Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development of a Predictive Model and Nomogram for Thyroid Dysfunction Following Thermal Ablation of Papillary Thyroid Carcinoma

Zhiyuan Chen
Zhejiang Chinese Medical University
Zheng Qiuqing
Zhejiang Cancer Hospital
Yang Zhang
Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences
Yao Jingcao
Zhejiang Chinese Medical University
Jiaheng Huang
Zhejiang Cancer Hospital
Yin Zheng
Zhejiang Chinese Medical University
Chen Chen
Zhejiang Cancer Hospital
Xinying Zhu
Zhejiang Cancer Hospital
Qian Li
Zhejiang Chinese Medical University
Hao Manman
Zhejiang Cancer Hospital
Yuqi Yan
Zhejiang Chinese Medical University
Zhiyan Jin
Zhejiang Cancer Hospital
Vicky Yang Wang
Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences
Dong Xu
xudong@zjcc.org.cn

Zhejiang Chinese Medical University

Research Article

Keywords: papillary thyroid carcinoma, thermal ablation, thyroid dysfunction, predictive model, nomogram

Posted Date: April 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9037631/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Thyroid dysfunction (TD) can occur after thermal ablation of papillary thyroid carcinoma (PTC), yet no dedicated clinical prediction tool has been available. To develop and internally validate a multivariable logistic regression model for predicting post-ablation TD.

Methods

In this single-center retrospective cohort, 295 patients with PTC who underwent ultrasound-guided thermal ablation and completed 12-month follow-up were randomly split 80:20 into training (n = 236) and validation (n = 59) sets. Nineteen candidate variables spanning demographics, serology, imaging, and procedural parameters were screened by univariable analysis; significant factors entered a multivariable logistic model via bidirectional stepwise selection (AIC). Discrimination, calibration, and clinical utility were assessed by AUC, calibration curves with the Hosmer–Lemeshow test, and decision curve analysis (DCA). The model was presented as a nomogram and a web-based calculator.

Results

Three independent predictors were retained: preoperative TSH (OR per mIU/mL, 1.91; 95% CI, 1.31–2.78), ablation time (OR per second, 1.01; 95% CI, 1.01–1.01), and nodule location (bilateral vs unilateral: OR, 3.85; 95% CI, 1.58–9.43; lobe+isthmus vs unilateral: OR, 7.95; 95% CI, 1.40–45.03). Overall, 83/295 (28.1%) patients developed TD within 12 months. Discrimination was good (AUC 0.72 training; 0.81 validation). Calibration was acceptable (Hosmer–Lemeshow P = 0.950 training; P = 0.251 validation). Across a wide range of threshold probabilities, DCA showed higher net benefit than treat-all or treat-none strategies. A nomogram and online calculator were derived for individualized risk estimation.

Conclusion

An internally validated multivariable model using preoperative TSH, ablation time, and ablation extent (bilateral or lobe+isthmus) provides individualized risk prediction of post-ablation TD and may support pre-procedure counseling, peri-procedural planning, and tailored follow-up. External, multicenter validation and prospective evaluation are warranted.

Introduction

Thyroid cancer is a common endocrine malignancy with a globally increasing incidence[1–3]. Papillary thyroid carcinoma (PTC) generally has a favorable prognosis, but a subset of patients present with cervical lymph node or distant metastases, and the presence of distant metastasis markedly worsens outcomes[4, 5].In recent years, thermal ablation has emerged as a minimally invasive alternative to surgery for carefully selected patients with low-risk PTC. Studies report post-ablation local tumor progression rates of approximately 1.25%–7.7%, with overall therapeutic efficacy comparable to surgery, supporting the safety and effectiveness of thermal ablation in PTC treatment[6]. However, the main complications of PTC thermal ablation include pain, voice changes, bleeding, and thyroid function dysfunction[7]. Thyroid dysfunction (TD)—especially hypothyroidism—often lacks specific symptoms in the early stage,, making clinical diagnosis challenging; thus, diagnosis relies primarily on measurements of serum TSH and free thyroxine levels[8]. Especially, TD has been linked to increased risks of coronary artery disease, heart failure, and stroke[9]. Therefore, timely evaluation and individualized prediction of TD risk after thermal ablation are essential for optimizing follow-up strategies and enabling early intervention.

Previous studies have investigated risk factors for TD after thyroid nodule ablation. For example, Song *et al.* identified pre-ablation TSH levels, Hashimoto's thyroiditis, and multiple nodules as risk factors for post-ablation TD in a large retrospective cohort[10]. However, that study did not derive a multivariable clinical prediction model, which limits its applicability for individualized risk assessment in routine practice. In addition, ablation-related technical and energy-related parameters were not considered. Building on this work, the present study aimed to incorporate a broader range of candidate predictors—including patient demographics, biochemical indices, imaging characteristics, and detailed ablation technical parameters—to identify risk factors for TD after PTC thermal ablation using multivariate logistic regression. On this basis, we developed and validated a clinical prediction model for post-ablation TD. Finally, we translated the model into an easy-to-use nomogram and a web-based calculator to visualize individualized TD risk, facilitate risk stratification, and support clinical decision-making.

