Geospatial Health 2022; volume 17:1030

Effects of geographical factors on reference values of the thyroid stimulating hormone in healthy adults in China and its clinical significance

Xinrui Pang,[1] Miao Ge,[1] Congxia Wang,[2] Jinwei He[3]

[1]School of Geographic Sciences and Tourism, Shaanxi Normal University, Xian; [2]Medical College of Xi'an Jiao tong University, Xian; [3]Medical College of Yan'an University, Shaanxi, Xian, China

Abstract

The thyroid stimulating hormone (TSH) plays an important regulatory role in maintaining normal function of the thyroid gland. The purpose of this study was to explore the geographical, spatial distribution of TSH normal values in healthy Chinese adults to be used for the formulation of a standard reference. TSH values of 9321 healthy adults from 120 cities in China were collected together with 24 topographic, climatic and soil variables and used for the determination of spatial, significant relationships between TSH and these geographical factors by correlation analysis. Eleven significant factors were extracted and subjected to ridge regression and construction of vector machine models. The predicted values were tested for normality, with the disjunctive Kriging interpolation method used for geographical distribution. The values found showed a spatial pattern of higher values in the North and west but lower in the South and east We concluded that ridge regression models are useful for this kind of investigations and that certain geographical factors determine the level of TSH in healthy adults in a large expanse of land where topography, climate and soil indices vary.

Correspondence: Miao Ge, School of Geographic Sciences and Tourism, Shaanxi Normal University, Xian 710119, China.
E-mail: gemiao@snnu.edu.cn

Key words: Thyroid stimulating hormone; geographical factors; correlation analysis; Ridge regression; support vector regression; geographical spatial distribution; China.

Funding information: this work was supported by grant 41761100 from the Nature Science Foundation of China. This work was supported by grant 2019JM-408 from the Natural Science basic Research Program of Shaanxi Province.

Contributions: the authors have read and agreed to the published version of the manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Availability of data and materials: the data can be requested from the authors.

Received for publication: 20 June 2021.
Revision received: 17 February 2022.
Accepted for publication: 17 February 2022.

Geospatial Health 2022; 17:1030
doi:10.4081/gh.2022.1030

Introduction

Located under the anterior cervical thyroid cartilage and on both sides of the trachea, the thyroid gland is the largest endocrine gland in the human body (Zheng and Sun, 2008). The hypothalamic-pituitary system controls the thyroid gland by the thyrotropin-releasing hormone, which stimulates the basophilic cells in the anterior lobe of the pituitary gland to secrete thyrotropin, *i.e.* the thyroid-stimulating hormone (TSH), which in turn makes the thyroid release thyroxine and triiodothyronine that govern all cell metabolism and synthesis in the body and further affect the expression of thyroid function (Ji *et al*., 2020).

TSH is closely related to the cardio-cerebro-vascular system as well as glucose and lipid metabolism and can be involved in the pathology of neurological diseases (Yin, 2016; Yuan and Peng, 2018; Li *et al*., 2020). Medical geography presents a long history of research on thyroid, mainly focusing on iodine-deficient goiter caused by local and regional iodine deficiency and thyroid dysfunction caused by excessive iodine (Dou *et al*., 2020).

Medical reference values of thyrotropin are measured in milli-internal units per litre (mIU/L) and used as standard to distinguish between patients with a thyroid malfunction and healthy people. In recent years, based on gender, age, population and iodine levels, different countries have carried out measurements of this hormone in their populations and set up their own references. For example, the 95% confidence interval (CI) for TSH was 0.34-4.66 mU/L in a survey of 6434 people aged between 18 and 85 years in northern Netherlands (Hoogendoorn *et al*., 2006). As iodine intake affect the serum TSH content, Danish scholars have proposed that monitoring and adjusting iodine intake of residents is an important way to prevent thyroid diseases (Laurberg *et al*., 2021) At the same time, Chinese scholars have noted that thyroid hormone levels of healthy people in different regions are affected by natural environment, dietary habits, economic development and other regional characteristics, which cannot be measured uniformly by the same method (Zhu and Luan, 2020). They investigated and studied the reference value of TSH in people of different ages and genders in Xinjiang to provide a specific standard for the diagnosis of thyroid diseases in this region, while Yin *et al*. (2021) formulated different TSH reference intervals for healthy people in Xiamen. Accordingly, it is necessary to develop a set of accurate and applicable TSH reference value reference range standard system in different regions of China.

The geographical environment includes variables, such as air, water, biology and minerals. The human body function is in a dynamic balance with the geographical environment with regard to material and energy exchange, and the relationship between the thyroid gland and the environment and is very close. For this reason, we decided to study the distribution of multiple geographical criteria and establish TSH reference values for healthy adults in the various parts of China.

Materials and methods

Collection of samples for determination of reference values

Through searching the Chinese periodical net, the China Excellent Master and Doctor Dissertation Full-text Database, the China 'Conference Full-text Database and hospital measurement data with the search terms 'reference value of thyrotropic hormone' and 'thyroid disease', we collected TSH-related data from 9, 321 healthy adults from 120 cities in China for physical examination. The data were mainly from Guangdong Province, Jiangsu Province and north-eastern China, and there were more data from eastern China than from the western part. We lacked data from Hong Kong, Macao and Taiwan completely and there were relatively scarce data from Lhasa, Xinjiang and other remote areas.

In order to ensure a good selection of study subjects, physical examination was carried out on male and female adults ranging in age from 18 to 80 years and with a healthy physiological condition. About 10 mL venous blood were collected from each study participant, who were asked to fast for 8~10 h before blood collection and not east in the morning before being bled. All blood samples were screened by the chemiluminescence microparticle immunoassay (CMI) in an automatic chemiluminescence analyser. Table 1 shows diseases and conditions that were excluded.

The geographical environment

We searched the literature and official sites to determine which relevant geographical factors to include and decided on 24 with special reference to geographic location, climate and various soil indices (Table 2).

Table 1. Exclusion criteria.

1.	Patients with hyperthyroidism or hypothyroidism, including those with subclinical hyperthyroidism, prior clear diagnosis or related treatment
2.	Patients with autoimmune diseases or diabetes mellitus
3.	Patients with a history of cerebro-vascular or haematological diseases; and malignant tumours
4.	Pregnant or breastfeeding women

Table 2. The geographical environment in the study area.

