Associations between taste preferences and chronic diseases: a population-based exploratory study in China

Hanqi Li[1] , Peng Jia[2,3] and Teng Fei[1],*

[1]School of Resource and Environmental Sciences, Wuhan University, Wuhan, Hubei 430079, China: [2]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong kong, China: [3]International Initiative on Spatial Lifecourse Epidemiology (ISLE), Hong Kong, China

Abstract

Objective: To explore the associations between dietary tastes and chronic diseases quantitatively.

Design: We used the Geodetector method to establish associations between seven tastes and a variety of chronic diseases from the perspective of spatial stratified heterogeneity and explained the effects of dietary tastes on the spatial distribution of chronic diseases.

Setting: We used crowdsourcing online recipe data to extract multiple taste information about cuisines, combined with point of interest data on categorised restaurant data in different regions, to quantitatively analyse the taste preferences of people in different regions.

Participants: Crowdsourcing online recipe data and restaurant data in different regions.

Results: The results showed that sixteen diseases were significantly associated with dietary tastes among the seventy-one types of chronic diseases. Compared with the effects of individual tastes, the interactions of tastes increased the risk of sixteen diseases, and many combinations of tastes produced nonlinear enhancement effects on the risk for diseases.

Conclusions: This study presents a quantitative study approach based on the crowdsourcing of data to explore potential health risk factors, which can be applied to the exploratory analysis of disease aetiology and help public health authorities to develop corresponding interventions.

Chronic diseases, also known as noncommunicable diseases, are the leading causes of disease and death worldwide. They cause 41 million deaths every year, equivalent to 71 % of global deaths[1,2]. In 2016, the mortality rate of chronic diseases among Chinese residents was 739/100 000, accounting for 89 % of total deaths, which was much higher than that caused by infectious diseases, traffic accidents, etc.[3]. Chronic diseases have devastating impacts on society, economy and public health[4,5]. The WHO has pointed out four modifiable behavioural risk factors for chronic diseases, including unhealthy diet, smoking, alcohol consumption and physical inactivity[1]. An unhealthy diet is of primary risk and is the most prevalent factor affecting the occurrence and development of chronic diseases[6,7]. The Global Burden of Disease Study 2017 showed the number of deaths attributed to unhealthy diet worldwide reached 10·9 million[8]. Therefore, a comprehensive estimation and comparison of the disease risk of dietary factors can help to develop specific and effective disease prevention and control strategies[9,10].

A large number of studies have analysed people's eating and drinking behaviours, including the time and frequency of eating, the collocation of meat and vegetables, the intake of dairy products, salt and oil, and the proportion of coarse and fine grains[11–14]. However, there were only a few studies related to dietary tastes, and their taste qualitative indexes were based on subjective evaluation, which was mainly due to the scarcity of systematically collected databases[14,15]. As a result, the quantitative studies on the relationship between taste preferences and chronic diseases have been extremely deficient.

Corresponding author: Email feiteng@whu.edu.en

It is true that the taste preference surveys based on individuals are costly and complex to conduct. In this respect, increasingly popular crowdsourcing data can provide a new perspective for quantitative taste research – taste preference quantisation based on population. In practice, due to the advantages of ample sources, sizable amounts and comprehensiveness, crowdsourcing data have been widely used in public health and medical fields, such as disease prediction, evidence-based public health decision-making, health management and monitoring, and personalised medicine[16–18]. The use of crowdsourcing data to mine the overall taste preferences of the residents in a region can not only eliminate the impact of undersampling and reduce the cost of data collection but also show the overall characteristics of regional taste preferences more effectively.

China has a vast territory, large population and abundant resources. Different products, climates, religions and customs have bred unique regional food culture, with each province boasting specialties that form completely separate cuisines[19–23]. Tastes, the core of Chinese food, differ from cuisine to cuisine. To some extent, the taste characteristics of the cuisine in a province can reflect the taste habits and preferences of the local people. Therefore, China is selected as our target area to carry out studies related to regional taste differences.

In this study, we extract and quantify people's taste preferences by combining the crowdsourcing online recipes and restaurant data of each province in China. The association between chronic diseases and dietary tastes was established through the Geodetector method[24]. The results provide basic data for exploring the impact of dietary taste preferences on the occurrence of chronic diseases. The present study addresses gaps in the quantitative studies related to tastes and chronic diseases, and it is a new attempt to apply crowdsourcing data, instead of individual surveys, to obtain research information in the field of public health.

Data and methods

Data

Crowdsourcing online recipes
Dietary taste data were obtained from Meishijie (http://www.meishij.net/), which provides structured recipe data and divides all dishes into a total of twenty cuisines according to their origins. In this study, the seasoning list of each recipe was used as the main data source for extracting dietary tastes. For example, the seasonings used in the Sichuan dish 'Poached Spicy Slices of Pork' include bean paste, light soy sauce, salt, dried chili, sugar, Chinese prickly ash, oil and garlic, as shown in Fig. 1.

Points of interests of restaurants
Points of interests (POI) of restaurants were retrieved from Amap (https://www.amap.com/), an electronic map website, in August 2017. More than 7 million POI have been obtained in thirty-four provincial administrative regions in China. Each POI records information, including the name, detailed location coordinates and cuisine of the restaurant (such as 'Sichuan cuisine'). Such information was used to analyse the proportion of each cuisine in each province.

Mortality of chronic diseases
Death cause surveillance served as one of the data sources of chronic disease[25]. The data used in this study came from the 'Atlas of Death Cause by Major Diseases of Residents in China', which was published in 2013 by the National Centre for Chronic and Noncommunicable Disease Control and Prevention (http://ncncd.chinacdc.cn/)[26]. Cause-specific mortality (unit: 1/100 000) refers to the number of deaths caused by a certain disease per 100 000 people in a given region, which reflects the degree of harm of various diseases and injuries to the residents' lives[27]. The data include multiple causes of death, covering thirty-three provincial administrative regions in China (no data available in Taiwan Province). In this study, seventy-one categories of chronic diseases were selected from all death causes as study subjects.

Methods
The data processing and research flow of this study are shown in Fig. 2. First, the seven taste indexes of each cuisine (Cuisine-Taste Matrix) were set up by analysing the recipe data. Through the statistics of the restaurant POI data, the proportions of various cuisines in each provincial region (Region-Cuisine Matrix) were obtained. Then, the seven taste indexes in each provincial region (Region-Taste Matrix) were calculated based on the above results. Finally, the Geodetector method was used to detect the association between the seven tastes and mortality of each chronic disease.

Extraction of cuisine taste
A web crawler was used to collect the recipe for each dish of every cuisine. Since there was no standardised measurements of the amounts of seasonings, we quantified the taste from the perspective of seasoning frequency of use.

