions (*unempl_i7*; regions with a common border (group A) and regions with a common border with regions from group A). The specifics of the Chinese labor market is that they are quite competitive and workers can change the region where they live, including in search of better working conditions. In this regard, the level of unemployment in neighboring regions influences employers' decisions about the level of wages sometimes no less than the conditions in the labor market within the region. If there is high unemployment within a region, this does not mean that local businesses can significantly affect real wages, for example, because workers may decide to leave for neighboring regions for more attractive working conditions. At the same time, if the level of real wages in neighboring regions decreases on average, this to a certain extent creates for employers of the region under consideration preconditions for a decrease in the level of wages. At the same time, this may be caused by the fact that workers from neighboring regions may increase the supply of labor in the region under consideration. On the other hand, workers in the region under consideration are less likely to choose to move to neighboring regions with high unemployment rates in response to lower wages in their region.

This reasoning suggests the relevance of the instrument in question. The test results are likewise consistent with this, e.g., the corresponding first-stage *F*-statistics for this variable is greater than 81, which is even greater than that of the instrument adopted in the literature in the form of unemployment in the province itself (Supplementary material Table A2). Potential threats to the exogeneity of this instrument may arise through two main channels. For example, the average unemployment rate in neighboring regions might, through the income variable of neighbors, or might even be hypothesized to, directly affect the average mortality rate in neighboring regions. The mortality rate in neighboring regions might in turn hypothetically affect the mortality rate in the region in question. In order to test the hypothesis about the existence of spatial relationship between mortality rates in different regions in Subsection 6.3 in robustness check Spatial autoregressive model (SAR) will be estimated.

The second potential channel of existence of connection between average unemployment rate in neighboring regions and mortality rate in the region under consideration is an influence of this variable on unemployment in the region under consideration and already through it on the mortality rate in the region under consideration. On the one hand, this will be true only if the exogeneity property for the unemployment variable itself (4.1) is violated, which is still a debatable topic. On the other hand, even assuming such an impact, the degree of confidence in the exogeneity of the newly introduced variable would still be higher, since it is the unemployment in neighboring regions that is at stake, the so-called potentially more mediated threat of violation of the exogeneity property. In Subsection 6.1 under robustness check 1, the variable Average Death Rate in neighboring regions (*death_rate_i7*) (regions with a common

border (group A) and regions with a common border with regions from group A) and unemployment in the region itself will be considered as potential control variables. Adding these controls and estimating changes in the corresponding regression coefficients may provide additional information for the discussion of potential threats to the exogeneity of the instrument in question. It may be noted that the exogeneity of the proposed instrument requires further investigation, taking into account the literature and our understanding of the interrelationships among the regions in question. However, at least in the short and medium term, we see it as promising to continue working with this instrumental variable.

Let us consider the algorithm of calculation of variables Average unemployment rate in neighboring regions (*unempl_i7*) and Average Death Rate in neighboring regions (*death_rate_i7*). In order to form these variables, it is necessary to determine which regions can be considered neighbors for the region in question. For example, the first option: for the region in question, neighboring regions can be understood as regions with a direct border with the region in question. The second option: for the region in question, neighboring regions can be understood as all other regions, but their influence will be less the greater the distance is between them (for this purpose, a matrix of squares of inverse distances between regions is calculated). The third option, neighboring regions can be understood as regions that directly share a border (group A) and regions that share a border with group A. For the construction of the instrument at the current stage, the third option was chosen, but in the future all options will be calculated to assess the stability of the results. The advantage of the third option is that, on the one hand, it does not take into account distant regions as neighbors, which allows to avoid loss of variation, on the other hand, a sufficient number of regions is considered as neighbors to avoid random deviations of variables. Thus, in order to calculate, for instance, Average unemployment rate in neighboring regions (*unempl_i7*) variable, it is necessary to identify neighbors for each region according to the above rule and find the average value. For the variable Average Death Rate in neighboring regions (*death_rate_i7*), the same is true. Examples of instrumental variable construction according to this technology can be found in studies assessing the impact of socio-economic indicators on mortality rate (Nagapetyan et al., 2023, Nagapetyan et al., 2024).

