


RESEARCH

Open Access



# High-resolution age-specific mapping of the two-week illness prevalence rate based on the National Health Services Survey and geostatistical analysis: a case study in Guangdong province, China

Chuchun Wen<sup>1</sup>, Xiaoliang Huang<sup>2</sup>, Lifeng Feng<sup>2</sup>, Long Chen<sup>2</sup>, Wei Hu<sup>2</sup>, Yingsi Lai<sup>1\*</sup>  and Yuantao Hao<sup>1</sup>

## Abstract

**Background:** The two-week illness prevalence rate is an important and comparable indicator of health service needs. High-spatial-resolution, age-specific risk mapping of this indicator can provide valuable information for health resource allocation. The age-prevalence relationships may be different among areas of the study region, but previous geostatistical models usually ignored the spatial-age interaction.

**Methods:** We took Guangdong province, the province with the largest population and economy in China, as a study case. We collected two-week illness data and other potential influencing predictors from the fifth National Health Services Survey in 2013 and other open-access databases. Bayesian geostatistical binary regression models were developed with spatial-age structured random effect, based on which, high-resolution, age-specific two-week illness prevalence rates, as well as number of people reporting two-week illness, were estimated. The equality of health resource distribution was further evaluated based on the two-week illness mapping results and the health supply data.

**Results:** The map across all age groups revealed that the highest risk was concentrated in the central (i.e., Pearl River Delta) and northern regions of the province. These areas had a two-week illness prevalence > 25.0%, compared with 10.0–20.0% in other areas. Age-specific maps revealed significant differences in prevalence between age groups, and the age-prevalence relationships also differed across locations. In most areas, the prevalence rates decrease from age 0 to age 20, and then increase gradually. Overall, the estimated age- and population-adjusted prevalence was 16.5% [95% Bayesian credible interval (BCI): 14.5–18.6%], and the estimated total number of people reporting illness within the two-week period was 17.5 million (95% BCI: 15.5–19.8 million) in Guangdong Province. The Lorenz curve and the Gini coefficient (resulted in 0.3526) showed a moderate level of inequality in health resource distribution.

**Conclusions:** We developed a Bayesian geostatistical modeling framework with spatial-age structured effect to produce age-specific, high-resolution maps of the two-week illness prevalence rate and the numbers of people reporting two-week illness in Guangdong province. The methodology developed in this study can be generalized to

\*Correspondence: laiy3@mail.sysu.edu.cn

<sup>1</sup> School of Public Health, Sun Yat-sen University, Guangzhou 510080, China

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

other global regions with available relevant survey data. The mapping results will support plans for health resource allocation.

**Keywords:** Two-week illness, Bayesian geostatistical model, High resolution mapping

## Background

Decision making in health services should be based on accurate information of supply and need, for optimization of health resource allocation. The accurate supply information (e.g., number of medical staffs, number of hospital beds) are usually recorded by health facilities and easy to obtain, while common methods to estimate the health need have certain limitations. Many researchers used population number as an indicator for health need [1, 2], which is simple, but too crude considering the heterogeneity of illness risk in different places. Others used the demand indicators such as the number of visits to hospital [3, 4], which is an important indicator for hospital planning [5, 6]. However, the demand indicators may not fully represent the need, as gaps usually exist due to barriers like lack of money of patients to hospital, low geographical accessibility and cultural differences between patients and health providers [7]. And there may be “supplier-introduced demand” [8]. In this regard, the prevalence of illness within a two-week period, known as the two-week illness prevalence rate, directly reflects the risk of illness, breaks through the above limitations, and becomes an important and comparable indicator for health service needs [9–11]. This indicator is usually estimated by the self-reported illness status in health surveys, and the time frame “two-week” is considered reasonable for accurate recall of information about the illness and treatment [12, 13]. Till now, two-week illness has been applied in various health surveys worldwide [10, 14–18].

As spatial difference of two-week illness prevalence rate may exist within administrative divisions, obtaining it at high-spatial resolution is helpful for local spatial-targeting allocation of health resources. With a limited number of survey locations, this indicator has often been reported at a low spatial resolution [e.g., cities or provinces), as it is difficult to estimate for smaller spatial units using common statistical approaches [19]. Today, a Bayesian geostatistical modeling framework can be applied to overcome this challenge. This framework can be used to map the disease risk in areas without observed data, via extrapolation based on available disease data from adjacent areas and information about potential influencing predictors [20]. In this way, explicit spatial information of the disease is obtained despite the limited availability of survey locations. Accordingly, the Bayesian geostatistical modeling framework plays an important

role in both the implementation of interventions and the study of supply-and-need relationships.

Disease occurrence can be affected by a wide range of variables, including demographic factors (e.g., age [21] and gender [21]), socioeconomic factors (e.g., ethnicity [22], marital status [23], education level [24], employment status [24], occupation [25], income [24] and urbanization level [26]) and climatic and environmental factors (e.g., temperature [27], precipitation [28], land-cover [29], elevation [30] and air pollution [27]). Particularly, age is a main factor related to disease occurrence; different age groups are susceptible to different diseases and have different levels of disease risk. Furthermore, a previous report showed that different age groups report different levels of satisfaction with health services, and suggested that services for children and the elderly should be strengthened to promote equity in health service utilization [31]. This suggests that an age-specific analysis of the distribution of illness prevalence is necessary to inform management measures specific to different age groups.

Most previous statistical models addressed the age effect on disease risk either using a linear or categorical form of age as a fixed effect, or applying spline methods [32–34]. However, such models are not able to deal with situations where the age-prevalence relationship differs across the study region. Even though the two-week illness prevalence usually shows a pattern of a higher risk in young children and elderly people and a lower risk in other age groups [35], different areas may not have the exact same age-prevalence relationship. Geostatistical models can, however, be developed to address spatial-age interaction issue, allowing to estimate the different age-prevalence relationships among different areas. Parsaeian et al. used area-specific exponential functions to account for the variation of disease burden with age [36], which may call for high computational burden when interaction with space at high resolution. Based on the Bayesian geostatistical modeling framework, setting and projecting mesh knots is a considerable method to lower the computational burden for space-factor interaction [34].

This study aimed to develop a geostatistical modeling framework to address spatial-age interaction of the two-week illness prevalence and to provide high-spatial-resolution, age-specific estimates, based on data from health services survey. We chose Guangdong province, China as a representative region for case study. To further show

the application potential of the outcomes, we evaluated the equality of health resource distribution based on the two-week illness mapping results and the health supply data.

## Methods

### Study area

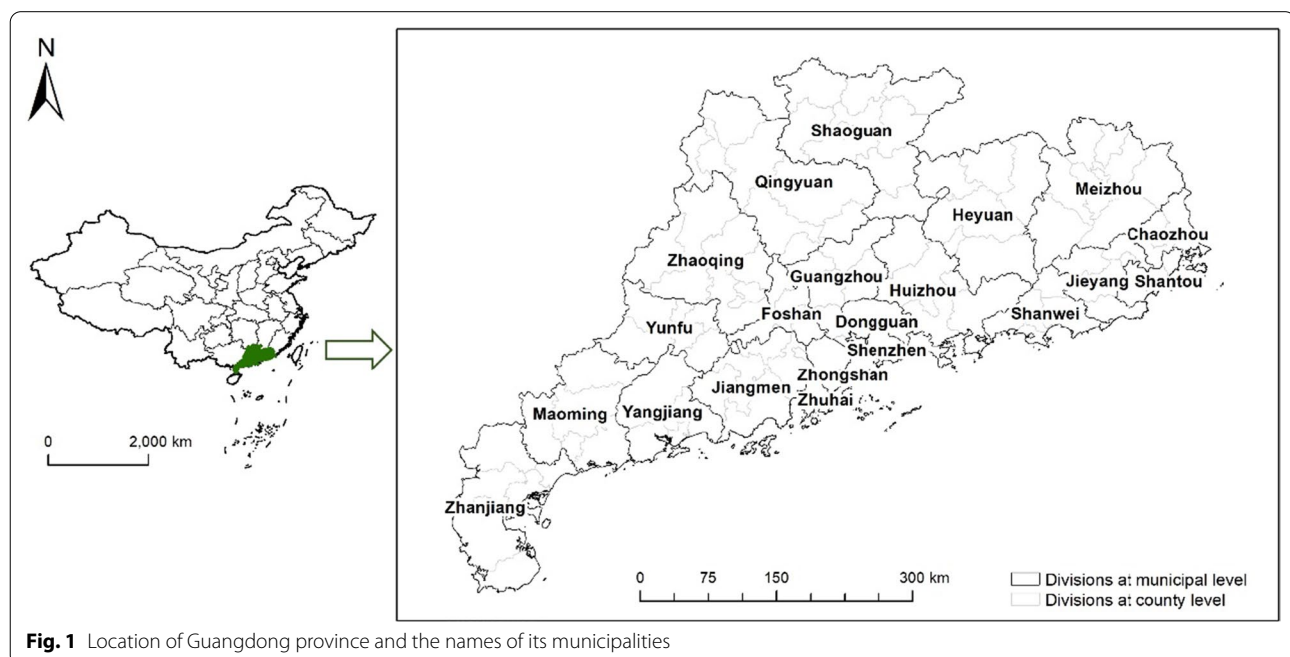
Guangdong province occupies a region of south China with a longitude ranging from 109° 39' to 117° 19' E and latitude from 20° 13' to 25° 31' N (Fig. 1), having a subtropical to tropical climate. The province is divided into 21 prefecture-level cities and further subdivided into 119 county-level divisions. In 2018, Guangdong had a population of 11.3 million and a gross domestic product (GDP) of 972.7 million yuan, making it the province with both the largest provincial population and economy in China [37]. The five most prevalent diseases in Guangdong province in 2013 were reported as hypertension, acute nasopharyngitis, diabetes, upper respiratory tract infection and gastroenteritis [38], and the raw prevalence are 6.09%, 2.15%, 1.27%, 1.86%, 0.78%, respectively, in the current study dataset, which contributes greatly to the two-week illness. Since 2005, Guangdong has faced the increasing pressure imposed by an aging population [39], bringing the age effect to forefront on illness and health services. The province shows large geographic heterogeneity in disease risk, demographic characteristics (e.g., age structure), socioeconomic characteristics (e.g., ethnic composition, income and urbanization level), environmental characteristics (e.g., temperature, landcover and

elevation), as well as health service provision [40]. To certain extent, it represents a typical large region to study the health service need.