	Geographical factor	Code	Unit	Source
Topographic index	Longitude	X_1	Degree (°)	State Bureau of Surveying and Mapping in China
	Latitude	X_2	Degree (°)	
	Altitude	X_3	Meter (m)	
Climate index	Annual sunshine duration	X_4	Hour (h)	China Meteorological Data Science Data Sharing Network
	Annual mean air temperature	X_5	Centigrade (°C)	
	Annual mean relative humidity	X_6	Percentage (%)	
	Annual precipitation	X_7	Millimetre (mm)	
	Annual range of air temperature	X_8	Centigrade (°C)	
	Annual mean wind speed	X_9	Meter (m) per-second (m/s)	
Soil index	Topsoil sand fraction	X_{10}	Weight percentage(% wt)	The World Data Soil Bank of the United Nations Food and Agriculture Organization (FAO); the International System for Applications (IIASA)
	Topsoil silt fraction	X_{11}	Weight percentage(%wt)	
	Topsoil clay fraction	X_{12}	Weight percentage(%wt)	
	Topsoil reference capacity	X_{13}	Kilogram per-cubic meter (kg/ m^3)	
	Topsoil bulk density	X_{14}	Kilogram per-cubic meter(kg/m^3)	
	Topsoil gravel content	X_{15}	Volume percentage (%vol)	
	Topsoil organic carbon	X_{16}	Weight percentage (%)	
	Topsoil pH	X_{17}		
	Topsoil cation exchange capacity-clay (X_{18})	X_{18}	Centimol per kilogram (cmol/kg)	
	Topsoil cation exchange capacity-silt (X_{19})	X_{19}	Centimol per kilogram (cmol/kg)	
	Topsoil base saturation	X_{20}	Percentage (%)	
	Topsoil total exchange capacity	X_{21}	Centimol per kilogram (cmol/kg)	
	Topsoil calcium carbonate	X_{22}	Centimol per kilogram (cmol/kg)	
	Topsoil calcium sulphate	X_{23}	Centimol per kilogram (cmol/kg)	
	Surface soil alkalinity	X_{24}	Centimol per kilogram (cmol/kg)	

Statistical analysis

Spatial autocorrelation

Autocorrelation reflects Tobler's First Law of Geography stating that '...Everything is related to everything else. But near things are more related than distant things'. This is an important aspect of our study and we conducted global spatial autocorrelation analysis in ArcGIS 10.2. (Redlands, CA, USA) as we were interested in the degree of the specific range of TSH values in all of China (Tang and Yang, 2006). We chose 0.01 as the desired level of confidence (p), which corresponds to Z=2.54 making Z≥2.54 the criterion for significant spatial autocorrelation in the region under study and Z≤2.54 the cut-off, *i.e.* for values that would not be significant but follow a random distribution.

Correlation analysis and collinearity diagnostics

Correlation analysis is a statistical method that investigates whether there is a interdependent relationship between two variables. Among them, the common correlation coefficients are Pearson, Spearman, Kendall, *etc*.

The distribution of the TSH sample data were investigated with the nonparametric analysis method K-S test and with the respect to the linearity of the data, Spearman correlation coefficient was selected to judge the monotone form among the geographical factors. The correlation coefficient r>0 indicates a positive correlation and r<0 a negative correlation, while the significance coefficient (p) judges the correlation between geographic factors and the TSH reference value where P≤0.01 indicates a significant correlation, 0.01<P≤0.05, a general correlation and P≥0.05, absence of a significant correlation. In SPSS19.0 software (Marr-Lyon *et al*., 2012), The TSH reference values of healthy adults in China were matched with 24 geographical indicators one by one, screening out those not related to TSH values of healthy adults. To further test the influence of collinearity among factors, a correlation heat map of geographical factors was drawn in Origin 2018b (Sun and Jia, 2015).

Ridge regression analysis

Ridge regression analysis was used to overcome the strong linear relationship within the regression independent variables and reduce the influence of the least squares estimator of the regression coefficient (Cong *et al*., 2014). In SAS software (https://www.sas.com), this is a method of biased estimation regression that essentially improves the least square estimation resulting in a more practical approach giving more accurate results. This study uses ridge regression analysis to solve the problem of collinear data and eliminate the interaction between geographic factors. When k=0, ridge regression equation is equal with the least square estimation, and as β depends on the size of the k value (mean square error), this is determined by ridge tracing to predict the TSH reference value of healthy adults nationwide. The paired sample t test was then applied to determine whether or not the predicted value and measured value would accord with the ridge regression analysis (Dong *et al*., 2017).

Support vector regression

SVR is a type of support vector machine(SVM), with training sample set as data object (Boser *et al*., 1992). The support vector regression model can overcome the collinearity and complexity of the geographical factors, reflecting the relationship between reference values and geographical factors suitably, and keeping a certain predicted accuracy. It analyses the quantitative relation between input variable from new samples which has the same distribution as training sample set (Li DC *et al*., 2012).

In this study, Clementine software was used to build SVR model (Yang *et al*., 2016). as one of the efficient data mining software, Clementine is featured with straightforward operation and power data-processing ability by using graphical language programming. Use the following three kernel functions to make nonlinear predictions and determine the most suitable kernel function:

POL (Polynomial kernel): K(x,y)＝[(x·y)＋c]d (1)

RBF (Radial basis function): K(x,y)＝exp[-γ‖x － y‖2] (2)

SIG (Linear kernel function): K(x,y)＝ tanh[γ(x·y)＋c，γ>0，c>0 (3)

The corresponding mean square error (MSE) and mean absolute error (MAE) of the SVM model can be used to evaluate the degree of data change. The smaller the value, the higher the accuracy of the prediction model to describe the experimental data. Similarly, MAE indicates the accuracy of the experimental data:

$$MSE = \frac{1}{N}\sum_{t-1}^{N}(X_i - x_i)^2 \quad MAE = \frac{1}{N}\sum_{t-1}^{N}|(X_i - x_i)| \tag{4}$$

where N is the number of samples; x_i the measured value; and x_i is the predicted value. Further establish the relationship model between geographical factors and medical health reference value TSH.

Evaluation results-paired sample t test

In the practical application, there existed a deviation between measured values and predicted values. In this study, the actual TSH values of 120 cities in China were tested in SPSS 19.0 to find out whether they were normally distributed. When used to prove whether the reference values of thyrotropin of healthy adults in China were consistent with the predicted values given by ridge regression analysis and support vector regression it allow ed choosing the most appropriate model.

