

## Article

# Risk Analysis of Thyroid Cancer in China: A Spatial Analysis

Yu Wang<sup>1</sup>, Wenhui Wang<sup>1</sup>, Peng Li<sup>1</sup>, Xin Qi<sup>1,\*</sup>  and Wenbiao Hu<sup>2</sup><sup>1</sup> School of Public Health, Xi'an Jiaotong University, Xi'an 710061, China<sup>2</sup> School of Public Health, Queensland University of Technology, Brisbane, QLD 4059, Australia

\* Correspondence: xin.qi@mail.xjtu.edu.cn; Tel.:86-18829447238

**Abstract:** Thyroid cancer (TC) is the fastest growing cancer in China and has lots of influencing factors which can be intervened to reduce its incidence. In this article, we aimed to identify the risk factors of TC. The regional TC data in 2016 were obtained from the China Cancer Registry Annual Report published by the National Cancer Center (NCC). Univariate correlation analysis and generalized linear Poisson regression analysis were used to determine risk factors for morbidity of TC from the provincial and prefecture levels. High urbanization rate (UR) (RR = 1.109, 95%CI: 1.084, 1.135), high GDP per capita (PGDP) (RR = 1.013, 95%CI: 1.007, 1.018), high aquatic products (RR = 1.047, 95%CI: 1.020, 1.075) and dry and fresh fruit consumption (RR = 1.024, 95%CI: 1.007, 1.040) can increase TC incidence. Therefore, high PGDP, high UR, high aquatic products and dry and fresh fruit consumption were all risk factors for TC incidence. Our results may be helpful for providing analytical ideas and methodological references for the regionalized prevention and control of TC in a targeted manner.

**Keywords:** thyroid cancer; spatial distribution; risk factors; ecological study



**Citation:** Wang, Y.; Wang, W.; Li, P.; Qi, X.; Hu, W. Risk Analysis of Thyroid Cancer in China: A Spatial Analysis. *Geographies* **2022**, *2*, 577–592. <https://doi.org/10.3390/geographies2040035>

Academic Editor: Stephen A. Matthews

Received: 21 August 2022

Accepted: 21 September 2022

Published: 25 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Thyroid cancer (TC) is one of the most common malignant tumors of the endocrine system, accounting for about 1% of systemic tumors. The main pathological types are follicular thyroid carcinoma (FTC), papillary thyroid carcinoma (PTC), anaplastic thyroid carcinoma and medullary thyroid carcinoma (MTC). Among them, PTC accounted for the majority of cases (92.98%), but this type of cancer has a low degree of malignancy and a good prognosis. Studies show approximately 2.7% of women and 0.7% of men worldwide suffer from TC [1]. According to global cancer statistics, the number of new cancer cases worldwide is about 18,078,957 every year, of which 4% are TC [2]. In 2016, there were 238,000 TC patients worldwide, yet the number of TC patients reached 586,000 in just four years, a 1.46-fold increase [3,4]. It is predicted that by 2030, TC will surpass colorectal cancer to become the fourth major malignant tumor after lung cancer, breast cancer and prostate cancer [5]. TC incidence in China has been increasing in recent years, similar to the global trend. According to the Annual Report of China Cancer Registry, the incidence of TC increased from 4.3/100,000 (2005) to 13.22/100,000 (2016). Researchers have predicted the incidence of TC in the mainland of China will continue increasing for at least 20 years, with more than 3.7 million new cases between 2028–2032 [6].

Previous studies discussed many risk factors associated with TC, e.g., ionizing radiation, weight, iodine intake as well as environment pollutants, diet, behaviors and economic level. Studies show that exposure to <sup>131</sup>I in childhood increased the risk of TC, exposure dose had a certain correlation with the degree of risk, and the two were linearly related when the radiation dose was 1.5–2.0 Gy [7]. The mechanism of obesity-induced TC was very complex, which was the result of the combined action of a variety of independent and synergistic factors. A cohort study of adults in the United Kingdom found that for every 5 kg/m<sup>2</sup> increase in BMI, the risk of TC increased by 1.09 times [8]. Iodine was one of the main raw materials for the synthesis of thyroid hormone. Its deficiency and excess both

could affect the synthesis and secretion of thyroid hormone, change the level of TSH, cause changes in the structure and function of the thyroid, so as to increase the incidence of PTC. Korean scholars found that PTC patients had higher iodine exposure [9]. Moreover, the exposure of PM<sub>2.5</sub>, NO<sub>2</sub> and PM<sub>10</sub> were all closely related to the occurrence of TC [10,11]. A Canadian study found that people with higher socioeconomic status were more likely to be diagnosed with TC ( $p < 0.001$ ) [12]. High monthly household income was also a risk factor for TC [13].

However, the current research on the risk factors of TC at home and abroad hardly include dietary behavior and socioeconomic factors at the same time. On the other hand, the research scope of ecological research on the risk factors of TC in China was relatively limited, and the research carried out on a national scale was comparatively scarce. Moreover, many studies identified the risk factors of TC through traditional epidemiological methods, such as case-control studies, without combining the spatial distribution analysis of TC at the same time. For example, a study in South Korea used the incidence data of the cancer center in the country and the data of the community health survey to determine the spatial cluster of TC incidence through spatial scanning analysis and to evaluate the influencing factors of TC incidence through the Besag–York–Mollie model [14]. However, dietary and environmental factors, such as iodine intake and PM<sub>2.5</sub>, were not considered in the analysis of associated factors, and only socioeconomic and behavioral factors, such as family income, obesity, drinking and smoking were included, with certain flaws in the analysis of variables. Fei’s research only focused on the Hangzhou prefecture and the factors in the research were mostly socioeconomic factors, such as factory density, building area ratio, and natural environment factors, such as average altitude, but did not take into account the impact of dietary behavior and other factors on TC [15]. A study in Iran that concentrated on the risk factors for TC in the province with the highest incidence in the southern country used a traditional case-control study and did not describe the underlying spatial distribution of TC incidence or death in that province [16].

Therefore, a preliminary description of the spatial distribution of TC incidence and an exploration of the risk factors of TC occurrence in China from two aspects—socioeconomic factors and dietary behavior factors at the same time—will be studied, so as to provide targeted ideas and methodological references for the regionalized prevention and control of TC.

## 2. Materials and Methods

### 2.1. Data Sources

The data of this study consisted of five parts, their sources and attributes are shown in the table below (Table 1).

**Table 1.** Data sources and attributes.

Category	Name	Source	Type	Attributes
TC data	TC cases, CIR, IR	2019 Annual Report of China Cancer Registry *	Quantitative data	-
Geographic information data	2021 geographic information layers of provinces, prefectures and nine-dash line, water towns	1:250,000 National Catalogue Service for Geographic Information ( <a href="https://www.webmap.cn/">https://www.webmap.cn/</a> , accessed on 20 September 2022)	Vector data	shape
Socioeconomic factors data	Sex ratio, UR, PGDP, beds, MIR, GR, education	Statistical yearbook, statistical bulletin, the sixth national census	Quantitative data	-

**Table 1.** *Cont.*

Category	Name	Source	Type	Attributes
Dietary behavior factors data	Aquatic product, grain, vegetable and edible mushroom, meat, egg, milk, dried and fresh fruits, daily salt intake, salt and water iodine, smoking rate, obesity rate	Statistical yearbook, local research and government active report	Quantitative data	-
PM <sub>2.5</sub> data	PM <sub>2.5</sub> concentration of year 2006, 2011, 2016	NASA, ( <a href="https://search.earthdata.nasa.gov/">https://search.earthdata.nasa.gov/</a> , accessed on 20 September 2022)	Grid data	-

Note: crude incidence rate (CIR), age-standardized incidence (IR), urbanization rate (UR), GDP per capita (PGDP), and number of beds per 1,000 people (beds), the coverage rates of basic medical insurance (MIR), green space coverage rates in urban built-up areas (GR), National Space Administration (NASA). Data on socioeconomic factors were at prefecture level and data on dietary behavior factors were at provincial level. Except for average years of education, which were from the sixth national census in 2010, all other data were 2016 data.\* indicated the report was published by NCC and there was a three-year lag in the year in the published data. Influencing factors and geographic information layers were connected by administrative codes.

