View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Exeley Inc.

STATISTICS IN TRANSITION new series, December 2018
Vol. 19, No. 4, pp. 645–670, DOI 10.21307/stattrans-2018-034

BAYESIAN SPATIAL ANALYSIS OF CHRONIC DISEASES IN ELDERLY CHINESE PEOPLE USING A STAR MODEL

Ping Gao[1], **Hikaru Hasegawa**[2]

ABSTRACT

Chronic diseases have become important factors affecting the health of elderly Chinese people. Because the prevalence of chronic diseases varies among the provinces, it is necessary to understand the spatial effects on these diseases, as well as their relationships with potential risk factors. This study applies a structured additive regression model and the `R2BayesX` package to conduct a Bayesian analysis. The data are taken from the 2000, 2006, and 2010 Chinese Urban and Rural Elderly Population Surveys. The findings are as follows: (1) the following covariates have considerable effects on chronic diseases in general, and on specific chronic diseases (hypertension and heart disease) (in descending order): census register (rural or urban), gender, smoking, drinking, province, time, age, cultural activities, years of education, and sports activities; (2) the effect of marital status is negligible; (3) province is a critical factor, with the highest spatial effect appearing in two types of provinces: economically developed provinces, and economically backward provinces; and (4) time also has considerable effects. Based on these findings, the government should further strengthen its investment in rural areas and economically backward provinces as a cost-effective intervention, and should educate the population on the harmful effects of smoking and drinking alcohol on health.

Key words: Bayesian analysis, Markov chain Monte Carlo (MCMC), `R2BayesX`, Spatial effect, Structured additive regression (STAR) models.

1. Introduction

The medical definition of a chronic disease is a disease that persists for a long time. For example, the U.S. National Center for Health Statistics defines a chronic disease as one that lasts for three months or more. In China, more than 70% of elderly people struggle with a chronic disease. In addition, the majority of the elderly suffer from multiple chronic diseases (Thorpe and Howard, 2006; Vogeli et al., 2007; Wolff et al., 2002). The prevalence of chronic diseases is related to many potential risk factors, such as age, lifestyle habits (smoking, drinking), a lack of exercise, and so on. In addition to these influencing factors, the spacial dimension is increasingly being considered as an independent factor, which can be examined using geostatistical methods.

[1]Graduate School of Economics and Business, Hokkaido University. E-mail: gaoping4069111@gmail.com

[2]Department of Economics, Hokkaido University. E-mail: hasegawa@econ.hokudai.ac.jp

Kriging is a common geostatistical method that produces a map of a quantity of interest over a geographical region. In addition, as an extension to kriging, universal kriging takes into account the linearity of the covariate (Cressie, 2015, pp.151–172). However, when the effect of the covariate is nonlinear, the universal kriging method is not appropriate. In this case, a geoadditive model may be a better choice. The geoadditive model was introduced by Kammann and Wand (2003), and accounts for non-linear covariate effects under the assumption of additivity. Today, many researchers apply this model in their studies (Basile et al., 2013; Geniaux and Napoléone, 2008; Sauleau et al., 2007; Wand et al., 2011).

Another powerful model used in spatial analyses is the generalized additive model (GAM). The GAM is suitable for modelling nonlinear effects of continuous covariates in regression models with non-Gaussian responses. Furthermore, structured additive regression (STAR) models extend GAM models by including spatial effects, the nonlinear effects of continuous factors, and linear or fixed effects in one model (Kneib, 2006). STAR models include generalized linear models and generalized additive models as special cases, but also allow for a wider class of effects, such as geographical or spatio-temporal effects (Fahrmeir et al., 2004, 2013; Umlauf et al., 2015).

With the rising prevalence of chronic diseases (Freedman and Martin, 2000; Thorpe and Howard, 2006) and the large elderly population (the population aged 60 and over in China was $2,308,600$ in 2016), the number of elderly Chinese people suffering from chronic diseases is very high. Because chronic diseases are essentially permanent, they introduce a heavy economic burden to families and society. In China, the most common cause of death is chronic diseases, rather than infectious diseases (Gu et al., 2009; He et al., 2005). For example, chronic disease-induced deaths accounted for 71.88% of all deaths among residents in Kunming (Yunnan provincial capital) in the period 2007–2010 (Li et al., 2012). The high number of elderly people with severe chronic conditions places a significant burden on medical care. Thus, the Chinese government faces enormous challenges in terms of medical investment. Numerous studies have examined chronic diseases, with many focusing on the factors influencing such diseases.

Fillingim et al. (2009) studied samples from different countries (China, France, Sweden, United States, etc.), and found that the prevalence of the most common forms of pain caused by chronic diseases is higher in women than in men. Furthermore, Zhen (2010) used data on Gansu province (in China), finding that the resistance of females to chronic disease pain is poor and that females' thresholds for discomfort are lower than those of men. As a result, women are more likely to visit a doctor and, thus, are more likely to be diagnosed with a chronic disease. Thus, we hypothesise that gender has a great influence on the reported prevalence of chronic diseases and of specific chronic diseases (e.g., hypertension and heart disease) in elderly Chinese people, and that elderly females are more likely to have chronic diseases and specific chronic diseases than are elderly males.

In this study, we use reported prevalence rather than prevalence, because the prevalence is not the true prevalence. Chronic diseases are usually non-fatal diseases and persist for a long time. Many elderly people may have chronic diseases, but may not be aware of this, in which case, they will report not having a disease, even though they do. Thus, the study can only obtain the reported prevalence.

Woolf et al. (2015) noted that people with a high economic status are more conscious of self-care, and that such individuals are more likely to be diagnosed with chronic diseases. Zhen (2010) found the reported prevalence of chronic diseases is significantly affected by access to health care. Income and medical security are greater among the urban elderly than among the rural elderly. Therefore, we hypothesise that the census register is a critical factor related to chronic diseases and to specific chronic diseases (e.g., hypertension and heart disease), and that the reported prevalence of chronic diseases and specific chronic diseases is higher among the urban elderly than among the rural elderly.

Chen (2005), Ye (2013), and Zhao et al. (2015) found that marital status also has an affect on chronic diseases in China. Furthermore, Ye (2013) analysed data on Jilin province, and found that those who are single have the lowest prevalence of chronic diseases, while the prevalence among divorced/widowed persons is the highest. Thus, we hypothesise that the reported prevalence is higher among divorced and widowed elderly people than among other types of elderly people. The WHO (2005) reported that chronic diseases among the elderly are predominantly attributable to unhealthy habits during youth, such as excessive smoking and drinking. In addition, using data on China, Chen (2005), Jiao et al. (2002), and Zhao et al. (2015) found similar results, namely, that cigarette smoking and alcohol usage are risk factors for chronic diseases. Therefore, we hypothesise that (cigarette) smoking and drinking (alcohol) have a considerable influence on the reported prevalence of chronic diseases and specific chronic diseases in China.

Numerous studies have confirmed that age has a considerable effect on chronic diseases in China (Chen, 2005; Jiao et al., 2002; Lin et al., 2002; Yin, 2011), with the prevalence increasing with age. Furthermore, Jiao et al. (2002), Ye (2013), and Yin (2011) pointed out that education and exercise have nonnegligible effects on chronic diseases, with the prevalence increasing for lower levels of education and less exercise. In this study, we divide exercise into two categories: sports activities and cultural activities. Based on the above studies, we hypothesise that the prevalence increases in older people who get less exercise. However, we hypothesise that people with higher levels of education are more likely to report having chronic diseases because they learn more about the dangers of such diseases and pay more attention to their health.

The above studies are limited to a single province or city. Furthermore, with the exception of some statistical descriptive reports, few studies consider the spatial dimension as an independent factor in chronic disease research in China. There are tremendous differences in economic development levels, medical conditions, and living conditions among the provinces in China, all of which can affect the diagnosis and treatment of a chronic disease. As a result, the prevalence of chronic diseases is quite different among the provinces. Thus, we hypothesise that the province is a critical factor affecting chronic diseases and specific chronic diseases. In addition, we hypothesise that the reported prevalence is similar to the case of the census register, in that it is higher in economically developed provinces.