Spatial statistics analysis

Kriging interpolation was carried out in ArcGIS10.2 and spatial trend analysis module was used for TSH reference value spatial trend analysis. Then the Kolmogorov-Smirnov (K-S) method was used to test the normality of predicted data (Paul *et al*., 2021). In SPSS 19.0, the K-S test shows that the TSH sample data show non-normal distribution. Therefore, an ordinary kring interpolation methods was used.

With the help of variogram theory (Liu *et al*., 2009) and structural analysis, the optimal interpolation points were calculated and the map interpolated to generate the geospatial distribution results of TSH reference value of healthy adults in China (Xu, 2002; Ge *et al*., 2009).

Results

Spatial autocorrelation

The P value is 0.005415, which is less than its judgment value of 0.01. It can be concluded that there is an extremely significant relationship between the TSH reference value of healthy adults in China and geographical factors in various place. The global autocorrelation coefficient Z value is 2.781, which is greater than the absolute value of the global autocorrelation coefficient in the 0.01 confidence interval of 2.540 (Figure 1). It is considered that the spatial autocorrelation of the TSH reference value of Chinese healthy adults in the spatial region is significant.

Correlation analysis

Based on SPSS v. 19.0, with TSH of healthy adults as dependent variables and the 24 geographical indicators in various regions of China as independent variables, the correlation coefficient (r) and the significance coefficient (*p*) (Table 3), 11 were correlated with the TSH value (P<0.05) in healthy adults: longitude

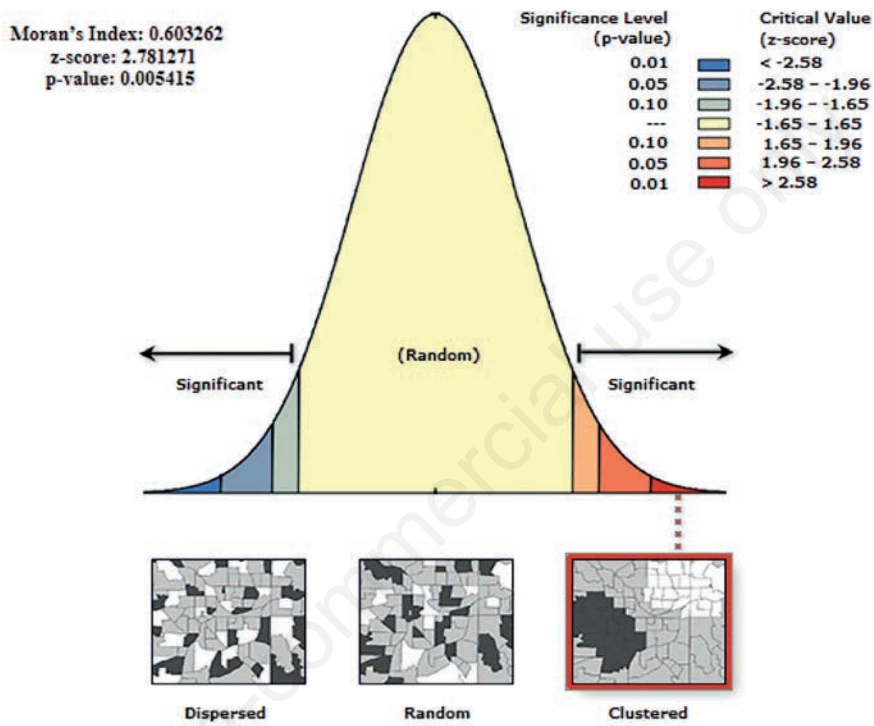

Figure 1. Spatial autocorrelation analysis of the thyroid stimulating hormone in healthy Chinese adults.

Table 3. The correlation between thyroid stimulating hormone levels and geographical indices in different areas in China.

Variable	Correlation coefficient (r)	P	Variable	Correlation coefficient (r)	P
Longitude (X_1)	0.222*	0.015	Topsoil reference capacity (X_{13})	0.215*	0.018
Latitude (X_2)	0.279**	0.002	Topsoil bulk density (X_{14})	0.137	0.135
Altitude (X_3)	–0.106	0.249	Topsoil gravel content (X_{15})	–0.143	0.120
Annual sunshine duration (X_4)	0.215*	0.018	Topsoil organic carbon (X_{16})	0.007	0.935
Annual mean air temperature (X_5)	–0.221*	0.015	Topsoil pH (X_{17})	0.137	0.134
Annual mean relative humidity (X_6)	–0.124	0.177	Topsoil cation exchange capacity in clay (X_{18})	0.011	0.908
Annual precipitation (X_7)	–0.191*	0.036	Topsoil cation exchange capacity in silt (X_{19})	0.225*	0.014
Annual range of air temperature (X_8)	0.247**	0.007	Topsoil base saturation (X_{20})	–0.039	0.670
Annual mean wind speed (X_9)	0.258**	0.004	Topsoil total exchange capacity (X_{21})	–0.003	0.975
Topsoil sand fraction (X_{10})	0.173	0.059	Topsoil calcium carbonate (X_{22})	0.109	0.234
Topsoil silt fraction (X_{11})	0.215*	0.018	Topsoil calcium sulphate (X_{23})	0.051	0.578
Topsoil clay fraction (X_{12})	–0.223*	0.015	Surface soil alkalinity (X_{24})	0.026	0.777

**Significant correlation within a 99% confidence interval; *significant correlation, within a 95% confidence interval.

(X_1); annual sunshine duration (X_4); annual mean temperature (X_5); annual precipitation (X_7); soil particle percentage (X_{11}); soil clay percentage (X_{12}); soil reference bulk density (X_{13}); soil (silt) cation exchange capacity (X_{19}); latitude (X_2); annual temperature range (X_8); and annual mean wind speed (X_9), 8 of which were significantly correlated ($P<0.05$): X_1; X_4; X_5; X_7; X_{11}; X_{12}; X_{13}; and X_{19}. To avoid the influence of correlation between the geographical factors and reduce the accuracy of the regression function after establishment, a heat map was developed to judge the degree of collinearity among the 11 geographical factors (Figure 2).

When the correlation is greater than 0.9, there is significant collinearity among the factors, so it cannot simply be predicted by a linear model (Adnan *et al.*, 2006). As can be seen in Table 4, latitude has a strong collinearity effect on annual mean temperature, annual precipitation and annual temperature range. Otherwise, the predicted result would not be consistent with real situation. In this case, ridge regression and support vector regression were chosen to build predictive models since these two methods could solve collinearity problem.