### 2.1.1. Data of TC

We entered the number of TC cases, crude incidence rate (CIR) and age-standardized incidence (standardized according to the world standard population, hereinafter referred to as incidence, IR) in 2016 and divided them into prefectures and provinces. If a prefecture was not tumor registration site, but there were two or more registration sites in its subordinate counties, then we used the average incidence of these counties to estimate its TC incidence. The data we obtained were called estimated prefecture level data. If there was only one county in a prefecture as a registration site, it was assumed that the incidence of TC in this county covered the entire prefecture, and then made an estimate (Equation (1)):

$$IR_{eci} = \frac{\sum_{i=1}^N IR_{aco}^i}{N} \tag{1}$$

In the formula,  $IR_{eci}$  was estimated prefecture level IR,  $IR_{aco}$  was actual county level IR under the prefecture, N was the number of county tumor registries in the prefecture.

Similarly, provincial TC data were estimated using data from subordinate prefectures. If a prefecture had both its own and county level TC incidence, only the prefecture level data would be used to estimate TC incidence of the province in which it was located. In other words, the IR of a province was equal to the average of the actual and estimated IR of its subordinate prefectures (Equation (2)):

$$IR_{epr} = \frac{\sum_{i=1}^N (IR_{eci}^\alpha + IR_{aci}^\beta)}{N}, \alpha, \beta = 1, 2, \dots, n, \alpha + \beta = N \tag{2}$$

In the formula,  $IR_{epr}$  was estimated provincial level IR,  $IR_{aci}$  was actual prefecture level IR under the province,  $\alpha, \beta$  were the number of estimated prefectures and actual prefectures,  $IR_{eci}$  and N were the same as the Equation (1).

### 2.1.2. Geographic Information Data

The spatial coordinate system adopted the 2000 National Geodetic Coordinate System, the 1985 National Elevation Datum, and the latitude and longitude coordinates. This study obtained geographic information data of provinces, prefectures, and counties across the country in 2021, including administrative boundaries and geographic coordinates. The administrative codes of 2019 were used for registered prefectures. The cropping and merging function of ArcGIS 10.8.1 software was used to obtain the geographic information of the national tumor registration prefectures (including the actual registered prefectures and estimated registered prefectures) in 2016.

### 2.1.3. Socioeconomic Factors Data

The details were shown in Table 1.

### 2.1.4. Dietary Behavior Factors Data

Except for salt iodine, data came from the “2011 National Iodine Deficiency Disease Surveillance” and water iodine data came from the “2017 National Survey on Iodine Content in Drinking Water”; all of the consumption data in each province were from the “2017 China Statistical Yearbook”. Data of daily salt intake per capita in Hebei Province were from the “2007 Survey on Chronic Diseases and Risk Factors of Adults in Hebei Province”, which was missing in Tibet. Apart from some provinces that came from local research, the smoking rates were all from active reports of each province. Among them, the smoking data of Inner Mongolia Autonomous Region were from 2021, Yunnan were from 2019, Jilin, Guizhou, Gansu were from 2013, and other provinces were from 2014–2018. The data for the above years were the latest data available to us and due to the availability of the data years, data of Tibet, Qinghai and Xinjiang were missing. Obesity rates data came from the “Chronic Diseases and Risk Factors Surveillance Data” report released by the Chinese Center for Disease Control and Prevention in 2017.

### 2.1.5. PM<sub>2.5</sub> Concentration

In our dataset, PM<sub>2.5</sub> concentration data were obtained by two satellites, MISR and Sea WiFS, and aerosol optical depth (AOD) obtained by the Mesoscale Imaging Spectroradiometer (MODIS). On this basis, the atmospheric chemical transport model was used to establish a mapping relationship with the ground-measured PM<sub>2.5</sub> concentration, and then a geographically weighted regression model was used to calibrate the estimated PM<sub>2.5</sub> concentration and data combined with the ground-measured data. The final output raster layer was at a resolution of  $0.01^\circ \times 0.01^\circ$ .

Through other studies [10,17], we selected compromised PM<sub>2.5</sub> lag years, which were 0, 5, and 10 years. Thus, this study obtained PM<sub>2.5</sub> concentration data from 2006, 2011 and 2016. We extracted the PM<sub>2.5</sub> concentrations of each tumor registry in three time periods through the spatial clipping, spatial analysis and spatial data extraction functions of ArcGIS 10.8.1 software.

## 2.2. Statistical Analysis

This study mainly studied TC from two aspects, spatial analysis and risk factor analysis. The specific details are shown in the figure below (Figure 1).

### 2.2.1. Descriptive and Geographical Analysis

The national and different gender TC incidence rates in 2016 were described. Moreover, based on the incidence data of the existing tumor registries, we drew the incidence map of TC nationwide in 2016 by ArcGIS 10.8.1 software according to the prefecture (including the actual registration prefecture and the estimated registration prefecture).

### 2.2.2. Spatial Autocorrelation Analysis

Spatial autocorrelation referred to spatial dependencies between adjacent spatial units of the same variable and showed correlations between variables across space [18].

It was divided into two categories: global autocorrelation and local autocorrelation. In this study, we used ArcGIS 10.8.1 to perform local autocorrelation analysis on the incidence of TC in 2016. Local autocorrelation analysis detected the spatial distribution of variables from a specific local area and can find the specific type of aggregation and the location of the aggregation area. It was precisely because the global autocorrelation cannot accurately indicate the location of the aggregation area and it was mainly measured by Local Moran's I coefficient (LISA). According to LISA, the specific aggregation type, location and corresponding hot and cold areas of the disease can be explored. Thus, we

used LISA to explore the high–high, high–low, low–high, and low–low clusters of TC incidence, further identifying high risk areas of TC incidence.