In summary, based on past studies and on China's tremendous geographic differences, we hypothesise that the following factors are important factors affecting the prevalence of chronic diseases and specific chronic diseases in elderly Chinese people: gender, census register (urban or rural), marital status, smoking, drinking, age, education years, sports activities and cultural activities, and province. In addition, because most of the samples between surveys in 2006 and 2010 are the same, we further take the effects of time into account.

Prior studies usually only consider the linear effects of the continuous covariates (such as education years) on a chronic disease, even though they may have nonlinear effects. Because STAR models can include spatial effects, the nonlinear effects of continuous factors, and linear or fixed effects in a single model, we apply a STAR model in our empirical study in order to determine which covariates have considerable effects on chronic diseases in China.

In applying this model, we use a fully Bayesian estimation based on Markov chain Monte Carlo (MCMC) simulations, as well as `BayesX`, a standalone software package used to fit general STAR models. Moreover, Umlauf et al. (2015) developed an interactive `R` interface for `BayesX`, called `R2BayesX`, which can be used to specify STAR models using `R`'s formula language. Furthermore, this package adds extensive graphics capabilities for visualizing fitted STAR models.

The rest of this paper is structured as follows. Section 2 introduces the STAR models, and their estimations. Section 3 explains the data. Section 4 presents the STAR model applied in this study, and describes the `R2BayesX` settings. Section 5 discusses the empirical results for chronic diseases, as well as for two specific chronic diseases (hypertension and heart disease) using the `R2BayesX` package. Section 6 presents our conclusions.

2. Estimation Methods

2.1. STAR models

STAR models were first introduced by Fahrmeir et al. (2004), and not only contain generalized linear effects, but also allow for nonlinear effects of continuous covariates and spatial effects.

For generalized linear models, the mean μ of the response variable y is linked to a linear predictor η by

$$\mu = h^{-1}(\eta),\ \eta = x'\gamma, \tag{1}$$

where h is a known link function and γ denotes unknown regression coefficients.

Following Fahrmeir et al. (2004, p.734), in STAR models, the linear predictor is replaced by the following structured additive predictor:

$$\eta = f_1(x_1) + \cdots + f_p(x_p) + w'\gamma, \tag{2}$$

where $x_1,\ldots,x_p$ are nonlinear covariates, and f_j are smooth functions, which can represent potentially non-linear effects of continuous covariates or spatially structured and unstructured effects.

Furthermore, when the response variable is binary, the link function becomes a logit function, and we can consider the following logistic STAR model:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \eta = f_1(x_1) + \cdots + f_p(x_p) + w'\gamma, \tag{3}$$

where p denotes the probability of a specific event occurring (such as the probability of a person having a chronic disease).

2.2. Estimation of STAR models

In this study, the STAR model is estimated using a Bayesian inference. For the Bayesian inference, all components of the STAR models must be supplemented with appropriate prior assumptions.

In the STAR model (2), $w'\gamma$ denotes the fixed effects. In general, a diffuse prior $p(\gamma) \propto const$ is assumed for the fixed effects parameter γ. At the same time, specific priors are given to the functions $f_j(\cdot)$, and depend on the type of the covariate.

For the nonlinear effects of continuous covariates $f_j(\cdot)$, Bayesian P-splines are utilized. P-splines are an improvement over B-splines, and introduce a penalty variable to prevent overfitting.

The basic idea behind P-splines is dividing the data interval into a relatively large number of sub-intervals, and an unknown smooth function f of a covariate x can be approximated by a linear combination of some basis functions. P-splines can be approximated by a polynomial spline of degree l, defined on a set of equally spaced knots $x^{\min} = \zeta_1 < \zeta_2 < \cdots < \zeta_m = x^{\max}$ within the domain of x. Following Fahrmeir et al. (2013, pp.426–431), a spline can be expressed by an adequate linear combination of $d = m+l-1$ B-spline basis functions:

$$f(x) = \sum_{j=1}^{d} \beta_j B_j(x), \tag{4}$$

where $B_j(x)$ of degree l is defined as follows: for $j = 1,\ldots,d-1$,

$$\begin{cases} B_j^0(x) = I(\zeta_j \le x < \zeta_{j+1}) & l = 0 \\ B_j^1(x) = \dfrac{x-\zeta_{j-1}}{\zeta_j - \zeta_{j-1}} I(\zeta_{j-1} \le x < \zeta_j) + \dfrac{\zeta_{j+1}-x}{\zeta_{j+1}-\zeta_j} I(\zeta_j \le x < \zeta_{j+1}) & l = 1 \\ B_j^l(x) = \dfrac{x-\zeta_{j-l}}{\zeta_j - \zeta_{j-l}} B_{j-1}^{l-1}(x) + \dfrac{\zeta_{j+1}-x}{\zeta_{j+1}-\zeta_{j+1-l}} B_j^{l-1}(x) & l \ge 2, \end{cases} \tag{5}$$

where $I(\cdot)$ is an indicator function.

The crucial choice for P-splines is the number of knots: too few knots may not be flexible enough, while choosing too many knots may overfit the data. To prevent overfitting, a penalty term is included. The penalty terms are expressed as in Eilers and Marx (1996):

$$P(\lambda) = \frac{1}{2}\lambda \sum_{j=r+1}^{d} (\Delta^r \beta_j)^2, \tag{6}$$

where λ is the smoothing parameter and Δ^r denotes the rth-order differences. In most applications, second-order differences are chosen, which are defined as $\Delta^2\beta_j = \Delta^1\Delta^1\beta_j = \Delta^1\beta_j - \Delta^1\beta_{j-1} = \beta_j - 2\beta_{j-1} + \beta_{j-2}$. Furthermore, when $\lambda \to \infty$, the function estimate for $f(x)$ is close to linear in the case of second-order differences.

In applications, a common choice for P-splines is B-splines of degree $l = 3$, with $m = 20$ equidistant knots. These setting ensure that the estimated function is twice continuously differentiable, which allows sufficient flexibility to capture the typical nonlinear mode.

The general form of the prior of β_j for a P-spline is given by the multivariate normal distribution:

$$p(\beta_j|\tau_j^2) \propto \exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right), \tag{7}$$

where K_j is a penalty matrix, and τ_j^2 is a prior variance, which determines the impact of the prior distribution on the function estimates. For the full Bayesian inference, weakly informative inverse Gamma hyperpriors $\tau_j^2 \sim \mathrm{IG}(a_j, b_j)$ are assigned to τ_j^2, with $a_j = b_j = 0.001$ as a general setting. More detailed information about the Bayesian P-splines can be found in Lang and Brezger (2004).

As we mentioned above, f_j denotes smooth functions that can be used to represent the potentially non-linear effects of continuous covariates or to represent a spatial effect. If f_j represents a spatial effect, it is expressed as $f_{spat}(\cdot)$.

The spatial effect in a STAR model is the effect of a spatial covariate. Usually, this is a proxy for unobserved influential factors, some of which may have a strong spatial correlation (structured), while others may be present only locally (unstructured). Thus, in order to distinguish between these two kinds of spatial effects, $f_{spat}(\cdot)$ is split into a spatially correlated (structured) part $f_{str}(\cdot)$ and spatially uncorrelated (unstructured) part $f_{unstr}(\cdot)$, i.e. $f_{spat}(\cdot) = f_{str}(\cdot) + f_{unstr}(\cdot)$. The structured spatial effects simply indicate that the spatial effects are correlated. There is no specific structure imposed on the spatial effects.

The spatially structured effect $f_{str}(\cdot)$ can be specified using stationary Gaussian random field (GRF) priors. When the place of residence is known exactly, given by geographical x-y coordinates, the spatial analysis can be conducted using a stationary GRF. The estimation of a GRF is based on the centroids of particular regions for geosplines and geokriging terms. More detailed information about stationary GRF priors can be found in Fahrmeir et al. (2013, pp.500–530). The spatially unstructured effect $f_{unstr}(\cdot)$ can be specified using simple Gaussian i.i.d. priors, and denotes the random effect of a covariate.