Methods

Study Design and Population

This retrospective study included PTC patients who underwent ultrasound-guided thermal ablation of thyroid nodules at Zhejiang Cancer Hospital between 2015 and 2024. All patients had at least 12 months of postoperative follow-up post-ablation data.

Inclusion criteria

(1) Age between 18 and 65 years; (2) Pathologically confirmed PTC (via fine-needle aspiration), clinical stage T1N0M0, suitable for thermal ablation; (3) Single PTC nodule. Patients with multiple nodules were included only if all other nodules were confirmed benign or low-risk by fine-needle aspiration (FNA) or imaging evidence and not considered malignant[11].

Exclusion criteria

(1) Severe comorbid conditions (e.g. heart failure or hepatic/renal failure); (2) Preoperative TD (any of T3, T4, or TSH outside the normal reference range, as defined in the Outcome section); (3) Use of thyroid hormone replacement or anti-thyroid drugs preoperatively; (4) History of thyroid surgery; (5) Missing key follow-up data; (6) Refusal to provide informed consent.

A total of 295 patients meeting the criteria were enrolled. The included patients were randomly divided into a training set and a validation set in an 8:2 ratio. **(**Fig. 1**)** illustrates the patient enrollment and allocation process.

Pre-ablation Assessment and Data Collection

Prior to ablation, each nodule was evaluated by ultrasound, including measurement of three orthogonal diameters (the maximal diameter and two perpendicular diameters), assessment of shape (tall vs. wide), location, calcifications, echogenicity, margins, vascularity, internal composition, and relationship to adjacent anatomical structures. All patients underwent comprehensive pre-ablation examinations, including complete blood count, thyroid function tests, coagulation profile, and additional imaging studies when indicated. Nodule volume was calculated as $V = (\pi/6)abc$, where a is the maximum diameter, b and c are the two perpendicular diameters[12]. Demographic data, laboratory results, ultrasound findings, and ablation technical parameters were retrieved from the hospital electronic medical record and imaging databases.

RFA procedure

A Voko color Doppler ultrasound diagnostic system equipped with an L741 high-frequency linear-array probe was used, with the probe frequency adjusted to 12 MHz. Thermal ablation was performed using a MedSphere ablation system (MedSphere International, Inc.). For all patients diagnosed with T1N0M0 PTC, Type L-121 disposable 18-gauge ablation needles were selected. Patients were positioned supine, with their shoulders elevated to facilitate head tilting backwards. After complete exposure, the neck was disinfected and sterile drapes were applied. Under ultrasound guidance, 2% lidocaine was infiltrated for local anesthesia. The use of hydro-dissection was determined according to the tumor's proximity to adjacent tissues and organs. When tumors were close to critical neck structures such as the carotid artery, trachea, esophagus, or recurrent laryngeal nerve (RLN), hydrodissection was performed to prevent thermal injury. An appropriate volume of normal saline or 5% glucose solution was injected to create a protective "hydrodissection zone" between the tumor and these structures. Under real-time ultrasound guidance, the ablation needle was advanced into the deepest portion of the tumor, and ablation was initiated after activating the generator. A moving-shot technique, usually via the thyroid isthmus, was used to ensure complete coverage of the lesion. To minimize residual tumor and local recurrence, an expanded ablation aimed to prevent tumor residue or recurrence, ensuring that perilesional echogenic changes extended beyond the tumor boundary by at least 3 mm. Once the perilesional echogenic changes completely encompassed the tumor and its surrounding area, ablation was concluded. Immediately after ablation, contrast-enhanced ultrasound (CEUS) examination was performed to assess the completeness of the procedure. If residual tumors were detected, supplementary ablation was promptly performed. After ablation, patients were observed in the treatment room for 1–2 h. Vital signs and potential procedure-related complications were closely monitored during and after the procedure. Patients were discharged after an additional 3–5 h of hospital observation, provided no abnormalities were detected.

Predictors Definitions

Nineteen candidate predictors were involved in the model construction, including patient demographics, biochemical markers, nodule characteristics, and ablation parameters, including: patient age; preoperative serum levels of free triiodothyronine (FT3), free thyroxine (FT4), TSH, thyroglobulin antibody (TGAb), and thyroid peroxidase antibody (TPOAb); maximum diameter of the thyroid nodule; ablation time; ablation power; total ablation energy; overall volume of the ablated nodules; ratio of total energy to nodule volume; patient sex; history of underlying thyroid disease; distribution of the ablated nodules (anatomical location); number of nodules ablated; use of intraoperative isolation fluid and whether epinephrine was added; and the ablation modality (radiofrequency vs. microwave ablation). For simplicity, some variables are abbreviated in the **(**Table 1**)**.

Table 1
Predictor variables and their definitions.