First, the ingredients and seasonings of all recipes were gathered and merged (if the same item had multiple names) to determine the types of tastes used in this study. Based on the existing research on taste science, and by combining the particular context of Chinese food culture, we selected seven dimensional taste or sensation measurements to establish the 'ingredients-taste' reference table. They were 'sour', 'sweet', 'umami', 'salty', 'fat', 'spicy' and 'pungent'[28–31]. Technically, 'spicy' and 'pungent' are better acknowledged as flavours but not tastes. But in China, people often think of them as two common tastes paralleling the other five tastes. So, in this study, these two sensations are also called tastes for convenience. The 'spicy' sensation here refers to the 'hot' feeling of capsicum, while the
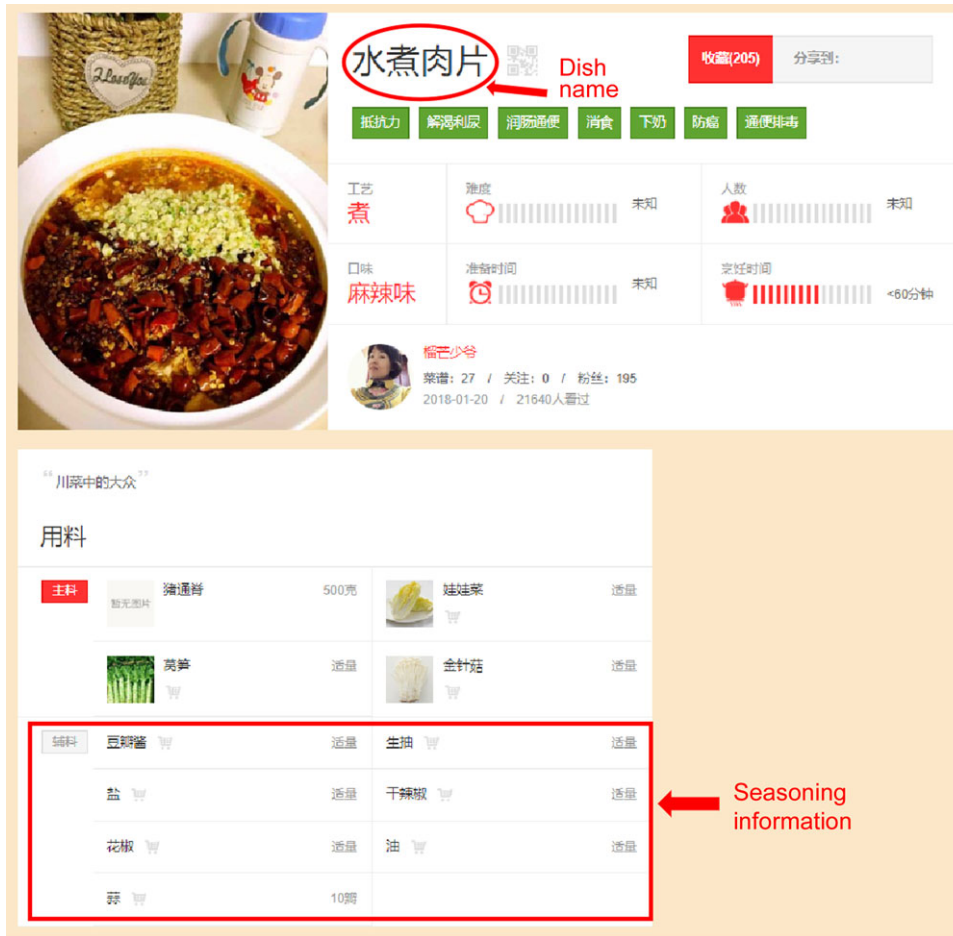
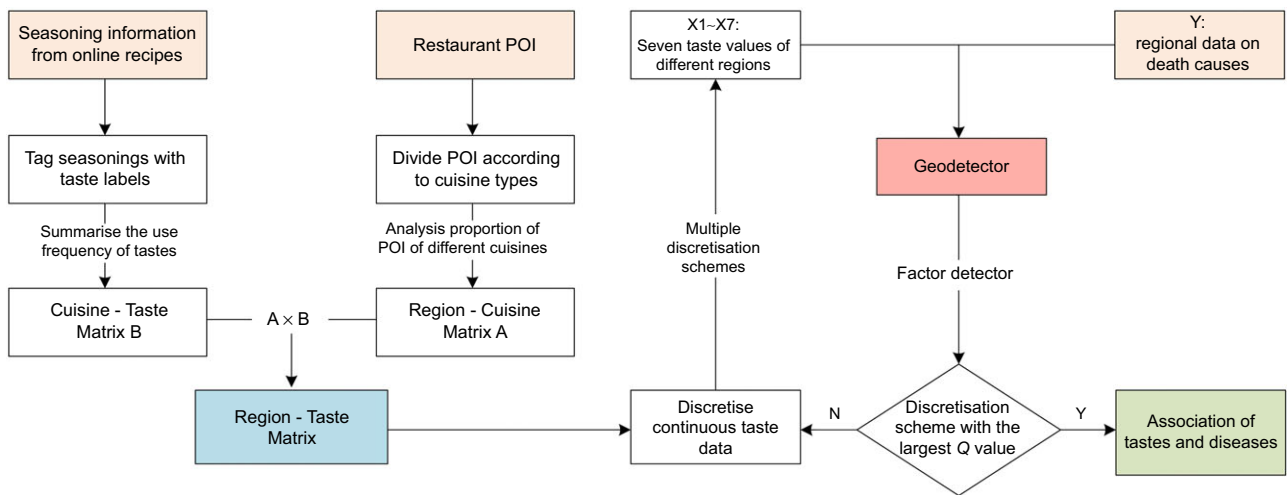**Fig. 1** Seasonings of a dish (captured from Meishijie.com)

Fig. 2 Data processing and research flow

'pungent' sensation refers to various peppery and irritating feelings of various sources other than capsicum, such as mustard, garlic and *Zanthoxylum bungeanum* Maxim.

Then, after deleting the tasteless ingredients, every ingredient or seasoning was labelled with one or more tastes. For example, the seasoning 'seafood sauce' was labelled with 'salty' and 'umami.' If some different seasonings with a certain taste were used in one dish at the same time, the dish would be labelled with that taste several times. The labels of seven tastes for every dish were

counted to calculate the frequency of use of these seven tastes for every dish.

Finally, through summarising all the dishes of each cuisine and taking the average of each taste's frequency of use in every dish, seven taste quantitative indexes for each cuisine were obtained.

Calculation of the taste indexes in each region

The sixteen cuisines that the POI used to classify the categories of restaurants are all included in the twenty cuisines covered by the recipe data. In this study, we used the POI corresponding to these sixteen cuisines that were shared by the two groups of data to extract regional tastes. There were 387 509 valid POI corresponding to the sixteen cuisines, with Guangdong and Macao having the largest and the least number of valid POI – 63 420 and seventy-one, respectively. The proportions of the POI of these sixteen cuisines to the total valid POI in the thirty-three provincial regions were analysed, and the values of the seven taste indexes of each provincial region were calculated according to the taste quantitative indexes of each cuisine. Taking 'spicy' in Sichuan Province as an example, the proportion of the POI of Sichuan cuisine to the total valid POI in Sichuan Province was taken as the weight and multiplied by the spicy index of Sichuan cuisine to obtain the contribution of Sichuan cuisine to the spicy taste in Sichuan Province. By analogy, every cuisine's contribution to the spicy taste in Sichuan Province was obtained and added together to obtain the spicy index in Sichuan Province.