Ridge regression analysis

In SAS software, 11 geographic factors related to the TSH reference values are used as independent variables. TSH reference values were used as the dependent variable to perform ridge

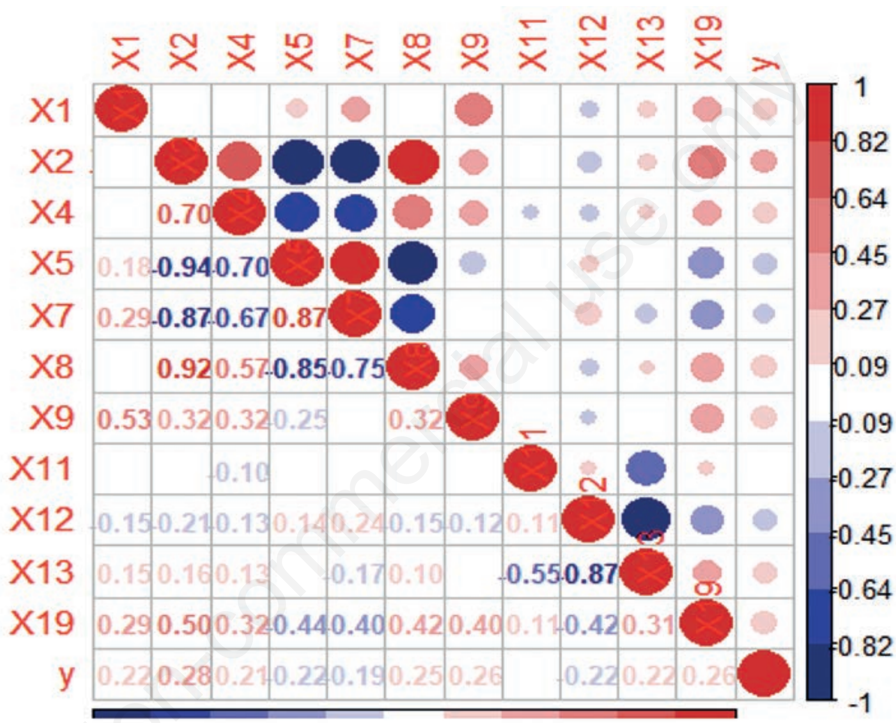

Figure 2. Heat table showing the correlation coefficients between the geographical factors. Red signifies different degrees of high correlation and blue the opposite.

Table 4. Correlation between 11 local different environmental factors.

	X_1	X_2	X_4	X_5	X_7	X_8	X_9	X_{11}	X_{12}	X_{13}	X_{19}
X_1	1	0.11	0.035	0.013	0.212	0.158	0.69	–0.053	–0.108	0.162	0.427
X_2	0.11	1	0.765	–0.941	–0.9	0.929	0.255	–0.102	–0.144	0.159	0.484
X_4	0.035	0.765	1	–0.794	–0.734	0.69	0.233	–0.101	–0.022	0.075	0.317
X_5	0.013	–0.941	–0.794	1	0.923	–0.841	–0.174	0.057	0.078	–0.108	–0.442
X_7	0.212	–0.9	–0.734	0.923	1	–0.799	–0.003	0.045	0.143	–0.131	–0.332
X_8	0.158	0.929	0.69	–0.841	–0.799	1	0.28	–0.046	–0.1	0.087	0.396
X_9	0.69	0.255	0.233	–0.174	–0.003	0.28	1	–0.064	–0.018	0.141	0.437
X_{11}	–0.053	–0.102	–0.101	0.057	0.045	–0.046	–0.064	1	0.029	–0.369	0.007
X_{12}	–0.108	–0.144	–0.022	0.078	0.143	–0.1	–0.018	0.029	1	–0.849	–0.389
X_{13}	0.162	0.159	0.075	–0.108	–0.131	0.087	0.141	–0.369	–0.849	1	0.431
X_{19}	0.427	0.484	0.317	–0.442	–0.332	0.396	0.437	0.007	–0.389	0.431	1

regression analysis to draw a ridge plot (Figure 3). In the ridge trace, the ridge parameter K is taken as the vertical axis, and the regression coefficient of each factor is taken as the vertical axis. When the ridge parameter K=0.1, the trend of the ridge trace becomes stable, so choose K=0.1 to get the regression equation:

$$\hat{Y}= -2.910+0.01891X_1+0.02010X_2+0.00005000X_4-0.006300X_5-0.00002000X_{7+}0.1006X_8+0.001040X_9+1.033X_{11}+0.4790X_{12}+0.0001300X_{13}+0.003180X_{19}\pm 0.9138 \quad (5)$$

of which $\hat{Y}$ is the predicted value of the TSH reference value of Chinese healthy adults, x_1-x_{19} the geographical element values, and 0.9138 is the remaining standard deviation.

Support vector regression

In Clementine12.0 software, the TSH reference value was output from source data, and 11 relevant geographical factors were used as input variables. Different kernel functions under the SVM module were selected for modelling analysis. Since the data were nonlinear, the following three nonlinear kernel functions were used (Table 5). Combined with the different values of MAE and MSE corresponding to each kernel function. Therefore, the SVM model of RBF kernel function is used as the optimal prediction model of TSH reference value.

Evaluation results

The two methods mentioned above were used to evaluate the prediction accuracy and model quality of each ridge regression model (Figure 4) and SVM model (Figure 5), respectively. The measured and predicted values of 12 cities in China such as Beijing, Guangzhou, Shanghai, Chengdu, Nanjing, Changsha, Lanzhou, Foshan, Shijiazhuang, Urumqi, Xi an, Rizhao were selected for comparison, and the mean difference of the two groups of data was compared to judge its significance according to the T-distribution theory.

In SPSS 19.0, K-S test found that the measured TSH values of healthy adults in China were normally distributed, so the paired sample t-test method could be used to prove whether the reference values of TSH of healthy adults in China were consistent with the predicted values under these two models.

In order to improve the accuracy of sample t test, the CI is 99%. The results are as follows: the significance of the lower ridge regression model of the two models was $P=0.008$ ($P<0.05$) and that of the SVM model is $P=0.003$ ($P<0.05$) indicating that there was a significant linear change between the measured and predicted values under the prediction of the two models, and the linear correlation is strong. The Ridge regression model's t-test $P=0.781$ was larger than the P (0.725) predicted by the support vector machine model, and the difference between the significance of the double-tail and the subtraction from 0 was large, indicating that there was a good fit between the predicted value and the measured value.