Shorthand variables	Full name variables
FT3, FT4, TSH, TGAb, TPOAb	Preoperative serum levels of free triiodothyronine, free thyroxine, thyroid-stimulating hormone, thyroglobulin antibody, and thyroid peroxidase antibody
Maximum longitude	Maximum diameter of the thyroid nodule
Time	Ablation time
Power	Ablation power
Energy	Total ablation energy
Total volume	Overall volume of the ablated nodules
Total energy-to-volume ratio	Ratio of total ablation energy to the overall nodule volume
Nodule location	Anatomical distribution of the ablated nodules
Number of nodules	Number of nodules subjected to ablation
Intraoperative isolation fluid	Use of isolation fluid during ablation and whether epinephrine was added
Ablation method	Type of thermal ablation applied

Outcome Definition

The primary outcome was TD occurring within 12 months after the procedure. Patients who met the biochemical criteria for any thyroid function abnormality at any of the scheduled follow-up visits (1, 3, 6, 9, or 12 months) were defined as a positive outcome (TD), whereas those with normal thyroid function throughout follow-up were defined as negative. TD included subclinical hyperthyroidism, overt hyperthyroidism, subclinical hypothyroidism, or overt hypothyroidism. Subclinical hyperthyroidism was defined as a subnormal thyroid-stimulating hormone (TSH) level with normal free thyroxine (FT4); overt hyperthyroidism as subnormal TSH with elevated FT4; subclinical hypothyroidism as elevated TSH with normal FT4; and overt hypothyroidism as elevated TSH with subnormal FT4[13].

Normal reference ranges for thyroid function tests in this study were: total triiodothyronine (T3) 0.8–2.0 ng/mL, free triiodothyronine (FT3) 2.0–4.4 pg/mL, total thyroxine (T4) 5.1–14.1 μg/dL, free thyroxine (FT4) 0.93–1.7 ng/dL, and TSH 0.27–4.20 mIU/L.

Model Development and Validation

All statistical analyses were performed using R software (v4.5.1) and IBM SPSS Statistics (v27.0.1). Continuous variables were expressed as mean ± standard deviation (SD) if normally distributed, or as median and interquartile range (25th–75th percentiles, IQR) if not normally distributed. For baseline comparisons on the training and validation sets, continuous variables with normal distribution were compared using independent-sample *t* tests, and non-normal variables were compared with Mann–Whitney U tests. Categorical variables were compared using the Chi-square test. $P < 0.05$ was considered statistically significant. Only cases with complete follow-up exams were included in the analysis to ensure reliability.

For the feature selection, univariate binary logistic regression was first performed in the training set for each candidate predictor. Variables with $P < 0.05$ were then applied to build a multivariate logistic regression using a bidirectional stepwise selection method. The selected model with the lowest Akaike information criterion (AIC) was chosen as the final model.

Discrimination was assessed by the area under the ROC curve (AUC), and Calibration was evaluated by calibration plots as well as the Hosmer–Lemeshow goodness-of-fit test. Clinical validity was assessed via decision curve analysis (DCA). These evaluations were performed on both the training set and the hold-out validation set.

To enhance the model interpretability, the nomogram was established for the visualization of the multivariable logistic regression model, which can provides a graphical representation of the model's predictions. Then web-based tools was developed for easy risk assessment in clinical practice.

Results

Study Population

A total of 647 patients were screened. Among these, 352 cases were excluded: 6 cases with severe underlying diseases, 35 cases with preoperative TD, 42 cases on thyroid medication, and 269 cases with incomplete follow-up. Then 295 patients were finally included in this study. Among these, 83 patients (28.1%) developed TD during the 12-month follow-up (positive cases), while 212 (71.9%) showed normal thyroid function (negative cases). The trends of various types of TD through follow-ups are shown in (Fig. 2). Overall, the incidences of all categories of TD decreased over time. The number of Subclinical and overt hyperthyroidism incidences showed a monotonically decreasing trend post-ablation, while subclinical and overt hypothyroidism incidence number showed slight fluctuations but an overall decline trend during the 12 months period.

The 295 patients were randomly divided into a training set (236 cases) and a validation set (59 cases) at an 8:2 ratio. (Table 2) summarizes the baseline characteristics of the total cohort, as well as the training and validation subsets, and the balance test between the two sets. The two sets were comparable in age, sex, thyroid function indices (FT3, FT4, TSH, TPOAb), history of thyroid disease, nodule location and number, ablation time, power, total energy, use of isolation fluid, and ablation modality (all $P > 0.05$). However, baseline TGAb levels, maximum nodule diameter, total nodule volume, and energy-to-volume ratio differed significantly between the training and validation sets (all $P < 0.05$).

Table 2
Balance test of training set and validation set.