3. Data

We define a chronic disease using the medical definition: a disease that lasts for a long time (more than three months) and cannot be cured.

The data for this study are taken from the 2000, 2006, and 2010 Chinese Urban and Rural Elderly Population Surveys, conducted by the China Research Center on Aging of the National Committee on Aging. The survey in 2000 only investigated whether people were suffering from chronic diseases, while those in 2006 and 2010 were expanded to include questions on specific chronic diseases. And most of the samples between surveys in 2006 and 2010 are the same. Thus, we also analyse two specific chronic diseases (hypertension and heart disease, both of which are common in China, with a prevalence of more than 10%) in 2006 and 2010.

Moreover, the surveys focused on the following 20 representative provinces, municipalities, and autonomous regions: the eastern region - Beijing, Shanghai, Hebei, Liaoning, Jiangsu, Zhejiang, Fujian, Shandong, and Guangdong; the central region - Heilongjiang, Anhui, Henan, Shanxi, Hubei, and Hunan; and the western region - Sichuan, Yunnan, Shaanxi, Xinjiang, and Guangxi. The selected provinces, municipalities, and autonomous regions are shown in Figure 1. In China, the degree of economic development is closely related to geographical location. In general, provinces in the eastern region are almost economically developed provinces, provinces in the central region are moderately developed provinces, and provinces in the western region tend to be economically backward provinces.

The data sampling method used is the same as that of the Fifth Population Census; based on the distribution of the population aged 60 and older, a quota from each of the regions is determined. Then, stratified sampling is used to confirm that the survey results represent the total elderly population in China (Gao and Li, 2016).

After the survey samples were determined, the interviewers conducted household surveys. Here, interviews were conducted by interviewers, who then completed the questionnaires on behalf of the interviewees, based on their responses. No questionnaires were completed by the interviewees. Then, the interviewers checked and verified the responses after the investigation. As a result, the data accuracy is high. The three surveys generated $20,256$ responses, $19,947$ responses, and $19,986$ responses, respectively.

4. Model

4.1. Model specification

Because the samples include data on whether people suffer from chronic diseases in 2000, 2006, and 2010, the first step is a Bayesian analysis of the geographic distribution of chronic diseases and their relationships with potential risk factors. Moreover, for 2006 and 2010, data are included on common chronic diseases (hypertension and heart disease). Thus, the second step is a Bayesian analysis of the geographic distribution of these two common chronic diseases and their relationships with potential risk factors. The second step is a refinement of the first step. Therefore, this paper describes how to implement the first step only.

Given a set of observations y_i, $1 \leq i \leq n$, y_i is a binary response for a chronic disease, such that

$$y_i = \begin{cases} 1 & \text{if one has a chronic disease} \\ 0 & \text{otherwise.} \end{cases}$$

Because the responses are binary, we consider a logistic STAR model to estimate the probability of an elderly person having a chronic disease ($y_i = 1$) versus the probability of an elderly person not having a chronic disease ($y_i = 0$):

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \eta_i = f_1(x_{i1}) + \cdots + f_p(x_{ip}) + w_i'\gamma, \ 1 \leq i \leq n, \tag{8}$$

where $p_i = \Pr(y_i = 1)$, $x_{i1}, \ldots, x_{ip}$ are p continuous covariates, f_j are smooth functions, $w_i = (w_{i1}, \ldots, w_{ir})'$ is a vector of r categorical covariates, and γ is an r-dimensional vector of unknown regression coefficients for the categorical covariates w_i. The response is distributed as a Bernoulli random variable, such that
$f(y_i|\eta_i) = p_i{}^{y_i}(1-p_i)^{(1-y_i)} = \exp[y_i\eta_i - \log(1+\exp(\eta_i))]$ for $y_i = 0, 1$,
where $\eta_i = \text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$.

Based on the previous studies mentioned in the Introduction, we analyse the linear effects of the following categorical covariates: gender of the elderly person (female or male), census register (urban or rural), marital status (live with spouse, live differently with spouse, widowed, divorce, and unmarried), smoking (smoked previously, currently smoke, and never smoke), and drinking (drank previously, currently drink, and never drink). In addition, we investigate the potential nonlinear effects of the following continuous covariates: the elderly's age (Age), education years (EY), number of sports activities (SA), and cultural activities (CA). Furthermore, because the observations on chronic diseases are associated with where a person lives, it is important to account for geographical/spatial differences in the analysis. Therefore, by taking spatial and nonlinear effects into account, the predictor shown in (8) is replaced by the following predictor:

$$\eta_i = f_1(Age_i) + f_2(EY_i) + f_3(SA_i) + f_4(CA_i) + f_{spat}(Province_i) + w_i'\gamma, \tag{9}$$

where $f_1(Age_i)$, $f_2(EY_i)$, $f_3(SA_i)$, and $f_4(CA_i)$ are nonlinear smooth effects of the continuous covariates, and $f_{spat}(Province)$ is the effect of the spatial covariate $Province_i$. Here, $Province_i \in \{1, \ldots, S\}$ is an integer, where S is the number of surveyed regions. Every integer indicates the province, municipality, or autonomous region in which the respondent is living. For example, $Province_i = 1$ means the respondent lives in Heilongjiang province, and $Province_i = 8$ denotes a respondent living in Beijing city. In this study, the number of surveyed provinces, municipalities, and autonomous regions is 20; thus $S = 20$ (i.e., $Province_i \in \{1,\ldots,20\}$).

In China, the provinces are connected in that they usually have some similarity and correlation, and so they are spatially correlated. Therefore, we consider the structured spatial effect $f_{str}(province)$ rather than the unstructured spatial effect $f_{unstr}(province)$.

Finally, we estimate the following STAR model:

$$\eta_i = \beta_0 + f_1(Age_i) + f_2(EY_i) + f_3(SA_i) + f_4(CA_i) + f_{str}(province_i) + w_i'\gamma, \tag{10}$$

where $f_{str}(province_i)$ is a structured spatial effect, and $w_i'\gamma$ are the linear effects of the following categorical covariates: gender of the elderly, census register, marital status, smoking, and drinking.

In this study, the estimation of the above STAR model is obtained using a Bayesian inference. In the STAR model (10), for the fixed effects $w_i'\gamma$, a diffuse prior $p(\gamma) \propto const$ is assumed for the parameter γ. For the nonlinear effects of the continuous covariates $f_1(Age)$, $f_2(EY)$, $f_3(SA)$, and $f_4(CA)$, Bayesian P-splines are utilized. In addition, we estimate the structured spatial effect $f_{str}(province)$ using the stationary Gaussian random field (GRF) approach, because the geographical x-y coordinates of every surveyed province in this study are known exactly.

4.2. Model implementation

For each data set (2000, 2006 and 2010), the STAR model (10) is fitted to a chronic disease.

We change all categorical covariates into dummy variates. For example, marital status has five categories: live with spouse; live differently with spouse; widowed; divorce; and unmarried. We use four dummy variates to represent this categorical covariate: marital statusA: live with spouse; marital statusB: live differently with spouse; marital statusC: widowed; marital statusD: divorce.

The model (10) can be implemented in `R2BayesX`, an open `R` package for STAR models. For this model, $25,000$ Markov chain Monte Carlo (MCMC) iterations were carried out after a burn in sample of $2,000$. In general, these random numbers are correlated. Thus, we store every 10th sampled parameter of the Markov chain. The posterior mean, posterior standard deviation, posterior median, and 90% and 95% credible intervals for all parameters, estimated from the posterior distributions, are used to assess the model fit.

When fitting $f_{str}(Province)$ in `R2BayesX`, a “map" argument is needed. It can be an object of class “SpatialPolygonsDataFrame" or an object of class “`bnd`." Spatial polygon data in China can be downloaded as shapefiles. Furthermore, using the function `shp2bnd()` of the `R` package – shapefiles package, the shapefiles can be changed to “`bnd`" objects, – `Chinabnd`. The class “`bnd`" is a `list()` of polygon matrices, with x- and y-coordinates of the boundary points in the first and second columns, respectively, which can be used to calculate the centroids of polygons to estimate the smooth bivariate effects of the resulting coordinates.