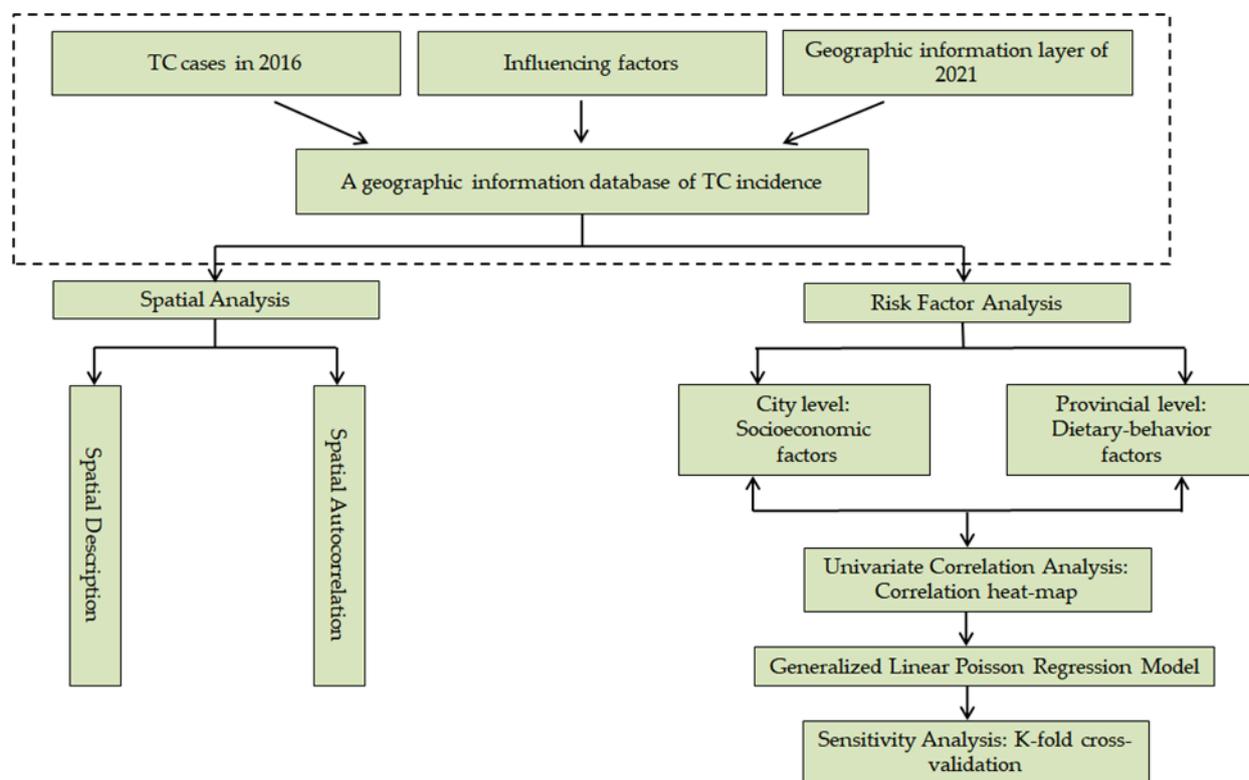


Figure 1. Technology roadmap for this study.

### 2.2.3. Analysis of Influencing Factors of TC

First, socioeconomic variables were available at the prefecture level, so we studied their impact on TC incidence at prefecture level. However, for dietary behavior variables, we can only obtain the data at the provincial level. As a result of their close association with the occurrence of TC, we had to explore their impact on TC at the provincial level.

#### (a) Univariate Correlation Analysis

In this study, the incidence of TC and the corresponding covariates did not meet the normal distribution at the same time, so we chose Spearman’s rank correlation in R4.2.0 for univariate correlation analysis.

At prefecture level: all socioeconomic factors were used as independent variables; IR of TC was used as dependent variable for analysis. At provincial level: dietary behavior factors were also used as independent variables; IR of TC in each province was used as dependent variable. The two-sided *p* value of less than 0.05 was considered statistically significant, and the “corrplot” package in R4.2.0 was further called to draw a correlation heat map.

#### (b) Generalized Linear Poisson Regression Model (GLM)

A generalized linear model was an extension of the general linear model, which established the relationship between the mathematical expectation of the response variable and the linearly combined predictor variable through a link function. In generalized linear models, independent variables can be continuous variables, binary variables, or rank variables. The dependent variable can obey the binomial, Poisson, negative binomial, gamma, inverse Gaussian distribution, etc. These distributions were collectively referred to as the exponential distribution family. The distribution of health outcome variables (such as morbidity, mortality, and hospital visits) was usually described by a Poisson distribution,

so the GLM was used to analyze the associated factors of TC. RR was used to describe the relationship between TC and various factors.

In this study, the variables with statistical significance after univariate correlation analysis were used as independent variables again at the prefecture and provincial levels, and the incidence of TC was used as the dependent variable. The fitting of GLM was performed in R4.2.0. At the same time, in order to exclude multi-collinearity between the included variables as much as possible, by calling the “leaps” package in R4.2.0 to perform full-subset regression to select the optimal and simplest GLM and eliminate the excessive deviance, we performed the Poisson regression again, which could help us identify risk factors for TC occurrence.

### (c) Sensitivity analysis

After obtaining the optimal and simplest GLM model, we used K-fold cross-validation to test its robustness. Cross-validation referred to selecting a certain proportion of samples as training samples, and another part of the samples as reserved samples, first using the training samples to obtain the regression equation, and then predicting on the reserved samples. Because the holdout sample did not participate in the model building process, it could be used to estimate the accuracy of the model. K-fold cross-validation meant that the samples were divided into k subsets, and k-1 sub-samples were taken as the training set in turn, and the other subset was used as the reserved set, and finally the average predicted value was obtained. The smaller the value of K, the larger the bias; the larger the value of K, the more unstable the result was, so k = 10 was usually recommended. In this study, K = 10.

## 3. Results

### 3.1. Descriptive Analysis

In 2016, the number of TC cases nationwide was 50,424, of which 12,240 (24.27%) were male patients and 38,184 (75.73%) were female patients, which was more than twice the number of male cases. The national crude incidence rate (CIR) of TC was 13.22/100,000, and the incidence rate (IR) was 9.70/100,000, both the CIR and IR of women were higher than men (Table 2).

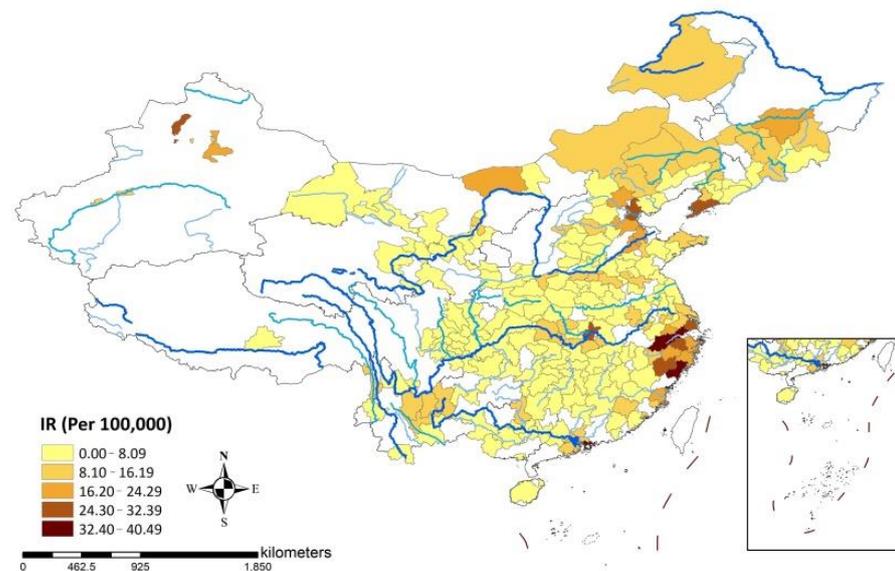
**Table 2.** Incidence (per 10,000) of TC nationwide and by gender in 2016.

Category	Cases (%)	CIR	IR
Male	12,240(24.27)	6.32	4.71
Female	38,184(75.73)	20.32	14.79
National	50,424	13.22	9.70

Note: crude incidence rate (CIR), age-standardized incidence rate according to the world standard population (IR).