### Data source

The data used in this study were obtained mainly from the fifth National Health Services Survey of Guangdong province, China. The original individual-level dataset is not publicly available due to the confidentiality required by the fifth National Health Services Survey. This large-scale survey was administered during September and October 2013 and was subdivided into three sub-surveys administered to household residents, medical staff and medical institutions. For the household survey, a multi-stage sampling design was applied to select representative participants randomly. Information was collected on various topics, including demographics; socioeconomic factors; the need, demand and utilization of health services; and the level of satisfaction of health services [38]. Qualified investigators visited the sampled households, inquired about all members of each household and completed the questionnaires [38]. Quality control measures were used throughout the investigation, and the survey was rated as being of good quality, with a Myer's index of 2.38 and Gini concentration ratio of 0.3048 [38].

The two-week illness status was used as the dependent variable in this study. Respondents were defined as having had an illness within the previous two-week period if they met one of the following three circumstances during the 14 days prior to the interview: (i)



**Fig. 1** Location of Guangdong province and the names of its municipalities

visiting a doctor because of illness or injury; (ii) receiving medical treatment because of illness or injury; or (iii) missing work or being bedridden because of illness or injury for at least 1 day [38]. Other demographic and socio-economic factors that might have influenced the two-week illness status (i.e., age, gender, ethnicity, marital status, education level, employment status, occupation, type of household registration and location of household registration) were also obtained from the survey and treated as independent variables in this study.

The coordinates of the survey villages/communities (Fig. 1) were extracted from BaiduMap and transformed to WGS84 using the `convBD2WGS` function in the `madlogos/aseshms` package on the R platform [41]. Other environmental and socio-economic factors with potential effects on the outcomes were extracted from different open-access sources (Additional file 1: Table S1, Figure S1). As communities/villages in China usually cover areas larger than 1 km<sup>2</sup> and many risk mapping studies usually took the spatial resolution of 5 × 5 km<sup>2</sup> [20, 33, 34], we adopted this resolution for this risk mapping study. We aligned the covariate data over a regular grid of 5 × 5 km<sup>2</sup> spatial resolution and further assigned to the survey villages/communities. For covariates with resolution higher than 5 × 5 km<sup>2</sup>, we linked each pixel of the grid with the aggregated value of the corresponding covariate within the square of 5 × 5 km<sup>2</sup> centered by the pixel. The arithmetic mean and the mode were used to calculate the aggregated value for continuous and categorical covariates, respectively. For covariates with resolution lower than 5 × 5 km<sup>2</sup>, we assigned each pixel of the 5 × 5 km<sup>2</sup> the value of the corresponding covariate closest to the pixel. Altogether, we considered 8 individual-level variables (age, gender, ethnicity, education level, marital status, occupation, type of household registration and location of household registration) and 19 location-level variables [living space per capita, proportion of no housing population, proportion of households with unimproved drinking water, proportion of households without sanitary toilets, family population, travel time to cities, GDP per capita, salary per capita, nighttime light, land-cover, land surface temperature (LST) in the daytime and at night, normalized difference vegetation index (NDVI), soil moisture, elevation, air temperature, precipitation, fire emissions indicators (total carbon content, TCC) and particulate matter < 2.5 μm in diameter (PM<sub>2.5</sub>)] as potential influencing factors in the subsequent geostatistical analysis. Furthermore, we used the numbers of health workers indicating the supply of health service, for the subsequent analysis of equality, which was obtained from the Guangdong Health Statistics Yearbook (2013) at county level [42] (Additional file 1: Figure S2).

### Statistical analysis

First, we conducted a univariate analysis to identify individual-level variables associated with the two-week illness status, and excluded those that did not differ significantly (significance level  $\alpha = 0.05$ ). Continuous variables were standardized, and the ratio variables were log transformed. To avoid collinearity, we dropped one continuous variable in each pair if the corresponding Pearson's correlation coefficient was > 0.8, and kept the other one which was more meaningful on the outcome variable or with better data quality. To identify the best functional forms (continuous or categorical) of continuous predictors, we converted them to categorical ones according to a preliminary exploratory graphical analysis. And we further built two univariate geostatistical models for each continuous variable, one with the corresponding variable as the only fixed effect variable in a continuous form, and the other with it in a categorical form (treated as dummy variables) [43, 44]. The functional form with the smallest log-scores was included in the multiple geostatistical models, and the best set of covariates was selected using the backward elimination approach [45].

Bayesian geostatistical binary regression models were developed to evaluate the correlations of potential influencing factors (covariates) with the two-week illness status. For the  $i^{th}$  individual belonging to the  $j^{th}$  location, we assumed an illness status  $Y_{ij}$  ( $Y_{ij} = 1$  indicating illness and  $Y_{ij} = 0$  indicating no illness within the previous two weeks) according to a Bernoulli distribution, where  $p_{ij}$  indicates the two-week illness prevalence rate:

$$Y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (1)$$

Here, the logit transformation of  $p_{ij}$  is assigned to a linear combination of effects:

$$\text{logit}(p_{ij}) = \beta_0 + X_{ij}^T \boldsymbol{\beta} + \varphi_{j(a)} + \omega_j \quad (2)$$

Here  $X_{ij}$ ,  $\beta_0$  and  $\boldsymbol{\beta}$  are denoted as the matrix of covariates, the intercept and the vector of regression coefficients, respectively. These make up the fixed effects (level-1) of the model. A spatial-age random effect  $\varphi_{j(a)}$  is introduced into the model to describe the age- and location-specific variations, with  $a$  the year of age. It is assumed to follow a zero-mean Gaussian process, whose covariance is defined as the Kronecker product of the spatial covariance matrix  $\mathbf{K}_s$  and the age covariance matrix  $\mathbf{K}_a$ :

$$\boldsymbol{\varphi} \sim GP(0, \mathbf{K}_s \otimes \mathbf{K}_a) \quad (3)$$

$\mathbf{K}_s$  is defined as the stationary Matérn covariance function, that is  $\mathbf{K}_s = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa \mathbf{D})^\nu K_\nu(\kappa \mathbf{D})$ , and  $\sigma^2 = \frac{1}{4\pi\kappa^{2\nu}\tau^2}$ . Here  $\mathbf{D}$  is the Euclidean distance



matrix,  $K_\nu$ , the modified Bessel function of the second kind with the smoothness parameter  $\nu$  fixed at 1,  $\kappa$  a scaling parameter for the range  $r = \sqrt{8\nu}/\kappa$ , and  $\tau$  the precision parameter for the spatial variance  $\sigma^2$ .  $\mathbf{K}_a$  is defined as autoregressive stochastic process with the first order (AR1). In order to reduce computational burden, the Gaussian Markov Random Fields (GMRF) are built on regular age knots, i.e.,  $\boldsymbol{\varphi} = (\varphi_{a=0}, \varphi_{a=10}, \dots, \varphi_{a=90}, \varphi_{a=\max(\text{age})})'$ , and the random fields of other ages are approximated by the projection of  $\boldsymbol{\varphi}$  using the B-spline basis function of degree two.  $\omega_j$  is the unstructured random effect, assumed to follow a zero-mean normal distribution:

$$\omega_j \sim \text{Normal}(0, \delta^2) \tag{4}$$

which describes the unexplained random variance separately from the variance explained by fixed effect predictors and spatial-age random effect.  $\varphi_{j(a)}$  and  $\omega_j$  make up the random effects (level-2) of the model.

The following less-informative priors were used for the parameters and hyperparameters:  $\beta_0, \beta_{ij}, \beta_j \sim \text{Normal}(0, 1000)$ ,  $\log(\tau) \sim \text{Normal}(1, 10)$ ,  $\log(\kappa) \sim \text{Normal}(1, 10)$ ,  $1/\delta^2 \sim \text{gamma}(1, 0.00005)$  and  $\log((1 + \rho)/(1 - \rho)) \sim N(0, 0.15)$ . Sensitivity analyses on priors of hyperparameters was done, by setting different priors for each hyperparameter and comparing the posterior estimates. The model was fitted using the integrated nested Laplace approximations (INLA) with the homonymous package on the R [46].

We calculated estimates over the grid of  $5 \times 5 \text{ km}^2$  resolution across Guangdong province, to produce the two-week illness prevalence maps at a high spatial resolution. Age-specific maps were developed first, and then a map of the overall two-week illness prevalence was produced by weighting the age-specific maps according to the population composition across the age groups. We also estimated the number of people who reported illness within the two-week period by multiplying the estimated two-week illness prevalence with the gridded population. The age- and population-adjusted two-week illness prevalence and the total number of people reporting illness within the two-week period were also estimated at the county- and municipality-levels and across the entire province, by combining the prevalence maps with demographic and administrative maps. In addition, we run separate models for the major types of diseases reported (i.e., endocrine and metabolic diseases, circulatory system diseases, respiratory diseases, and digestive system diseases), under the same modeling framework.