Table 5. Variable importance of each model with different kernel types.

Kernel	MSE	MAE
RBF	0.390	0.357
POL	0.650	0.518
SIG	17.935	2.748

MSE, mean square error; MAE, mean absolute error; RBF, radial basis function; POL, polynomial kernel; SIG, linear kernel function.

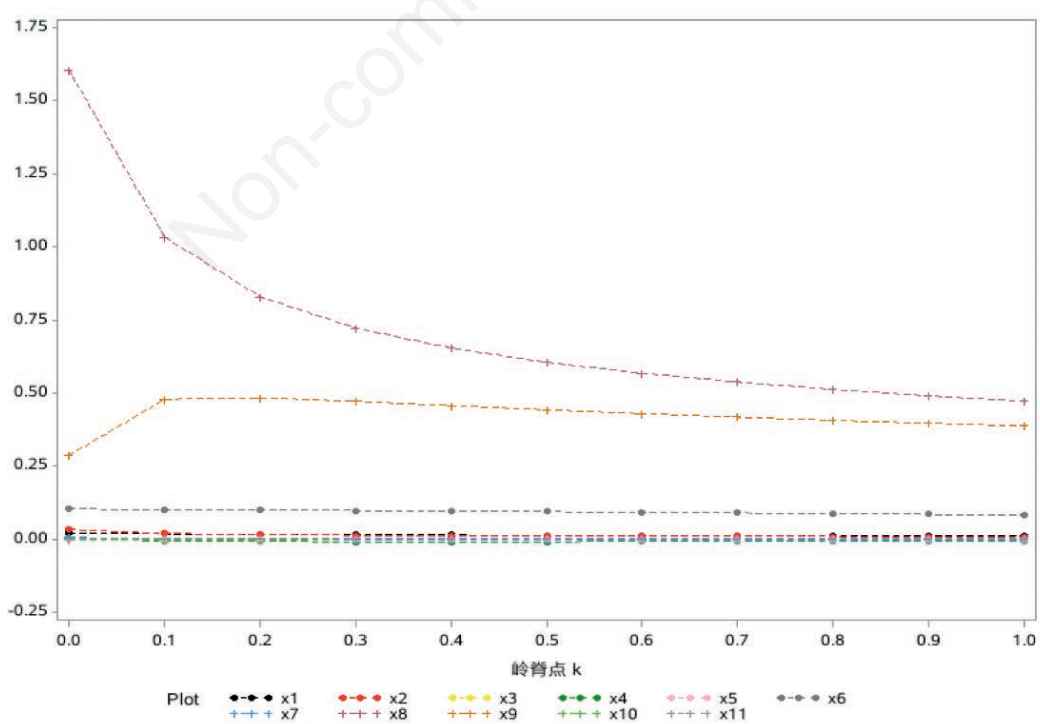

Figure 3. Ridge tracing of the thyroid stimulating hormone levels in healthy adults.

Therefore, the Ridge regression method was selected to predict the TSH reference value of 2322 data points across the country.

Statistical analysis

The predicted TSH values in 2322 counties and cities in China were imported into the vectorized map of China. The histogram analysis was carried out first, and the mean value was used to replace the same position to obtain the reference value of TSH of healthy adults in China, which presented a normal distribution. The trend distribution chart of TSH in China was drawn (Figure 6). In the (Y axis), the fitting curve showed an obvious trend of decreasing from north to south, while in the (X axis) east-west direction, it slightly decreased and then increased. The fitting curve showed an obvious trend of concave. In order to further observe the changes of reference values in different geographical regions, the geospatial distribution map of TSH reference values is also needed. The K-S test found that the TSH reference values of healthy adults in China were normally distributed, so the ordinary Kriging interpolation under geostatistical analysis was used.

The reference value of TSH for healthy adults in China ranges from 1.59-3.55 m U/L, in which the value in Northeast China (Heilongjiang, Liaoning and Jilin provinces) is higher and is represented by the red area. Southwest China (Sichuan Province, Guizhou Province, Xizang Province) and South China (Hainan Province, Guangxi Province) have low values, which are represented by green areas. The rest of the region is the median, shown

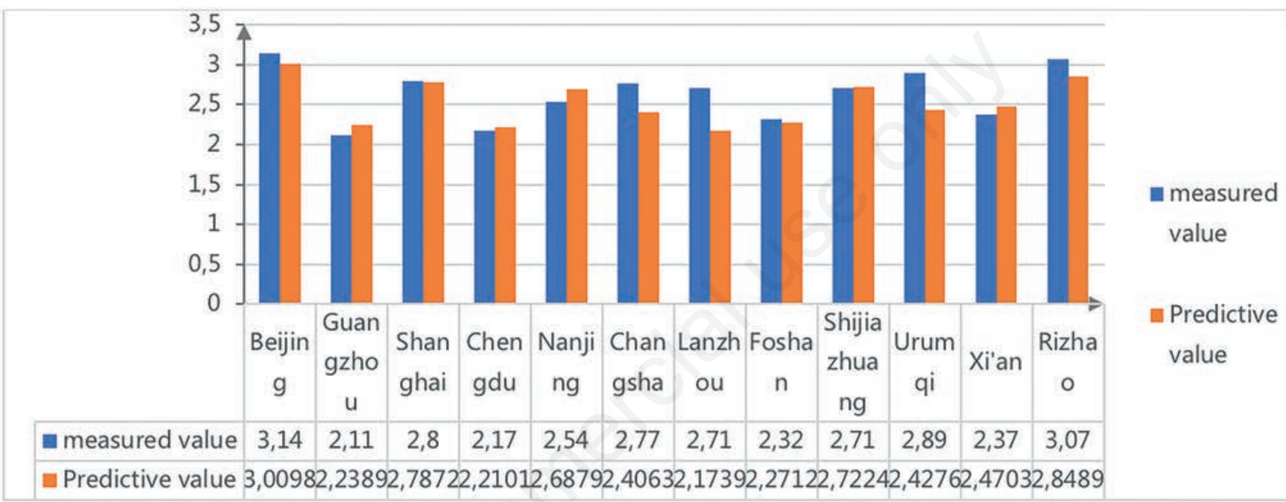

Figure 4. Ridge regression model of some urban thyroid-stimulating hormone values and forecast value comparisons.

