Large-Scale Comparative Analysis Reveals A Simple Model To Predict The Prevalence Of Thyroid Nodules

This article was published in the following Dove Press journal:
Risk Management and Healthcare Policy

Genpeng Li1,*
Lijuan Wang1,*
Jianyong Lei1
Linlin Song1
Huairong Tang2
Zhihui Li1
Rixiang Gong1
Jingqiang Zhu1
1Thyroid and Parathyroid Surgery Center, West China Hospital of Sichuan University, Chengdu 610041, People’s Republic of China; 2Health and Management Center, West China Hospital of Sichuan University, Chengdu 610041, People’s Republic of China
*These authors contributed equally to this work

Background: The frequency of thyroid nodules (TNs) has increased rapidly in recent decades. We established a simple, effective model to predict TN occurrence by a large-scale comparative analysis of the Chinese population.

Subjects and methods: A total of 13,307 subjects were included to screen risk factors between the case group (TN+) and the control group (TN-) by univariate and multivariate analyses. A simple prediction model formed by a subset of five factors was established. The optimal index points for predicting the prevalence of TNs, the model discriminatory power and the association of the model with the size of TNs were assessed.

Results: Age at diagnosis ≥45 years, female sex, body mass index (BMI) ≥24 kg/m², hypertension, antithyroglobulin antibody (TGAB) ≥16.5 IU/mL were independently related to an increased risk of TNs. The cutoff value of 4 points was found to be the best prediction for the prevalence of TNs in the current study, and the model had better discriminatory power than other single independent predictors. In addition, a positive correlation was also found between the index points and the diameter of TNs.

Conclusion: Based on our prediction model, thyroid high-resolution ultrasound and associated laboratory tests may be necessary for patients with index points ≥4 due to a higher prevalence of thyroid nodules.

Keywords: thyroid nodules, large-scale, risk factors, prediction model

Introduction
Thyroid nodules (TNs), one of the most common diseases of the endocrine system, develop as a consequence of the interplay among genetic, environmental, and endogenous factors.1,2 The incidence of TNs detected by high-resolution ultrasonography among randomly selected individuals ranges from 19–68%, and only 3–7% of TNs can be detected by palpation, while 5–15% of TNs are diagnosed as malignant tumors in a subsequent examination.2,3 Although TNs have no clinical manifestations in the majority of patients, they can be associated with a variety of disorders that affect quality of life and life expectancy, such as endocrine dysregulation, autoimmune thyroid disease,4–6 and even breathing and/or swallowing difficulties.7

Due to annually increasing trends worldwide, TNs have received much attention in the medical field. These trends cannot be explained by traditional known risk factors, such as exposure to radiation or chemicals or a family history of thyroid nodules or cancer, so numerous studies have attempted to determine whether other
demographic parameters and clinical histories have a profound effect on the formation and growth of TNs. Age and sex were associated with the increasing prevalence of TNs in the study by Akushevich. With regard to lifestyle, smoking was identified by multiple studies as predisposing the study population to TNs. Meanwhile, studies by Diez and Guo noted an increased incidence of TNs in individuals with clinically diagnosed hypertension, diabetes and abnormal thyroid hormone levels, in particular high thyroid-stimulating hormone (TSH) levels; however, there is still no consensus regarding which individuals should closely monitor the health of their thyroid in populations without definite known risk factors, as mentioned above.

The aim of the present study was to determine the association of different clinical parameters or lifestyle-associated factors with TNs in populations without definite known risk factors and to establish a simple multivariable prediction model for the prevalence of TNs via a large-scale comparative analysis in the Chinese population.