Variables	Total (n = 295)	test (n = 59)	train (n = 236)	Statistic	*P*
Age, Mean ± SD	41.41 ± 9.69	42.36 ± 10.22	41.18 ± 9.56	t = 0.83	0.404
FT3(pg/mL), Mean ± SD	3.25 ± 0.34	3.24 ± 0.30	3.26 ± 0.35	t=-0.36	0.719
FT4(ng/dL), M (Q_1, Q_3)	1.24 (1.15, 1.37)	1.24 (1.17, 1.36)	1.25 (1.15, 1.37)	Z=-0.18	0.858
TSH(μIU/mL), M (Q_1, Q_3)	1.53 (1.13, 2.07)	1.51 (1.25, 1.98)	1.54 (1.11, 2.16)	Z=-0.27	0.785
TGAb(IU/mL), M (Q_1, Q_3)	15.00 (15.00, 25.35)	15.00 (10.10, 18.35)	15.00 (15.00, 31.40)	Z=-2.84	0.005
TPOAb(IU/mL), M (Q_1, Q_3)	30.70 (28.00, 47.30)	32.00 (28.00, 47.30)	30.55 (28.00, 46.57)	Z=-0.11	0.912
Maximum longitude(mm), M (Q_1, Q_3)	6.00 (4.45, 8.00)	5.00 (4.00, 6.00)	6.00 (4.88, 8.00)	Z=-2.85	0.004
Time(s), M (Q_1, Q_3)	168.00 (120.00, 236.00)	155.00 (120.00, 210.00)	170.00 (120.00, 240.00)	Z=-0.93	0.352
Power(W), M (Q_1, Q_3)	25.00 (25.00, 27.50)	25.00 (25.00, 30.00)	25.00 (25.00, 27.50)	Z=-1.06	0.289
Energy(KJ), M (Q_1, Q_3)	4.20 (3.00, 6.00)	4.14 (3.00, 6.00)	4.20 (3.00, 6.00)	Z=-0.35	0.724
Total volume(mm^3), M (Q_1, Q_3)	66.20 (31.40, 175.90)	52.40 (32.25, 94.20)	78.50 (31.40, 190.05)	Z=-2.23	0.025
Total energy to volume ratio(KJ/mm^3), M (Q_1, Q_3)	0.06 (0.03, 0.11)	0.08 (0.05, 0.13)	0.06 (0.03, 0.10)	Z=-2.77	0.006
Gender, n(%)				χ^2=0.27	0.605
Male	83 (28.14)	15 (25.42)	68 (28.81)		
Female	212 (71.86)	44 (74.58)	168 (71.19)		
History of underlying thyroid disease, n(%)				-	0.875
No other underlying thyroid disease	243 (82.37)	48 (81.36)	195 (82.63)		
History of hyperthyroidism	2 (0.68)	0 (0.00)	2 (0.85)		
History of subacute thyroiditis	2 (0.68)	0 (0.00)	2 (0.85)		
History of hashimoto's thyroiditis	48 (16.27)	11 (18.64)	37 (15.68)		
Nodule location, n(%)				-	1.000
Unilateral lobe of the thyroid gland	239 (81.02)	48 (81.36)	191 (80.93)		
Bilateral lobes of the thyroid gland	35 (11.86)	7 (11.86)	28 (11.86)		
Isthmus of the thyroid gland	10 (3.39)	2 (3.39)	8 (3.39)		
Thyroid lobe and isthmus	11 (3.73)	2 (3.39)	9 (3.81)		
Number of nodules, n(%)				-	0.764
1	214 (72.54)	46 (77.97)	168 (71.19)		
2	65 (22.03)	11 (18.64)	54 (22.88)		
3	11 (3.73)	1 (1.69)	10 (4.24)		
4	5 (1.69)	1 (1.69)	4 (1.69)		
Intraoperative isolation fluid, n(%)				χ^2=0.78	0.678
No	10 (3.39)	3 (5.08)	7 (2.97)		
Used intraoperative isolation fluid	219 (74.24)	42 (71.19)	177 (75.00)		
Used intraoperative isolation fluid and adrenaline	66 (22.37)	14 (23.73)	52 (22.03)		
Ablation method, n(%)				χ^2=1.33	0.249
SD: standard deviation, M: Median, Q_1: 1st Quartile, Q_3: 3st Quartile 100 + 46.6*3 + 40+33.3*3					

Variables	Total (n = 295)	test (n = 59)	train (n = 236)	Statistic	*P*
RFA	223 (75.59)	48 (81.36)	175 (74.15)		
MWA	72 (24.41)	11 (18.64)	61 (25.85)		
Overall thyroid function, n(%)				χ^2=0.27	0.605
Negative result	212 (71.86)	44 (74.58)	168 (71.19)		
Positive result	83 (28.14)	15 (25.42)	68 (28.81)		
t: t-test, Z: Mann-Whitney test, χ^2: Chi-square test, -: Fisher exact					
SD: standard deviation, M: Median, Q_1: 1st Quartile, Q_3: 3st Quartile 100 + 46.6*3 + 40+33.3*3					