Pre-processing of taste data

Discrete taste explanatory variables are required when using the Geodetector method, while the taste indexes of regions are continuous. Therefore, taste indexes need to be discretised to meet the precondition of using this analysis method. In this study, we adopt the discretisation method introduced by Cao *et al.*[32], and the number of provincial regions within each level of each discretisation method was guaranteed to be at least two[33]. Finally, thirty-five discretisation schemes were formed by sequentially combining the five discretisation algorithms with the seven hierarchical approaches. These five discretisation algorithms were equal interval, geometrical interval, natural break, quantile and SD. The seven hierarchical approaches referred to dividing the taste indexes into two to eight levels, and thirty-five discretisation schemes were tested on each taste individually.

Using the above approach, the discretisation of a certain taste index was performed, which was equivalent to apply the hierarchical clustering to thirty-three provincial regions according to the frequency of use of that taste, and the entire study area was divided into sub-regions with different taste levels.

Geodetector

The Geodetector was proposed in 2010 to measure the spatial stratified heterogeneity of geographical entities, as well as to reveal the driving forces behind spatial stratified heterogeneity[33]. This method was first applied to assess the environmental risks of an endemic disease, such as neural tube defects, and has been widely used in landscape science, social science, land use, environmental science and human health in recent years[24,34]. In this study, the Geodetector was introduced to explore the association between dietary taste preferences and the occurrences of chronic diseases in different regions.

The Geodetector includes three detectors. The factor detector identifies factors that are responsible for the risk and the power of determining the risk factors for the disease, which is expressed by the *q*-value shown in equation (1). The interaction detector reveals whether the risk factors interact or lead to disease independently. Lastly, the risk detector indicates whether there are significant differences in disease occurrences between sub-regions with different taste levels.

$$q = 1 - \frac{\text{SSW}}{\text{SST}} = 1 - \frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi} - \overline{Y_h}\right)}{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)} \qquad (1)$$

In equation (1), SSW represents the sum of the variances of the disease occurrences for the regions with different taste levels, SST represents the total variance of the disease occurrences for all regions, *N* represents the number of regions and *h* represents the level of a factor (*h*=1,2,…,*L*). In this study, the discretisation scheme with the largest *q*-value was preferred[32], and the optimal number of levels is equal to *L*; $\overline{Y}$ and $\overline{Y_h}$ represent the mean disease occurrence for all regions and for the regions with the taste level *h*, respectively; Y_i and Y_{hi} represent the disease occurrence for the *i*-th region of all regions and for the *i*-th region with taste level *h*, respectively. Note that the range of *q*-values is (0, 1). *q* = 1 means that the geographical division of the factor completely explains the spatial pattern of the disease, whereas *q* = 0 implies a completely random spatial occurrence of the disease.

In this study, there are thirty-three provincial regions (*N* = 33). The disease occurrences of thirty-three provincial regions were used as response variables, and the discrete indexes of seven tastes for all regions were used as independent variables to conduct the Geodetector. For a certain taste, the higher *q*-value represents that this taste factor has a stronger ability to explain the spatial distribution pattern of the disease and can better reveal the effect of taste on the occurrence of the disease.

Results

Cuisine taste measurement

The numbers of the recipes of twenty cuisines obtained by web crawling in this study are shown in Table 1. For each cuisine, seven quantitative taste indexes were visualised by a colour chart. By comparing the seven tastes, this colour

Table 1 Numbers of recipes and cuisine taste indexes

Unique ID	Usage	Cuisine Type	Number of recipes	Sour	Sweet	Umami	Salty	Fat	Spicy	Pungent
1	Sixteen cuisines contained in both recipes and POI data	Chuan	1008	0.27	0.54	0.51	1.75	0.72	0.95	2.23
2		Dongbei	896	0.25	0.43	0.64	1.61	0.68	0.26	1.34
3		Gangtai	67	0.15	0.78	0.6	1.66	0.84	0.34	1.79
4		Hubei	233	0.22	0.48	0.85	1.59	0.89	0.26	2.35
5		Hu	328	0.19	0.64	0.91	1.43	0.97	0.06	1.38
6		Hui	327	0.17	0.6	0.69	1.18	0.77	0.13	1.16
7		Jing	360	0.21	0.58	0.57	1.26	0.68	0.11	1.26
8		Lu	757	0.18	0.36	0.57	1.24	0.68	0.16	1.14
9		Min	661	0.17	0.47	0.83	1.13	0.74	0.10	1.14
10		Qingzhen	141	0.28	0.61	0.94	1.68	0.92	0.38	2.01
11		Su	449	0.15	0.51	0.58	1.26	0.69	0.09	1.17
12		Xibei	339	0.21	0.46	0.92	1.58	0.91	0.34	2.01
13		Xiang	720	0.15	0.32	0.67	1.63	0.82	0.63	1.32
14		Yue	1007	0.13	0.48	0.45	1.28	0.54	0.12	1.01
15		Yungui	187	0.25	0.52	0.74	1.68	0.92	0.66	1.90
16		Zhe	570	0.16	0.47	0.7	1.26	0.74	0.09	1.09
17	Only four cuisines contained in recipes	Guangxi	98	0.35	0.53	1.01	1.91	0.69	0.39	1.84
18		Jiangxi	87	0.16	0.25	0.77	1.24	0.76	0.59	1.48
19		Shanxi	175	0.27	0.43	0.75	1.43	1.03	0.19	2.22
20		Yu	167	0.21	0.53	1.01	1.43	0.91	0.17	1.61

POI, points of interest.

chart showed that the usage of the 'pungent' taste was the most frequent, followed by 'salty', while 'spicy' and 'sour' tastes were used in low frequency. Comparison of twenty cuisines showed that the dishes in Hong Kong and Taiwan had a 'sweet' emphasis, the dishes in Sichuan leaned towards 'spicy' and the dishes in Hubei were tended towards 'pungent'. These results were basically in line with people's general perception.

Taste indexes of each region

After data collation and analysis, a total of 387 509 restaurant POI categorised by sixteen cuisines in the thirty-three provincial regions were obtained, with their proportion shown in Fig. 3. In most regions, the POI for Chuan cuisine and Qingzhen cuisine accounted for greater proportions. Besides, the cuisine accounted for a higher proportion in its origin region than in other regions. For example, Hui cuisine was the most abundant in Anhui Province and Zhe was the most popular cuisine in Zhejiang Province.

To facilitate the comparison of tastes in the entire study area, the taste index in each provincial region was standardised by the mean taste index of all regions. As shown in Fig. 4, beyond 'sweet', the other six tastes in Macao and Hong Kong were mild; Gansu, Ningxia and Qinghai generally preferred strong tastes; across all provincial regions, Sichuan food was the spiciest; and Ningxia used the most 'umami' taste seasonings.