We applied fivefold out-of-sample cross-validation to evaluate the model performance. The observed and

estimated prevalence ( $\hat{\pi}_j$ ) in the validation set (with  $N$  locations) were compared. Indicators were calculated, namely, the mean error ( $ME = \frac{1}{N} \sum_{j=1}^N (\pi_j - \hat{\pi}_j)$ ), the mean absolute error ( $MAE = \frac{1}{N} \sum_{j=1}^N |\pi_j - \hat{\pi}_j|$ ), the proportion of locations with observed prevalence rates within the 95% Bayesian credible intervals (BCIs), and the area under the receiver operating characteristic (ROC) curve (AUC) [47].

The Lorenz curve and Gini coefficient were used to evaluate the equality of health resource distribution, based on the estimated number of people reporting illness (need) and the number of health workers (supply) at county level:

$$G = 1 - \sum_{K=1}^{n-1} (F_{K+1} - F_K)(\phi_{K+1} + \phi_K) \tag{5}$$

$$F_K = \sum_{k=1}^K \frac{m_k}{M} \tag{6}$$

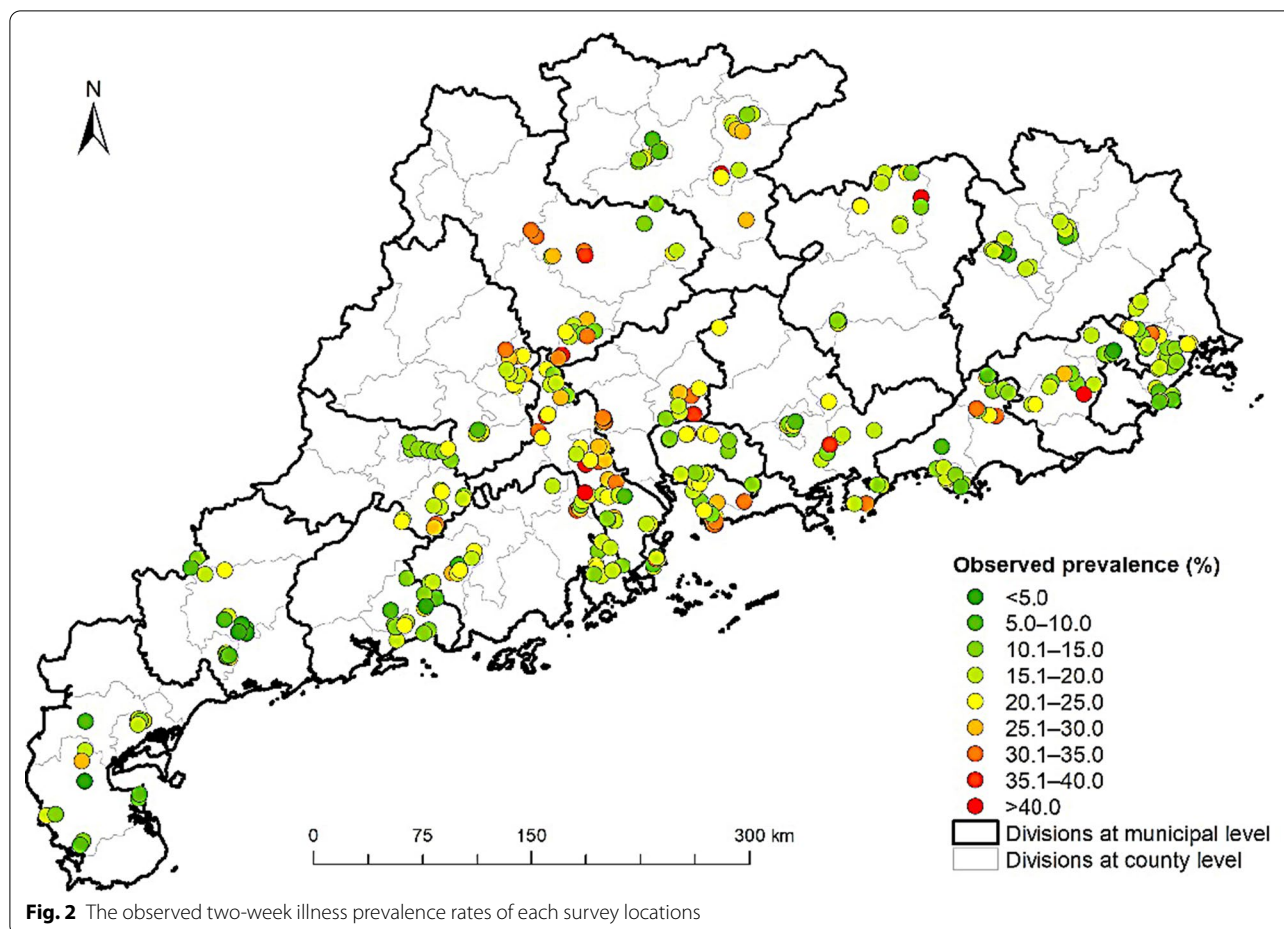
$$\phi_K = \sum_{k=1}^K \frac{h_k}{H} \tag{7}$$

Here  $m_k$  is denoted be the number of people reporting two-week illness at each county  $k$  (the sum of which is  $M$ ) and  $h_k$  the number of health workers at the county  $k$ . All the counties were ranked according to the supply-need ratio  $h_k/m_k$ . Next, the cumulative proportion of the patients up to county  $K$  ( $F_K$ ) and the cumulative proportion of health workers ( $\phi_K$ ) is calculated (define  $F_0 = \phi_0 = 0$ ). The Gini coefficient ( $G$ ) was obtained by formula (5), and the Lorenz curve was graphed with  $(F_K, \phi_K)$ . A map of the supply-need ratio was also produced to identify counties with relatively less health resources.

All statistical analyses were conducted using R software (version 3.6.1; The R Project for Statistical Computing, Vienna, Austria), and the maps were visualized using ArcGIS (version 10.2; ESRI, Redlands, CA, U.S.A.).

### Results

The survey investigated 81,859 people in 24,129 households from 400 villages/communities in 40 counties across Guangdong province. The raw two-week illness prevalence was 18.5% (number of individuals who reported illness within the two-week period divided by the total number of investigated individuals). The observed prevalence for each investigated location is shown in Fig. 2. The results of the univariate analysis of individual-level variables are summarized in Table 1. We found only two variables with a correlation coefficient larger than 0.8. These were the proportion of households with unimproved drinking water and the proportion of households without sanitary toilets. These two variables



came from the same data sources with similar data quality. Nevertheless, the latter variable may reflect more of the socio-economic status of households than the former one, as building sanitary toilets are usually decided by household itself, and improvement of drinking water are usually influenced by projects of local governments. Therefore, we kept the proportion of households without sanitary toilets for the subsequent analysis. 14 covariates were included in the final geostatistical model after variable selection (Table 2). People with education above the primary level had a lower illness risk than those with no education. Married people had lower risk than the unmarried ones, and people who were widowed had the highest risk. Furthermore, a higher risk was found among people who were unemployed or retired than those who were employed. The living space per capita and elevation was correlated positively with the two-week illness risk, whereas the NDVI was correlated negatively. People with higher salary per capita and locations with higher proportion of households without sanitary toilets showed high risk. Model validation revealed that our final model could correctly estimate 94.28% of the locations within

the 95% BCI, and the AUC was 0.775, indicating a reasonable predictive performance. However, the positive ME of 0.61% suggested that a slightly underestimated prevalence might exist. Sensitivity analysis shows that different priors for hyperparameters had little effect on the final outcomes (Additional file 1: Table S2, Figure S5–6).

The map of two-week illness prevalence (Fig. 3a) across all age groups showed notable spatial differences. The areas with the highest risk, denoted by an estimated median prevalence > 25.0%, were concentrated in the central (i.e., Pearl River Delta) and northern regions of the province, whereas many areas in southwestern and eastern regions showed the rates ranges from 10.0 to 20.0%. The highest estimated numbers of people reporting two-week illness were distributed in the Pearl River Delta and eastern coastal areas (> 10,000 people/25 km<sup>2</sup>) and the numbers in other areas were relatively low (mostly < 5000 people/25 km<sup>2</sup>) (Fig. 3c). In most areas, the standard deviation of the estimated error (Fig. 3b) ranged between 5.0 and 10.0%. Maps of random effect posteriors are shown in Additional file 1: Figure S3–4.

**Table 1** Univariate analysis of individual-level covariates

Factors	Reported two-week illness (n = 15,178)	No reported illness (n = 66,681)	p-value
Age (years)			
< 15	1353 (8.9%)	11,574 (17.4%)	< 0.001 <sup>#,*</sup>
15–29	898 (5.9%)	16,148 (24.2%)	
30–44	1738 (11.5%)	15,492 (23.2%)	
45–59	4262 (28.1%)	14,453 (21.7%)	
≥ 60	6927 (45.6%)	9014 (13.5%)	
Gender			
Male	7160 (47.2%)	34,427 (51.6%)	< 0.001 <sup>*</sup>
Female	8018 (52.8%)	32,254 (48.4%)	
Ethnicity			
Han	15,092 (99.4%)	66,265 (99.4%)	0.427
Others	86 (0.6%)	416 (0.6%)	
Education level			
No education	1914 (12.6%)	2811 (4.2%)	< 0.001 <sup>#,*</sup>
Primary	5120 (33.7%)	11,615 (17.4%)	
Junior	4000 (26.4%)	20,811 (31.2%)	
Senior	1992 (13.1%)	13,158 (19.7%)	
Bachelor or above	2152 (14.2%)	18,286 (27.4%)	
Marital status			
Unmarried	926 (6.1%)	13,118 (19.7%)	< 0.001 <sup>*</sup>
Married	10,758 (70.9%)	39,010 (58.5%)	
Widowed	2006 (13.2%)	2093 (3.1%)	
Divorced	101 (0.7%)	440 (0.7%)	
Other	1387 (9.1%)	12,020 (18.0%)	
Employment status			
Employed	5915 (39.0%)	36,876 (55.3%)	< 0.001 <sup>*</sup>
Unemployed	3052 (20.1%)	3749 (5.6%)	
Student	230 (1.5%)	4904 (7.4%)	
Retired	5981 (39.4%)	21,152 (31.7%)	
Occupation <sup>§</sup>			
Type 1	517 (3.4%)	1894 (2.8%)	< 0.001 <sup>*</sup>
Type 2	896 (5.9%)	4369 (6.6%)	
Type 3	625 (4.1%)	2217 (3.3%)	
Type 4	1452 (9.6%)	8555 (12.8%)	
Type 5	2633 (17.3%)	10,470 (15.7%)	
Type 6	734 (4.8%)	2693 (4.0%)	
Type 7	8321 (54.8%)	36,483 (54.7%)	
Type of household registration			
Rural	9446 (62.2%)	43,909 (65.8%)	< 0.001 <sup>*</sup>
Urban	5732 (37.8%)	22,772 (34.2%)	
Location of household registration			
In this county	14,199 (93.5%)	62,650 (94.0%)	0.048 <sup>*</sup>
Out of this county but in this province	397 (2.6%)	1613 (2.4%)	
Out of this province	542 (3.6%)	2186 (3.3%)	
Undetermined	40 (0.3%)	232 (0.3%)	