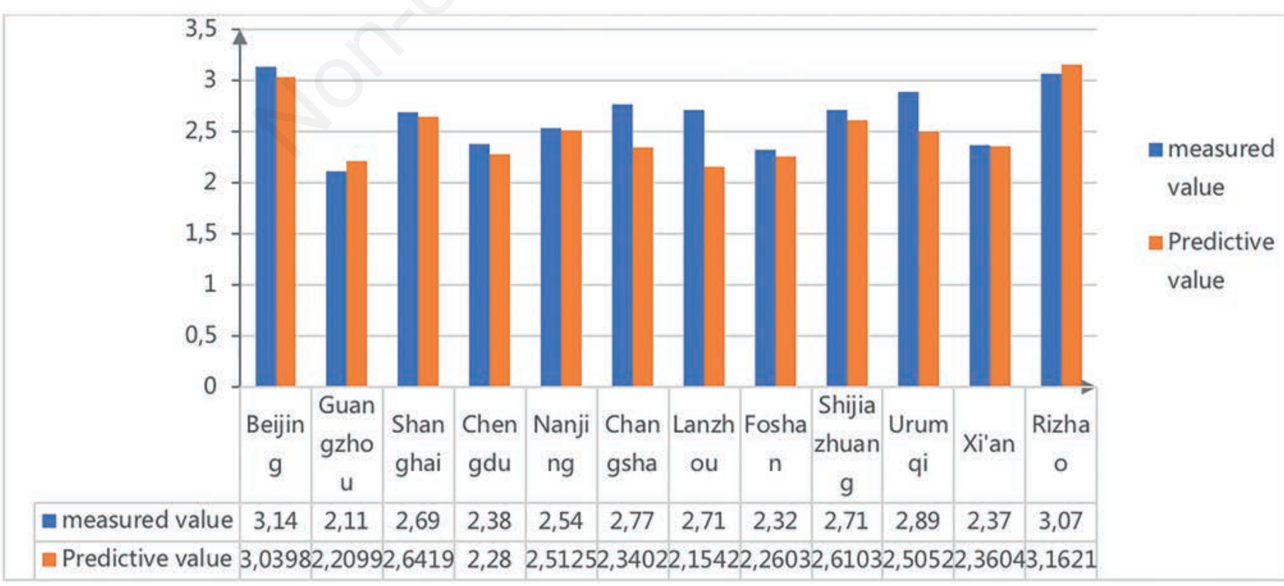

Figure 5. The support vector machine model for comparison of measured and predicted thyroid-stimulating hormone values in selected Chinese cities.

in yellow. Areas with the same colour have strong similarity, and the distribution of values is significantly different in latitude. The huge difference in hue between the south and the North indicates a significant difference in TSH reference values (Figure 7). The overall spatial pattern is high in the north and low in the south, high in the west and low in the east.

Discussion

Thyroid autoregulation is a sensitive and strict regulatory feedback control system. According to the 'Theory of Three Causes' proposed Wei *et al*. (2020) time, person and place differ according to variations in the natural geographical environments. Likewise,

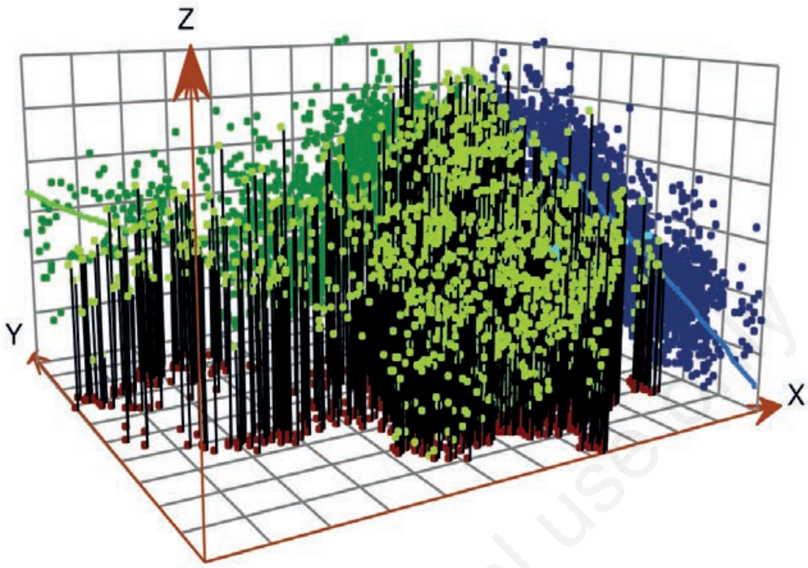

Figure 6. Trend surface analysis of thyroid stimulating hormone reference value of Chinese healthy adults.

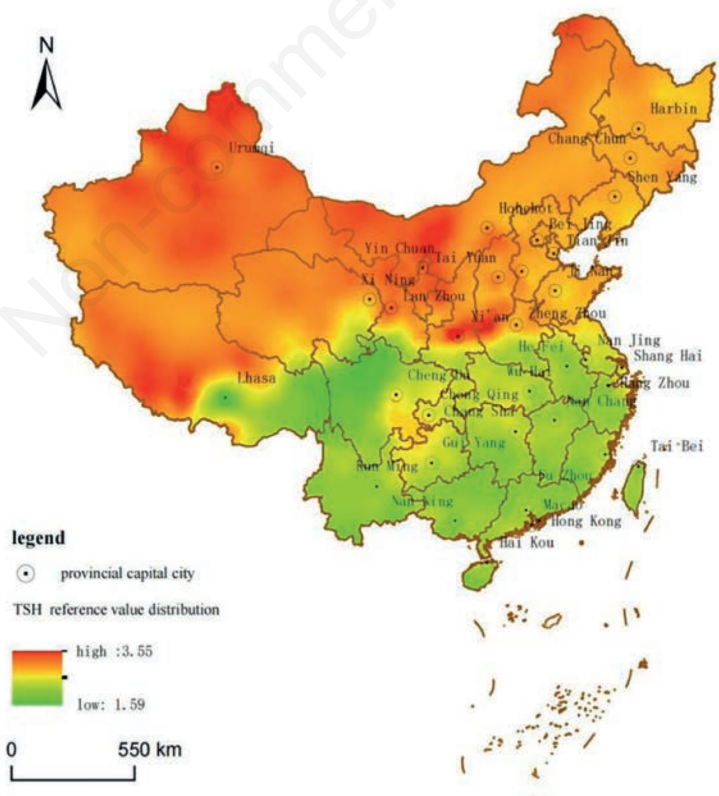

Figure 7. Geographic distribution of thyroid stimulating hormone reference value of Chinese healthy adults.

warm, hot and cold climates influence people's physiological indexes and the metabolic functions become different (Zimmermann and Boehlert, 2015). Indeed, annual sunshine hours, annual average temperature, annual precipitation, annual temperature and the annual average wind speed were found to be non-negligible factors affecting the TSH value of healthy people.