Geodetector results

Factor detector results

Given a significance level of 1 %, the q-values for the factor detectors significantly associated with taste-disease combinations are shown in Table 2. There were sixteen kinds of diseases that emerged out of the seventy-one chronic diseases that exhibited significant correlations with one or more tastes.

According to the number of diseases with significant association, seven taste factors were ranked: salty (9) > pungent (7) = sour (7) > fat (4) > spicy (3) > sweet (2) = umami (2).

From the view of the number of significant tastes corresponding to the disease, most chronic diseases had significant associations with only one taste factor. Whereas for a few chronic diseases, such as nasopharyngeal cancer, ovarian cancer and colorectal cancer, multiple tastes showed significant associations.

Interaction detector results

There were a total of twenty-one pairwise interactions for the seven tastes. As shown in Table 3, except for the interaction of the 'salty' and 'pungent' tastes, all the other taste interactions corresponded to one or more diseases. According to the interaction detector results for these diseases, seventy-five groups of 'tastes pairwise-disease' combinations had nonlinear enhancement outcomes, which means taste interactions can better explain the occurrence of the sixteen diseases than a single taste, at least in terms of spatial distribution.

As far as single taste factors, according to the number of 'taste pairwise-disease' combinations with significant results that they participated in, they were ranked in descending order as umami (37) > spicy (24) > sweet (22) > sour (20) > fat (19) > salty (15) > pungent (13). This result showed a trend generally opposite to the frequency of occurrence of each taste compared to the factor detector results (Table 2). Taking 'umami' as an example, there were
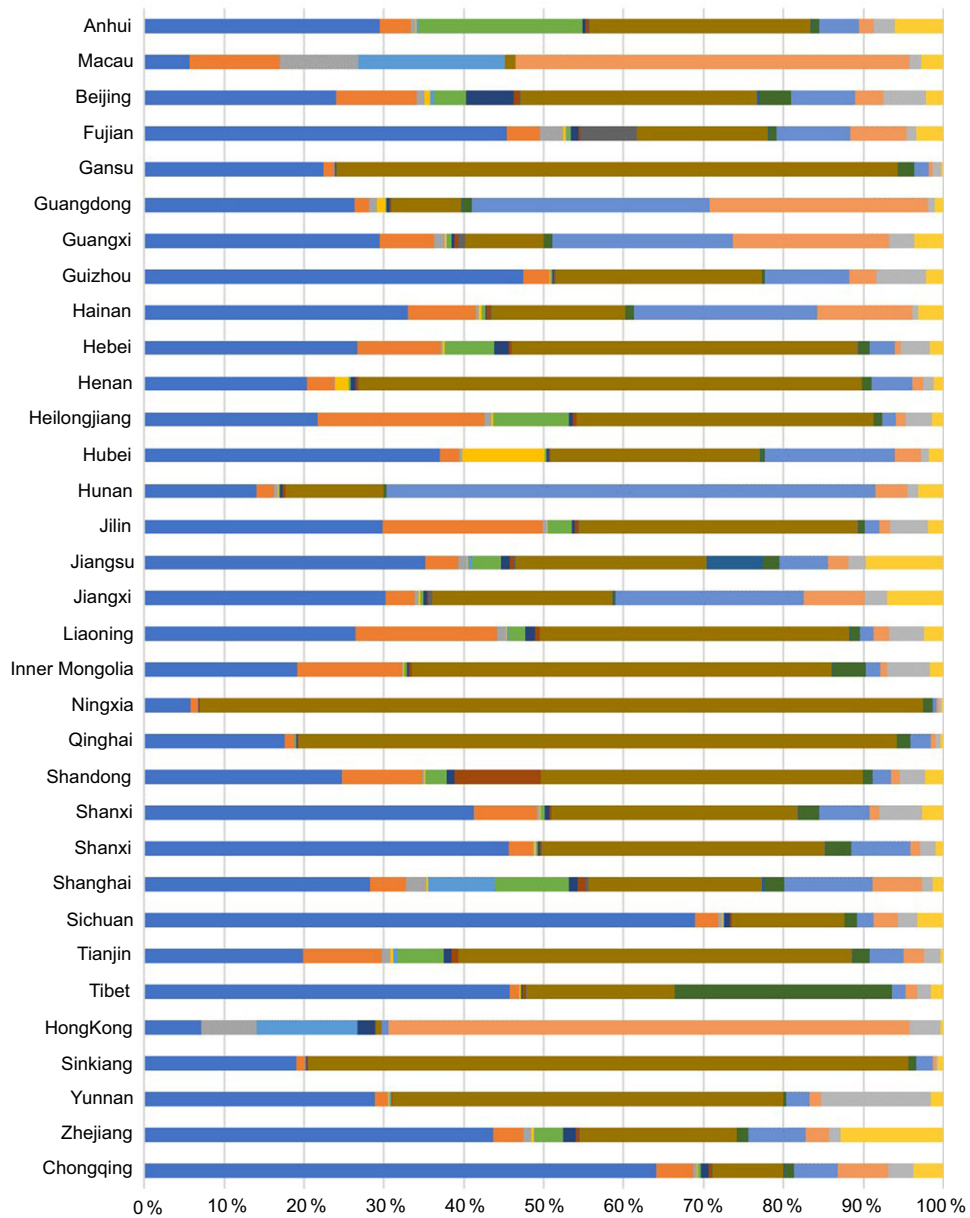
Fig. 3 Proportions of various types of POI in provincial regions. ■, Chuan; ■, Dongbei; ■, Gangtai; ■, Hubei; ■, Hu; ■, Hui; ■, Jing; ■, Lu; ■, Min; ■, Qingzhen; ■, Su; ■, Xibei; ■, Xiang; ■, Yue; ■, Yungui; ■, Zhe. POI, points of interest

only two diseases with significant results corresponding to this single taste factor, which was the lowest among the seven tastes. In contrast, when the taste 'umami' was interacting with other taste factors, the number of diseases having a nonlinear enhancement effect of the pathogenic risk was 37, which was the highest among the seven tastes.

Risk detector results
Combining the common perception of healthy eating behaviours and the popular perception of common disease risk, we chose the disease-taste combination of 'haemorrhagic stroke-salty' as an example to illustrate the results of the risk detector, which had a significant result.

The spatial distribution of the standardised mortality of haemorrhagic stroke is shown in Fig. 5. The western region as a whole was a high-value region, the value of the northeastern region was intermediate and the low-value region was continuously distributed along the southeastern coast. The value of the south-central region gradually increased from southeast to northwest, without any obvious abrupt changes in this area; however, the areas are distinguished hierarchically.