<sup>#</sup> Wilcoxon rank-sum test. All other analyses were conducted using the  $\chi^2$  test

<sup>\*</sup> Significant

<sup>§</sup> Occupation types classified in the survey: type 1, heads of organizations,

**Table 1** (continued)

enterprises or institutions; type 2, professionals or technicians; type 3, civil servants; type 4, business/service workers; type 5, agricultural, forestry, animal husbandry, fishery or water conservancy production workers; type 6, production and transportation equipment operators; and type 7, soldiers or others

The age-specific maps of the two-week illness prevalence (Fig. 4) had patterns similar to that of the map across all age groups with respect to spatial differences. High risk is shown in the central and northern regions, and lower risk in the southwestern and eastern regions. However, the prevalence rates differed significantly between age groups, and the age-prevalence relationships also differed across locations (Fig. 5). In most areas, the prevalence rates decrease from age 0 to age 20, and then increase gradually. The age-specific estimated numbers of people reporting illness also show spatial differences (Fig. 6), with the largest estimates in the Pearl River Delta and eastern coastal regions (> 100 people/25 km<sup>2</sup> in many areas).

The estimated county-level two-week illness prevalence rates ranged from 6.4 to 32.1%. Counties with prevalence rates > 30.0% included Shunde in the central region of the province, and Liannan and Lianzhou in the northern region (Additional file 1: Table S3). Counties with larger estimated numbers of people reporting an illness within the two-week period (> 1 million) included downtown Guangzhou and uptown Shenzhen in the central region. Municipalities with higher illness prevalence rates (> 20.0%) included Foshan, Guangzhou, Qingyuan and Shenzhen, and those with larger numbers of people reporting an illness within the two-week period (> 1 million) included Foshan, Guangzhou and Shenzhen (Additional file 1: Table S4). Overall, the estimated age- and population-adjusted two-week illness prevalence in Guangdong province in 2013 was 16.5% (95% BCI: 14.5–18.6), and the total number of people reporting an illness within the two-week period was 17.5 million (95% BCI: 15.5–19.8 million).

The Lorenz curve (Fig. 7a) and the Gini coefficient (resulted in 0.3526) showed a moderate level of inequality in health resource distribution. The supply-need ratio map called for more health resources in some counties of the northern part (Fig. 7b).

The mapping results of the major types of diseases reported show a high risk of endocrine and metabolic diseases risk in the central (i.e., Pearl River Delta) and eastern areas, a higher risk of circulatory system diseases around the central areas, a higher risk of respiratory diseases in the northern areas, and a higher risk of digestive system diseases in the central and northern regions of the study region (Additional file 1: Figure S7).

**Table 2** Posterior summaries of the model parameters

Parameters	Estimated median (95% BCI)
Gender (male) <sup>#</sup>	
Female	0.0380 (− 0.0043, 0.0803)
Education level (no education)	
Primary	0.0035 (− 0.0772, 0.0844)
Junior	− 0.0968 (− 0.1871, − 0.0063) <sup>*</sup>
Senior	− 0.1544 (− 0.2582, − 0.0506) <sup>*</sup>
Bachelor or above	− 0.1508 (− 0.2860, − 0.0160) <sup>*</sup>
Marital status (unmarried)	
Married	− 0.1253 (− 0.2346, − 0.0152) <sup>*</sup>
Widowed	0.1518 (0.0177, 0.2864) <sup>*</sup>
Divorced	− 0.0993 (− 0.3620, 0.1551)
Other	− 0.6890 (− 0.9716, − 0.4076) <sup>*</sup>
Employment status (employed)	
Unemployed	0.5886 (0.5009, 0.6763) <sup>*</sup>
Student	0.0153 (− 0.1935, 0.2224)
Retired	0.5290 (0.4504, 0.6078) <sup>*</sup>
Occupation (type 1) <sup>5</sup>	
Type 2	0.0228 (− 0.1183, 0.1644)
Type 3	0.1444 (− 0.0086, 0.2976)
Type 4	0.0410 (− 0.0934, 0.1763)
Type 5	0.0911 (− 0.0457, 0.2290)
Type 6	0.0648 (− 0.0888, 0.2187)
Type 7	− 0.0328 (− 0.1627, 0.0982)
Type of household registration (rural)	
Urban	0.0453 (− 0.0356, 0.1259)
Location of household registration (in this county)	
Out of this county but in this province	0.0111 (− 0.1318, 0.1519)
Out of this province	− 0.0117 (− 0.1530, 0.1283)
Undetermined	− 0.1043 (− 0.4821, 0.2466)
Living space per capita	0.2439 (0.0870, 0.4007) <sup>*</sup>
No-housing ratio (< 0.9%)	
0.9–2.4	0.2979 (− 0.0153, 0.6110)
2.4–4.6	0.0639 (− 0.3150, 0.4426)
≥ 4.6	− 0.0220 (− 0.3756, 0.3312)
Proportion of households without sanitary toilets (< 0.3%)	
0.3–0.7	0.0228 (− 0.2849, 0.3302)
0.7–1.6	0.2597 (− 0.0484, 0.5676)
≥ 1.6	0.4634 (0.0331, 0.8935) <sup>*</sup>
Travel time to cities	− 0.0767 (− 0.1687, 0.0151)
Salary per capita (< ¥ 4.0 × 10 <sup>4</sup> /year)	
4.0–4.2	0.0511 (− 0.3374, 0.4393)
4.2–4.7	0.1760 (− 0.2763, 0.6281)
≥ 4.7	0.6969 (0.3395, 1.0540) <sup>*</sup>
NDVI <sup>6</sup>	− 0.1074 (− 0.2088, − 0.0061) <sup>*</sup>
Elevation	0.1330 (0.0242, 0.2416) <sup>*</sup>
Range (km)	219.4650 (172.1405, 280.8523)
Spatial variance	1.5884 (1.0445, 2.4707)
Age correlation coefficient $\rho$	0.8409 (0.7579, 0.8999)
Non-spatial variance $\delta^2$	0.1983 (0.1623, 0.2422)