4.3. *The division of PRC provinces into groups depending on the level of socio-economic development*

There are generally accepted approaches in the literature to divide PRC provinces into groups according to the level of development on the basis of geography, history, and administrative division. Researchers single out coastal areas as the most developed, and the western mainland areas as the least developed. Likewise, in a separate group are allocated medium mainland regions, characterized by the average level of development (Sahabudhee et al., 2023; Xiang et al., 2020). The provinces of China were divided into three groups depending on the level of socio-economic development (Categorical variable, 1—low level of social development, 2—average level of social development, 3—high of social development).

The social development of a region is understood as the quality of social infrastructure available to its inhabitants (quality of health care, education (qualifica-

tions and competencies of persons providing such services), etc.). In the literature it is customary to carry out this division into more socially developed and less socially developed, based on geography, because:

1.1. It was in the coastal territories that the first special economic zones began to open (since 1980), and at the moment where the main high-tech companies and highly qualified workers, including foreign specialists, are concentrated. The more developed level of social infrastructure in these territories is determined primarily because numerous enterprises are interested in this, in order to attract and retain highly qualified personnel, both from all over the world and from the PRC.

1.2. Administrative division coincides with geographical division, which correspondingly affects the specifics of the work of state and public institutions, and, what is even more important, the possibility of easier access of residents of some regions to the infrastructure of neighboring regions. *Separation based on geography is important because even if a person lives in a coastal area with a relatively (most developed coastal areas) low level of social infrastructure development, he is more likely to benefit from the infrastructure of a neighboring region, within the same administrative division compared to residents of geographically distant regions of the PRC mainland.*

1.3. It is the location of the regions that leads to the fact that there are innovative companies (foreigners, investors, innovation, human resources, etc.) and hence the high level of social infrastructure and high GRP per capita.

1.4. Given the peculiarities of the PRC economy, a high level of GRP per capita is not always accompanied by a high level of income, since in many regions GDP differs significantly from GNP. That is, a region may be more “developed” in the sense of high quality social infrastructure available to its inhabitants for geographical reasons, but the level of income of the population may be low.

5. Results

In models without the instrumental variable method, the coefficient characterizing the effect of income on mortality is negative and statistically significant. For example, according to model (2) (Fixed effects model) an increase in income by 1% leads to a decrease in the mortality rate by 0.84% (variables in logarithms). The application of the instrumental variable method leads to a significant increase in the value of the corresponding coefficient, and a statistically significant result is obtained by using as an instrumental variable the factor proposed in the paper—Average unemployment rate in neighboring regions (*unempl_i7_*). The use of the instrumental variable of unemployment in the province itself, although leading to an increase in the value of the coefficient, is not statistically significant. This may be, among other things, a consequence of the lack of variation due to the inclusion of a large number of control variables. In some modifications that included fewer control variables, the coefficient was significant (not presented in the study). If we trust the results of the model (4), we can state the underestimation of the value in the basic model almost twice. That is, a more plausible estimate seems to be a 2% decrease in the mortality rate (deaths per 1,000 person) with a 1% increase in income. One of the reasons for the underestimation of the coefficient we see in the problem of missing variables, in particular accurate data on the concentration in the territory of industries characterized by a high

degree of environmental pollution. This factor can both positively influence the value of the income of the population, and lead to an increase in the mortality rate in the territory due to the negative impact of the level of pollution on their health. Hence, higher values of incomes can be accompanied by higher values of mortality, which ultimately leads to the underestimation of the true value of the coefficient. There are other channels as well. For example, cultural and historical features, because of which on the one hand people may have a tendency to work hard without taking care of their health, which can in addition allow an increase in income and lead to an increase in mortality. Conversely, a more sedentary lifestyle can lead to lower incomes, but furthermore to the absence of the negative effects of labor in the form of body deterioration.