5. Empirical Results

We use a logistic STAR model to estimate the probability of an elderly person having a chronic disease $(y_i = 1)$ versus the probability of an elderly person not having a chronic disease $(y_i = 0)$. It is well known that the coefficients in a logistic regression model do not represent marginal effects, but rather log odds. It is difficult to interpret the coefficients, and this problem becomes even more complex when considering non-linear effects. Therefore, we focus on which group has a greater impact, not the extent of the effect. For example, we examine whether the prevalence of a chronic disease in elderly females is higher than that of elderly males, not the marginal effects of the prevalence of a chronic disease in elderly females and males.

From (8), the structured additive predictor η_i and probability $p_i = \Pr(y_i = 1)$ are positively related. If the coefficient is positive, η_i of the experimental group is larger than η_i of the control group. Then, $p_i = \Pr(y_i = 1)$ of the experimental group is larger, which means the experimental group is more likely to have chronic diseases. Otherwise, if the coefficient is negative, the control group is more likely to have chronic diseases. Furthermore, when all other coefficients remain unchanged, a certain coefficient becomes larger than η_i becomes larger, and the probability $p_i = \Pr(y_i = 1)$ becomes larger, that is, a larger coefficient denotes a greater effect on a chronic disease.

Table 1 displays the variables used in the models and gives their meanings and values. Table 2 compares the hypotheses and the empirical results on the reported prevalence of a chronic disease, hypertension, and heart disease.

5.1. Empirical results for chronic diseases

In Table 2, for the covariates of gender and census register, zero is not included in the 95% credible intervals in 2000, 2006, and 2010. Therefore, we find that these covariates do affect chronic diseases. In addition, the posterior means of gender and census register are positive. This indicates that in comparing elderly females and elderly males (female is 1, male is 0), and urban elderly people and rural elderly people (urban is 1, rural is 0), the reported prevalence of the former groups is higher than that of the latter groups. These results are also consistent with the hypotheses.

In general, marital status may affect the health of elderly people (Kiecolt-Glaser and Newton, 2001). However, being married does not guarantee health benefits. A decline in the quality of marriage has a negative effect on mental and physical health (Wickrama et al., 1997). We find that, marital status has a negligible effect on a chronic disease, because the 95% and 90% credible intervals of marital status include zero in all three years. Thus, we reject the hypothesis of marital status.

For the covariates smoking and drinking, zero was not included in the 95% credible intervals of smokingA (smoked previously), drinkingA (drank previously), and drinkingB (currently drink) in all three years. Therefore, we hold the hypotheses that smoking and drinking do affect chronic diseases. However, smoking and drinking have two different kinds of effects on a chronic disease. The results of smokingA (smoked previously)/drinkingA (drank previously) are consistent with the following hypotheses: the reported prevalence in elderly people who smoked/drank previously, but no longer do so, is higher than that in elderly people who never smoke/drink. However, the results of smokingB (currently smoke) and drinkingB (currently drink) seem to be counter-intuitive: the reported prevalence in elderly people who currently smoke/drink is lower than that in those who never smoke/drink. One possible reason for this apparent contradiction in the case of drinking may be the following: moderate drinking has little effect on health, and some reports even show that moderate drinking is beneficial to our health (Fillmore et al., 2006).

In addition, as explained above, a larger coefficient denotes a greater effect on chronic diseases. From Table 2, for the fixed effects of the categorical covariates on chronic diseases, the census register has the greatest effect on chronic diseases, followed by gender, smoking, and drinking.

The nonlinear effects of the continuous covariates (age, years of education, sports activities, and cultural activities) are considerable, but they are all very small (mean values are less than 0.012) in the three years. Furthermore, the posterior means are all positive, which indicates a greater reported prevalence for older people with a higher level of education, and who do more sports activities and cultural activities. Age and education are consistent with the hypotheses, but sports activities and cultural activities are contrary to the hypotheses. One possible reason may be as follows: people who do more sports activities and cultural activities pay more attention to their health, and so are more likely to go to the hospital for an examination, and more likely to report having a chronic disease. Figure 2 gives the detailed nonlinear effects of these continuous covariates on a chronic disease, with 95% credible bands in all three years. The tails of all continuous covariates are wide because the numbers of observations in these parts are all very small.

Since a larger coefficient denotes a greater effect on chronic diseases, and the coefficient of province on chronic diseases is relatively large, thus, the spatial effect of province is critical.

In 2000, Xinjiang province's structured spatial effect is the highest, followed by Anhui, Shaanxi, Sichuan, Guangxi, Beijing, and Shanxi provinces. Shandong province's structured spatial effect is the lowest (see Figure 3(a)). In 2006, Xinjiang province's structured spatial effect is still the highest, followed by Anhui, Hubei, and Hunan provinces. Guangdong province's structured spatial effect is the lowest, and Shandong, Guangxi and Fujian provinces' structured spatial effects are very low (see Figure 3(b)). In 2010, Zhejiang province's structured spatial effect is the highest, Hunan province's structured spatial effect is the lowest. In addition, Anhui, Fujian, and Xinjiang provinces' structured spatial effects are relatively high, and Shandong and Liaoning provinces' structured spatial effects are relatively low (see Figure 3(c)).

In conclusion, we find that the province is a critical factor affecting chronic diseases, but that the high reported prevalence of chronic diseases is not only in economically developed provinces (such as Zhejiang, Beijing, Fujian), but also in the economically backward provinces with complex terrain (such as Xinjiang, Guangxi), as shown in Figure 3. In addition, we should pay special attention to Xinjiang and Anhui provinces, because their structured spatial effects are quite high in all three years.

The high reported prevalence does not necessarily indicate that the health conditions in these provinces are bad. On the contrary, the high reported prevalence may indicate that elderly people pay more attention to their health. However, a high reported prevalence in economically backward provinces with complex terrain may indicate that the health conditions in these provinces are poor, for example, in Xinjiang province.

5.2. Empirical results for hypertension

As shown in Table 2, for the covariates of gender, census register, smokingB (currently smoke), drinkingA (drank previously), drinkingB (currently drink), age, education years, sports activities, cultural activities, and province, zero is not included in the 95% credible intervals in 2006 and 2010. Therefore, we find that they have considerable effects on hypertension, but we reject the hypothesis for marital status.

The empirical results for the fixed effects of the categorical covariates and for the nonlinear effects of the continuous covariates on hypertension are very similar to the results for chronic diseases. Thus, we do not repeat them here. The differences between general chronic diseases and hypertension are mainly reflected in the spatial effects.

Since a larger coefficient denotes a greater effect, and the coefficient of province on hypertension is relatively large, thus, the spatial effect of province on hypertension is critical.

In 2006, Hebei, Beijing, and Zhejiang provinces' structured spatial effects on hypertension are the highest, followed by Jiangsu, Shanghai and Heilongjiang. Guangxi, Yunnan, and Sichuan provinces' structured spatial effects are the lowest, although Guangdong and Liaoning provinces' structured spatial effects are also relatively low (see Figure 4(a)). In 2010, Yunnan province's structured spatial effect is the highest (see Figure 4(b)), showing a marked increase over the low level in 2006.

In conclusion, we find that the province is a critical factor affecting hypertension, but that the highest reported prevalence occurs mainly in economically developed provinces (e.g., Zhejiang, Beijing, and Guangdong), and not in economically backward provinces, as shown in Figure 4. In addition, note that Zhejiang province's structured spatial effects are quite high in both years.

The high reported prevalence in these economically developed provinces simply indicates that the elderly pay more attention to their health. However, the high reported prevalence in Yunnan province in 2010 indicates that the health conditions in this province are poor. Reports in 2013 revealed that for every 10 Kunming (Yunnan provincial capital) residents, two suffer from hypertension, but that about 70% of patients are not aware of this.