Result analysis

In spatial autocorrelation analysis, the significance level P value was 0.005412. The results suggests that TSH reference value of Chinese healthy adult is significantly correlated with geographical environment; accordingly, the TSH reference values vary with spatial attributes, it is feasible to explore the relationship between them and build predictive models.

According to correlation analysis, there were eleven geographical indices. In the case, the Ridge regression and SVM, which could effectively resolve collinearity problem, were used to construct predictive model, respectively. After constructing predictive model, selecting optimal model, prediction, and interpolation, the spatial distribution map was obtained. As shown in the Figure 7, the TSH reference value is high in the north and low in the south, high in the west and low in the east. These results show that the latitude is the most significant factor, which is consistent with the result of correlation analysis. From the perspective of single-phase correlation coefficient combined with geographic factors, the discussion can be carried out from the following three aspects: topographic indicators, climate indicators and soil indicators.

Topographic and geographic impact

China has a vast territory and abundant resources, which has a wide span of longitude and latitude. Latitude directly affects the distribution and climate types of temperature zones in the North and South of China. China's climate types are complex and diverse, with significant characteristics of continental monsoon climate, temperature decreases from South to North, and there are large temperature differences. Along all longitudes, the TSH value decreases as we move from high latitudes to low latitude (Gao and Li, 2019) noted that the winter is cold and long in north-eastern China and the summer is warm and short. It can be argued that people's brains in cold areas release impulses that stimulate the hypothalamus to release thyrotropin (Wang and Shan, 2021). Our results support the idea that that thermogenesis increases when the metabolism is raised eventually leading to TSH increase, which is consistent with the spatial pattern of the geographic spatial TSH distribution map, which showed higher values in the North and lower in the South. Annual sunshine duration is another closely related to latitude and altitude, and these geographical factors are interrelated and inseparable. Some surveys have shown that human serum cholesterol increases in migrants and people living in high-altitude areas, suggesting that it is consistent with hypothyroidism and the value of thyrotropic hormone is low as shown in Figure 7 in Lhasa, Qinghai Province (Gao, 2005). Studies have pointed out that the hypothalamic-pituitary-thyroid system has a chronic adaptation mechanism to the hypoxia environment at high altitude, and the reduced thyroid function has an adaptive protective effect on maintaining a low metabolic rate and avoiding excessive metabolic consumption (Peng *et al*., 2006). High altitude hypoxia can significantly reduce the ability of mitochondrial ATP synthesis in various oxidized substrate systems *in vitro*, resulting in reduced ATP synthesis and iodine uptake by iodine pump, leading to hypothyroidism.

Climate influence

Changes in ambient temperature can cause increased thyrotropin secretion leading to higher serum TSH concentration and metabolism (Zhou *et al*., 2019). selected the TSH reference data of 206 486 healthy people in Beijing Union Medical College Hospital and found that there was a significant seasonal and temperature dependence. The TSH reference value showed a peak value in winter and a low value in summer, which is consistent with the annual average temperature and annual difference of temperature as relevant factors. Foreign scholars (Das *et al*., 2018) measured 401 patients with normal thyroid and found that there was an inverse relationship between thyrotropin secretion and seasonal changes.

The average wind speed also has a significant impact, with the wind speed in winter and spring resulting in more obvious effects than that in summer and autumn (Nakayama and Yoshimura, 2017). In low temperature and high humidity environments, excessive wind speed will lead to excessive heat dissipation of human body, increasing the basic metabolism resulting in increased level of the thyroid stimulating hormone. Overall, the annual mean wind speed is relatively high in the north-eastern part of China, particularly in Inner Mongolia and the northern part of Xinjiang, and the value of thyrotropic hormone is also relatively high in these areas, which is consistent with the trend in the geographical distribution chart. Annual precipitation and humidity are two related climate factors, and some studies have pointed out that humidity and other factors can also affect the level of human thyroid hormone. Excessive humidity will stimulate the pineal gland to secret pineal hormone, resulting in a relatively lower TSH concentration in the body leading to people showing fatigue and listlessness

Influence of soil constituents

The TSH values are mostly indirectly related to soil factors such as silt percentage, clay-particle percentage, reference capacity of surface soil and cation exchange capacity of surface soil (clay), rather than factors such as longitude, latitude, temperature and humidity that directly affect the indices. The soil types in China have provincial and regional rules (Ma, 1992), presenting alternative changes under different landforms of plains, basins and plateaus in China. Soil types in different areas produce different crops (Cai and Ma, 1988), *e.g.*, with wheat in the North and rice in the South. This fact affects the eating habits in China, and the diet reflects the level of iodine intake to a certain extent. For example, in hot and dry areas, people eat more fruits and vegetables, which do not contain much iodine content. Sweat glands regulate body temperature by secreting sweat, and when the thyroid function is reduced, the thyroid stimulating hormone is less secreted. The cation exchange capacity of topsoil is an important index for evaluating soil fertility. High soil fertility will result in a higher content of protein, cellulose and trace elements, such as iodide. Thus, agricultural products produce a more balanced human diet, promoting the development of organs such as thyroid gland, and thereby affecting the secretion of thyroid stimulating hormone (Jiang, 2004). The spatial distribution of health has become one of the important contents of medical geography research. With the rise of interdisciplinary disciplines such as environmental health science, medical geology and human phenology, healthy people have gradually become the focus of medical geography. A series of studies have proved that the geographical environment may affect the human body (Yang *et al*., 2013). The reference value of TSH in this study does not include diet. Human and geographical factors such as nutrition need to be improved and discussed in future research. Due to the complexity of human environment and geographical environment, the influence mechanism between them also needs further research and analysis.

Conclusions

By the correlation analysis, it can be concluded that TSH reference values of healthy Chinese adults has a significant correlation with several environmental variables: longitude, latitude, annual sunshine duration, annual mean temperature, annual precipitation, annual temperature range, annual mean wind speed, soil particle percentage, soil clay percentage, soil reference bulk density and soil (silt) cation exchange capacity. Extraction of relevant geographical factors to build models shows that the SVM model is better than the ridge regression model for predicting TSH values, but there is no significant difference between measured values and predicted ones in the SVM model. By the spatial distribution of TSH values (Figure 8), we can see that the values are equal. However, the faster the environmental variables change, the greater the TSH values adjust. The trend of TSH is decreasing from North to South.