According to the results obtained in the model (7), income has a greater negative effect on mortality in more socially developed regions. The results allow us to confirm the hypothesis that the effect of income on mortality in more socially developed provinces of the PRC has a greater negative value compared to less socially developed areas. At the same time, the detected difference is statistically significant at the 1% significance level and is not less than 10%. The point is that in the group of provinces with a high level of social development (coastal areas) a 1% increase in income reduces mortality by 0.22% more compared to a similar reduction in mortality in the group of regions with a low level of social development. Moreover, the group of provinces with an average level of social development likewise shows a statistically significant greater negative effect of income on mortality of 0.07% at the 10% level of significance (p -value = 0.103)

This is consistent with the previously cited evidence from the literature, according to which the negative impact of income on mortality is most often recorded in more socially developed areas. Continuing to understand under the level of social development the quality of social infrastructure available to its inhabitants (quality of health care, education (qualification and competence of persons providing such services), etc.) we claim that these characteristics directly affect the possibility and need for the population to use income to improve their chances for survival. The internal mechanism explaining these results may be that it is not enough to have a high income, one must be able to spend it in such a way as to be able to influence one's probability of survival. In other words, residents of PRC provinces with more developed social infrastructure, given the advantages and opportunities that they can take advantage of, may be more likely to influence the reduction of mortality, especially if they have a high income, than residents of territories with less developed social infrastructure.

6. Robustness checks

6.1. Additional controls

Earlier in Subsection 4.2 we identified two potential channels of influence of the instrumental variable proposed in this paper (Average unemployment rate in neighboring regions) on the mortality rates in the regions under consideration, which could cast doubt on the exogeneity property of this instrument. As will be discussed further in Subsection 6.3, we are talking about the following channel: Average unemployment rate in neighboring regions \rightarrow average mortality rate in

neighboring regions → mortality rate in the region under consideration (further, in Subsection 6.3, this hypothesis about the existence of a relationship between mortality rates in neighboring regions will be tested with the help of spatial autoregressive model).

The second channel is: Average unemployment rate in neighboring regions → the unemployment rate in the region in question → the mortality rate in the region in question. Even if we assume that unemployment in the region (under consideration) is not exogenous and affects the mortality rate in the region (under consideration) through channels other than income; the impact through this mechanism of the average unemployment rate in neighboring regions will be indirect. In other words, the exogeneity of the proposed variable in the form of the average unemployment rate in neighboring regions (*unempl_i7_*) in this part will be higher than the unemployment rate in the region itself (*unempl_*). Within the framework of robustness check 1, it is proposed to consider Average Death Rate in neighboring regions (*death_rate_i7_*; regions with a common border (group A) and regions with a common border with regions from group A) and unemployment in the region itself as potential control variables. Of course, this approach cannot solve the problem of exogeneity, but based on the response of the corresponding regression coefficients we can make additional conclusions about the extent to which we can trust this instrument.

The results in Supplementary material Table A3 show that the obtained coefficients, in particular in model (7) do not statistically differ from the results obtained in the basic version of the regression. Although this does not completely eliminate doubts about the possibility of using this instrumental variable, to some extent due to the statistical stability of the results it lends it a certain confidence.

6.2. Removing outliers

In Section 3, a preliminary analysis of the data revealed and described outliers (e.g., income). We are talking about such provinces as Beijing, Shanghai. This result is quite intuitive, because these regions are among the most developed in China. In order to check the robustness of the obtained results, we will consider all the main regressions calculated earlier (with additional controls) under the condition of removing the outliers. According to the data in Supplementary material Table A4, practically all previously obtained key results are confirmed. Moreover, in the group of provinces with a high level of social development (coastal territories) with an increase in income by 1% mortality decreases by 0.16% more compared to a similar decrease in mortality in the group of regions with a low level of social development. The result is still virtually significant at the 10% significance level of p -value = 0.107, although this would be expected given that when the data on the most developed areas are removed from the group of socially developed areas, this group becomes more similar to the low level of social development group.

6.3. Testing interspatial relationships for the mortality rate variable

As was demonstrated in Subsection 4.2, one of the potential risks to the exogeneity of the instrumental variable used in this paper (Average unemployment rate in neighboring regions; *unempl_i7_*) in estimating the effect of income on

mortality is the hypothetical possibility that the average unemployment rate in neighboring regions might affect the average mortality rate in neighboring regions, and if it is true that average mortality rates in neighboring regions might affect mortality rates in the region under consideration, then there is a risk of a certain degree of violation of the exogeneity property to some extent.