Model Construction

Univariate Analysis Results In the training set, higher preoperative TSH was associated with an increased risk of post-ablation (TD) (OR per mIU/mL, 1.56; 95% CI, 1.10–2.19; P = 0.011). Longer ablation time was also associated with higher risk (OR per second, 1.01; 95% CI, 1.01–1.01; P < 0.001). Greater total energy delivered was associated with higher risk (OR per kJ, 1.19; 95% CI, 1.08–1.31; P < 0.001). Patients with more ablated nodules were at higher risk than those with a single nodule (P < 0.05 for 2–4 vs 1). Nodule location was significant: compared with unilateral-lobe ablation, bilateral ablation had an OR of 4.59 (95% CI, 2.02–10.44; P < 0.01), and lobe-plus-isthmus ablation had an OR of 12.05 (95% CI, 2.41–60.13; P < 0.01). No other variables reached statistical significance in univariable analyses.

Multivariate Analysis Results Variables significant in univariable analysis were entered into a multivariable logistic regression with AIC-minimizing stepwise selection. The final model retained three independent predictors of TD: preoperative TSH (OR, 1.91; 95% CI, 1.31–2.78; P < 0.001), ablation time (OR per second, 1.01; 95% CI, 1.01–1.01; P = 0.013), and nodule location (bilateral vs unilateral: OR, 3.85; 95% CI, 1.58–9.43; P = 0.003; lobe+isthmus vs unilateral: OR, 7.95; 95% CI, 1.40–45.03; P = 0.019). These findings indicate that higher TSH, longer ablation duration, and more extensive ablation (involving both lobes and/or the isthmus) are associated with increased TD risk.

The coefficients of the final multivariate logistic regression model are presented in **(**Table 3**)**. The final logistic regression equation (logit *P*) incorporates the three predictors above (TSH, time, and nodule location with dummy variables, using unilateral lobe as the reference category). For reference, the odds ratios from both univariate and multivariate analyses are also provided in **(**Table 3**)**.

Table 3
Univariate and multivariate logistic regression analysis results for predictors of post-ablation TD.

Variables	Univariate logistic regression					Multivariate logistic regression				
	β	S.E	Z	*P*	OR (95%CI)	β	S.E	Z	*P*	OR (95%CI)
Age	-0.00	0.02	-0.27	0.785	1.00 (0.97 ~ 1.03)					
Ft3(pg/mL),	-0.48	0.42	-1.15	0.252	0.62 (0.27 ~ 1.41)					
Ft4(ng/dL),	-0.29	0.91	-0.31	0.755	0.75 (0.13 ~ 4.49)					
TSH(μIU/mL)	0.44	0.17	2.53	0.011	1.56 (1.10 ~ 2.19)	0.64	0.19	3.34	< .001	1.91 (1.31 ~ 2.78)
TG-Ab(IU/mL)	0.00	0.00	1.46	0.145	1.00 (1.00 ~ 1.01)					
TPO-Ab(IU/mL)	0.00	0.00	1.10	0.270	1.00 (1.00 ~ 1.00)					
Maximum longitude(mm)	0.05	0.03	1.69	0.091	1.05 (0.99 ~ 1.12)					
Time(s)	0.01	0.00	3.80	< .001	1.01 (1.01 ~ 1.01)	0.01	0.00	2.49	0.013	1.01 (1.01 ~ 1.01)
Power(W)	-0.00	0.04	-0.03	0.973	1.00 (0.93 ~ 1.07)					
Energy(KJ)	0.18	0.05	3.63	< .001	1.19 (1.08 ~ 1.31)					
Total volume(mm^3)	0.00	0.00	0.71	0.478	1.00 (1.00 ~ 1.00)					
Total energy to volume ratio(KJ/mm^3)	-0.65	1.90	-0.34	0.731	0.52 (0.01 ~ 21.48)					
Number of nodules										
1					1.00 (Reference)					
2	0.92	0.33	2.76	0.006	2.52 (1.31 ~ 4.86)					
3	2.15	0.72	3.00	0.003	8.56 (2.11 ~ 34.76)					
4	2.40	1.17	2.05	0.040	11.00 (1.11 ~ 108.95)					
Gender										
Male					1.00 (Reference)					
Female	0.49	0.34	1.45	0.147	1.63 (0.84 ~ 3.15)					
History of underlying thyroid disease										
No other underlying thyroid disease					1.00 (Reference)					
History of hyperthyroidism	0.96	1.42	0.67	0.500	2.61 (0.16 ~ 42.49)					
History of subacute thyroiditis	-14.61	1029.12	-0.01	0.989	0.00 (0.00 ~ Inf)					
OR: Odds Ratio, CI: Confidence Interval										