### 3.2. Geographical Analysis

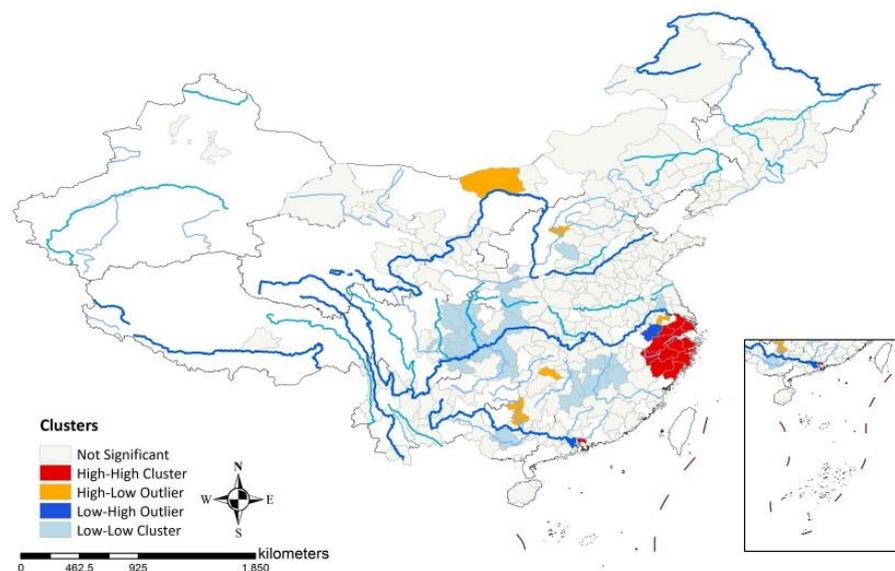
On the whole, the incidence of TC in most areas of China was at a low level, the high incidence areas were mainly in the coastal areas, especially in Zhejiang Province in the southeast coast, which had the highest incidence of TC in China. Hangzhou had the highest incidence (37.87/100,000) nationwide. The incidence of TC in Zhoushan, Jiaxing, Wenzhou, Shaoxing, Lishui, Ningbo and other prefectures in Zhejiang Province was also at a high level. In addition, Shenzhen, Dalian and Shanghai also had high incidence levels, which were 33.97/100,000, 31.49/100,000, and 31.49/100,000, respectively (Figure 2).



**Figure 2.** Map of TC incidence (per 100,000) in 2016. Note: The blank areas had no available registration data. The blue lines of different shades and thicknesses represented the water towns. The darker and the thicker the line was, the higher the stream order. The first order stream was the highest order.

### 3.3. Spatial Autocorrelation Analysis

There was a positive spatial autocorrelation in the incidence of TC nationwide in 2016 ( $Z$  score = 8.03,  $p < 0.001$ ). Furthermore, there were 13 high–high clusters of TC, including nine prefectures in Zhejiang Province (Hangzhou, Ningbo, Wenzhou, Jiaxing, Huzhou, Jinhua, Quzhou, Lishui, and Taizhou). Moreover, five prefectures, including Bayan Naoer, Taiyuan, Changzhou, Liuzhou and Loudi were high–low clusters, Xuancheng was a low–high cluster. Finally, the light blue areas in the map including Yangzhou, Nanchang, Chongqing and the other 24 prefectures were all low–low clusters (Figure 3).



**Figure 3.** Spatial autocorrelation map of TC incidence in 2016. Note: The blank areas had no available registration data. The blue lines of different shades and thicknesses represented the water system. The darker and the thicker the line was, the higher the stream order. The first order stream was the highest order.

### 3.4. Factors Associated with TC

#### 3.4.1. Univariate Correlation Analysis

Through the K–S test, we found that the incidence of TC did not meet the normal distribution ( $p < 0.001$ ), so Spearman correlation analysis was used to explore the factors associated with TC incidence.

##### (a) Prefecture Level—Socioeconomic Factors

Factors related to the incidence of TC were sex ratio, UR, PGDP, average years of education and MIR. Among them, sex ratio was negatively correlated with the incidence of TC ( $r = -0.350$ ,  $p = 0.010$ ), but UR ( $r = 0.448$ ,  $p < 0.001$ ), PGDP ( $r = 0.503$ ,  $p < 0.001$ ), average years of education ( $r = 0.370$ ,  $p = 0.011$ ), and MIC ( $r = 0.345$ ,  $p = 0.009$ ) were positively correlated with the incidence of TC (Table 3).

**Table 3.** Spearman correlation analysis.

	Variables	<i>p</i> -Value	<i>r</i>
IR (per 100,000)	Sex ratio	0.010	−0.350 *
	UR (%)	<0.001	0.448 *
	PGDP (CNY ten thousand)	<0.001	0.503 *
	Education (year)	0.011	0.370 *
	MIC (%)	0.009	0.345 *
	Beds	0.523	0.054
	2006 PM2.5 ( $\mu\text{g}/\text{m}^3$ )	0.073	−0.144
	2011 PM2.5 ( $\mu\text{g}/\text{m}^3$ )	0.085	−0.114
	2016 PM2.5 ( $\mu\text{g}/\text{m}^3$ )	0.126	−0.075
	GR2006 (%)	0.303	0.083
	GR2011 (%)	0.390	0.100
	GR2016 (%)	0.307	0.118

Note: \* indicated that the correlation coefficients were of statistical significance.

##### (b) Provincial Level—Dietary Behavior Factors

Salt iodine, grain consumption per capita, and smoking rate were negatively correlated with the incidence of TC, their correlation coefficients were  $-0.555$  ( $p < 0.001$ ),  $-0.383$  ( $p = 0.004$ ) and  $-0.355$  ( $p < 0.001$ ), respectively. Egg, aquatic products, dry and fresh fruit consumption per capita, and obesity rate were positively correlated with the incidence of TC, and their correlation coefficients were  $0.416$  ( $p = 0.001$ ),  $0.551$  ( $p = 0.005$ ),  $0.484$  ( $p = 0.008$ ), and  $0.396$  ( $p = 0.034$ ) (Table 4).

**Table 4.** Spearman correlation analysis.

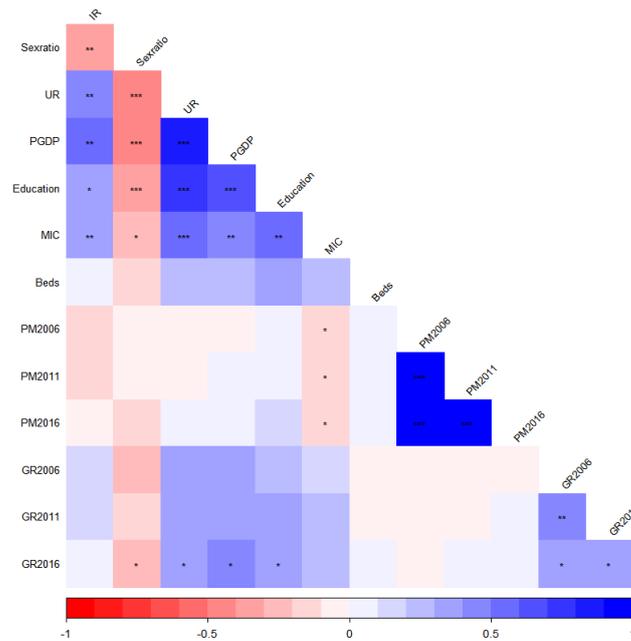
	Variables	<i>p</i> -Value	<i>r</i>
IR (per 100,000)	Water iodine ( $\mu\text{g}/\text{L}$ )	0.054	0.261
	Salt iodine (mg/kg)	<0.001	−0.555 *
	Salt (kg)	0.379	−0.138
	Grain (kg)	0.004	−0.363 *
	Veg and mushrooms (kg)	0.078	0.340
	Meat (kg)	0.296	0.025
	Egg (kg)	0.001	0.416 *
	Aquatic products (kg)	0.005	0.551 *
	Milk (kg)	0.094	0.197
	Fruit (kg)	0.008	0.484 *
	Obesity rate (%)	0.034	0.396 *
	Smoking rate (%)	<0.001	−0.355 *

Note: \* indicated that the correlation coefficients were of statistical significance.