'Salty' was the only taste risk factor significantly associated with haemorrhagic stroke (Table 2). The abscissa in Fig. 6 represents the six levels at which the taste 'salty' is divided, with higher levels representing greater saltiness. The ordinate represents the mean disease mortality rate
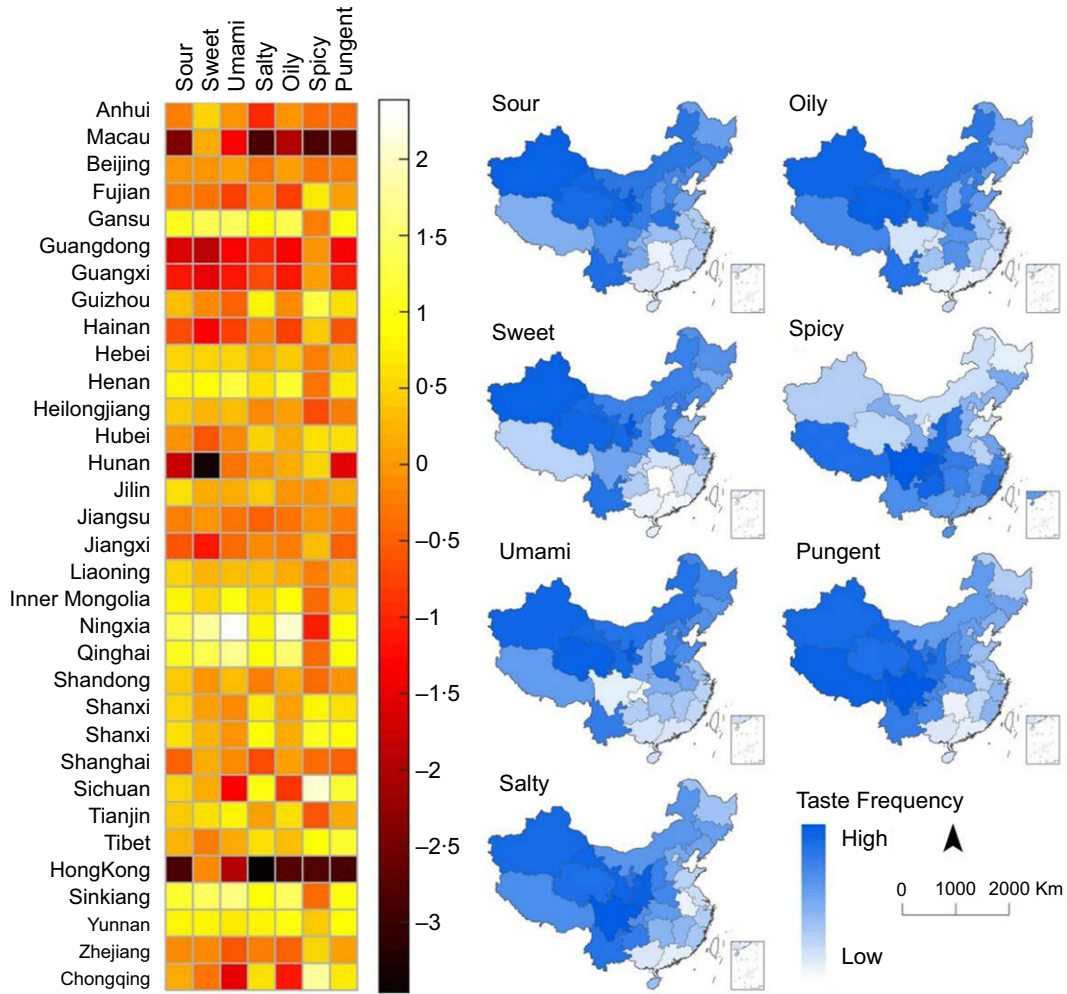
Fig. 4 Standardised indexes of seven tastes in provincial regions

Table 2 Q-statistics for the factor detector*

Disease ID	Disease name	Sour	Sweet	Umami	Salty	Fat	Spicy	Pungent
1	Aortic aneurysm	0·682	–	–	–	–	–	0·741
2	CVD	0·543	–	–	–	–	–	–
3	Cerebrovascular disease	0·579	–	–	0·560	–	–	0·498
4	Chronic obstructive pulmonary disease	–	–	–	0·590	–	–	–
5	Chronic respiratory diseases	–	–	–	0·584	–	–	–
6	Colorectal cancer	0·810	–	–	0·801	0·853	0·793	0·808
7	Haemorrhagic stroke	–	–	–	0·499	–	–	–
8	IHD	–	–	0·575	–	–	–	–
9	Meningitis	–	–	–	–	–	–	0·573
10	Nasopharyngeal cancer	0·789	0·617	0·728	0·612	0·692	–	0·741
11	Ovarian cancer	0·752	–	–	0·781	0·753	0·779	0·790
12	Pancreatic cancer	–	0·628	–	–	–	–	–
13	Pneumoconiosis	–	–	–	0·601	–	–	–
14	Prostate cancer	0·770	–	–	0·713	0·716	–	–
15	Rheumatic heart disease	–	–	–	–	–	0·613	–
16	Upper respiratory infections	–	–	–	–	–	–	0·559

*Association between disease and taste is not significant and represented by en dash.