Variables	Univariate logistic regression					Multivariate logistic regression				
	β	S.E	Z	*P*	OR (95%CI)	β	S.E	Z	*P*	OR (95%CI)
History of hashimoto's thyroiditis	0.35	0.38	0.91	0.361	1.41 (0.67 ~ 2.98)					
Nodule location										
Unilateral lobe of the thyroid gland					1.00 (Reference)					1.00 (Reference)
Bilateral lobes of the thyroid gland	1.52	0.42	3.63	< .001	4.59 (2.02 ~ 10.44)	1.35	0.46	2.96	0.003	3.85 (1.58 ~ 9.43)
Isthmus of the thyroid gland	0.14	0.83	0.16	0.869	1.15 (0.22 ~ 5.89)	0.61	0.86	0.70	0.481	1.83 (0.34 ~ 9.88)
Thyroid lobe and isthmus	2.49	0.82	3.03	0.002	12.05 (2.41 ~ 60.13)	2.07	0.88	2.34	0.019	7.95 (1.40 ~ 45.03)
Intraoperative isolation fluid										
No					1.00 (Reference)					
Used intraoperative isolation fluid	0.83	1.09	0.76	0.447	2.30 (0.27 ~ 19.57)					
Used intraoperative isolation fluid and adrenaline	1.16	1.12	1.03	0.302	3.18 (0.35 ~ 28.46)					
Ablation method										
RFA					1.00 (Reference)					
MWA	0.15	0.32	0.47	0.640	1.16 (0.62 ~ 2.19)					
OR: Odds Ratio, CI: Confidence Interval										

The final logistic regression model can be expressed as:

$$\begin{aligned} logit(P) = -3.13 + 0.64 \times TSH + 0.01 \times Time + 1.35 \times I(Nodule\ location \\ = Bilateral\ lobes\ of\ the\ thyroid\ gland) + 0.61 \times I(Nodule\ location \\ = Isthmus\ of\ the\ thyroid\ gland) + 2.07 \times I(Nodule\ location \\ = Thyroid\ lobe\ and\ isthmus) \end{aligned}$$

where $\varvec{P}$ is the probability of TD, $\varvec{Time}$ is ablation time in seconds, and $\varvec{Nodule\ location}$ is anatomical distribution of the ablated nodules. The model coefficients correspond to the multivariate β values in Table 3. For example, holding other factors constant, a 1-second increase in ablation time corresponds to an increase in log-odds of dysfunction by 0.009 (OR ~ 1.009). Model coefficients and ORs for all variables are provided in Table 3.

Model Performance Evaluation

Discrimination The prediction model demonstrated good discrimination in both the training and validation sets (Fig. 3). The AUC was 0.72 (95% CI: 0.65–0.80) for the training set and 0.81 (95% CI: 0.66–0.95) for the validation set, indicating moderate to excellent ability to distinguish between patients with and without post-ablation TD. Using a probability cutoff of 0.347 (chosen by Youden's index) as the decision threshold, the model achieved a sensitivity of 0.85 and specificity of 0.51 in the training set (accuracy 0.75), and a sensitivity of 0.86 and specificity of 0.67 in the validation set (accuracy 0.81). This suggests the model has a high true positive rate (sensitivity) in identifying patients at risk, with a moderate false positive rate given the trade-off in specificity.

Calibration Calibration curves for the model in the training and validation sets are shown in (Fig. 4). In the training set, the predicted probabilities closely matched the observed outcomes: the calibration curve lay near the diagonal line (ideal calibration), and the bias-corrected curve was also close to the ideal line, indicating good model fit. The Hosmer–Lemeshow test in the training set yielded *P* = 0.950, indicating no

significant difference between predicted and observed frequencies (i.e. no lack of fit). Similarly, the validation set's calibration curve showed good agreement between predicted risk and actual incidence of dysfunction, and the Hosmer–Lemeshow test $P = 0.251$, suggesting the model remained well-calibrated in the independent validation cohort.

Clinical Utility Decision curve analysis for the model is presented in **(**Fig. 5**)**. In both the training and validation sets, the model's decision curve (net benefit vs. threshold probability) was above the "treat-all" (gray diagonal line) and "treat-none" (black horizontal line) curves across the majority of threshold probabilities. This indicates that using the model to guide intervention decisions would achieve a greater net benefit than either intervening on all patients or on no patients, over a wide range of risk thresholds. In other words, the model provides meaningful clinical net benefit and could assist in decision-making.

Model Presentation

Nomogram To provide an intuitive tool for risk estimation, we constructed a nomogram incorporating the three independent predictors (nodule location, TSH, and ablation time) based on the multivariate logistic regression model **(**Fig. 6). Each predictor is assigned points proportional to its regression coefficient (the reference category receives 0 points; coefficients are linearly mapped to the point scale). By summing the scores for all predictors, a total score is obtained, which corresponds to a predicted probability of TD on the bottom scale.

Web-Based Risk Calculator To facilitate clinical use of the model, we developed two web-based tools: one tool for predicting the risk of TD after ablation, and another tool for estimating the maximum acceptable ablation time given a user-specified acceptable risk level. The risk prediction tool is available at https://8u88u.github.io/Predict-Probability-P-/, and the ablation time tool at https://8u88u.github.io/Required-Ablation-Time-for-Target-P/. By inputting a patient's risk factor values (TSH, planned ablation time, nodule location, etc.), the first tool calculates the estimated probability of post-ablation TD. Conversely, by inputting an acceptable risk threshold, the second tool provides the upper limit of ablation time that keeps the predicted risk below that threshold. (Fig. 7**)** shows a screenshot of the web-based tools' interface.