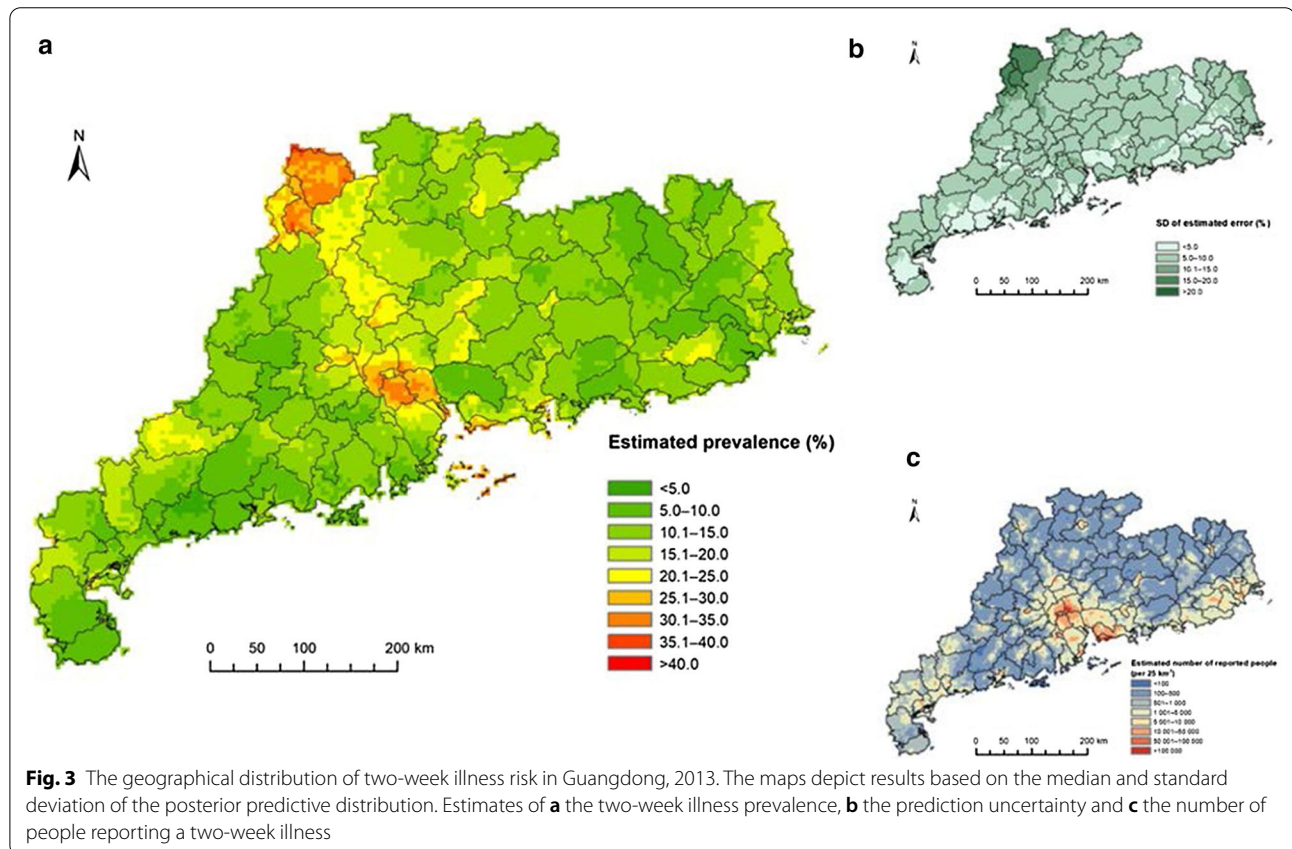


**Table 2** (continued)

# Baseline reference are reported in brackets

\* Significant, determined by no inclusion of zero in the 95%BCIs

§ Occupation types classified in the survey: type 1, heads of organizations, enterprises or institutions; type 2, professionals or technicians; type 3, civil servants; type 4, business/service workers; type 5, agricultural, forestry, animal husbandry, fishery or water conservancy production workers; type 6, production and transportation equipment operators; and type 7, soldiers or others. §Normalized difference vegetation index



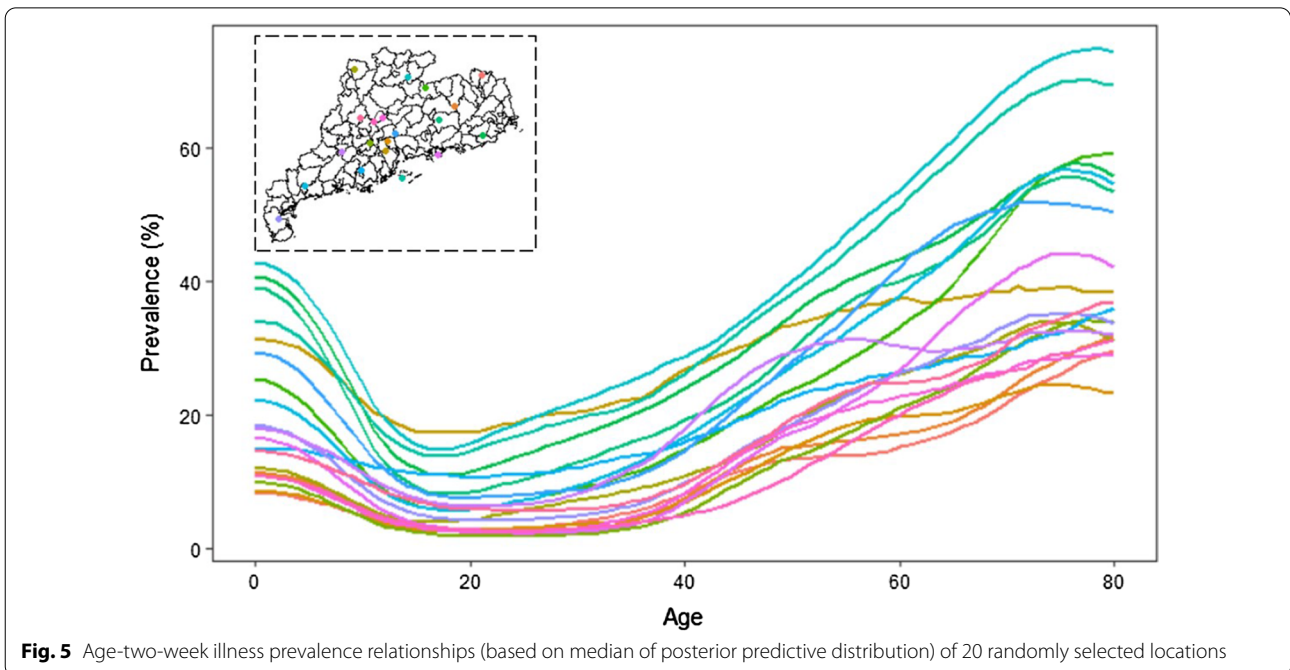
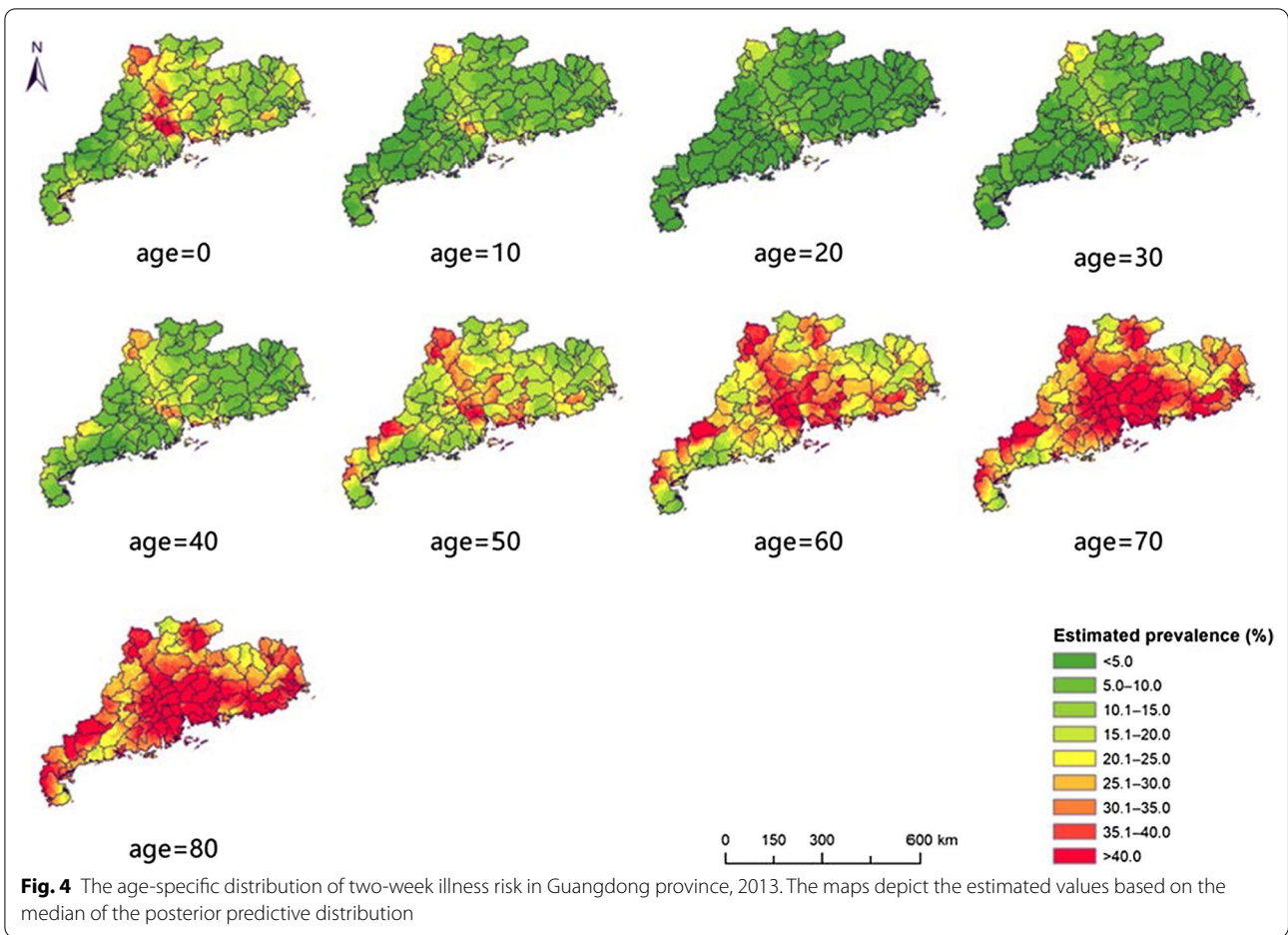
**Fig. 3** The geographical distribution of two-week illness risk in Guangdong, 2013. The maps depict results based on the median and standard deviation of the posterior predictive distribution. Estimates of **a** the two-week illness prevalence, **b** the prediction uncertainty and **c** the number of people reporting a two-week illness