### 3.4.2. Generalized Linear Poisson Regression Model

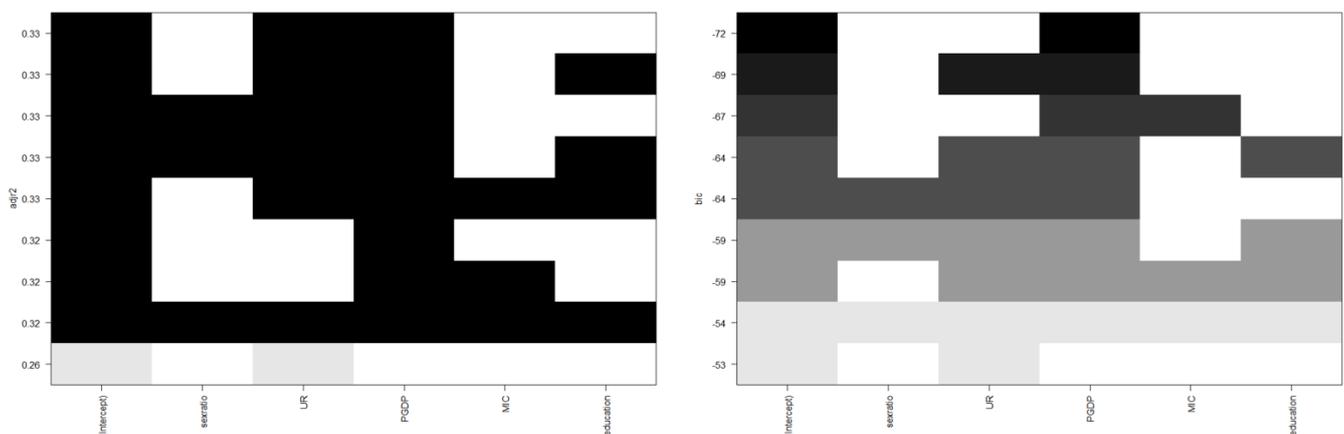
#### (a) Prefecture Level—Socioeconomic Factors

Some socioeconomic factors associated with TC morbidity were correlated with each other (Figure 4). To exclude the multi-collinearity between various factors, we called the regsubsets function in the R language “leaps” package to perform full subset regression on the incidence and influencing factors data of TC at the prefecture level. The optimal and simplest GLM was selected according to the principle that the larger the adjusted R<sup>2</sup> was, the smaller the BIC value was.



**Figure 4.** Heatmap of associations between socioeconomic factors and TC incidence. Note: PM (PM<sub>2.5</sub>),  $p < 0.05$  \*,  $p < 0.01$  \*\*,  $p < 0.001$  \*\*\*.

When the included independent variables were UR and PGDP, the adjusted R<sup>2</sup> of the model was the largest of 0.33, and its BIC was −69, the second-to-last smallest. Therefore, the model that incorporated UR and PGDP was the optimal model (Figure 5). We performed Poisson regression on it and called the qcc.overdispersion.test function in the “qcc” package in R4.2.0 to detect the excessive deviation. We observed an excessive deviation ( $p < 0.05$ ), so Quasi-Poisson regression was used to refit the regression model.



**Figure 5.** R<sup>2</sup> (left) and BIC (right) values for models with different socioeconomic factors.

The model results showed that the higher the UR, the higher the incidence of TC (RR = 1.109, 95%CI: 1.084 to 1.135). That is, for every 1% increase in the UR, the incidence of TC increased by 1.109 times. The higher the PGDP, the higher the incidence of TC (RR = 1.013, 95%CI: 1.007 to 1.018). That is, for every CNY 10,000 increase in PGDP, the incidence of TC increased by 1.013 times (Table 5).

Table 5. Generalized linear model results.

Dependent Variable	Covariate	RR	95%CI	
			Upper Limit	Lower Limit
IR (per 100,000)	UR (%)	1.109	1.084	1.135
	PGDP (CNY ten thousand)	1.013	1.007	1.018

(b) Provincial Level—Dietary Behavior Factors

Similarly, some dietary behavioral factors associated with TC incidence were correlated with each other (Figure 6). In order to exclude the mutual influence of them, we still called the regsubsets function in the “leaps” package of the R language to perform full subset regression on variables that were statistically significant after univariate correlation analysis. The optimal model selection principle was the same as the prefecture level. Excessive deviation detection was also required.

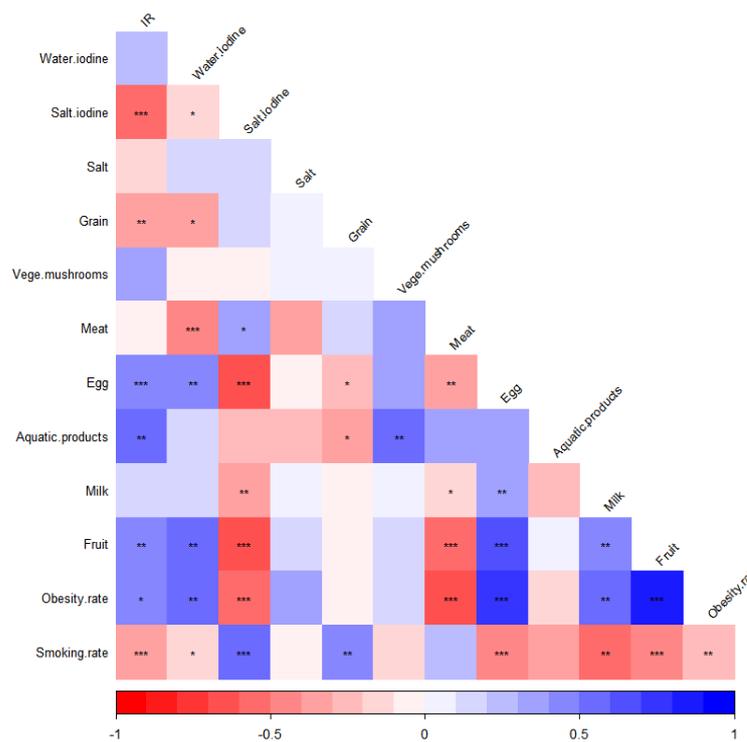
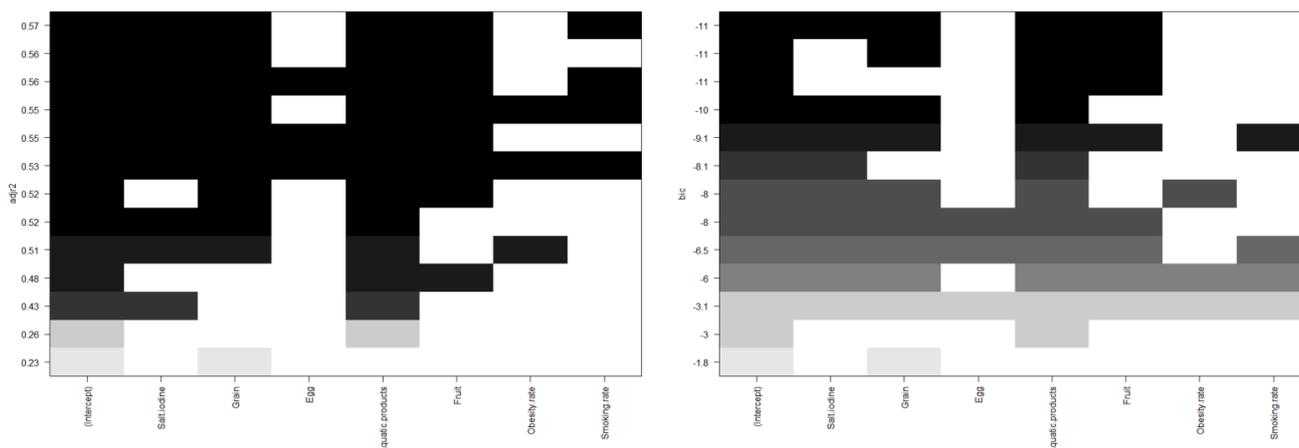


Figure 6. Heatmap of associations between dietary behavior factors and TC incidence. Note:  $p < 0.05$  \*,  $p < 0.01$  \*\*,  $p < 0.001$  \*\*\*.