Discussion

This study developed a prediction model for TD after PTC thermal ablation based on 19 candidate variables, and demonstrated good discrimination (AUC 0.72 in training, 0.81 in validation), with acceptable calibration and clinical net benefit in internal validation. The model may provide a practical aid in early clinical prevention and management of TD post-ablation.

To our knowledge, no prior clinical prediction model specifically targeting TD after PTC ablation has been previously reported. We combined demographic, serological, imaging, and ablation technical factors through multivariate logistic regression to establish an early prediction tool. This work expands the research landscape in this field and provides fundamental data and a modeling framework for future multicenter validation studies. Currently, clinicians clinicians lack effective tools for individualized prediction of TD following PTC ablation. These tools enable rapid identification of high-risk patients and risk stratification, facilitating timely clinical intervention.

This study observed that post-ablation TD primarily manifested as hyperthyroidism and hypothyroidism. Post-ablation hypothyroidism may result from direct thyroid tissue destruction and the release of thyroid antigens during ablation, triggering an immune-mediated response[14]. Some patients experience transient hyperthyroidism, primarily due to destruction of thyroid follicles, which results in the release of stored thyroid hormone into the bloodstream and short-term thyrotoxicosis[15, 16] In our study, hyperthyroidism most often developed between one and three months after ablation, with only one case experiencing persistent hyperthyroidism at 12 months, supporting the hypothesis that hyperthyroidism is a transient phenomenon. Independent risk factor analysis demonstrated that higher preoperative TSH, longer ablation time, and greater ablation extent (bilateral or lobe+isthmus, vs unilateral lobe) independently increased TD risk of post-ablation TD. Especially, maximum diameter and volume did not remain independent predictors after adjustment, suggesting extent of ablation rather than simple size metrics is more relevant. The CEM43 thermal dose concept aligns with these observations: longer ablation translates to higher cumulative heat and wider tissue injury[17]. When the ablation zone involves both lobes or the lobe plus the isthmus, overall tissue damage is more significant, further increasing the risk of post-ablation dysfunction. Elevated preoperative TSH levels often indicate diminished thyroid reserve or insufficiency, resulting in poor recovery and an increased risk of hypothyroidism.

Our findings differ from those of Qiu et al. Qiu et al. focused on the ablation of large benign nodules and found that low preoperative TSH was a risk factor for postoperative TD, whereas our study found that high preoperative TSH was an independent risk factor for postoperative TD. Furthermore, unlike the findings of Qiu et al., which showed a correlation with nodule size, we found that the maximum diameter and volume of the nodule had no significant impact in patients with PTC[18]. Given that Qiu et al. focused on the ablation of large benign nodules, these differences may be related to the different study populations and the different ablation methods used[19, 20]. We also observed that ablation power was not an independent predictor of TD, in line with experimental data indicating a non-linear power–efficiency relationship: moderate power can optimize energy deposition, whereas excessively low or high settings may produce uneven heating and suboptimal efficacy. Thus, power should be individualized rather than interpreted as a monotonic risk driver[21, 22].

Preoperative TSH, anticipated ablation duration, and planned extent (unilateral vs bilateral or lobe+isthmus) can inform pre-procedure counseling, peri-procedural planning, and tailored follow-up. Even though power did not independently predict TD in our model, power settings still warrant case-by-case optimization to balance efficacy and safety, particularly near critical anatomy (e.g., tracheoesophageal groove, recurrent laryngeal nerve paths), where lower power, moving-shot techniques, hydrodissection, and neuromonitoring may mitigate risk[23–25].

Despite its strengths, this study has several limitations that should be acknowledged. First, the sample size is relatively modest for predictive modeling, which may affect the model's robustness and generalizability. Second, the retrospective design carries an inherent risk of selection bias, particularly because patients with incomplete follow-up had to be excluded despite predefined inclusion criteria. Third, we used a traditional multivariate logistic regression approach; although this method is transparent and clinically interpretable, more complex machine learning or deep learning models might achieve higher predictive accuracy, particularly in larger datasets. Fourth, our follow-up duration was limited to 12 months, which may not capture long-term thyroid function dynamics; longer follow-up is needed to assess whether late-onset dysfunction occurs.

Future research should focus on a few key areas. First, external validation of the model in larger, multicenter cohorts is needed to evaluate generalizability and robustness, and to mitigate any single-center bias. Second, prospective cohort studies or randomized controlled trials would strengthen causal inference between risk factors and outcomes. Third, incorporating machine learning or deep learning methods to build predictive models and comparing them with the traditional logistic model could identify further improvements. Fourth, integrating the model into electronic health records or clinical decision support systems could enable real-time risk assessment and prompt clinicians toward personalized management and early intervention. Finally, exploring new predictive factors such as radiomic features, molecular markers, or lifestyle factors may further enhance the model's performance and clinical applicability.