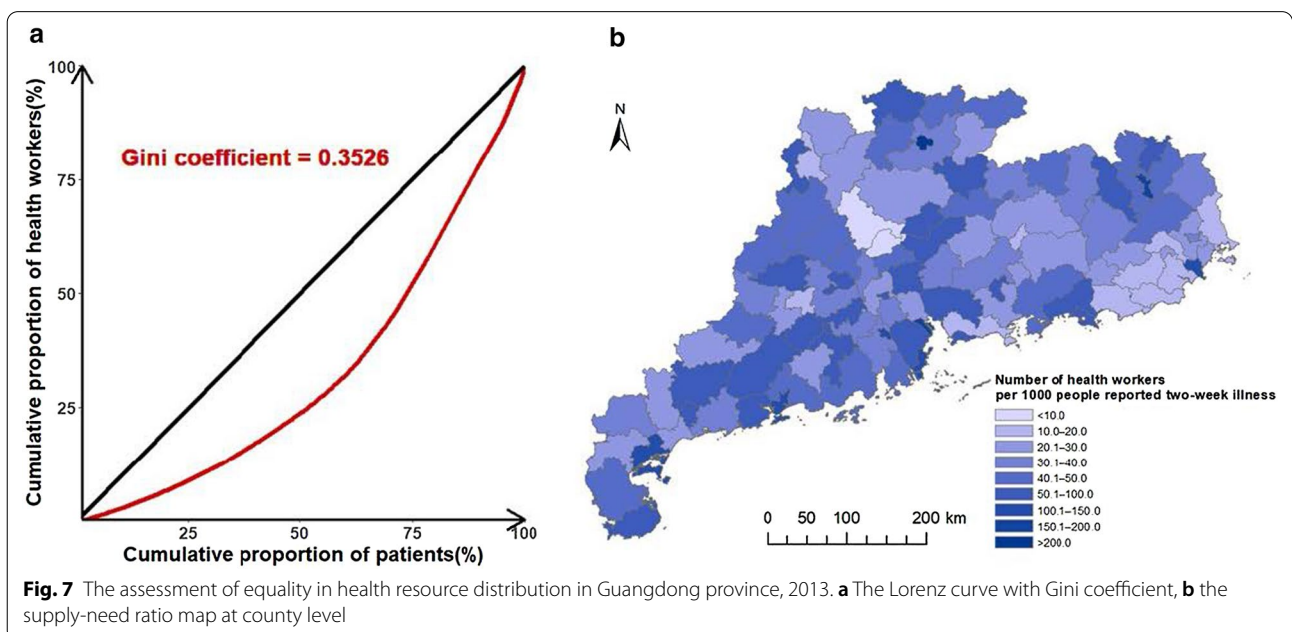
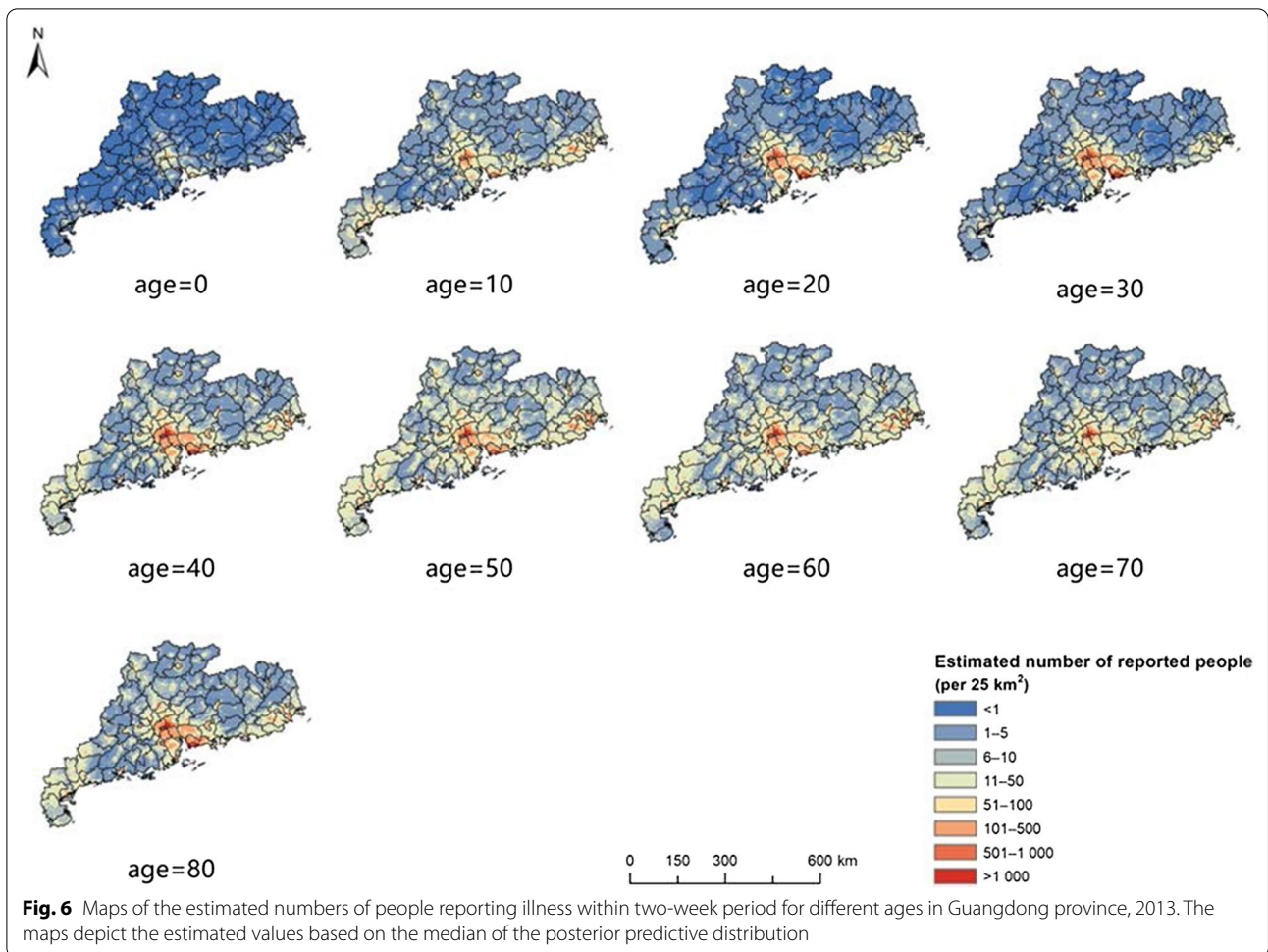
**Discussion**

This study provided age-specific, high-spatial-resolution estimates of the two-week illness prevalence, an important indicator of the health levels of residents and health service needs, in Guangdong province, China. The introduction of spatial-age random effect in the Bayesian geostatistical model allows to study the specific age-prevalence relationships among different areas. Risk mapping of illness at high resolution provides information on how the risk distributed within administrative divisions, thus is useful for local policy makers. And maps at a lower resolution (e.g., county- or city-level) can be further obtained from high-resolution maps weighted by age and population, thus improving the accuracy compared to raw observed rates, and ready for use by policy makers of different levels. Moreover, information of

two-week illness serves as indicator for health need, thus can be further adopted to evaluate the equality of health resource distribution. The versatile and efficient methodology can be applied to other regions if appropriate survey data are available.

To overcome the “big n problem”, we used INLA instead of the MCMC approaches for Bayesian inferences, for its fast speed and good accuracy [48]. The additional advantages include the allowance for greater automation and parallel implementation [46]. Though the application of INLA is mostly restricted to class of latent Gaussian models [48], it’s ready to use widely as most common statistical models belong to this class. Furthermore, we introduced spatial-age structure random effect to assess the space-age interaction. Separated by year, number of age groups is large. To decrease the computation burden,







we built the GMRFs on equally spaced age knots and approximate the latent fields of other age groups by projection of the knot fields. It is a trade-off between calculation accuracy and calculation speed. Nevertheless, the model performance is good.

Our model was able to correctly estimate 94.28% of the locations within 95% BCI, and had an AUC of 0.775, suggesting a reasonable model performance. On one hand, the spatial-age random effect introduced in the model allows to describe age- and location-specific variations across the study region. Thus, we are able to not only acquire the age-specific and the overall prevalence rates, but also estimate the age-prevalence relationships among different areas. Compared with common surveillance reports, national health surveys usually contain specific age information on participants, which provides the chance to improve the estimation by making full use of the individual age information. Our model is ready to apply on risk estimates with similar data structure. On the other hand, we made use of both various data from the national health survey and data from other open-access databases, considering both individual- and location-level potential predictors, which may improve the prediction, compared to models only with location-specific predictors or only based on data from single source.

Our final geostatistical model identified that socio-economic predictors, such as education level, marital status, employment status, proportion of households without sanitary toilets, living space per capita and salary per capita were significantly related to the two-week illness risk. Previous studies identified low education level as a risk factor for cardiovascular diseases and metabolic syndrome [49, 50], the major endemic diseases in Guangdong province [38]. Studies show that widowhood a risk factor for dementia and Alzheimer's disease [51, 52], unemployment could increase the risk of common mental disorders [53], and retirement may increase the risk of mental disorders and chronic diseases [54, 55]. Lower proportion of households without sanitary toilets, higher living place per capita and higher salary per capita reflects partially a better socio-economic status of households, people with which may be more sensitive to illness and likely to report the status [56, 57]. Studies also found income associated with metabolic syndrome related disorders such as obesity and type 2 diabetes [58, 59]. Taken together, these previous findings might support the corresponding associations of socio-economic factors identified in this study with the two-week illness risk. We additionally found environmental predictors, that is elevation and NDVI, with significant relationships with the two-week illness prevalence. These factors may interact health status in more complex ways, directly or indirectly [60, 61]. Our study did not identify gender, occupation,

type of household registration or location of household registration as significant risk factors in the final model, and this might be partially due to the existing correlations between the covariates (Additional file 1: Table S5).

The maps of the estimated age-specific and overall two-week illness prevalence distributions showed a greater concentration of illness prevalence in the central (the Pearl River Delta) and northern regions of the province, suggesting relatively greater need of health service in these areas. The province government should fully consider the various of health need across the study region, together with the current health supply, for a more reasonable allocation of new health resources. Furthermore, as the risk maps at high spatial resolutions show the difference of health need within administrative divisions, local governments should fully consider such difference for future planning of health supply distribution within the divisions. Reasons for geographical differences in two-week illness prevalence might result from large diversity of socio-economy, behaviors and climate across the study region. As higher risk of certain chronic diseases (e.g., metabolic syndrome) was found associated with higher income [58, 59], the higher prevalence of these diseases in the Pearl River Delta, the most developed region of the province, may contribute greatly to the high risk of the two-week illness in the region. In the northern region, the cold and humid climate and poor economic conditions may encourage people to burn wet firewood for cooking, which could increase the risk of smog-related diseases [62]. Risk maps of the typical types of diseases (Additional file 1: Figure S7) were consistent with the findings. Nevertheless, more detailed investigations are needed, particularly in vulnerable areas, to further clarify the reasons of geographical differences and to guide the specific disease control management.