References

Adnan N, Ahmad MH, Adnan R, 2006. A comparative study on some methods for handing multicollinearity problems. Mat 22:109-19.

Boser BE, Gyon IM, Vapnik VN, 1992. A training algorithm for optimal margin classifiers. In COLT '92: Proceedings of the fifth annual workshop on Computational learning theory. Assoc Comput Mach 144-52.

Cai CZ, Ma YJ, 1988. Relationship between soil organic matter and soil cation exchange capacity. Prog Soil Sci 3:10-5.

Das G, Taylor PN, Javaid H, Tennant BP, Geen J, Aldridge A, Okosieme O, 2018. Seasonal variation of vitamin d and serum thyrotropin levels and its relationship in a euthyroid Caucasian population. Endocr Pract 24:3-59.

Dong XG, Zhao L, Lin SM, 2017. Analysis of factors of agricultural output value under the return of Ridge and the return of the main components. J Changchun Univ Technol 38:1-7.

Dou YG, Mu XD, Sun W, Cao Y, Wang Y, Liang K, 2020. Clinical characteristics analysis of thyroid disease patients in Lanzhou City, 2008-2018. Chin J Endemiol 39:602-5.

Gao H, Li CJ, 2019. Advances in the study of thyroid-stimulating hormones and cardiovascular diseases. Adv Cardiovasc Dis 40:191-4.

Gao YQ, 2005. Highland military medicine. Chongqing: Chongqing Press, China.

Ge M, Zhang YP, Wang X, 2009. Relationship between reference value of red blood cell count and geographical factors in male infants. J Jilin Univ Med Ed 35:751-4.

Hoogendoorn EH, Hermus AR, De Vegt F, Ross HA, Verbeek AL, Kiemeney LA, Swinkels DW, Sweep FC, den Heijer M, 2006. Thyroid function and prevalence of anti-thioperoxides antibodies in a population with borderline sufficient iodine intake: influences of age and sex. Clin Chem 52:104-11.

Ji LB, Shen C, Yang GR, 2020. Relationship between iodine and thyroid diseases. Chin J Control Endem Dis 35:120-3.

Jiang JQ, 2004. The relationship between environmental geochemistry and human health and agricultural production. Geophys Geochem Explor 284:3300-22.

Laurberg P, Jørgensen T, Perrild H, Ovesen L, Knudsen N, Pedersen IB, Rasmussen LB, Carlé A, Vejbjerg P, 2021. The Danish investigation on iodine intake and thyroid disease, Anther: status and perspectives. Eur J Endocinol 155:219-28.

Li DC, Chen WC, Liu CW, 2012. A non-linear quality improvement model using SVR for manufacturing TFT-LCDs. J Intell Manuf 23:835-44.

Li JJ, Zhang L, Li J, 2020. The relationship between serum BDNF, GABA and TSH levels in patients with bipolar depression and anaphase depression. Chin J Labor Diagn 24:1781-4.

Liu XZ, Zhang AD, Li JZ, 2009. Mathematical methods in geography, Beijing: Science Press, 32pp.

Ma JH, 1992. Re-understanding of soil distribution laws. J Henan Univ 1:1-5.

Marr-Lyon LR, Gupchup GV, Anderson JR, 2012. An evaluation of the psychometric properties of the Purdue Pharmacist Directive Guidance Scale using SPSS and R software packages. Res Soc Admin Pharm 8:166-71.

Nakayama T, Yoshimura T, 2017. Seasonal rhythms: the role of thyrotropin and thyroid hormones. Thyroid 28:4-10.

Paul B, De SK, Ghosh AK, 2021. Some clustering-based exact distribution-free k-sample tests applicable to high dimension, low sample size data. J Multivar Anal 104897.

Peng H, Li SZ, Wang HB, 2006. Survey of normal thyroid hormone in healthy adults in highland areas. Gaoyuan Yixue Zazhi 3:45-6.

Sun MZ, Jia YM, 2015. Research on application of origin software in experimental data processing. Res Explor Lab 10:96-97.

Tang GA, Yang X, 2006. ArcGIS spatial analysis tutorial. Beijing: Science Press, 203 pp.

Wang XC, Shan ZY, 2021. [Role of serum thyroid-stimulating Hormone in the diagnosis of subclinical hypothyroidism and its influencing factors]. Chin J Intern Med 60:925-8. [In Chinese].

Wei DZ, Ge M, Wang CX, Lin QY, Li MJ, Li P, 2016. [Geographical distribution of blood creatinine reference values in healthy adults]. J South Med Univ 36:1555-60. [In Chinese].

Xu JH, 2002. Mathematical methods in modern geography. Beijing: Higher Education Press, 48pp.

Yang QS, Kevin M Mwenda, Ge M, 2013. Incorporating geographical factors with artificial neural networks to predict reference values of erythrocyte sedimentation rate. Int J Health Geogr 12:1-7.

Yang YH, Liu BX, Wan YC, 2016. A decision tree-based mobile communication customer churn analysis method. Chin Manage Inf Technol 19:70-1.

Yin QH, 2016. Study on the correlation between TSH and TPOAb levels and blood pressure and glycolipid metabolism in people with normal thyroid function. J Anhui Medical Univ 03:4-6.

Yin XF, Huang S, Wang Y, 2021. Different thyroid hormone reference intervals were applied to the clinical value of healthy adults in Xiamen area. Exp Lab Med 39:118-20.

Yuan LL, Peng NC, 2018. Advances in the study of the correlation between TSH and cardiovascular disease and risk factors. The Med Forum 22:987-9.

Zheng ZL, Sun H, Gai BD, 2008. Hotspots and difficulties in surgical treatment of thyroid diseases. Chin J Bas Clin Gen Surg 5:307-10.

Zhou P, Zhang SH, Wei X, 2019. Effects of seasonal changes on thyroid hormones. Int J Endocrinol Metab 5:349-52.

Zhu YZ, Luan ZY, 2020. The establishment of the reference range for thyroid hormone levels among Uighur population in Xinjiang region. Chin Pract Med 15:166-8.

Zimmermann MB, Boehlert K, 2015. Iodine deficiency and thyroid disorders. Lancet Diabetes Endocrinol 13:286-95.