**Subjects And Methods**

From October 2014 to October 2017, the medical records of 19,108 patients who received health examinations at the Health and Management Center of the West China Hospital of Sichuan University were reviewed. The inclusion criteria were subjects from 18 to 80 years of age whose thyroid ultrasound and associated laboratory test results (performed in our center) were available and who had adequate medical histories and person without definite known risk factors, such as exposure to radiation or chemicals or a family history of thyroid nodules or cancer. In our study, pregnant women and subjects taking contraceptive agents or estrogen were not included. Those who suffered from severe diseases, such as chronic renal failure, hepatic cirrhosis, evident cardiac insufficiency, or abdominal ascites, were excluded. Individuals who received medications that influence thyroid function, such as amiodarone and iodine, or hormones, such as glucocorticoid and somatostatin, were also excluded. Based on the inclusion and exclusion criteria, 13,307 subjects were eligible for this retrospective analysis. Data on the subjects’ clinical features, such as age at diagnosis, sex, body mass index (BMI), size of TNs, blood pressure, prediabetes or diabetes, smoking, alcohol consumption, antithyroid peroxidase antibody (TPOAB, normal reference <34 IU/mL) and antithyroglobulin antibody (TGAB, normal reference <115 IU/mL), were extracted from electronic medical records. TNs referred to any detected discrete lesion within the thyroid gland that is radiologically distinct from the surrounding thyroid parenchyma. Receiver operating characteristic (ROC) curve studies were performed to identify the optimal cutoff point for predicting TNs in continuous variables, such as age, TPOAB and TGAB. Univariate and multivariate analyses were performed with a focus on risk factors associated with TNs. Thus, a multivariable prediction model of TNs was constructed in accordance with these independent risk factors, and ROC curve analyses were conducted to identify the optimal cutoff point and compare discriminatory power. Additionally, the relationship between the prediction model and the diameter of thyroid nodules was identified. The study protocol was approved by the Institutional Review Board of West China Hospital of Sichuan University, and patient approval or informed consent was required for review of the patients’ medical records.

Statistical analyses were performed using SPSS software, version 20.0 (SPSS, Chicago, IL, USA). Continuous variables with normal distributions are expressed as the mean ± standard deviation, and continuous variables without normal distributions are expressed as medians. Categorical variables are presented as numbers. Continuous variables with normal distributions were compared using Student’s t-test or analysis of variance (ANOVA). Categorical variables were compared using the chi-square test or Fisher’s exact chi-square test. Independent predictors were determined by a multivariable logistic regression model based on the risk factors related to TNs. According to multiple logistic regression analyses, features that were independent factors were assigned different points based on the OR to develop a simple model. Significance was set at a level of 5%.

**Results**

**Associations Between Clinical Parameters Or Lifestyle Factors And Subjects With TNs**

The subjects were divided into two groups according to their thyroid gland status (case group: TNs detected by ultrasound; control group: TNs not detected by ultrasound). Of the enrolled subjects, 5534 and 7773 were in the case group and control group, respectively. The baseline demographics and lifestyle-associated factors that were investigated in the present study included the following: age, sex, BMI, hypertension, prediabetes or diabetes, smoking, alcohol consumption, thyroid function, and...
presence of autoimmune antibody. No significant differences were found in the thyroid-stimulating hormone (TSH) level between the cases and controls, as shown in Table 1 (P=0.115); however, 13 clinical and lifestyle factors did have correlations with the prevalence of TNs in the univariate analysis (positive correlations: 9 factors; negative correlations: 4 factors). As shown in Figure 1, ROC curve analysis was conducted to predict the cutoff point of continuous variables with positive correlations with the prevalence of TNs (age ≥45 years; FT3 <3.60 pmol/L; FT4 <12.0 pmol/L; TPOAB ≥16.0 IU/mL; TGAB ≥16.5 IU/mL). However, thyroxine (T4) and triiodothyronine (T3) did not exhibit differences in the ROC curve analysis (all P>0.05).

**Multivariable Risk Assessment And Establishment Of A Prediction Model**

In the univariate and ROC curve analyses, the following 9 factors significantly increased the risk of TNs: age at diagnosis ≥45 years, female sex, BMI ≥24 kg/m², hypertension, prediabetes or diabetes, FT4 <12.0 pmol/L, FT3 <3.60 pmol/L, TPOAB ≥16.0 IU/mL and TGAB ≥16.5 IU/mL (all P<0.05).

A multivariate analysis was performed to compare these factors with the risk potential as indicated by univariate and ROC curve analyses; prediabetes or diabetes, FT4 <12.0 pmol/L, FT3 <3.60 pmol/L and TPOAB ≥16.0 IU/mL were not significantly associated with the prevalence of TNs in the multivariable analysis (all P>0.05). The factors that remained independently associated with the prevalence of TNs were age at diagnosis ≥45 years (OR 2.026, 95% CI 1.841–2.230, P=0.000), female sex (OR 1.772, 95% CI 1.598–1.965, P=0.000), BMI ≥24 kg/m² (OR 1.326, 95% CI 1.105–1.548, P=0.024), hypertension (OR 1.187, 95% CI 1.018–1.385, P=0.029); and TGAB ≥16.5 IU/mL (OR 1.198, 95% CI 1.005–1.294, P=0.003) (Table 2). Based on the multivariable logistic regression analysis, a simple prediction model formed by a subset of five factors was established. The evaluated cutoff points for each characteristic are shown in Table 2 (total 7 points), and the sum of the points was evaluated to distinguish between subjects with and those without TNs.