A study showed that in 2013, the Gini coefficient of health worker number versus the population in Guangdong province was 0.1800, drawing the conclusion of absolute equality in health resource distribution in Guangdong [63]. In contrast, our study worked out a Gini coefficient of 0.3526, concluding the moderate level of inequality in health resource distribution. Since both studies use the same indicator (i.e., number of health workers) with the same data source for the supply [42], but different indicators (i.e., population and the estimated number of people reporting two-week illness in the previous and the current studies, respectively) for the need, the discrepancy conclusions reveal the decisive effect of the selection of need indicators. The moderate level of inequality resulted in our study calls for further optimization of health resource allocation in Guangdong, and the map of supply-and-need ratio may provide useful supporting information.

The key data used in this study were obtained through the fifth National Health Services Survey in 2013, which covered 31 provinces, 156 counties, 93,613 families and 273,688 residents in China [64]. Apart from the two-week illness status, the survey collected information on a wide range of indicators, such as prevalence of different types of diseases, two-week hospital visitation rate, annual hospitalization rate and travel time to the nearest medical institution [64]. Thus, it provides a wealth of information about the needs, demands and utilization of health services nationwide, which enables follow-up studies focusing on mapping of other indicators of interest. Running separate models of acute and chronic diseases would improve the analysis considering confounders like age and help inform needs for different illness types. However, based on the questionnaire, it is difficult to make a complete distinction between those who had acute and chronic diseases. As a complement, we run separate models of the major types of diseases reported (i.e., endocrine and metabolic diseases, circulatory system diseases, respiratory diseases, and digestive system diseases), to show the potential of the methodology to study different types of illness. The risk mapping of different types of diseases informs the disease-specific health needs (Additional file 1: Figure S7).

Beyond China, many other countries have conducted health surveys that consider health service needs, demands and utilization [14, 17, 65, 66]. These surveys usually include data on the illness prevalence in a limited number of sampled locations. Accordingly, analogous studies could be conducted based on the Bayesian geostatistical modeling framework, and these analyses would provide high-resolution, age-specific estimates of the illness prevalence throughout any region of interest. We provided the corresponding R codes as Additional file 2.

Our study had several limitations. First, the self-reported illness status in the two weeks prior to the survey might have been subject to both reporting bias and recall bias. Both types of bias are inevitable because of the pre-specified survey design and outcome indicators. However, a previous study noted that self-reported data can be reasonably reliable [67]. Second, we intended to consider as many potential influencing factors as possible. To meet this goal, however, we sacrificed some of the spatial resolution. For example, some of the covariates were collected merely at the county level, and their inclusion might reduce the credibility of the estimations. Nevertheless, the model validation showed a reasonable predictive performance. Third, when this article was written, the sixth National Health Services Survey had been completed, but we were unable to obtain data at the village/community level. We will update our study accordingly when these data are available.

## Conclusions

In conclusion, we developed a Bayesian geostatistical modeling framework with spatial-age structured effect to analyze health services survey data. High-resolution and age-specific maps of the two-week illness prevalence rate and the estimated number of people reporting illness within a two-week period are provided in Guangdong province, China. The versatile and efficient methodology can be applied in any regions where the appropriate survey data are available, and results will help to support plans for health resource allocation.

## Abbreviations

GDP: Gross domestic product; LST: Land surface temperature; NDVI: Normalized difference vegetation index; TCC: Total carbon content; PM<sub>2.5</sub>: Particulate matter < 2.5 µm in diameter; NCEI: National Centers for Environmental Information; MODIS: Moderate resolution imaging spectroradiometer; PSL: Physical Sciences Laboratory; SRTM: Shuttle Radar Topographic Mission; REDCP: Resource and Environment Data Cloud Platform; SEDAC: Socioeconomic Data and Applications Center; ENVF: Environmental Central Facility; GMRF: Gaussian Markov Random Fields; INLA: Integrated nested Laplace approximations; BCI: Bayesian credible interval; ME: Mean error; MAE: Mean absolute error; AUC: Area under the curve; ROC: Receiver operating characteristic.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12942-021-00273-1>.

**Additional file 1.** Supplementary materials for the methods, results and discussion.

**Additional file 2.** The R codes of the study.

## Acknowledgements

Not applicable.

## Authors' contributions

Conceptualization, CW, YL and YH; Data curation, CW, XH, LF, LC, WH and YH; Methodology, CW and YL; Writing—original draft, CW and YL; Writing—review and editing, XH, LF, LC, WH and YH. All authors read and approved the final manuscript.

## Funding

This research was funded by grants from the following agencies: China Medical Board (Grant ID 17-274), National Natural Science Foundation of China (Grant IDs 82073665 and 81703320) and Sanming Project of Medicine in Shenzhen (Grant ID SZSM201803061).

## Availability of data and materials

The original individual-level dataset of the health services survey is not publicly available due to the confidentiality required by the fifth National Health Services Survey. All other data are available from the open-access databases.

## Declarations

### Ethics approval and consent to participate

All subjects provided informed consent for inclusion prior to participating in the fifth National Health Services Survey in Guangdong province. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Medical Ethics Committee of the School of Public Health, Sun Yat-sen University (No. 130 in 2020).



**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>School of Public Health, Sun Yat-sen University, Guangzhou 510080, China.

<sup>2</sup>Department of Statistics, Government Affairs Service Center of Health Commission of Guangdong Province, Guangzhou 510060, China.