Table 3 Results of the interaction detector

| Taste combination	Sour	Sweet	Umami	Salty	Fat	Spicy
Sweet	• Aortic aneurysm • Rheumatic heart disease					
Umami	• Chronic obstructive pulmonary disease • Chronic respiratory diseases • Haemorrhagic stroke • Pancreatic cancer • Rheumatic heart disease • Upper respiratory infections	• Aortic aneurysm • Cerebrovascular disease • Haemorrhagic stroke • Meningitis • Ovarian cancer • Pancreatic cancer • Rheumatic heart disease • Upper respiratory infections				
Salty	• Pancreatic cancer • Rheumatic heart disease	• Aortic aneurysm • Meningitis • Rheumatic heart disease	• Haemorrhagic stroke • Meningitis • Pancreatic cancer • Rheumatic heart disease • Upper respiratory infections			
Fat	• Haemorrhagic stroke • Pancreatic cancer • Rheumatic heart disease	• Aortic aneurysm • Haemorrhagic stroke • Meningitis	• Haemorrhagic stroke • Meningitis • Pancreatic cancer • Upper respiratory infections	• Haemorrhagic stroke • Meningitis • Rheumatic heart disease		
Spicy	• Chronic obstructive pulmonary disease • Chronic respiratory diseases • Pancreatic cancer • Pneumoconiosis • Rheumatic heart disease	• CVD • Aortic aneurysm • Meningitis • Pneumoconiosis	• Chronic obstructive pulmonary disease • Chronic respiratory diseases • Haemorrhagic stroke • Meningitis • Pancreatic cancer • Pneumoconiosis • Rheumatic heart disease • Upper respiratory infections	• Meningitis • Pneumoconiosis	• Meningitis • Pancreatic cancer • Pneumoconiosis • Rheumatic heart disease	
Pungent	• Chronic obstructive pulmonary disease • Rheumatic heart disease	• Aortic aneurysm • Pneumoconiosis	• Chronic obstructive pulmonary disease • Haemorrhagic stroke • Meningitis • Pancreatic cancer • Rheumatic heart disease • Upper respiratory infections	–	• Pancreatic cancer • Rheumatic heart disease	• Pneumoconiosis
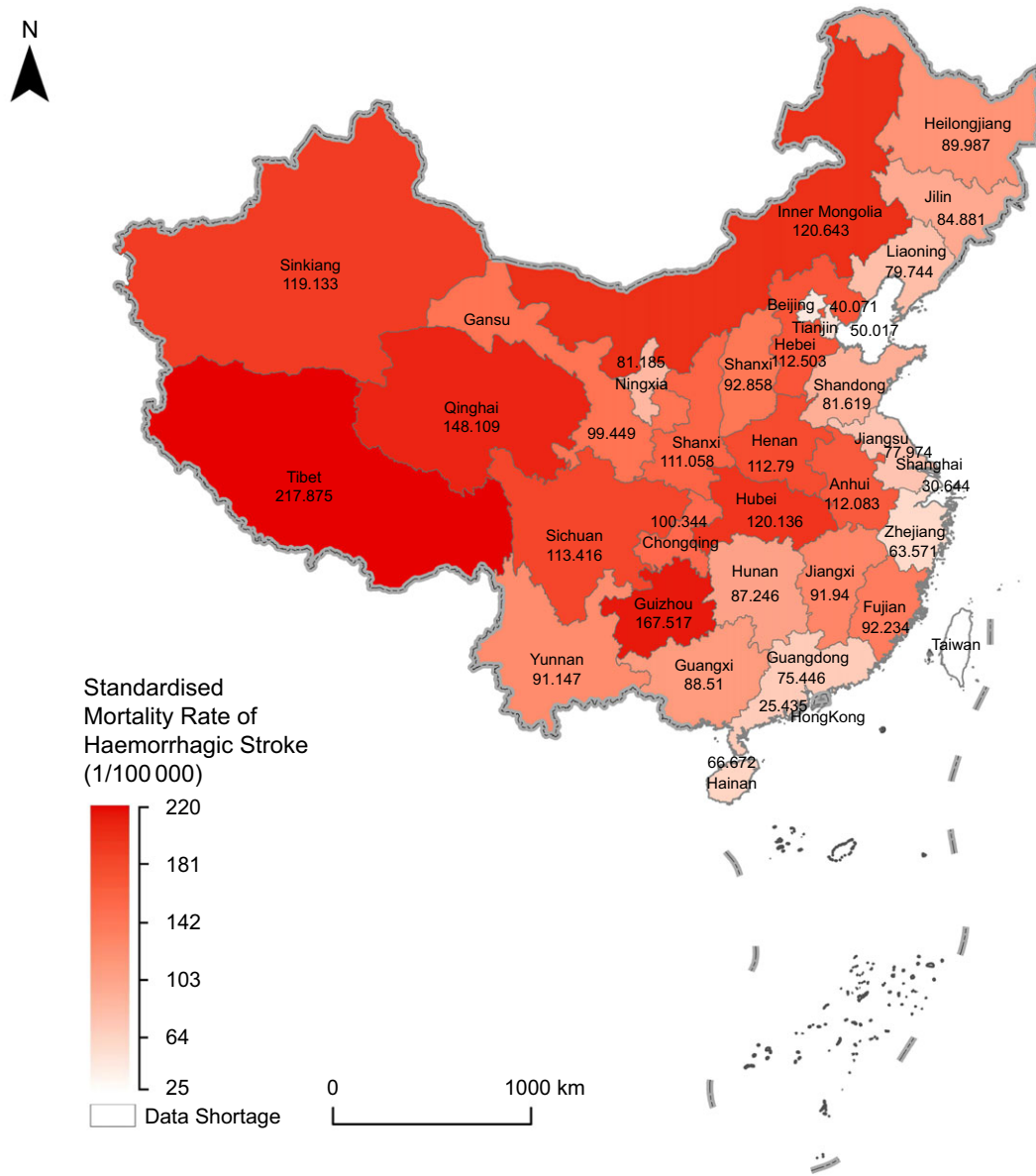
Fig. 5 Standardised mortality of haemorrhagic stroke in provincial regions

of the provincial regions with respective saltiness. It can be seen that the risk of haemorrhagic stroke goes up with the increase in the 'saltiness' of the dietary taste.

Due to the monotonically increasing trend shown in Fig. 6, this association can be more intuitively observed in spatial correlation by comparing the small maps of the spatial distributions of the seven tastes in Figures 4 and 5. In other words, the spatial distribution of the frequency of use of the taste 'salty' most closely approximates the description of the distribution of the above diseases, which is more in line with people's universal cognition. In addition, upon reviewing the medical literature related to
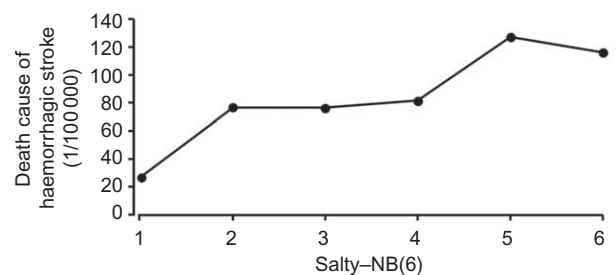
Fig. 6 Risk detector results for haemorrhagic stroke (NB(6): natural break method, six levels)

haemorrhagic stroke, the finding that increased saltiness increased the risk of haemorrhagic stroke was confirmed to be consistent with existing studies[35,36].

Discussion

Taste is a form of chemoreception, which occurs in specialised taste receptors in the mouth. According to previous research, there are five different types of taste these receptors can detect – sour, sweet, umami, salty and bitter[37]. Because bitter foods are generally found unpleasant, bitter is not common in diets. As a result, we neglected this kind of taste in our study. Besides the five basic tastes, there is evidence for a sixth taste that senses fatty substances. In 2015, researchers suggested a new basic taste of fatty acids called fat taste. Olive oil, fresh butter and various kinds of vegetable and nut oils are typical foods that offer fat taste[38–40]. The tongue can also feel other sensations not generally included in the basic tastes. These are largely detected by the somatosensory system. In Asian countries, within the sphere of mainly Chinese and Indian cultural influence, 'spicy' and 'pungent' have traditionally been considered two representative tastes besides the basic tastes. Although not tastes, these two kinds of sensations fall into a category called chemesthesis, and we called them tastes as well paralleling the other basic tastes for convenience in this study. Therefore, seven tastes were settled on in this study.

To date, there have been many cell biology studies on taste mechanisms demonstrating that taste has an important role in human evolution. For example, taste is responsible for evaluating the nutritional content of foods; taste preferences may have specific correlations with nutritional deficiencies and different tastes take on the role of recognising different chemicals to help identify and distinguish key dietary components and prevent the intake of toxic substances[41–43]. When the body suffers from disease or discomfort, the function of taste will be subject to interference, thereby affecting the body's intake of food[44]. An improved understanding of taste biology and genetics may prevent or influence the development of chronic disease risk[45]. All of the above findings have been obtained through the experimental study of individuals. Although such a method enforces a more rigorous scientific basis, the research cost is higher and research method less transplantable. This study is the first to present a quantitative approach to the study of potential health risk factors based on crowdsourcing data mining. A unique feature of our study is that it starts from the spatial distribution pattern of the disease on the one hand, extracts the spatial distribution pattern of risk factor from crowdsourcing data on the other hand and spatially correlates the two. The results obtained are more universal and exploratory. Therefore, they can be used for exploration of associations in further public health research.