**The Quality Of Prediction Model For The Prevalence Of Thyroid Nodules**

The score distributions of the two groups are presented in Figure 2A. The number of control and case groups all showed an N-shaped curve, which peaked at the score of three. The mean index score was 2.1 ± 1.5 in the control group and 3.3 ± 1.8 in the case group, which was a significant

---

**Table 1** Comparison Of The Baseline Demographics And Clinical Features Of Patients With TN+ And TN−

<table>
<thead>
<tr>
<th></th>
<th>Cases Group (TN+)</th>
<th>Control Group (TN−)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis (mean ±SD, years)</td>
<td>45.5 ± 8.7</td>
<td>41.9 ± 8.8</td>
<td>0.000*</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>2674/2860</td>
<td>4858/2915</td>
<td>0.000*</td>
</tr>
<tr>
<td>BMI ≥24 kg/m² (yes/no)</td>
<td>2257/3277</td>
<td>3498/4275</td>
<td>0.000*</td>
</tr>
<tr>
<td>Hypertension (yes/no)</td>
<td>655/4879</td>
<td>920/6853</td>
<td>0.000*</td>
</tr>
<tr>
<td>Prediabetes or diabetes (yes/no)</td>
<td>272/5262</td>
<td>268/7505</td>
<td>0.000*</td>
</tr>
<tr>
<td>Smoking (yes/no)</td>
<td>1237/4297</td>
<td>2137/5636</td>
<td>0.000*</td>
</tr>
<tr>
<td>Alcohol consumption (yes/no)</td>
<td>864/4670</td>
<td>1543/6230</td>
<td>0.000*</td>
</tr>
<tr>
<td>TSH level (mU/L)</td>
<td>3.0 ± 2.7</td>
<td>2.9 ± 3.2</td>
<td>0.115</td>
</tr>
<tr>
<td>FT4 level (pmol/L)</td>
<td>99.8 ± 19.1</td>
<td>97.7 ± 19.4</td>
<td>0.000*</td>
</tr>
<tr>
<td>FT3 level (pmol/L)</td>
<td>16.5 ± 2.7</td>
<td>16.7 ± 2.9</td>
<td>0.013*</td>
</tr>
<tr>
<td>T3 level (pmol/L)</td>
<td>1.8 ± 0.4</td>
<td>1.8 ± 0.3</td>
<td>0.016</td>
</tr>
<tr>
<td>TPOAB (IU/mL)</td>
<td>28.8 ± 66.4</td>
<td>24.4 ± 59.1</td>
<td>0.000*</td>
</tr>
<tr>
<td>TGAB (IU/mL)</td>
<td>79.0 ± 288.2</td>
<td>60.9 ± 190.1</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

Notes: *Statistically significantly differences; **Negative correlations.

Abbreviations: TN+, thyroid nodules; TN−, non-thyroid nodules; BMI, body mass index; TSH, thyroid stimulating hormone; T4/FT4, thyroxine/free thyroxine; T3/FT3, triiodothyronine/free triiodothyronine; TPOAB, anti-thyroid peroxidase antibody; TGAB, anti-thyroglobulin antibody.
In China, studies have shown that the prevalence of TNs ranges from 67.1% in North America, 27.3% in Finland, 19.5% in Belgium, and 17.7% in Brazil in 2013. In China, studies have reported that in Hangzhou, the prevalence in men and women was 24.1% and 34.7%, respectively, in 2013, and among people in Nanjing aged >40 years, the prevalence reached 46.6% in 2014. The results of this survey showed that the prevalence of TNs and the occurrence of TNs was 49.5% in females, 67.1% in North America, 27.3% in Finland, and 19.5% in Belgium. The prevalence was found in iodine-deficient areas, whereas the association is less pronounced in iodine-replete areas. Our results revealed that high thyroglobulin antibody levels (TGAB >16.5 IU/mL) also contribute to the prevalence of TNs. This finding was expected, as experimental studies have reported thiocyanate to mediate the goitrogenic effect of tobacco smoke acting as a competitive inhibitor of iodine uptake. In addition, a low occurrence of goiter among alcohol consumers has been found, but the mechanism underlying this association remains unknown.

