Science of the Total Environment 487 (2014) 57–64

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A study of air pollutants influencing life expectancy and longevity from spatial perspective in China

Li Wang [a,b], BingganWei [b], Yonghua Li [b,*], Hairong Li [b], Fengying Zhang [a,c], Mark Rosenberg [d], Linsheng Yang [b], Jixia Huang [b], Thomas Krafft [a,b,e], Wuyi Wang [b]

[a] *Department of International Health, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands*
[b] *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[c] *China National Environmental Monitoring Center, Beijing 100012, China*
[d] *Department of Geography, Queen's University, Kingston K7L 3N6, Canada*
[e] *Institute of Environmental Education and Research, Bharati Vidyapeeth University, Pune, India*

HIGHLIGHTS

- Using GWR to investigate the spatial correlations between health and air pollutants.
- Difference of 10 μg/m3 in SO_2 can cause adjusted 0.28 year in life expectancy.
- Difference of 10 μg/m3 in PM_{10} can lead to a difference of 2.23 in longevity ratio.
- Per capita GDP was positively associating with life expectancy in China.

ARTICLE INFO

Article history:
Received 24 January 2014
Received in revised form 28 March 2014
Accepted 31 March 2014
Available online 23 April 2014

Editor name: P. Kassomenos

Keywords:
Life expectancy
Longevity
Air pollutants
Per capita GDP
Geographically Weighted Regression
Stepwise Regression

ABSTRACT

Life expectancy and longevity are influenced by air pollutants and socioeconomic status, but the extend and significance are still unclear. Better understanding how the spatial differences of life expectancy and longevity are affected by air pollutants is needed for generating public health and environmental strategies since the whole of China is now threatened by deteriorated air quality. 85 major city regions were chosen as research areas. Geographically Weighted Regression (GWR) and Stepwise Regression (SR) were used to find the spatial correlations between health indicators and air pollutants, adjusted by per capita GDP[1]. The results were, regions with higher life expectancy were mainly located in the east area and areas with good air quality, a regional difference of 10 $\mu g/m^3$ in ambient air SO_2[2] could cause adjusted 0.28 year's difference in life expectancy, a regional difference of 10 $\mu g/m^3$ in ambient air PM_{10}[3] could lead to a longevity ratio difference of 2.23, and per capita GDP was positively associating with life expectancy but not longevity ratio, with a regional difference of 10,000 RMB[4] associating with adjusted 0.49 year's difference in life expectancy. This research also showed the evidences that there exist spatially differences for ambient air PM_{10} and SO_2 influencing life expectancy and longevity in China, and this influences were clearer in south China.

1. Background

Life expectancy is considered as one of the three major parameters for calculating the Human Development Index (HDI), which is used by the United Nations Development Program to rank human development levels of countries (UNDP, 2011). It has also been used to assess the health impact of air pollution (Chen et al., 2013; Pope et al., 2009b; Pope et al., 2013; Wang et al., 2013). Life expectancy is affected by multiple factors, and the social environment is considered as one of the most important factors (Blum, 1974). Ambient air SO_2, PM, and NO_x have been proved to lead to multiple diseases and diminished life expectancy (Chen et al., 2013; Cao et al., 2011; Wang et al., 2013; Pope et al., 2002, 2003, 2009a,b, 2013). The elderly population is found to be more vulnerable to air pollutants except ozone and more sensitive to air pollutants because of their depressed immune systems, existing diseases, and the accumulation of toxic agents in their bodies (Fischer et al., 2003; Sun and Gu, 2008). Most of the studies on air pollution affecting the elderly focused on the population of 65 or 75 years old and

---

* Corresponding author at: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11 A Datun Road, Beijing 100101, China. Tel.: +86 10 64889198; fax: +86 10 64856504.
E-mail address: yhli@igsnrr.ac.cn (Y. Li).

[1] GDP, gross domestic product.
[2] SO_2, sulfur dioxide.
[3] PM_{10}, particular matter with diameter <10 μm.
[4] RMB, renminbi.

http://dx.doi.org/10.1016/j.scitotenv.2014.03.142

over (Fischer et al., 2003; Wen and Gu, 2012). Due to the toxic accumulation effect, the 100 years old and over population is more sensitive to air pollutants. A clear understanding of the effects of air pollutants to the longevity group (people living to be 100 and over) is critically needed for the county with the increasing elderly population. The negative associations between air pollutants and health outcomes have been largely proved through time-series or cohort based studies (Zhang et al., 2011; Cao et al., 2011; Gouveia and Fletcher, 2000). But there was no study considering the regional differences of population's sensitivity to air pollutants. A research regarding spatial differences of health outcomes associating with air pollutants is urgently needed for formulating regional healthy coping strategies against air pollutants.

In this paper, we used life expectancy and longevity as public health outcomes. Socioeconomic represented by per capita GDP (gross domestic product), and air pollutants, including PM_{10} and SO_2, which are regarded as the major ambient air pollutants challenging China's public health, were analyzed to find out how they influence life expectancy and longevity. We hypothesized that spatial differences of PM_{10} and SO_2 were associated with changes in life expectancy and longevity. Regional differences in life expectancy and longevity could also be partly attributed to socioeconomic status. This research was conducted at the prefecture-city level.