When the included independent variables were salt iodine, aquatic product consumption, dry and fresh fruit consumption and grain consumption, the BIC value was the smallest of -11, its corresponding adjusted  $R^2$  value was 0.56, which was only 0.01 less than the maximum adjusted  $R^2$  0.57. Therefore, the model including salt iodine, aquatic products, dried and fresh fruits and grain was the optimal model (Figure 7). Similarly, it can be seen that there was an excessive deviation phenomenon ( $p < 0.05$ ), so Quasi-Poisson regression was used for refitting.



**Figure 7.** R<sup>2</sup> (left) and BIC (right) values for models with different dietary behavioral factors.

The higher the aquatic product consumption per capita, the higher the incidence of TC (RR = 1.047, 95%CI: 1.020 to 1.075). That was, for every 1kg increase in aquatic product consumption per capita, the incidence of TC increased by 1.047 times. The higher the consumption of fresh and dried fruits per capita, the higher the incidence of TC (RR = 1.024, 95%CI: 1.007 to 1.040). The incidence of TC increased by 1.024 times for every 1kg increase in the consumption of fresh and dried fruits. However, salt iodine and grain consumption per capita were not associated with TC incidence, as their RR confidence intervals included 1 (Table 6).

**Table 6.** Generalized linear model results.

Dependent Variable	Covariate	RR	95%CI	
			Upper Limit	Lower Limit
IR (per 100,000)	Salt iodine (µg/L)	0.876	0.750	1.025
	Aquatic product (kg)	1.047	1.020	1.075
	Dry and fresh fruit (kg)	1.024	1.007	1.040
	Grain (kg)	0.995	0.984	1.006

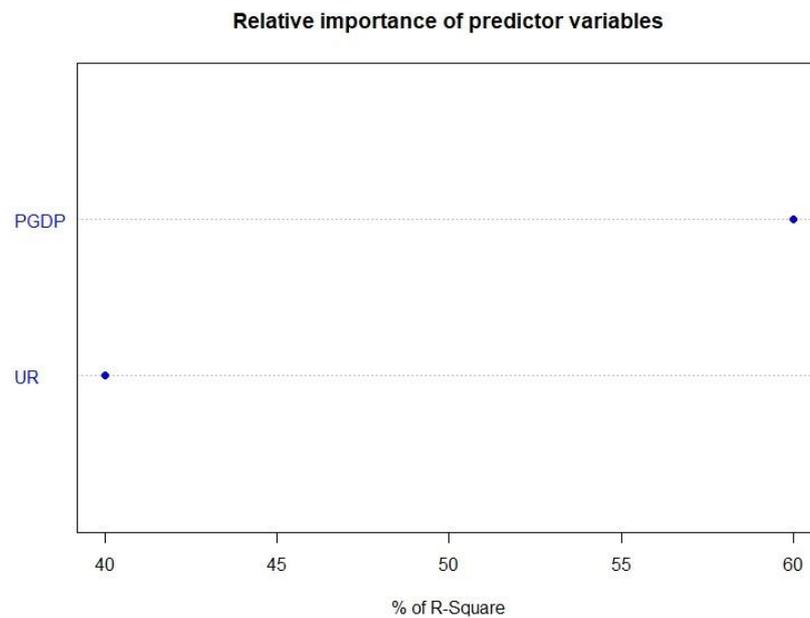
### 3.4.3. Sensitivity Analysis

#### (a) Prefecture Level—Socioeconomic Factors

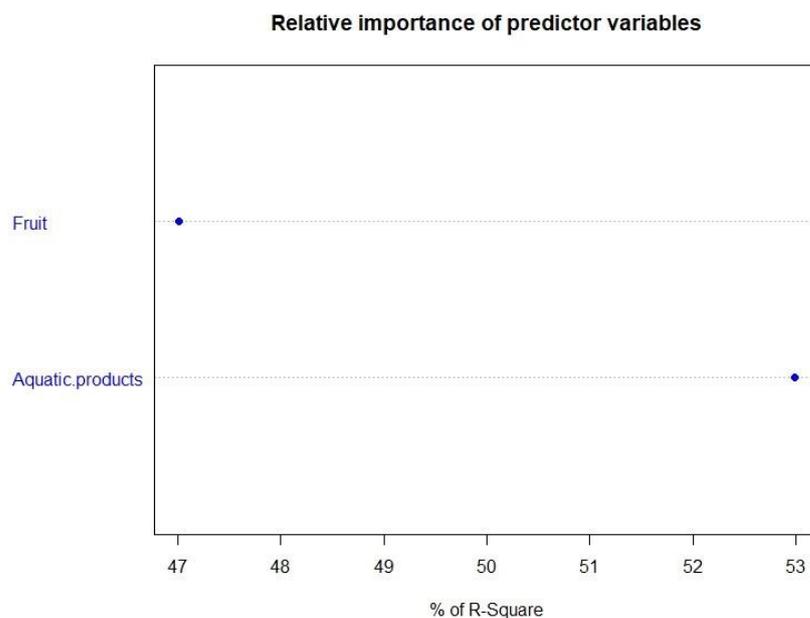
The results of the 10-fold cross-validation showed that the actual explained amount of incidence was 0.29 for the model we finally obtained; the variability (equivalent to error) was very small at 0.02. The relative importance of the variables is shown in the figure below (Figure 8).

#### (b) Provincial Level—Dietary Behavior Factors

The 10-fold cross-validation results showed that when the two variables, salt iodine and grain consumption, which were not statistically significant, were included in the model, the variability of the model was 0.40. When only the consumption of aquatic products and dried and fresh fruits and vegetables were included, the variability of the model was reduced to 0.22. At this time, the robustness of the model was better. The relative importance of the variables is shown in the figure below (Figure 9).



**Figure 8.** The relative importance of predictor variables.



**Figure 9.** The relative importance of predictor variables.

#### 4. Discussion

There were significant regional differences in the incidence of TC, e.g., eastern and coastal areas were higher than inland areas, urban areas were higher than rural areas, socio-economically developed areas were higher than less-developed areas [19]. This study found that coastal areas had higher TC incidence, especially in southeastern coastal areas, which may be related to the economic development in those regions. With the development of the regional economy, the examination of thyroid-related diseases had been included in the routine physical examination items of residents, which greatly improved the lead time of TC detection. In addition, the more developed the economy, the higher the accessibility of local health resources and the more advanced the medical technology. At the same time, with the higher education level of residents, people had more knowledge about

health and paid more attention to disease prevention. The high level of the above factors can strengthen people's health awareness and disease screening awareness, prompting people to seek medical treatment in the early stage of the disease. Importantly, due to the emergence of high-resolution B-ultrasound and the application of fine needle aspiration technology, more early TCs were found, resulting in lead time bias and over-diagnosis bias, which caused high TC incidence in economically developed areas in China. A study found that from 2008 to 2012, 85% of female TC patients in Shanghai were over-diagnosed, and 82.8% of male patients were over-diagnosed. Similarly, female over-diagnosed patients in Jiaxing accounted for 93.2% of all female TC patients, followed by Hangzhou, which was 92.9% [20]. Moreover, we also found TC incidence was at a high level for most prefectures in Zhejiang Province. In addition to the over-diagnosis, it may also be related to the dietary structure of local residents. The aquatic products consumption per capita in Zhejiang Province was 23.3 kg, ranking fourth nationwide and much higher than the national average level of 11.4 kg [21]. As we all know, aquatic products are rich in iodine, which is one of the main raw materials for the synthesis of thyroid hormones. Its deficiency can lead to decreased thyroid hormone synthesis and secretion, resulting in increased thyroid-stimulating hormone (TSH) levels. Long-term chronic stimulation of TSH can lead to hypertrophy of thyroid follicles, and even further formation of nodules or cancer. A diet high in iodine can also cause changes in the structure and function of the thyroid gland. A study found that the incidence of TC in Shanghai increased significantly after 5 and 8 years of iodization in salt [22].