5.3. Empirical results for heart disease

As shown in Table 2, for the covariates of gender, census register, smokingA (smoked previously), drinkingB (currently drink), age, education years, sports activities, cultural activities, and province, zero is not included in the 95% and 90% credible intervals in 2006 and 2010. Therefore, we find that they have considerable effects on heart disease, but we reject the hypothesis for marital status.

The empirical results for the fixed effects of the categorical covariates and the nonlinear effects of the continuous covariates on heart disease are very similar to the results for chronic diseases. Thus, we do not repeat them here. The differences between general chronic diseases and heart disease are still mainly reflected in the spatial effects.

As explained above, a larger coefficient denotes a greater effect, and the coefficient of province on heart disease is relatively large, thus, the spatial effect of province on heart disease is critical.

In 2006, Heilongjiang province's structured spatial effect was the highest, followed by Liaoning, Beijing, Hebei, and Xinjiang. Guangdong, Guangxi, Yunnan, and Sichuan provinces' structured spatial effects were the lowest (see Figure 4(c)). In 2010, Shaanxi province's structured spatial effect was the highest, followed by Shanxi and Anhui. Shandong province's structured spatial effect was the lowest, although Xinjiang, Hunan, and Guangxi were relatively low (see Figure 4(d)).

In conclusion, we find that the province is a critical factor affecting heart disease. As shown in Figure 4, in 2006, the highest reported prevalence appeared not only in economically backward provinces (e.g., Xinjiang), but also in the economically developed provinces (e.g., Hebei, Beijing) and the moderately developed provinces (e.g., Heilongjiang, Liaoning). In contrast, in 2010, high spatial effects appeared in moderately developed provinces only (e.g., Shaanxi, Shanxi, Anhui).

5.4. Empirical results of adding time factor

Because most of the samples between surveys in 2006 and 2010 are the same, we further take the effects of time into account. And the estimation results are shown in Table 3.

Table 3 gives the posterior means, posterior standard deviations, and posterior medians for all covariates (including time) of a chronic disease, hypertension, and heart disease. In Table 3, for the covariate of time, zero is not included in the 95% credible intervals. Therefore, time do affect chronic diseases. In addition, the posterior means of time are positive. This indicates that in comparing 2006 and 2010 (2010 is 1, 2006 is 0), the reported prevalence in 2010 is higher than that in 2006.

The higher reported prevalence in 2010 does not necessarily indicate that the health conditions of the elderly people are worse. On the contrary, the high reported prevalence may indicate that elderly people pay more attention to their health with the time going.

6. Conclusions

This study applied a STAR model to determine which covariates have considerable effects on chronic diseases and specific chronic diseases (hypertension and heart disease). STAR models combine spatial effects, nonlinear effects of continuous factors, and the linear or fixed effects into a single model.

The findings are as follows: (1) the following covariates have considerable effects on chronic diseases and specific chronic diseases (hypertension and heart disease) (in descending order): census register (rural or urban), gender, smoking, drinking, province, age, cultural activities, years of education, sports activities; (2) because the 95% and 90% credible intervals of marital status include zero in all three years, thus, we reject the hypothesis for marital status; that is, the effect of marriage is negligible; (3) elderly females, urban elderly people and the elderly who smoke and drink are more likely to report having chronic diseases. In addition, the reported prevalence increases with age, education level, and participation in sports activities and cultural activities; (4) a high reported prevalence may indicate that the health conditions are bad, but may also indicate that the elderly pay more attention to their health; (5) a larger coefficient denotes a greater effect on chronic diseases, thus, province is a critical factor, but the high reported prevalence is not restricted to economically developed provinces. For chronic diseases, a high reported prevalence occurs in economically developed provinces and in economically backward provinces with complex terrain; and (6) because most of the samples between surveys in 2006 and 2010 are the same, we further take the effects of time into account, and find that time also has considerable effects.

Based on the above findings, the government should further strengthen its investment in rural areas and in economically underdeveloped provinces, such as Xinjiang province and Anhui province, as a cost-effective intervention. In addition, the government should educate the population on the harmful effects of smoking and drinking on health. Furthermore, economically developed provinces' highest structured spatial effects on hypertension do not mean that the elderly in these provinces are more likely to have hypertension. However, the government should strengthen its investment in the promotion and diagnosis of hypertension in economically backward areas (e.g., Yunnan province). In the case of heart disease, the government should strengthen its investment in provinces such as Heilongjiang province and Liaoning province.

The nonlinear effects in this study are considerable, but very small. Thus, we should build a better method to reconsider the nonlinear effects. Furthermore, we have data for three years, and conduct separate estimates for the model in each year. In future research, we will include time as a feasible covariate in the regression model to consider the impact of time on chronic diseases.

Acknowledgements

The authors appreciate the comments of anonymous referees, which improve the article greatly. They thank the China Research Center on Aging for their support in data collection and management. The first author would like to thank Hokkaido University President's Fellowship for its financial support. The work of the second author was supported in part by a Grant-in-Aid for Scientific Research (No.16K03589) from the Japan Society for the Promotion of Science (JSPS).

References

AASM, (2017). Sleep disorders affect men and women differently: Women are more likely to feel tired and depressed than men. https://www.sciencedaily.com/releases/2017/05/170523081838.htm. Accessed 23 May 2017.

BASILE, R., BENFRATELLO, L., CASTELLANI, D., (2013). Geoadditive models for regional count data: An application to industrial location. Geographical Analysis 45, pp. 28–48.

CHEN, J. J., (2005). The prevalence study on main chronic diseases and the associated behavioral risk factors among the middle and old aged people in some areas of Hubei province. Master's thesis, Wuhan University (In Chinese).

CRESSIE, N., (2015). Statistics for spatial data. Wiley, New York.

EILERS, P. H. C., MARX, B. D., (1996). Flexible smoothing with B-splines and penalties. Statistical Science 11(2), pp. 89–102.

FAHRMEIR, L., KNEIB, T., LANG, S., MARX, B., (2013). Regression: models, methods and applications. Springer, New York.

FAHRMEIR, L., KNEIB, T., LANG, S., (2004). Penalized additive regression for space-time data: A Bayesian perspective. Statistica Sinica 14, pp. 731–761.

FILLINGIM, R. B., KING, C. D., RIBEIRO-DASILVA, M. C., RAHIM-WILLIAMS, B., RILEY, J. L., (2009) Sex, gender, and pain: A review of recent clinical and experimental findings. Journal of Pain 10, pp. 447–485.

FILLMORE, K. M., KERR, W. C., STOCKWELL, T., CHIKRITZHS, T., BOSTROM, A., (2006). Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies. Addiction Research and Theory 14 (2), pp. 101–132.

FREEDMAN, V. A., MARTIN, L. G., (2000). Contribution of chronic conditions to aggregate changes in old-age functioning. American Journal of Public Health 90, pp. 1755–1760.

GAO, P., LI, H. D., (2016). New characteristics of active life expectancy of the elderly in China. Advances in Aging Research 5, pp. 27–39.

GENIAUX, G., NAPOLÉONE, C., (2008). Semi-parametric tools for spatial hedonic models: An introduction to mixed geographically weighted regression and geoadditive models. In: Baranzini A, Ramirez J, Schaerer C, Thalmann P (eds) Hedonic Methods in Housing Markets. Springer, New York, pp. 101–127.

GU, Z. L., ZHANG, H, HUANG, J. P., MI, Y. P., (2009) Analysis on causes of main chronic diseases and death from 2004 to 2006 in Gangza district of Nantong city. Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases 17, pp. 89–90 (In Chinese).

HE, J., GU, D., WU, X., REYNOLDS, K., DUAN, X., YAO, C., WANG, J., CHEN, C. S., CHEN, J., WILDMAN, R. P., KLAG, M. J., WHELTON, P. K., (2005). Major causes of death among men and women in China. The New England Journal of Medicine 353, pp. 1124–1134.