As an exploratory attempt to quantitatively analyse the taste risk factors of diseases using crowdsourcing data, the uncertainty of the research and applicability of the results still need further discussion and reflection:

(1) As to the method, on the one hand, since the chronic diseases were associated with dietary tastes by the Geodetector simply from the spatial statistical perspective, the diseases unselected by the Geodetector did not mean that they were not related to dietary tastes. For example, there was no significant association between sweet and diabetes in this study. Also, the diseases selected by the Geodetector did not necessarily mean that they could be completely explained by the related tastes. On the other hand, compared with simple bivariate correlation analysis, the advantage of the Geodetector is to detect the complex association between the taste and disease that is not limited to a linear relationship. By discretising the taste data, the nonlinear correlation can be more intuitively presented by clustering. Furthermore, if a significant association can be further confirmed in medical experiments, a discrete taste level will be easier in practice when developing public health interventions.

(2) As for the results, haemorrhagic stroke, as an example, showed our finding was consistent with prior knowledge, which preliminarily confirms the quantitative approach presented in this study is useful to the study of potential health risk factors and the findings are referable. Besides, there are some interesting results to note. In the results for risk detectors, a higher sour level corresponds to a lower risk of colorectal cancer, and a higher spicy level corresponds to a lower risk of ovarian cancer. Some pathological research has provided a possible explanation for such results. For example, the principal active component of spicy food is capsaicin. Researchers have found that capsaicin has a beneficial role in various cancers[46]. Garlic is a common pungent seasoning, which has an antibacterial effect, and it has the potential to reduce the occurrence of colorectal cancer[47]. But all these conjectures need to be verified by rigorous clinical experiments.

We have noted the active expression of umami in our results of the interaction detector. The taste umami acts as a 'catalyst' when it interacts with other tastes. The taste umami itself is not strongly related to most chronic diseases; however, when combined with other tastes, especially sour and sweet, it forms compound tastes such as umami-sour and umami-sweet, both of which have statistically significant associations with many chronic diseases. This joint effect of multiple tastes as a health factor has been suggested by many studies[48–51]. For example, the combination of ginger (pungent) and vinegar (sour) has a synergistic antibacterial effect[52]. This indirect clue derived

from our results may be beneficial to explain the association between this taste pairwise and chronic obstructive pulmonary disease or rheumatic heart disease.

It should be noted that the results of this study are data-driven. Due to the deficiency of the research on the relationship between chronic diseases and dietary tastes, only a small part of the results shown in Tables 2 and 3 can be related to the existing research. Most findings are presented for the first time, and their rationality may lack support from existing research findings or may have not even been studied yet. These results need to be interpreted or overturned more reasonably through pathological experiments. Although these results have not yet been confirmed, the value of these exploratory findings lies in providing basic data for scientific research in related fields and pointing out new study directions.

(3) This study was restricted by its data sources – the taste data obtained by using the POI of restaurants in 2017 were used to try to interpret the cause of death in 2013. As the effects of dietary tastes on chronic diseases exhibit hysteresis, the time design of the data in the experiment was not logically optimal; instead, the data's accessibility alone became the priority. But, considering eating habits in a region are the result of accumulated behaviours following a long period of practice, dietary tastes will not change substantially in a short period of time. Therefore, in this study, we used 1-year restaurant POI data as a microcosm of the regional food categories to reflect the local taste characteristics, which is of scientific significance. However, if there are longer periods of regional restaurant data and chronic disease data collected thereafter and they are analysed from the perspective of spatial lifecourse epidemiology, the impact of the time effect described above can be maximally eliminated, allowing for a more scientific conclusion to be drawn[53]. Also, this research method would have higher reference value and can be applied to the exploratory analysis of the causes of more diseases.

(4) From the perspective of dietary culture geography[19], dietary cultures of different regions are integrated with each other. Differences in dietary taste do not have obvious geographical divisions, and choosing different spatial scales will yield different sensitivities to the problem, which is affected by the modifiable areal unit problem. Therefore, it should be emphasised that this study is limited by disease data sources. The experiment was conducted on the provincial spatial scale. The conclusions obtained are specific to the national scale and are not necessarily applicable to other spatial scales. Therefore, the conclusions have a spatial scale-effect problem. In addition, the conclusions from this experiment may be mainly applicable to China where the food environment is arranged and affects chronic disease risk in a unique way[54–57]. For

different countries or regions, the experiment needs to be adapted to local data. If the data on chronic diseases can be obtained with higher spatial resolution, the methodology of this experiment can be expanded at different scales of research to detect chronic disease risk factors that are more regionally specific.

Acknowledgements

References

1. World Health Organization (2018) *Noncommunicable Diseases.* http://wwwwhoint/en/news-room/fact-sheets/detail/noncommunicable-diseases (accessed June 2018).
2. World Health Organization (2013) Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020. https://www.who.int/nmh/events/ncd_action_plan/en/ (accessed June 2019).
3. World Health Organization (2018) Noncommunicable Diseases Country Profiles 2018. https://www.who.int/nmh/countries/en/ (accessed June 2019).
4. Yang G, Wang Y, Zeng Y *et al.* (2013) Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* **381**, 1987–2015.
5. World Health Organization (2014) *Global Status Report on Noncommunicable Diseases 2014.* Geneva: WHO.
6. Majid E & Elio R (2012) Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals. *Science* **337**, 1482–1487.
7. Majid E & Elio R (2013) Behavioral and dietary risk factors for noncommunicable diseases. *N Engl J Med* **369**, 954–964.
8. Gakidou E, Afshin A, Abajobir AA *et al.* (2017) Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic

analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1345–1422.

9. Sun XDLY & Li LM (2008) Prevalence level of main risk factors of chronic diseases and development of prevention strategies. *Chin J Prev Control Chron Non-commun Dis* **16**, 538–540.

10. Li Y, Liu S, Wang L *et al.* (2015) Burden of disease attributable to main behavioral risk factor of chronic disease inactivity in China, 1990 and 2010. *Zhonghua Yu Fang Yi Xue Za Zhi* **49**, 303–308.

11. Stookey JD, Wang Y, Ge K *et al.* (2000) Measuring diet quality in china: the INFH-UNC-CH diet quality index. *Eur J Clin Nutr* **54**, 811.

12. Xu JW, Chen XR, Li YC *et al.* (2015) Analysis of Chinese adults' eating behavior in 2010. *Chin J PrevControl Chron Dis* **23**, 684–686.