Received: 21 October 2020 Accepted: 23 April 2021

Published online: 03 May 2021

**References**

- Papanicolas I, Woskie LR, Jha AK. Health care spending in the United States and other high-income countries. *J Am Med Assoc*. 2018;319(10):1024–39.
- Barrera-Algarin E, Estepa-Maestre F, Sarasola-Sanchez-Serrano JL, Vallejo-Andrada A. COVID-19, neoliberalism and health systems in 30 European countries: relationship to decesses. *Rev Esp Salud Public*. 2020;94:e202010140.
- Vidondo B, Oberrreich J, Brockmann SO, Duerr HP, Schwehm M, Eichner M. Effects of interventions on the demand for hospital services in an influenza pandemic: a sensitivity analysis. *Swiss Med Wkly*. 2009;139(35–36):505–10.
- Zhang XL, Yu Y, Xiong F, Luo L. Prediction of daily blood sampling room visits based on ARIMA and SES Model. *Comput Math Method M*. 2020;2020:1720134.
- Higginson I, Whyatt J, Silvester K. Demand and capacity planning in the emergency department: how to do it. *Emerg Med J*. 2011;28(2):128–35.
- Wild C, Narath M. Evaluating and planning ICUs: methods and approaches to differentiate between need and demand. *Health Policy*. 2005;71(3):289–301.
- Elorza ME, Moscoso NS, Blanco AM, Gentili JO. Estimating need, demand and supply in primary health care services: a local application in Argentina. *MEDICC Rev*. 2018;20(3):36–44.
- Petrelli A, Picariello R, Costa G. Toward a needs based mechanism for capitation purposes in Italy: the role of socioeconomic level in explaining differences in the use of health services. *Int J Health Care Fi*. 2010;10(1):29–42.
- Li YN, Nong DX, Wei B, Feng QM, Luo HY. The impact of predisposing, enabling, and need factors in utilization of health services among rural residents in Guangxi, China. *BMC Health Serv Res*. 2016;16(1):592.
- Lu CL, Liu K, Li LL, Yang YH. Sensitivity of measuring the progress in financial risk protection to survey design and its socioeconomic and demographic determinants: a case study in Rwanda. *Soc Sci Med*. 2017;178:11–8.
- Ryan T, Ingleton C, Gardiner C, Parker C, Gott M, Noble B. Symptom burden, palliative care need and predictors of physical and psychological discomfort in two UK hospitals. *Bmc Palliat Care*. 2013;12:11.
- WHO. Manual for the Household Survey to Measure Access and Use of Medicines 2008. [https://www.who.int/medicines/areas/coordination/household\\_manual\\_february\\_2008.pdf](https://www.who.int/medicines/areas/coordination/household_manual_february_2008.pdf).
- WHO. WORLD HEALTH SURVEY GUIDE TO ADMINISTRATION AND QUESTION BY QUESTION SPECIFICATIONS-Short Questionnaire 2002. <https://www.who.int/healthinfo/survey/whshortversionguide.pdf?ua=1>.
- Benova L, Campbell OMR, Ploubidis GB. Socio-economic inequalities in curative health-seeking for children in Egypt: analysis of the 2008 Demographic and Health Survey. *BMC Health Serv Res*. 2015;15:482.
- Liu Y, Collins C, Wang KS, Xie X, Bie RH. The prevalence and trend of depression among veterans in the United States. *J Affect Disord*. 2019;245:724–7.
- Taneja S, Dalpath S, Bhandari N, Kaur J, Mazumder S, Chowdhury R, et al. Operationalising integrated community case management of childhood illnesses by community health workers in rural Haryana. *Acta Paediatr*. 2018;107:80–8.
- Srinivasan R, Mohan VR, Venugopal S, Kang G. Utilization of preventive and curative services in five rural blocks in Vellore. *India Indian Pediatr*. 2017;54(9):777–8.
- Abegunde D, Orobato N, Bassi A, Oguntunde O, Bamidele M, Abdulkrim M, et al. The impact of integrated community case management of childhood diseases interventions to prevent malaria fever in children less than five years old in Bauchi State of Nigeria. *PLoS ONE*. 2016;11(2):e0148586.
- Gomez-Rubio V, Best N, Richardson S, Li G, Clarke P. Bayesian Statistics Small Area Estimation. [Technical Report]. In press 2010.
- Lai YS, Zhou XN, Pan ZH, Utzinger J, Vounatsou P. Risk mapping of clonorchiasis in the People's Republic of China: A systematic review and Bayesian geostatistical analysis. *PLoS Negl Trop Dis*. 2017;11(3):e0005239.
- Fowkes FGR, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *Lancet*. 2013;382(9901):1329–40.
- Dehlin M, Jacobsson L, Roddy E. Global epidemiology of gout: prevalence, incidence, treatment patterns and risk factors. *Nat Rev Rheumatol*. 2020;16(7):380–90.
- Schultz WM, Hayek SS, Tahhan AS, Ko YA, Sandesara P, Awad M, et al. Marital status and outcomes in patients with cardiovascular disease. *J Am Heart Assoc*. 2017;6(12):e005890.
- Schultz WM, Kelli HM, Lisko JC, Varghese T, Shen J, Sandesara P, et al. Socioeconomic status and cardiovascular outcomes challenges and interventions. *Circulation*. 2018;137(20):2166–78.
- Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet*. 2007;370(9589):765–73.
- Gorrell S, Trainor C, Le Grange D. The impact of urbanization on risk for eating disorders. *Curr Opin Psychiatr*. 2019;32(3):242–7.
- Hansel NN, McCormack MC, Kim V. The effects of air pollution and temperature on COPD. *COPD*. 2016;13(3):372–9.
- Soneja S, Jiang CS, Upperman CR, Murtugudde R, Mitchell CS, Blythe D, et al. Extreme precipitation events and increased risk of campylobacteriosis in Maryland, USA. *Environ Res*. 2016;149:216–21.
- Messier KP, Jackson LE, White JL, Hilborn ED. Landscape risk factors for Lyme disease in the eastern broadleaf forest province of the Hudson River valley and the effect of explanatory data classification resolution. *Spat Spatio-Temporal*. 2015;12:9–17.
- Zhang LF, Wang ZW, Chen YD, Wang X, Chen Z, Feng B, et al. Prevalence and risk factors associated with chronic kidney disease in adults living in 3 different altitude regions in the Tibetan Plateau. *Clin Chim Acta*. 2018;481:212–7.
- Dong X, Liu L, Cao S, Yang H, Song F, Yang C, et al. Focus on vulnerable populations and promoting equity in health service utilization—an analysis of visitor characteristics and service utilization of the Chinese community health service. *BMC Public Health*. 2014;14:503.
- Uwiringiyimana V, Veldkamp A, Amer S. Stunting spatial pattern in Rwanda: an examination of the demographic, socio-economic and environmental determinants. *Geospat Health*. 2019;14(2):329–39.
- Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Jessica C, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine*. 2018;36(12):1583–91.
- Local Burden of Disease Neglected Tropical Diseases. The global distribution of lymphatic filariasis, 2008–18: a geospatial analysis. *Lancet Glob Health*. 2020;8(9):e1186.
- Tian DP, Sun L, Zhang LL, Zhang W, Li L, et al. Large urban-rural disparity in the severity of two-week illness: updated results based on the first health service survey of Hunan Province, China. *Int J Equity Health*. 2016;15:37.
- Parsaeian M, Khaledi MJ, Farzadfar F, Mahdavi M, Zeraati H, Mahmoudi M, et al. Spatio-temporal analysis of misaligned burden of disease data using a geo-statistical approach. *Stat Med*. 2021;40(4):1021–33.
- National Bureau of Statistics of China. National Data. <http://data.stats.gov.cn/easyquery.htm?cn=E0103>.
- Administrative Service Center of Guangdong Health Department, School of Public Health SY-sU. An analysis report of National Health Services Survey in Guangdong province, 2013. Guangzhou: Yangcheng Evening News Press; 2016. p. 513.
- Wang T. Research on the development trend and countermeasures of population aging in Guangdong. *Prod Res*. 2011;8:134–5.

40. Xiu-fang C, Li Z. An analysis of the difference between areas in Guangdong Province in the allocation of Public Health Expenditure and Resources. *J Huizhou Univ*. 2012;32(2):49–53.
41. convBD2WGS: Transform BD-09 (Baidu) coordinates to WGS-84 (Global Coord). <https://rdr.io/github/madlogos/recharts2/man/convBD2WGS.html>.
42. Guangdong Provincial Health and Family Planning Commission. Guangdong Health Statistics Yearbook (2013). Guangzhou: Huacheng Press; 2013.
43. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc*. 2007;102(477):359–78.
44. Held L, Schrödle B, Rue H. Posterior and cross-validated predictive checks: a comparison of MCMC and INLA. In: Kneib T, Tutz G, editors. *Statistical modelling and regression structures*. Heidelberg: Physica-Verlag; 2010.
45. ME K. Selection of the Best Regression Equation by sorting out Variables <http://www.angelfire.com/ab5/get5/selreg.pdf>.
46. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc B*. 2009;71:319–92.
47. Brooker S, Hay SI, Issae W, Hall A, Kihamia CM, Lwambo NJ, et al. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Trop Med Int Health*. 2001;6(12):998–1007.
48. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: a review. *Annu Rev Stat Appl*. 2017;4:395–421.
49. Kubota Y, Heiss G, MacLehose RF, Roetker NS, Folsom AR. Association of educational attainment with lifetime risk of cardiovascular disease: The Atherosclerosis Risk in Communities Study. *JAMA Intern Med*. 2017;177(8):1165–72.
50. Kim I, Song YM, Ko H, Sung J, Lee K, Shin J, et al. Educational disparities in risk for metabolic syndrome. *Metab Syndr Relat Disord*. 2018;16(8):416–24.
51. Gerritsen L, Wang HX, Reynolds CA, Fratiglioni L, Gatz M, Pedersen NL. Influence of negative life events and widowhood on risk for dementia. *Am J Geriatr Psychiatry*. 2017;25(7):766–78.
52. Hatch DJ, Schwartz S, Norton MC. Depression and antidepressant use moderate association between widowhood and Alzheimer's disease. *Int J Geriatr Psychiatry*. 2015;30(3):292–9.
53. Fu TS, Lee CS, Gunnell D, Lee WC, Cheng AT. Changing trends in the prevalence of common mental disorders in Taiwan: a 20-year repeated cross-sectional survey. *Lancet*. 2013;381(9862):235–41.
54. Heller-Sahlgren G. Retirement blues. *J Health Econ*. 2017;54:66–78.
55. Li Ranzi T, d'Errico A, Costa G. Association between chronic morbidity and early retirement in Italy. *Int Arch Occup Environ Health*. 2013;86(3):295–303.
56. Cai L, Dong J, Shu ZK, Lu YC, Tao J. Socioeconomic differences in diabetes prevalence, awareness, and treatment in rural southwest China. *Trop Med Int Health*. 2011;16(9):1070–6.
57. Porapakkham Y, Pattaraarchachai J, Aekplakorn W. Prevalence, awareness, treatment and control of hypertension and diabetes mellitus among the elderly: the 2004 National Health Examination Survey III. *Thailand Singap Med J*. 2008;49(11):868–73.
58. Dong Y, Jan C, Ma Y, Dong B, Zou Z, Yang Y, et al. Economic development and the nutritional status of Chinese school-aged children and adolescents from 1995 to 2014: an analysis of five successive national surveys. *Lancet Diabetes Endocrinol*. 2019;7(4):288–99.
59. Meo SA, Usmani AM, Qalbani E. Prevalence of type 2 diabetes in the Arab world: impact of GDP and energy consumption. *Eur Rev Med Pharmacol Sci*. 2017;21(6):1303–12.
60. Yang BY, Liu KK, Markevych I, Knibbs LD, Bloom MS, Dharmage SC, et al. Association between residential greenness and metabolic syndrome in Chinese adults. *Environ Int*. 2020;135:105388.
61. Tischer C, Gascon M, Fernández-Somoano A, Tardón A, Lertxundi Materola A, Ibarluzea J, et al. Urban green and grey space in relation to respiratory health in children. *Eur Respir J*. 2017;49(6):1502112.
62. Wang X, Zhou Y, Zeng X, Liu S, Qiu R, Xie J, et al. Study on the prevalence rate of chronic obstructive pulmonary disease in northern part of Guangdong province. *Chin J Epidemiol*. 2005;26(3):63–5.
63. Li X, Zhang Q. Analysis on the equity of health resources allocation based on Gini coefficient and index of dissimilarity in Guangdong. *Modern Prev Med*. 2019;46(4):658–86.
64. Center for Health Statistics and Information. An Analysis Report of National Health Services Survey in China, 2013. [http://www.nhc.gov.cn/mohwsbwstjxxzx/s8211/new\\_list.shtml](http://www.nhc.gov.cn/mohwsbwstjxxzx/s8211/new_list.shtml).
65. Schiller JS, Lucas JW, Peregoy JA. Summary health statistics for U.S. adults: National Health Interview Survey, 2011. *Vital and health statistics Series 10. Data Natl Health Surv*. 2012;256:1–218.
66. Ghandour RM, Jones JR, Lebrun-Harris LA, Minnaert J, Blumberg SJ, Fields J, et al. The design and implementation of the 2016 National Survey of Children's Health. *Matern Child Health J*. 2018;22(8):1093–102.
67. Lo NC, Heft-Neal S, Coulibaly JT, Leonard L, Bendavid E, Addiss DG. State of deworming coverage and equity in low-income and middle-income countries using household health surveys: a spatiotemporal cross-sectional study. *Lancet Glob Health*. 2019;7(11):e1511–20.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