The major objectives of this paper were: (1) to demonstrate the spatial distribution of life expectancy and longevity at the prefecture-city level; (2) to analyze the spatial relationship between air pollutants and the two health indicators; (3) and to analyze the socioeconomic effect on life expectancy and longevity. The research findings lead to policy recommendations for decision-makers to use in developing health strategies to help the elderly population cope with the deteriorated air quality and to find a balance between environmental protection and socioeconomic development from a public health perspective.

2. Method and data

2.1. Method and explanatory variables for models

Using ArcGIS software, we constructed a geographic and attribute database of the longevity group using an administrative map and the sixth national population census of China. Geographic distribution maps of the longevity group, life expectancy, PM_{10}, SO_2, per capita GDP were generated on prefecture-city level. In order to find out the possible effect of pollutants and socioeconomic factors on life expectancy and longevity, we collected data on PM_{10}, SO_2, and per capita GDP for our analysis. Life expectancy and longevity ratio were analyzed against air pollutants and socioeconomic indicator using the Stepwise Regression (SR) by SPSS 18.0 and Geographically Weighted Regression (GWR) by ArcGIS 10.1.

In SR, PM_{10}, SO_2 and per capita GDP were used as independent variables, while life expectancy and longevity ratio were dependent variables, respectively. After we conducted SR, we used the variables included by SR for GWR. GWR generates a separate regression equation for every feature analyzed in a sample dataset as a means to address spatial variation. A general version of the model can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{z=1}^{n} \beta_z(u_i, v_i) x_{iz} + \varepsilon_i \tag{1}$$

Where y_i denotes the dependent variable, in this case the life expectancy or longevity i at location i, $\beta_0(u_i,v_i)$ denotes the intercept coefficient at location i, x_{iz} is the value of the zth explanatory variable at location i and $\beta_z(u_i,v_i)$ is the location regression coefficient for the zth explanatory variable. Furthermore, (u_i,v_i) denotes Cartesian x and y point coordinates and ε_i is the random location specific error term. When GWR was used, the parameters can be estimated by solving:

$$\beta_f(u_i, v_i) = \left(X^T W(u_i, v_i) X\right)^{-1} X^T W(u_i, v_i) y \tag{2}$$

where $\beta_f(u_i,v_i)$ is the estimate of the location-specific parameter, $W(u_i, v_i)$ is an n by n spatial weight matrix whose off-diagonal elements are zero and the diagonal elements denote the geographical weights of observed data a location i. The geographic weight structure (u_i,v_i) is based on a Gaussian Kernel function such that the influence of data points in the proximity of i is given larger weights in the estimation.

This paper used an adaptive bi-square function to generate the geographic weights. An adaptive function fitted the demographic data analyzed in this paper since the research points clustered in some regions. The spatial context (the Gaussian kernel) is a function of a specified number of neighbors. Where feature distribution is dense, the spatial context is smaller; where feature distribution is sparse, the spatial context is larger (Charlton et al., 2009).

The bandwidth may be either defined by a given distance, or a fixed number of nearest neighbors from the analysis location. In this case we used AICs, that the optimal number of nearest neighbors was determined through selecting the model with the lowest Akaike Information Criterion (AIC) score (Hurvich et al., 1998), given as:

$$\mathrm{AICc} = 2n\ln(\sigma) + n\ln(2\pi) + n\left\{\frac{n + tr(s)}{n - 2 - tr(s)}\right\}. \tag{3}$$

Here tr(s) is the trace of the hat matrix. The AIC method can be used to select between a number of competing models by taking into account differences in model complexity (Fotheringham et al., 2002).

2.2. Data

Life expectancy data was calculated from demographic data, which were obtained from the demographic database of the six national population census of China (National Bureau of Statistics of China, 2010b). The formula for calculating life expectancy was illustrated in Table 1.

From left to right, where, x represents age. l(x) is “the survivorship function”: the number of persons alive at age x. For example of the original 100,000 people in the hypothetical cohort, l(14–19) = 98,989 (or 98.989%) live to age 14–19. These values are computed recursively from the d(x) values using the formula l(x + i) = l(x) − d(x), with l(0), the “radix” of the table, arbitrarily set to 100,000. For example, l(1–4) = l(0) − d(0) = 100,000 − 550 = 99,450. d(x) is the number of deaths in the interval (x,x + i) for persons alive at age x, computed as d(x) = q(x) ∗ I(x). For example, the I(10–14) = 99,127 persons alive at age 10–14, d(10–14) = 0.00139 ∗ 99,127 = 138 died prior to age 10–14. q(x) is probability of dying at age x. Also known as the (age-specific) risk of death. Generally these are derived using the formula q(x) = 1 − exp[−m(x)], under the assumption that the instantaneous mortality rate, or force of mortality, remains constant throughout the age interval from x to x + i(here i = 5). m(x): the mortality rate at age x. Generally these quantities are obtained between P(x) and D(x). By construction, m(x) = D(x) / P(x). D(x) and P(x) are death and live number of population at x year, which are the inputs in the life table. The data are obtained from national census. L(x) is total number of person–years alive by the cohort from age x to x + I (here i = 5). This is the sum of the years lived by the l(x) and I(x + i) persons who survive the interval. L(x) = i / 2 [I(x) + I(x + i)]. For I(0), this is the sum of the years lived by the L(1) persons who survive, and the d(x) persons who die during the interval, L(0) = I(1) + 0.5 ∗ d(0).T(x) is the total number of person–years lived by the cohort from age x until all members of the cohort have died. This is the sum of numbers in the L(x) column from age x to the last row in the table. e(x): the (remaining) life expectancy of persons alive at age x, computed as e(x) = T(x) / l(x).