JIAO, S. F., YIN, X. J., WANG, Y., XIE, J., GUO, B., (2002). Analysis on behaviors risk factors of chronic disease in residents of Beijing and study its countermeasure. China Public Health 18 (2), pp. 197–198 (In Chinese).

KAMMANN, E., WAND, M. P., (2003). Geoadditive models. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52, pp. 1–18.

KIECOLT-GLASER, J. K., NEWTON, T. L., (2001). Marriage and health: His and hers. Psychological Bulletin 127, pp. 472–503.

KNEIB, T., (2006). Mixed model based inference in structured additive regression. Dissertation, Ludwig-Maximilians-Universität.

LANG, S., BREZGER, A., (2004). Bayesian P-splines. Journal of Computational and Graphical Statistics 13 (1), pp. 183–212.

LI, Z. K., ZHANG, M. R., YANG, Z., SHEN, L., MA, Y., GONG, H. Q., TIAN, R., YAN, F., (2012). The epidemiological characteristics and PYLL analysis of death of chronic diseases among residents in Kunming city in 2007-2010. Chinese Primary Health Care 26, pp. 84–86 (In Chinese).

LIN, H., ZHANG, T. H., YANG, H., WANG, C. B., DUAN, L. R., ZHANG, X. P., GONG, G. C., (2002). Analysis of the chronic disease and its influencing factors among 895 elders in Beijing. Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases 10 (6), pp. 270–272 (In Chinese).

SAULEAU, E. A., HENNERFEIND, A., BUEMI, A., HELD, L., (2007). Age, period and cohort effects in Bayesian smoothing of spatial cancer survival with geoadditive models. Statistics in Medicine 26, pp. 212–229.

THORPE, K. E., HOWARD, D. H., (2006). The rise in spending among medicare beneficiaries: The role of chronic disease prevalence and changes in treatment intensity. Health Affairs - Web Exclusive 25, pp. w378–w388.

UMLAUF, N., ADLER, D., KNEIB, T., LANG, S., ZEILEIS, A., (2015). Structured additive regression models: An R interface to BayesX. Journal of Statistical Software 63, pp. 1–46.

VOGELI, C., SHIELDS, A. E., LEE, T. A., GIBSON, T. B., MARDER, W. D., WEISS, K. B., BLUMENTHAL, D., (2007). Multiple chronic conditions: Prevalence, health consequences, and implications for quality, care management and costs. Journal of General Internal Medicine 22, pp. 391–395.

WAND, H., WHITAKER, C., RAMJEE, G., (2011). Geoadditive models to assess spatial variation of HIV infections among women in local communities of Durban, South Africa. International Journal of Health Geographics 10, p. 28.

WHO, (2005). Preventing chronic diseases: A vital investment. Technical report, World Health Organization.

WICKRAMA, K., LORENZ, F. O., CONGER, R. D., ELDER JR, G. H., (1997). Marital quality and physical illness: A latent growth curve analysis. Journal of Marriage and the Family 59, pp. 143–155.

WOLFF, J. L., STARFIELD, B., ANDERSON, G., (2002). Prevalence, expenditures and complications of multiple chronic conditions in the elderly. Archives of Internal Medicine 162, pp. 2269–2276.

WOOLF, S. H., ARON, L., DUBAY, L., SIMON, S. M., ZIMMERMAN, E., LUK, K. X., (2015). How are income and wealth linked to health and longevity? Urban Institute.

YE, S. M., (2013). Analysis on non-communicable chronic diseases status and its influencing factors among adults in Gong Zhu-ling of Jilin province in 2012. Master's thesis, Jilin University (In Chinese).

YIN, S. Q., (2011). Analysis of chronic disease prevalence and influencing factors of the elderly in Chongwen district of Beijing. Chinese Primary Health Care 25 (4): 43–46 (In Chinese).

ZHAO, J., WAN, R., LIU, Z., WAN, Q. Q., ZHANG, Q., RUAN, Y., (2015). Prevalence and influencing factors of chronic diseases in residents aged 15 years and above in urban and rural area of Yunnan province, 2010-2012. Chinese Journal of Public Health 31 (7), pp. 878–882 (In Chinese).

ZHEN, S., (2010). Study of the current situation and influencing factors of chronic illnesses in Gansu. Master's thesis, Lanzhou University (In Chinese).

APPENDIX

Table 1: Variables in the models

Variable	Description	Value
Chronic Disease	Whether one has a chronic disease or not.	Having chronic disease is 1, else is 0.
Hypertension	Whether one has hypertension or not.	Having hypertension is 1, else is 0.
Heart Disease	Whether one has heart disease or not.	Having heart disease is 1, else is 0.
Time	Time of investigation.	2010 is 1; 2006 is 0.
Gender	Gender of the elderly person with categories 'male' and 'female'.	Female is 1; Male is 0.
Census Register	Census register with categories 'urban' and 'rural'.	Urban is 1; Rural is 0.
Marital Status	Marital status with categories 'live with spouse', 'live differently with spouse', 'widowed', 'divorce' and 'unmarried'	Using four dummy variables: Marital StatusA: live with spouse; Marital StatusB: live differently with spouse; Marital StatusC: widowed; Marital StatusD: divorce. Yes is 1, no is 0.
Smoking	Smoking condition with categories 'smoked previously', 'currently smoke' and 'never smoke'.	Using two dummy variables: SmokingA: smoked previously; SmokingB: currently smoke; Yes is 1, no is 0.
Drinking	Drinking condition with categories 'drank previously', 'currently drink' and 'never drink'.	Using two dummy variables: DrinkingA: drank previously; DrinkingB: currently drink; Yes is 1, no is 0.
Age	Age of the elderly people in years.	Continuous covariate, minimum value is 60.
Education Years	One's education years.	Continuous covariate, minimum value is 0.
Sports Activities	Number of sports activities one takes part in.	Continuous covariate, value changes from 0 to 5.
Cultural Activities	Number of cultural activities one takes part in.	Continuous covariate, value changes from 0 to 10.
Province	Province where one lives in.	Spatial covariate, integer, value changes from 1 to 20.

Table 2: Comparison of Hypotheses and Empirical Results on the Reported Prevalence

Variable	Hypotheses	Empirical Results (Posterior Mean)						
		Chronic Disease			Hypertension		Heart Disease	
		2000	2006	2010	2006	2010	2006	2010
Gender	considerable influence females > males	0.340**	0.344**	0.247**	0.158**	0.093**	0.489**	0.372**
Census Register	critical influence urban > rural	0.642**	0.631**	0.544**	0.599**	0.462**	0.872**	0.368**
Marital StatusA	considerable influence	-0.037	0.115	-0.100	-0.036	-0.033	0.342	0.111
Marital StatusB	divorced/widowed >	-0.101	0.258	-0.378	-0.045	-0.392	0.268	-0.024
Marital StatusC	other types	-0.034	0.098	-0.213	-0.028	-0.018	0.258	0.056
Marital StatusD		-0.031	0.181	-0.085	-0.024	-0.325	0.193	-0.058
SmokingA	considerable influence	0.430**	0.261**	0.267**	-0.051	-0.009	0.116*	0.117*
SmokingB	smoke > never smoke	0.037	-0.122**	-0.022	-0.237**	-0.178**	-0.003	-0.180**
DrinkingA	considerable influence	0.233**	0.274**	0.181**	0.173**	0.221**	0.018	0.057
DrinkingB	drink > never drink	-0.294**	-0.175**	-0.239**	-0.156**	-0.087**	-0.306**	-0.253**
Age	non-negligible influence older > younger	0.005**	0.005**	0.011**	0.006**	0.003**	0.010**	0.014**
Education Years	non-negligible influence longer > shorter	0.003**	0.005**	0.002**	0.004**	0.002**	0.003**	0.002**
Sports Activities	non-negligible influence more > less	0.003**	0.003**	0.003**	0.003**	0.003**	0.003**	0.004**
Cultural Activities	non-negligible influence more > less	0.004**	0.003**	0.004**	0.005**	0.005**	0.003**	0.003**
Province	critical influence developed > backward	0.416**	0.375**	0.054**	0.144**	0.201**	0.307**	1.975**

[1] "$**$" and "$*$" denote that zero is not included in the 95% and 90% credible intervals, respectively.