In order to test the relationship between mortality rates in neighboring regions, it is proposed to estimate the following modification of the spatial autoregressive model:

$$\begin{aligned} \log(\text{death_rate}_{it}) = & \alpha_i + \rho \times W \times \log(\text{death_rate}_{it}) + \beta_1 \times \log(\text{income_real}_{it}) + \\ & + \beta_2 \times \text{sex}_{it} + \beta_3 \times \text{old}_{it} + \beta_4 \times \text{beds_p}_{it} + \beta_5 \times \text{city_p}_{it} + \beta_6 \times \text{educ_high_p}_{it} + \\ & + \beta_7 \times \text{doctors_lic_p}_{it} + \beta_8 \times \text{birth_rate}_{it} + \beta_9 \times \text{marriage_no_p}_{it} + \\ & + \beta_{10} \times \text{divorce_p}_{it} + \beta_{11} \times \text{gdp_p_r}_{it} + \beta_{12} \times \text{air_pollut}_{it} + \beta_{13} \times \text{visits_p}_{it} + \\ & + \beta_{14} \times \text{factor}(\text{year}_t) + \varepsilon_{it}, \end{aligned} \quad (8)$$

where ρ —coefficient characterizing the spatial relationship between the mortality rates of neighboring territories: W —spatial matrix (matrix of squares of inverse distances between provinces).

More details about spatial matrix were discussed in Subsection 4.2. Results of model estimation (8) do not allow for rejecting the null hypothesis about absence of interspatial connections between mortality rates in neighboring regions. The coefficient of Spatial rho in absolute size is very close to zero and statistically insignificant. To a certain extent, these results increase the confidence in the proposed instrument (Average unemployment rate in neighboring regions (*unempl_i7_*); see Supplementary material Table A5).

6.4. Clarification of causal relationships based on panel data structure

The potential limitations of the instrumental variables used in the study, as demonstrated in Section 4, lead to the need to propose an alternative strategy for confirming causality, in particular to address the problem of reverse causality. A detailed description of current approaches to explain the causality based on panel data structure is given in the study by Leszczensky and Wolbring (2022). For this purpose, we consider an approach involving the use of the generalized method of moments to correctly estimate model (9) (Arellano and Bond, 1991; Hansen, 1982).

$$\begin{aligned} \log(\text{death_rate}_{it}) = & \alpha_i + \beta_1 \times \log(\text{income_real}_{it}) + \beta_2 \times \text{sex}_{it} + \beta_3 \times \text{old}_{it} + \\ & + \beta_4 \times \text{beds_p}_{it} + \beta_5 \times \text{city_p}_{it} + \beta_6 \times \text{educ_high_p}_{it} + \beta_7 \times \text{doctors_lic_p}_{it} + \\ & + \beta_8 \times \text{birth_rate}_{it} + \beta_9 \times \text{marriage_no_p}_{it} + \beta_{10} \times \text{divorce_p}_{it} + \\ & + \beta_{11} \times \text{gdp_p_r}_{it} + \beta_{12} \times \text{air_pollut}_{it} + \beta_{13} \times \text{visits_p}_{it} + \beta_{14} \times \text{factor}(\text{year}_t) + \\ & + \beta_{15} \times \log(\text{income_real}_{it})_{\text{sector3}(=2)} + \\ & + \beta_{16} \times \log(\text{income_real}_{it})_{\text{sector3}(=3)} + \beta_{17} \times \log(\text{death_rate}_{it-1}) + \\ & + \beta_{17} \times \log(\text{death_rate}_{it-2}) + \beta_{18} \times \log(\text{death_rate}_{it-3}) + \varepsilon_{it}, \end{aligned} \quad (9)$$

Columns (1) and (2) in Supplementary material Table A6 repeat the results of columns (2) and (5) from Table 4 for ease of comparison with the new results. Columns (3) and (4) in Table 8 discuss the results obtained by applying the generalized method of moments to estimate the coefficients using only GMM instruments (all allowed lags of *death_rate_ln* and *income_real_ln* variables) for model (9), respectively without and with partitioning into territorial sectors. Thus, in this case we do not use instrumental variables such as unemployment (*unempl_*) and average unemployment rate in neighboring regions (*unempl_i7_*). The results obtained allow us to confirm the key conclusions. According to the results in column 3 of Table 8, income does have a negative effect on mortality, and as with the instrumental variable method, the use of GMM also indicates that income has a greater negative effect on mortality in more socially developed regions. The results also allow us to confirm the hypothesis that the effect of income on mortality in more socially developed provinces of the PRC has a greater negative value compared to less socially developed areas. The independent validation of the IV results through the use of GMM allows us to increase our confidence in the evidence obtained in this study.