13. Ma GS, Hu XQ, Luan DC *et al.* (2005) The eating practice of people in China. *Acta Nutrimenta Sin* **27**, 272–275.

14. Su YP, Yang K, Liu XT *et al.* (2018) Relationship between the dietary behavior and the risk of diabetes in adults. *Chin J PrevControl Chron Dis* **7**, 481–485.

15. Shen M, Shi LY, Li LC *et al.* (2002) A case-control study on urinary calculi and dietary factors. *Zhonghua Liu Xing Bing Xue Za Zhi* **23**, 134–137.

16. Howe J (2006) The rise of crowdsourcing. *Wired Magazine* **14**, 1–4.

17. Zhang YN & Hu JY (2013) Thinking about big data's application in education. *China Electric Power Educ* **2013**, 5–7.

18. Meng RT, Luo Y, Yu CH *et al.* (2015) Application and challenges of healthy big data in the field of public health. *Chin Gen Pract* **18**, 4388–4392.

19. Cheng CK (1994) The culture of Chinese diet: regional differentiation and developing trends. *Acta Geogr Sin* **49**, 226–235.

20. Wang WW (2011) Regional differences in Chinese diet culture. *Food Life* **1**, 46–47.

21. Zhu YX, Huang JM, Zhang ZK *et al.* (2013) Geography and similarity of regional cuisines in china. *PLos One* **8**, e79161.

22. Zhu J, Xu Y, Fang Z *et al.* (2018) Geographic prevalence and mix of regional cuisines in Chinese cities. *ISPRS Int J Geo-Inf* **7**, 183.

23. Wang WW & Qiu S (2007) Sweet, sour, bitter and salty-the regional difference of Chinese food culture. *Chin Food* **11**, 14–15.

24. Wang JF, Li XH, Christakos G *et al.* (2010) Geographical detectors-based health risk assessment and its application in the Neural Tube Defects Study of the Heshun Region, China. *Int J Geogr Inf Sci* **24**, 107–127.

25. Gu JF (2016) An interpretation of report on the status of nutrition and chronic diseases in Chinese residents (2015). *J Nutr* **38**, 525–529.

26. Abubakar I, Tillmann T & Banerjee A (2015) Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **385**, 117–171.

27. Kirkwood B & Sterne J (2003) *Essential Medical Statistics*, 2nd ed. London: Blackwell Science.

28. Zhang YL (2018) The Name and Diagram of 1000 Kinds of Spices in Kitchen. http://www.360doc.com/content/18/0811/17/58294334_777493365.shtml (accessed August 2018).

29. Lan Y (2001) The reasons and distribution of pungent flavour districts in China's dietetics. *Geogr Res* **21**, 35–42.

30. Running CA, Craig BA & Mattes RD (2015) Oleogustus: the unique taste of fat. *Chem Senses* **40**, 507–516.

31. Ikeda K (2002) New seasonings. *Chem Senses* **27**, 847.

32. Cao F, Ge Y & Wang JF (2013) Optimal discretization for geographical detectors-based risk assessment. *Mapp Sci Remote Sens* **50**, 78–92.

33. Wang JF & Xu CD (2017) Geodetector: principle and prospective. *Acta Geogr Sin* **72**, 116–134.

34. Xu QR & Zheng XQ (2015) Analysis of influencing mechanism of urban growth using geographical detector. *Acta Geod Cartogr Sin* **44**, 96–101.

35. Chisato N, Naoyoshi T, Natsuki S *et al.* (2004) Sodium intake and risk of death from stroke in Japanese men and women. *Stroke* **35**, 1543–1547.

36. Yu CH, Luo LS, Li M *et al.* (2016) From the global views to understand the seriousness of the burden of stroke in China. *J Public Health Prev Med* **27**, 1–5.

37. Pmhdev (2012) *How Does Our Sense of Taste Work?* Cologne: Institute for Quality and Efficiency in Health Care (IQWiG).

38. Keast RS & Costanzo A (2015) Is fat the sixth taste primary? Evidence and implications. *Flavour* **4**, 5.

39. Laugerette F, Passilly-Degrace P, Patris B *et al.* (2005) CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions. *J Clin Invest* **115**, 3177–3184.

40. DiPatrizio NV (2014) Is fat taste ready for primetime? *Physiol Behav* **136**, 145–154.

41. Sherman PW & Billing J (1999) Darwinian gastronomy: why we use spices. *Bioscience* **49**, 453–463.

42. Jayaram C, Hoon MA, Ryba NJP *et al.* (2006) The receptors and cells for mammalian taste. *Nature* **444**, 288–294.

43. Chaudhari N & Roper SD (2010) The cell biology of taste. *J Cell Biol* **190**, 285–296.

44. Li GL, Chen ML & Zhang GH (2012) The cell biology of mammalian taste. *Chem Life* **32**, 456–459.

45. Chamoun E, Mutch DM, Allen-Vercoe E *et al.* (2018) A review of the associations between single nucleotide polymorphisms in taste receptors, eating behaviours, and health. *Crit Rev Food Sci Nutr* **58**, 194–207.

46. Sharma SK, Vij AS & Sharma M (2013) Mechanisms and clinical uses of capsaicin. *Eur J Pharmacol* **720**, 55–62.

47. Tsai C-W, Chen H-W, Sheen L-Y *et al.* (2012) Garlic: health benefits and actions. *BioMedicine* **2**, 17–29.

48. Bartoshuk LM (1975) Taste mixtures: is mixture suppression related to compression? **14**, 643–649.

49. Schifferstein HNJ (1992) *Sensory Interactions in Mixtures of Tastants*. Wageningen: Prof.

50. Keast RSJ & Breslin PAS (2003) An overview of binary taste – taste interactions. *Food Qual Preference* **14**, 111–124.

51. Capitanio A & Lucci G & Tommasi L (2011) Mixing taste illusions: the effect of miraculin on binary and trinary mixtures. *J Sens Stud* **26**, 54–61.

52. Guan HQ, Li Tian QX, Da Huang TSZ *et al.* (2003) Study on synergistic antibacterial effect of ginger, salt and acetic acid on food contaminating bacteria. *J Tradit Chin Med* **21**, 119–120.

53. Jia P (2019) Spatial lifecourse epidemiology. *Lancet Planet Health* **3**, e57–e59.

54. Jia P, Xue H, Cheng X *et al.* (2019) Effects of school neighborhood food environments on childhood obesity at multiple scales: a longitudinal kindergarten cohort study in the USA. *BMC Med* **17**, 99.

55. Zhang X, Zhang M, Zhao Z *et al.* (2020) Obesogenic environmental factors of adult obesity in China: a nationally representative cross-sectional study. *Environ Res Lett* **15**, 044009.

56. Jia P, Luo M, Li Y *et al.* (2019) Fast-food restaurant, unhealthy eating, and childhood obesity: a systematic review and meta analysis. *Obes Rev* **40**, 1–27.

57. Wang Y, Jia P, Cheng X *et al.* (2019) Improvement in food environments may help prevent childhood obesity: evidence from a 9-year cohort study. *Pediatr Obes* **14**, e12536.