Table 3: Posterior Estimates of the Parameters for Chronic Disease (Add Time)

	Chronic Disease			Hypertension			Heart Disease		
Variable	Mean	Sd	Median	Mean	Sd	Median	Mean	Sd	Median
Fixed effects:									
(Intercept)	0.789	0.173	0.784	-0.911	0.191	-0.906	-2.209	0.271	-2.191
Time	0.129	0.027	0.129**	0.329	0.027	0.330**	0.088	0.030	0.088**
Gender	0.292	0.033	0.292**	0.129	0.034	0.129**	0.431	0.036	0.431**
CR	0.590	0.033	0.590**	0.549	0.031	0.549**	0.721	0.035	0.721**
MSA	0.015	0.105	0.017	-0.054	0.123	-0.056	0.193	0.162	0.192
MSB	0.006	0.150	0.003	-0.280	0.166	-0.281*	-0.031	0.211	-0.029
MSC	-0.048	0.107	-0.047	-0.051	0.125	-0.050	0.120	0.163	0.119
MSD	0.067	0.166	0.065	-0.204	0.174	-0.201	0.053	0.219	0.052
SmokingA	0.260	0.045	0.261**	-0.020	0.043	-0.021	0.113	0.047	0.114**
SmokingB	-0.076	0.035	-0.077**	-0.207	0.037	-0.206**	-0.090	0.042	-0.089**
DrinkingA	0.226	0.045	0.225**	0.181	0.042	0.181**	0.022	0.046	0.022
DrinkingB	-0.201	0.034	-0.201**	-0.125	0.036	-0.124**	-0.280	0.041	-0.280**
Nonlinear effects:									
sx(Age)	0.003	0.004	0.002**	0.004	0.005	0.002**	0.011	0.012	0.007**
sx(EY)	0.003	0.006	0.002**	0.002	0.003	0.001**	0.002	0.002	0.001**
sx(CA)	0.003	0.006	0.002**	0.002	0.003	0.002**	0.002	0.003	0.002**
sx(SA)	0.002	0.003	0.001**	0.003	0.007	0.001**	0.002	0.004	0.001**
sx(Province)	0.118	0.057	0.105**	0.106	0.052	0.094**	0.356	0.181	0.312**

[1] sx($\cdot$) corresponds to the smooth functions in STAR models.

[2] CR, MSA, MSB, MSC, MSD, CA, EY and SA are the abbreviation of Census Register, Marital StatusA, Marital StatusB, Marital StatusC, Marital StatusD, Cultural Activities, Education Years and Sports Activities, respectively.

[3] "$**$" and "$*$" denote that zero is not included in the 95% and 90% credible intervals, respectively. Mean, Sd and Median are posterior mean, posterior standard deviation and posterior median, respectively.

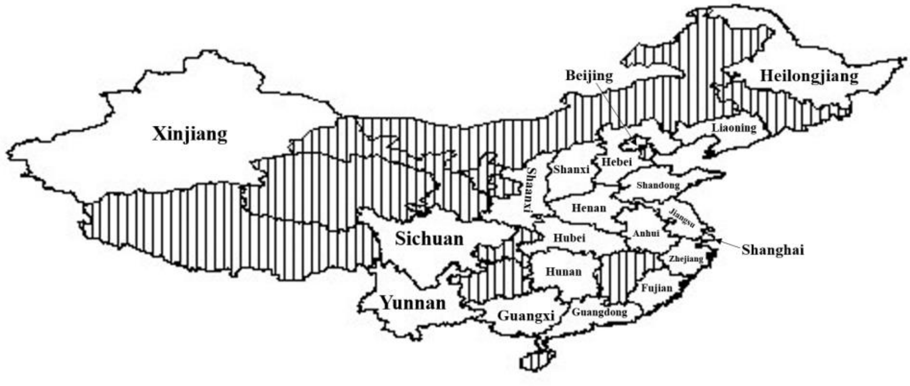
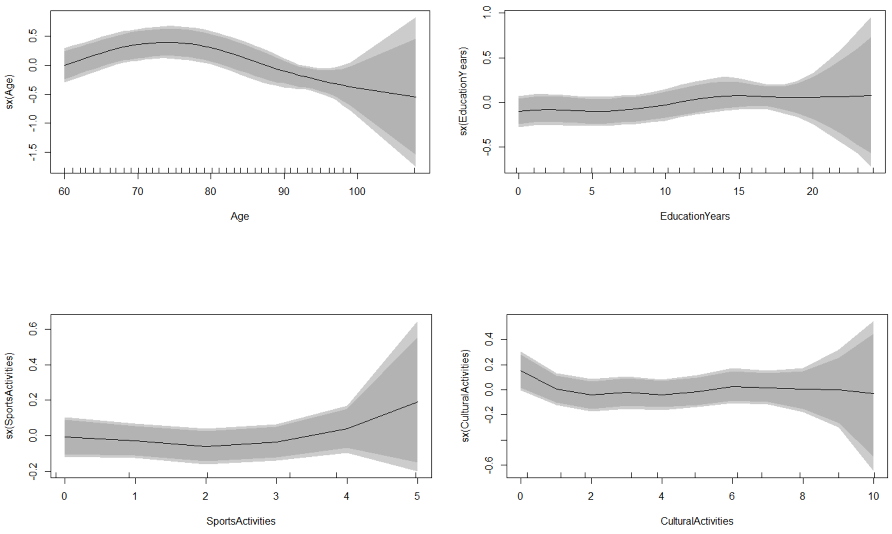
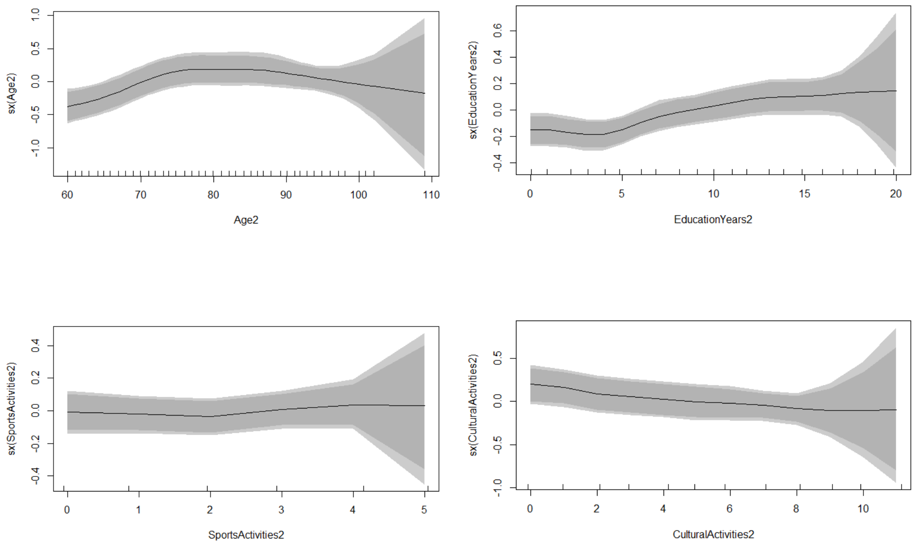


Figure 1. 20 selected provinces, municipalities and autonomous regions of China mainland