### Discussion

In recent years, the prevalence of TNs has significantly increased. According to the literature, the prevalence of TNs was 67.1% in North America, 27.3% in Finland, 19.5% in Belgium, and 17.7% in Brazil in 2013. In China, studies have reported that in Hangzhou, the prevalence in men and women was 24.1% and 34.7%, respectively, in 2013, and among people in Nanjing aged >40 years, the prevalence reached 46.6% in 2014. The results of this survey showed that the prevalence of TNs was 41.6% in the Chinese population and that the occurrence of TNs was 49.5% in females, which was significantly higher than the incidence of 35.5% among males. Although TNs are the most common thyroid disorder in the adult population, the etiology is not completely understood. The high prevalence of TNs may be partly due to advancements in diagnostic technologies, but it still cannot be explained by traditional known risk factors such as iodine intake, exposure to radiation or chemicals, or a family history of thyroid nodules or cancer. The presentation of TNs ranges from mild changes in the thyroid structure without clinical symptoms to severe symptoms, such as breathing and/or swallowing difficulties. In this cross-sectional study, we conducted a large-scale comparative analysis that aimed to further identify the risk factors and construct a prediction model for the occurrence of TNs.

Our data suggested that age ≥45 years, female sex, BMI ≥24 kg/m², and hypertension were independently associated with a higher risk of TNs, which is consistent with the results of many previous studies. In 2015, Qin reported that high thyroglobulin antibody levels will increase the risk of differentiated thyroid carcinoma. Our results revealed that high thyroglobulin antibody levels (TGAB ≥16.5 IU/mL) also contribute to the prevalence of TNs. TSH plays a major role in the regulation of thyroid cell growth and differentiation and may play a direct role in nodule formation. Zhu and Ayturk revealed that increased serum TSH was an independent risk factor for increased thyroid volume in patients; however, TSH levels were not found to be associated with the formation and growth of TNs in the present study. Interestingly, smoking and alcohol were not identified as risk factors for the frequency of TNs. In a study by Knudsen, a strong association between tobacco smoking and goiter prevalence was found in iodine-deficient areas, whereas the association is less pronounced in iodine-replete areas. This finding was expected, as experimental studies have reported thiocyanate to mediate the goitrogenic effect of tobacco smoke acting as a competitive inhibitor of iodine uptake. In addition, a low occurrence of goiter among alcohol consumers has been found, but the mechanism underlying this association remains unknown.

### Table 2 Multivariate Analyses Of Factors Associated With The Prevalence Of Thyroid Nodules

<table>
<thead>
<tr>
<th>Variables</th>
<th>Odd Ratio</th>
<th>95% CI</th>
<th>P value</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis ≥45 years</td>
<td>2.026</td>
<td>1.841–2.230</td>
<td>0.000*</td>
<td>2</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>1.772</td>
<td>1.598–1.965</td>
<td>0.000*</td>
<td>2</td>
</tr>
<tr>
<td>BMI ≥24 kg/m²</td>
<td>1.326</td>
<td>1.015–1.548</td>
<td>0.024*</td>
<td>1</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1.187</td>
<td>1.018–1.385</td>
<td>0.029*</td>
<td>1</td>
</tr>
<tr>
<td>Prediabetes or diabetes</td>
<td>1.257</td>
<td>0.991–1.595</td>
<td>0.060</td>
<td>0</td>
</tr>
<tr>
<td>FT4 &lt;12.0 pmol/L</td>
<td>1.330</td>
<td>0.899–1.967</td>
<td>0.153</td>
<td>1</td>
</tr>
<tr>
<td>FT3 &lt;3.60 pmol/L</td>
<td>1.391</td>
<td>0.907–2.134</td>
<td>0.131</td>
<td>0</td>
</tr>
<tr>
<td>TPOAB ≥16.0 IU/mL</td>
<td>1.097</td>
<td>0.982–1.227</td>
<td>0.102</td>
<td>0</td>
</tr>
<tr>
<td>TGAB ≥16.5 IU/mL</td>
<td>1.198</td>
<td>1.055–1.294</td>
<td>0.003*</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