7. Conclusions

The effect of income on mortality in more socially developed provinces of PRC has a greater negative value compared to less socially developed areas, adjusted for the correctness of the selection of the relevant groups of provinces on the basis of the approaches generally accepted in the literature based on the geographical, historical, administrative and other features of these territories. At the same time, the difference found is statistically significant at 1% level of significance and is not less than 10%. The point is that in the group of provinces with a high level of social development (coastal territories) with a 1% increase in income the mortality rate decreases by 0.22% more (or 0.44% if relying on GMM results) compared with a similar decrease in mortality in the group of regions with a low level of social development. At the same time, a similar, though smaller result (−0.07%) is observed for the groups of provinces with the average level of social development.

Another research result is the proposal of a new approach to the construction of an instrumental variable in the form of Average unemployment rate in neighboring regions. The relevance of this instrument was demonstrated on the basis of economic theory and the corresponding test (first-stage). At the same time, the value of the *F*-statistic is almost twice as high as a similar indicator of another instrument accepted in the literature—the unemployment rate of the region in question. Despite identification of potential channels of violation of exogeneity property in the study based on robustness checks, including direct testing of the hypothesis about existence of spatial correlations between mortality rates of neighboring regions, as well as other approaches, potential areas of criticism of the proposed instrument have been shown to be debatable. Among other things, the use of this instrumental variable has demonstrated that in the territory in question a 1% increase in personal income leads to a reduction in the mortality rate by more than 2%, while the estimated impact is less than 1% when the instrumental variable is not used, which leads to a risk of underestimating it by more than

a factor of two. The application of GMM permits the enhancement of confidence in the evidence derived from the IV results.

Our results suggest that in order to realize a negative association of income level on mortality rates it is not enough to have a high income; one must still be able to spend it in such a way as to be able to influence one's probability of survival. The latter is determined, among other things, by the level of social development, quality of social infrastructure available to its inhabitants (quality of health care, education (qualification and competence of persons providing such services), etc.). This is one of the most important channels to explain why income has a greater negative impact on mortality in more socially developed regions. It is, however, important to note that there may be other mechanisms to explain the findings. In particular, the disparate impact of income on mortality across geographically and socially distinct regions can be attributed to the varying structures of mortality causes and the differing stages of epidemiological transition in these regions. A more detailed specification of the mechanisms that may explain the findings may be a topic for further research.

The results obtained and their further development can have a significant impact on the implementation of effective political governance measures aimed at reducing mortality in certain territories and mortality in general. On the one hand, more accurate estimates of the impact of income on mortality allow for a fairer assessment and prediction of the impact of certain programs and measures to combat mortality. On the other hand, the results generated shed light on the reasons for the variation in the literature in the results characterizing the impact of income on mortality in different territories. Contributing to a partial explanation of this variation provides guidance for policymakers in combating mortality inequality to shape more targeted and informed management decisions.

References

- Adda, J., Banks, J., & von Gaudecker, H.-M. (2009). The impact of income shocks on health: Evidence from cohort data. *Journal of the European Economic Association*, 7(6), 1361–1399. <https://doi.org/10.1162/JEEA.2009.7.6.1361>
- Ahammer, A., Horvath, T., & Winter-Ebmer, R. (2016). The effect of income on mortality: New evidence for the absence of a causal link. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(3), 793–816. <https://doi.org/10.1111/rssa.12240>
- Arceo, E., Hanna, R., & Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *Economic Journal*, 126(591), 257–280. <https://doi.org/10.1111/ecoj.12273>
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277–297. <https://doi.org/10.2307/2297968>
- Backlund, E., Sorlie, P. D., & Johnson, N. J. (1996). The shape of the relationship between income and mortality in the United States. *Annals of Epidemiology*, 6(1), 12–20. [https://doi.org/10.1016/1047-2797\(95\)00090-9](https://doi.org/10.1016/1047-2797(95)00090-9)
- Backlund, E., Sorlie, P. D., & Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: The national longitudinal mortality study. *Social Science & Medicine*, 49(10), 1373–1384. [https://doi.org/10.1016/S0277-9536\(99\)00209-9](https://doi.org/10.1016/S0277-9536(99)00209-9)
- Blakely, T., Kawachi, I., Atkinson, J., & Fawcett, J. (2004). Income and mortality: The shape of the association and confounding New Zealand Census–Mortality Study, 1981–1999. *International Journal of Epidemiology*, 33(4), 874–883. <https://doi.org/10.1093/ije/dyh156>