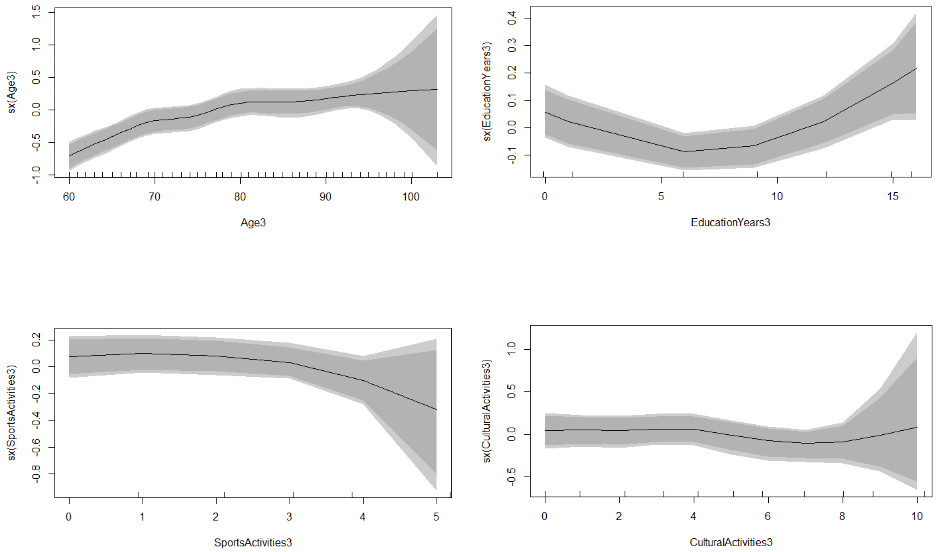


(a) 2000

Figure. 2. Effects of smooth terms on a chronic disease with 90% and 95% credible bands in 2000, 2006 and 2010

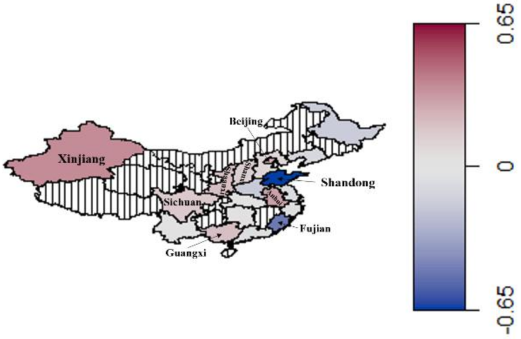


(b) 2006

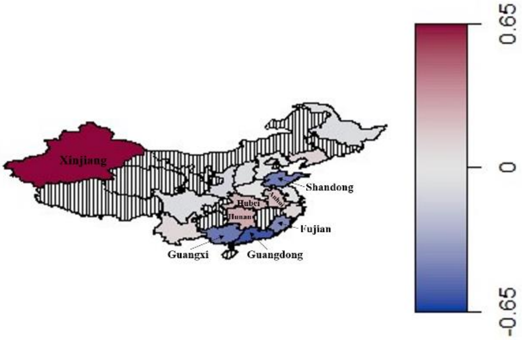


(c) 2010

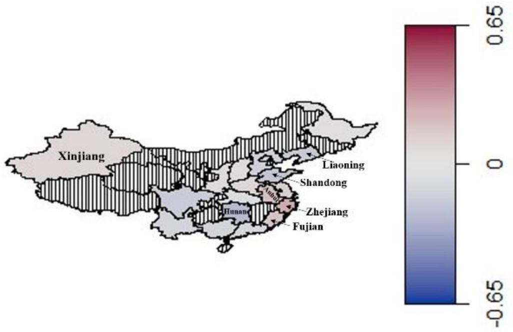
Figure. 2. Effects of smooth terms on a chronic disease with 90% and 95% credible bands in 2000, 2006 and 2010 (cont.)



(a) 2000

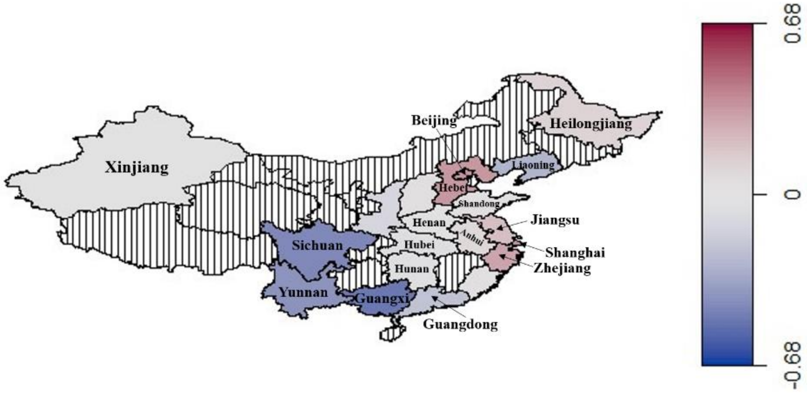


(b) 2006

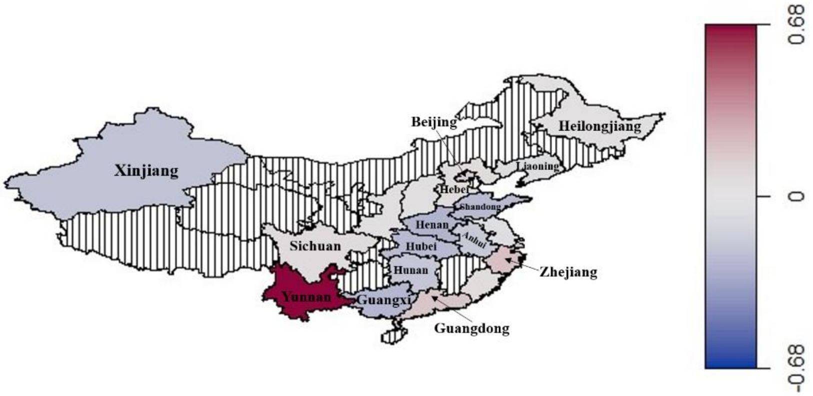


(c) 2010

Figure 3. Structured spatial effect on a chronic disease in 2000, 2006 and 2010

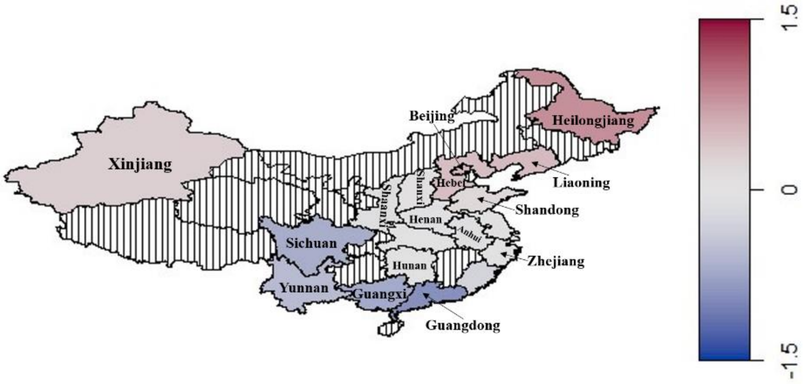


(a) Hypertension-2006

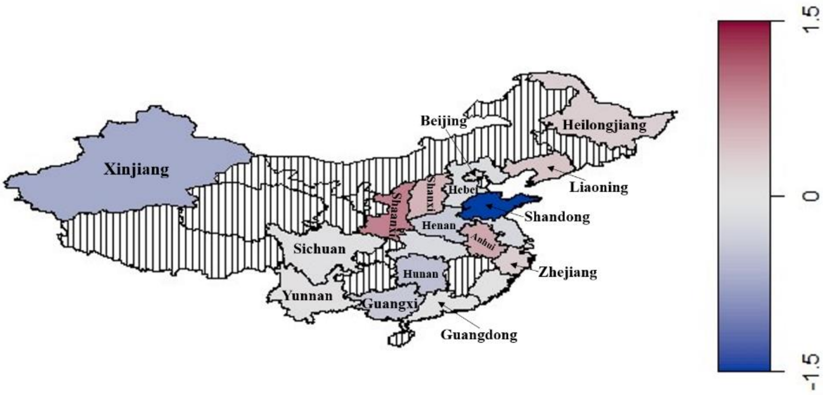


(b) Hypertension-2010

Figure 4. Structured spatial effect on hypertension and heart disease in 2006 and 2010



(c) Heart disease-2006



(d) Heart disease-2010

Figure 4. Structured spatial effect on hypertension and heart disease in 2006 and 2010 (cont.)