In this study, we found that high UR and PGDP, high aquatic products and dried and fresh fruit consumption per capita were risk factors for TC. As mentioned above, the impact of high UR and PGDP on the incidence of TC was due to the development of medical technology and medical methods caused by economic development, resulting in over-diagnosis bias. The effect of high aquatic product consumption may be related to iodine intake. Studies have shown that eating more marine animal foods (OR = 9.484) and eating too much seafood (OR = 1.933) are risk factors for TC [23]. As for the effect of high consumption of dried and fresh fruits per capita on TC, there was no research to explore its association with the occurrence of TC. Most of the existing studies have explored only the effect of fresh fruit intake on TC. For example, a study in South Korea found that high consumption of persimmons and oranges may reduce the risk of TC to a certain extent and play a preventive role in early TC [24]. This protective effect of fruits might be related to their vitamin and antioxidant content [25]. Another study found no link between fruit intake and TC [26]. The above results were different from ours, probably because the consumption of fruits in our study included dried fruits, such as peanuts, walnuts, and cashews, most of which were rich in protein, and a related study found that excessive intake of protein was related to the incidence of DTC [27]. Moreover, we found that water iodine content, salt iodine content and daily salt intake per capita were not associated with the occurrence of TC, possibly because they did not represent the actual iodine intake of the population. Firstly, the main source of iodine intake of the population was salt; salt iodine content and salt intake influenced each other. Secondly, the degree of salt iodization varied in different regions.

We also found that PM<sub>2.5</sub> concentration had no statistical association with the occurrence of TC, which was different from previous studies. For example, a study showed that an increase in the incidence of PTC was significantly associated with an increase in PM<sub>2.5</sub> concentrations with a lag of 2 and 3 years [10]. The abnormal role of PM<sub>2.5</sub> in our study may be related to other air pollutants as well as people's health behavior and awareness. Firstly, our study only included PM<sub>2.5</sub>, without adjusting the effects of other pollutants, such as PM<sub>10</sub> and SO<sub>2</sub>, on the incidence of TC. Secondly, the PM<sub>2.5</sub> concentration level in our study only referred to the external exposure level rather than the actual internal dose. The higher the concentration of PM<sub>2.5</sub> in a certain area, the more likely people would take corresponding measures, such as wearing masks and reducing going out as much as

possible to reduce their exposure level. The above factors might lead to the abnormal effect of PM<sub>2.5</sub> to a certain extent.

Finally, a large number of studies have shown that greening rate and green space are closely related to human health. For example, studies found that people with diabetes who lived in areas with relatively high levels of greenery had a significantly lower risk of cardiovascular disease [28]. Increasing residential green exposure levels was significantly associated with a decrease in all-cause mortality [29]. Studies have also found that urban greenness was inversely associated with childhood respiratory outcomes, especially among children exposed to high levels of tobacco smoke pollution [30]. Therefore, we used GR as an influencing factor for the first time to explore its impact on the incidence of TC. However, we found no association between them. On the one hand, it might be because it was indeed not an influencing factor of TC. On the other hand, the GR in this study referred to the green space coverage rate of urban built-up areas. Generally speaking, the high green space coverage rate of urban built-up areas reflected a high development level of prefecture to a certain extent, and the development of prefecture must be accompanied by acceleration of the industrialization process, which produced a series of environmental problems, thereby masking or even reversing the protective effect of green space coverage on TC.

There were strengths in this study. First of all, our study was a nationwide study, which was relatively representative in describing patterns of TC in China. Secondly, this study examined the impact of socioeconomic factors and dietary behavioral factors on TC at the prefecture level and provincial level, respectively.

In addition, this study had inevitable limitations. Firstly, the incidence data of TC in some prefectures and all provinces were estimated data, and their accuracy was limited by the number of tumor registration sites in their subordinate administrative regions. Secondly, this study was an ecological study, individual level data cannot be obtained, which resulted in ecological fallacies. Thirdly, due to data availability (although we have tried our best to keep the time consistent), behavioral factors, such as smoking rates and obesity rates, were from different years and came from local research or active reports from provinces. The quality and comparability of the data were difficult to guarantee. Furthermore, due to the inability to obtain relevant data, we did not take migration (all the predictors may be quite substantial in China between rural areas and cities and across regions) into account, which will potentially undermine the spatial cross-sectional analysis, although maybe the results might not change too much (optimistically).

## 5. Conclusions

High UR, high PGDP, high consumption of aquatic products, and dried and fresh fruits were risk factors for TC. By exploring the factors associated with TC incidence in China, we found that the incidence of TC presented a special phenomenon of “economic differentiation”, which meant that the more developed the regional economy, the higher the incidence of TC to a certain extent, and the reason behind this was over-diagnosis as a result of advanced medical technology and perfect medical systems in economically developed areas. In order to eliminate this differentiation, the state must adjust and unify the diagnostic criteria for TC at a reasonable level, improve the medical and health system and policy nationwide as much as possible. For example, screening of thyroid-related diseases should be included into the routine physical examination items of residents, especially in less-developed regions, such as the western areas. We should also carry out targeted health education policies according to people’s dietary habits in high risk areas of TC occurrence, correct their bad eating habits and improve their health awareness, so as to reduce the incidence of TC in high risk areas. Finally, individual level studies and more precise data should be used to further explore the risk factors of TC in the future.