- Chen, W. (2019). Income and effort: An instrumental variables approach. *Atlantic Economic Journal*, 47, 485–497. <https://doi.org/10.1007/s11293-019-09644-3>
- Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6), 2087–2120. <https://doi.org/10.1257/aer.103.6.2087>
- Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, 15(1), 67–85. [https://doi.org/10.1016/0167-6296\(95\)00032-1](https://doi.org/10.1016/0167-6296(95)00032-1)
- Fichera, E., & Savage, D. (2015). Income and health in Tanzania. An instrumental variable approach. *World Development*, 66, 500–515. <https://doi.org/10.1016/j.worlddev.2014.09.016>
- Flegal, K., Kit, B., Orpana, H., & Graubard, B. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis. *JAMA*, 309(1), 71–82. <https://doi.org/10.1001/jama.2012.113905>
- GlobalSurg Collaborative (2016). Mortality of emergency abdominal surgery in high-, middle- and low-income countries. *British Journal of Surgery*, 103(8), 971–988. <https://doi.org/10.1002/bjs.10151>
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054. <https://doi.org/10.2307/1912775>
- Koch, E., Romero, T., Romero, C. X., Aguilera, H., Paredes, M., Vargas, M., & Ahumada, C. (2010a). Early life and adult socioeconomic influences on mortality risk: Preliminary report of a “pauper rich” paradox in a Chilean adult cohort. *Annals of Epidemiology*, 20(6), 487–492. <https://doi.org/10.1016/j.annepidem.2010.03.009>
- Koch, E., Romero, T., Romero, C. X., Akel, C., Manríquez, L., Paredes, M., Román, C., Taylor, A., Vargas, M., & Kirschbaum, A. (2010b). Impact of education, income and chronic disease risk factors on mortality of adults: Does “a pauper-rich paradox” exist in Latin American societies?. *Public Health*, 124(1), 39–48. <https://doi.org/10.1016/j.puhe.2009.11.008>
- Kondo, N., Sembajwe, G., Kawachi, I., van Dam, R. M., Subramanian, S. V., & Yamagata, Z. (2009). Income inequality, mortality, and self rated health: Meta-analysis of multilevel studies. *BMJ*, 339, b4471. <https://doi.org/10.1136/bmj.b4471>
- Kuehnl, D. (2014). The causal effect of family income on child health in the U.K. *Journal of Health Economics*, 36, 137–150. <https://doi.org/10.1016/j.jhealeco.2014.03.011>
- Leszczensky, L., & Wolbring, T. (2022). How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study. *Sociological Methods & Research*, 51 (2), 837–865. <https://doi.org/10.1177/0049124119882473>
- Lincoff, A. M., Brown-Frandsen, K., Colhoun, H. M., Deanfield, J., Emerson, S. S., Esbjerg, S., Hardt-Lindberg, S., Hovingh, G. K., Kahn, S. E., Kushner, R. F., Lingvay, I., Oral, T. K., Michelsen, M. M., Plutzky, J., Tornøe, C. W., & Ryan, D. H. (2023). Semaglutide and cardiovascular outcomes in obesity without diabetes. *New England Journal of Medicine*, 389(24), 2221–2232. <https://doi.org/10.1056/NEJMoa2307563>
- Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, 40(1), 144–168. <https://doi.org/10.3368/jhr.XL.1.144>
- Lindsay, A. C., Greaney, M. L., Wallington, S. F., Mesa, T., & Salas, C. F. (2017). A review of early influences on physical activity and sedentary behaviors of preschool-age children in high-income countries. *Journal for Specialists in Pediatric Nursing*, 22(3), e12182. <https://doi.org/10.1111/jspn.12182>
- Martikainen, P., Mäkelä, P., Koskinen, S., & Valkonen, T. (2001). Income differences in mortality: A register-based follow-up study of three million men and women. *International Journal of Epidemiology*, 30(6), 1397–405. <https://doi.org/10.1093/ije/30.6.1397>
- Nagapetyan, A., Drozd, A., & Subbotovsky, D. (2023). How to determine the optimal number of cardiologists in a region? *Mathematics*, 11(21), 4422. <https://doi.org/10.3390/math11214422>
- Nagapetyan, A., Subbotovsky, D., Pavlova, T., & Jun, Li (2024). How does the level of morbidity affect the indicator of real income of the population in the regions of the Russian Federation? *Terra Economicus*, 22(2), 77–95 (in Russian). <https://doi.org/10.18522/2073-6606-2024-22-2-77-95>
- O’Hare, B., Makuta, I., Chiwaula, L., & Bar-Zeev, N. (2013). Income and child mortality in developing countries: A systematic review and meta-analysis. *Journal of the Royal Society of Medicine*, 106(10), 408–414. <https://doi.org/10.1177/0141076813489680>