**Note:** *Statistically significantly different.

**Abbreviations:** BMI, body mass index; TSH, thyroid stimulating hormone; T4/FT4, thyroxine/free thyroxine; T3/FT3, triiodothyronine/free triiodothyronine; TPOAB, anti-thyroid peroxidase antibody; TGAB, anti-thyroglobulin antibody.
Figure 2 (A) The distribution of cases in the two groups in different points according to the prediction model. (B) The mean index points in the cases group and control group according to the prediction model.
simple model for the risk factors that predict TNs may be another approach that can be used to guide physical examinations. High-resolution ultrasound and associated laboratory tests may be necessary for patients with index scores ≥4 points due to the higher risk of TN occurrence.

In 2006, Davies and Welch found a mismatch between trends in the incidence and mortality of thyroid cancer; their study was the first to comment on the problem of thyroid cancer overdiagnosis and overtreatment and postoperative management of thyroid cancer is also a focus of attention. One of the activities leading to overdiagnosis of thyroid cancer is the workup of incidental thyroid nodules detected on imaging. To mitigate the problem of overdiagnosis and overtreatment, the most suitable population for our prediction model may be individuals who warrant a thyroid examination. Based on our model, a brief assessment can be performed before conducting the examination. High-resolution ultrasound and associated laboratory tests for thyroid can be considered for patients with index scores ≥4 points, while associated examinations may not be necessary for these persons with index scores <4 points, which may help save considerable healthcare resources and reduce thyroid nodule overdiagnosis and overtreatment.

In this study, we conducted a comprehensive analysis of the risk factors for the prevalence of TNs and established a multivariable prediction model for TNs. Future work is needed to determine whether the characteristics of this study population influenced these observational results. This study also has several limitations. First, some tiny TNs may not be detected by high-resolution ultrasound, which could cause slight bias during grouping. Second, the lack of repeated laboratory tests in the subjects is a possible limitation. Third, there was no dose-response analysis of factors such as blood glucose and blood pressure. The large sample size, however, was a major strength of the present study.

Table 3 Predictive Value For The Prediction Model And Independent Risk Factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Youden Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis ≥45 years</td>
<td>52.3</td>
<td>67.5</td>
<td>0.198</td>
</tr>
<tr>
<td>Gender (female)</td>
<td>56.1</td>
<td>62.9</td>
<td>0.190</td>
</tr>
<tr>
<td>BMI ≥24 kg/m²</td>
<td>63.0</td>
<td>47.0</td>
<td>0.100</td>
</tr>
<tr>
<td>Hypertension</td>
<td>12.3</td>
<td>90.0</td>
<td>0.023</td>
</tr>
<tr>
<td>TGAB ≥16.5 IU/mL</td>
<td>46.6</td>
<td>61.8</td>
<td>0.084</td>
</tr>
<tr>
<td>Model</td>
<td>65.9</td>
<td>77.8</td>
<td>0.437</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; TGAB, anti-thyroglobulin antibody.

Conclusions

This work is one of the largest investigations of risk factors affecting the occurrence of TNs. A simple model was established based on five identified independent risk factors. Patients with index scores ≥4 points could be considered at higher risk for the occurrence of TNs; these results provide some insights for clinical practice that may help improve preventive and therapeutic strategies for TNs.

Acknowledgments

This study was supported by grants from the National Key R&D Program of China (2017YFO907504), the National Natural Science Foundation (81702646), the National Post-Doctoral Research Project (2018M631086, 2019T120846), the Sichuan Providence Science and Technology Project of China (2017SZ0139), the Sichuan University for Youth Fund (2017SCU11016), the Health and Family Planning Commission of Sichuan Province (17PJ398), the Postdoctoral Sustentation Fund of Sichuan University (2017SCU12035), and the Post-Doctoral Research Project, West China Hospital, Sichuan.
Figure 4 (A) The occurrence of thyroid nodules increased with increasing score. (B) Diameter of thyroid nodules was correlated with the index points in the cases group.
University (2019HXBH043). Li Genpeng and Wang Lijuan are co-first authors.

**Author Contributions**
All authors made substantial contributions to the study conception and design and data acquisition, analysis and interpretation; were involved in drafting the article or revising it critically for important intellectual content; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work.

**Disclosure**
The authors report no conflicts of interest in this work.

**References**