**Author Contributions:** Conceptualization, X.Q. and Y.W.; methodology, P.L. and W.H.; software, Y.W.; validation, X.Q., Y.W. and W.W.; formal analysis, Y.W.; data curation, Y.W. and W.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W. and X.Q.; visualization, Y.W.; supervision, X.Q. and W.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Ministry of Education, Humanities and Social Sciences Research Youth Foundation (Grant No. 20YJC840014), and Key R & D Program of Shaanxi Province (Grant No. 2018ZDCXL-SF-02-03-01).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fei, X.; Chen, W.; Zhang, S.; Liu, Q.; Zhang, Z.; Pei, Q. The spatio-temporal distribution and risk factors of thyroid cancer during rapid urbanization—A case study in China. *Sci. Total Environ.* **2018**, *630*, 1436–1445. [[CrossRef](#)] [[PubMed](#)]
2. Salamat, F.; Aryannia, A.; Rajaei, S.; Naeimi-Tabiei, M.; Afghani, R.; Sedaghat, S.M.; Naeimi-Tabiei, E.; Salamat, F.; Hosseinpour, R.; Fazel, A.; et al. Temporal and Geographical Trends of Incidence of Thyroid Cancer in Golestan, Iran, 2004–2013. *Arch. Iran. Med.* **2021**, *24*, 1–6. [[CrossRef](#)] [[PubMed](#)]
3. Fitzmaurice, C.; Akinyemiju, T.F.; Al Lami, F.H.; Alam, T.; Alizadeh-Navaei, R.; Allen, C.; Alsharif, U.; Alvis-Guzman, N.; Amini, E.; Anderson, B.O.; et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* **2018**, *4*, 1553–1568. [[PubMed](#)]
4. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
5. Rahib, L.; Smith, B.D.; Aizenberg, R.; Rosenzweig, A.B.; Fleshman, J.M.; Matrisian, L.M. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **2014**, *74*, 2913–2921. [[CrossRef](#)]
6. Li, M.; Pei, J.; Xu, M.; Shu, T.; Qin, C.; Hu, M.; Zhang, Y.; Jiang, M.; Zhu, C. Changing incidence and projections of thyroid cancer in mainland China, 1983–2032: Evidence from Cancer Incidence in Five Continents. *Cancer Causes Control* **2021**, *32*, 1095–1105. [[CrossRef](#)]
7. Cardis, E.; Kesminiene, A.; Ivanov, V.; Malakhova, I.; Shibata, Y.; Khrouch, V.; Drozdovitch, V.; Maceika, E.; Zvonova, I.; Vlassov, O.; et al. Risk of Thyroid Cancer After Exposure to 131 I in Childhood. *J. Natl. Cancer Inst.* **2005**, *97*, 724–732. [[CrossRef](#)]
8. Bhaskaran, K.; Douglas, I.; Forbes, H.; dos-Santos-Silva, I.; Leon, D.A.; Smeeth, L. Body-mass index and risk of 22 specific cancers: A population-based cohort study of 5.24 million UK adults. *Lancet* **2014**, *384*, 755–765. [[CrossRef](#)]
9. Lee, J.-H.; Hwang, Y.; Song, R.-Y.; Yi, J.W.; Yu, H.W.; Kim, S.-J.; Chai, Y.J.; Choi, J.Y.; Lee, K.E.; Park, S.K. Relationship between iodine levels and papillary thyroid carcinoma: A systematic review and meta-analysis. *Head Neck* **2017**, *39*, 1711–1718. [[CrossRef](#)]
10. Karzai, S.; Zhang, Z.; Sutton, W.; Prescott, J.; Segev, D.L.; McAdams-DeMarco, M.; Biswal, S.S.; Ramanathan, M., Jr.; Mathur, A. Ambient particulate matter air pollution is associated with increased risk of papillary thyroid cancer. *Surgery* **2022**, *171*, 212–218. [[CrossRef](#)]
11. Park, S.J.; Min, C.; Yoo, D.M.; Choi, H.G. National cohort and meteorological data based nested case–control study on the association between air pollution exposure and thyroid cancer. *Sci. Rep.* **2021**, *11*, 21562. [[CrossRef](#)] [[PubMed](#)]
12. Siu, S.; McDonald, J.T.; Rajaraman, M.; Franklin, J.; Paul, T.; Rachinsky, I.; Morrison, D.; Imran, S.A.; Burrell, S.; Hart, R.; et al. Is Lower Socioeconomic Status Associated with More Advanced Thyroid Cancer Stage at Presentation? A Study in Two Canadian Centers. *Thyroid* **2014**, *24*, 545–551. [[CrossRef](#)] [[PubMed](#)]
13. Choi, S.-W.; Ryu, S.-Y.; Han, M.-A.; Park, J. The Association Between the Socioeconomic Status and Thyroid Cancer Prevalence; Based on the Korean National Health and Nutrition Examination Survey 2010–2011. *J. Korean Med Sci.* **2013**, *28*, 1734–1740. [[CrossRef](#)] [[PubMed](#)]
14. Jang, J.; Yoo, D.-S.; Chun, B.C. Spatial distribution and determinants of thyroid cancer incidence from 1999 to 2013 in Korea. *Sci. Rep.* **2021**, *11*, 22474. [[CrossRef](#)]
15. Fei, X.; Lou, Z.; Christakos, G.; Liu, Q.; Ren, Y.; Wu, J. Contribution of industrial density and socioeconomic status to the spatial distribution of thyroid cancer risk in Hangzhou, China. *Sci. Total Environ.* **2018**, *613*, 679–686. [[CrossRef](#)]
16. Parad, M.T.; Fararouei, M.; Mirahmadizadeh, A.R.; Afrashteh, S. Thyroid cancer and its associated factors: A population-based case-control study. *Int. J. Cancer* **2021**, *149*, 514–521. [[CrossRef](#)]

17. Giannoula, E.; Melidis, C.; Frangos, S.; Papadopoulos, N.; Koutsouki, G.; Iakovou, I. Ecological Study on Thyroid Cancer Incidence and Mortality in Association with European Union Member States' Air Pollution. *Int. J. Environ. Res. Public Health* **2020**, *18*, 153. [[CrossRef](#)]
18. Getis, A. Reflections on spatial autocorrelation. *Reg. Sci. Urban Econ.* **2007**, *37*, 491–496. [[CrossRef](#)]
19. Han, J.; Kang, Y. Prevalence and influencing factors of thyroid cancer. *Pract. Prev. Med.* **2018**, *25*, 894–897.
20. Li, M.; Zheng, R.; Maso, L.D.; Zhang, S.; Wei, W.; Vaccarella, S. Mapping overdiagnosis of thyroid cancer in China. *Lancet Diabetes Endocrinol.* **2021**, *9*, 330–332. [[CrossRef](#)]
21. Liu, P.; SU, X.; Shen, H.; Meng, F.; Fan, L.; Liu, S.; Sun, D. National iodine deficiency disorders: An analysis of surveillance data in 2011. *Chin. J. Endem.* **2015**, *34*, 181–185.
22. Wang, Y.; Wang, W. Increasing Incidence of Thyroid Cancer in Shanghai, China, 1983–2007. *Asia-Pac. J. Public Health* **2015**, *27*, 223–229. [[CrossRef](#)] [[PubMed](#)]
23. Chen, Y.C.; Gao, X.H.; Cao, M.Q.; Li, X.F.; Ma, L. Risk factors of thyroid cancer: A case-control study. *Chin. J. Public Health* **2014**, *30*, 955–957.
24. Jung, S.K.; Kim, K.; Tae, K.; Kong, G.; Kim, M.K. The effect of raw vegetable and fruit intake on thyroid cancer risk among women: A case-control study in South Korea. *Br. J. Nutr.* **2012**, *109*, 118–128. [[CrossRef](#)] [[PubMed](#)]
25. Barrea, L.; Pugliese, G.; Frias-Toral, E.; Laudisio, D.; Rodriguez, D.; Vitale, G.; Colombo, C.; Colao, A.; Savastano, S.; Muscogiuri, G. Diet as a possible influencing factor in thyroid cancer incidence: The point of view of the nutritionist. *Panminerva Med.* **2021**, *63*, 349–360. [[CrossRef](#)] [[PubMed](#)]
26. Zamora-Ros, R.; Béraud, V.; Franceschi, S.; Cayssials, V.; Tsilidis, K.K.; Boutron-Ruault, M.C.; Weiderpass, E.; Overvad, K.; Tjønneland, A.; Eriksen, A.K.; et al. Consumption of fruits, vegetables and fruit juices and differentiated thyroid carcinoma risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Int. J. Cancer* **2018**, *142*, 449–459. [[CrossRef](#)]
27. Marcello, M.; Sampaio, A.C.; Geloneze, B.; Vasques, A.C.J.; Assumpção, L.V.M.; Ward, L.S. Obesity and Excess Protein and Carbohydrate Consumption Are Risk Factors for Thyroid Cancer. *Nutr. Cancer* **2012**, *64*, 1190–1195. [[CrossRef](#)]
28. Anna, P.; Jordi, B.; Lia, A.; Ruth, M.; Marc, C.; Marta, C.; Mark, N.; Mar, G.M.D.; Rafel, R. Impact of residential greenness on myocardial infarction in population with diabetes: A sex-dependent association? *Environ. Res.* **2021**, *205*, 112449.
29. Kua, K.; Lee, S. The influence of residential greenness on mortality in the Asia-Pacific region: A systematic review and meta-analysis. *Perspect. Public Health* **2021**, *141*, 342–353. [[CrossRef](#)]
30. Eldeirawi, K.; Kunzweiler, C.; Zennk, S.; Finn, P.; Nyenhuis, S.; Rosenberg, N.; Persky, V. Associations of urban greenness with asthma and respiratory symptoms in Mexican American children. *Ann. Allergy Asthma Immunol.* **2018**, *122*, 289–295. [[CrossRef](#)]