- Ren, Y., Castro Campos, B., Loy, J.-P., & Brosig, S. (2019) Low-income and overweight in China: Evidence from a life-course utility model. *Journal of Integrative Agriculture*, 18(8), 1753–1767. [https://doi.org/10.1016/S2095-3119\(19\)62691-2](https://doi.org/10.1016/S2095-3119(19)62691-2)
- Sahabudhee, A., Rao, C. R., Chandrasekaran, B., & Pedersen, S. J. (2023). Dose-response effects of periodic physical activity breaks on the chronic inflammatory risk associated with sedentary behavior in high- and upper-middle income countries: A systematic review and meta-analysis. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 17(3), 102730. <https://doi.org/10.1016/j.dsx.2023.102730>
- Wei, L., & Feeny, D. (2019). The dynamics of the gradient between child's health and family income: Evidence from Canada. *Social Science & Medicine*, 226, 182–189. <https://doi.org/10.1016/j.socscimed.2019.02.033>
- Xiang, L., Stillwell, J., Burns, L., & Heppenstall, A. (2020). Measuring and assessing regional education inequalities in China under changing policy regimes. *Applied Spatial Analysis and Policy*, 13(1), 91–112. <https://doi.org/10.1007/s12061-019-09293-8>
- Xu, X. (2013). The business cycle and health behaviors. *Social Science & Medicine*, 77, 126–136. <https://doi.org/10.1016/j.socscimed.2012.11.016>
- Zhang, L., Wang, Z., Wang, X., Chen, Z., Shao, L., Tian, Y., Zheng, C., Li, S., Zhu, M., & Gao, R. (2020). Prevalence of overweight and obesity in China: Results from a cross-sectional study of 441 thousand adults, 2012–2015. *Obesity research & clinical practice*, 14(2), 119–126. <https://doi.org/10.1016/j.orcp.2020.02.005>
- Zhang, X., Zhao, L., Cui, Z., & Wang, Y. (2015). Study on equity and efficiency of health resources and services based on key indicators in China. *PLoS ONE*, 10(12), e0144809. <https://doi.org/10.1371/journal.pone.0144809>

Acknowledgements

The study was funded by the Ministry of Science and High Education of the Russian Federation, project No. FZNS-2023-0016 “Sustainable regional development: Efficient economic mechanisms for organising markets and entrepreneurial competencies of the population under uncertainty (balancing security and risk).”

Supplementary material

Descriptive statistics and additional calculations

Authors: Artur R. Nagapetyan, Tatiana I. Pavlova, Jun Li

Data type: Text

Explanation note: The supplementary material contains statistical data for the provinces of the People's Republic of China to reproduce the main findings of the study. It also provides information on additional regressions including robustness checks to increase the level of confidence in the results.

This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.32609/j.ruje.10.125107.suppl>