Conclusion

Using routinely available clinical and procedural data, we developed an internally validated multivariable logistic regression model that predicts the 12-month risk of TD after PTC thermal ablation. Higher preoperative TSH, longer ablation time, and more extensive ablation (bilateral or lobe+isthmus) independently increased risk. The model showed good discrimination, acceptable calibration, and favorable clinical net benefit, and is delivered as a nomogram and web-based tool to facilitate point-of-care use. These findings support individualized pre-procedure counseling and post-procedure surveillance; nonetheless, external multicenter validation and prospective studies are needed before broad clinical deployment.

Abbreviations

TD, thyroid dysfunction; PTC, papillary thyroid carcinoma; T1N0M0, T1N0M0 stage of papillary thyroid carcinoma; TSH, thyroid-stimulating hormone; T3, triiodothyronine; T4, thyroxine; FT3, free triiodothyronine; FT4, free thyroxine; TGAb, thyroglobulin antibody; TPOAb, thyroid peroxidase antibody; FNA, fine-needle aspiration; RFA, radiofrequency ablation; MWA, microwave ablation; CEUS, contrast-enhanced ultrasound; RLN, recurrent laryngeal nerve; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic; DCA, decision curve analysis; SD, standard deviation; IQR, interquartile range; OR, odds ratio; CI, confidence interval; CEM43, cumulative equivalent minutes at 43 °C.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Zhejiang Cancer Hospital (Approval No.: IRB-2025-387 (IIT)). The committee approved the study protocol and granted a waiver of informed consent for retrospective data collection. All patients had previously provided written informed consent for the thermal ablation procedure. The study was conducted in accordance with relevant national regulations and the principles of the Declaration of Helsinki.

Consent for publication

The authors affirm that all individuals included in this manuscript have provided their written informed consent for publication. Additionally, any personal data or images included in the manuscript have been published with the consent of the individuals involved.

Availability of data and materials

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at Lanzhou University Second Hospital.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

This work was supported by the National Natural Science Foundation of China (82471990) and the “Pioneer” and “Leading Goose” R&D Program of Zhejiang Province (2023C04039).

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Zhiyuan Chen, Zheng Qiuqing, and Yang Zhang. The first draft of the manuscript was written by Zhiyuan Chen, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Special thanks to the National Natural Science Foundation of China (Grant No. 82471990) and the “Pioneer” and “Leading Goose” R&D Program of Zhejiang Province (2023C04039) for their financial support.

References

1. J. Zhao, W. Zhang, D. Lu, C. Shao, Y. Chen, X. Huang, et al., Clinical prognostic risk assessment of different pathological subtypes of papillary thyroid cancer: a systematic review and network meta-analysis, Langenbecks Arch Surg 410 (2025) 251, https://doi.org/10.1007/s00423-025-03841-2.
2. J. Zhang, S. Xu, High aggressiveness of papillary thyroid cancer: from clinical evidence to regulatory cellular networks, Cell Death Discov 10 (2024) 378, https://doi.org/10.1038/s41420-024-02157-2.
3. A. Forma, K. Kłodnicka, W. Pająk, J. Flieger, B. Teresińska, J. Januszewski, et al., Thyroid cancer: epidemiology, classification, risk factors, diagnostic and prognostic markers, and current treatment strategies, Int J Mol Sci 26 (2025) 5173, https://doi.org/10.3390/ijms26115173.
4. M.D. Ringel, J.A. Sosa, Z. Baloch, L. Bischoff, G. Bloom, G.A. Brent, et al., 2025 American Thyroid Association management guidelines for adult patients with differentiated thyroid cancer, Thyroid 35 (2025) 841–985, https://doi.org/10.1177/10507256251363120.
5. Y. Ito, A. Miyauchi, M. Kihara, M. Fukushima, T. Higashiyama, A. Miya, Overall survival of papillary thyroid carcinoma patients: a single-institution long-term follow-up of 5897 patients, World J Surg 42 (2018) 615–622, https://doi.org/10.1007/s00268-018-4479-z.
6. R. Li, L. Yang, M. Xu, B. Wu, Q. Liu, Q. An, et al., Current evidence and strategies for preventing tumor recurrence following thermal ablation of papillary thyroid carcinoma, Cancer Imaging 25 (2025) 88, https://doi.org/10.1186/s40644-025-00908-7.
7. D. Ou, C. Chen, T. Jiang, D. Xu, Research review of thermal ablation in the treatment of papillary thyroid carcinoma, Front Oncol 12 (2022) 859396, https://doi.org/10.3389/fonc.2022.859396.
8. S.A. Wilson, L.A. Stem, R.D. Bruehlman, Hypothyroidism: diagnosis and treatment, Am Fam Physician 103 (2021) 605–613, https://pubmed.ncbi.nlm.nih.gov/33983002/.
9. L. Chaker, C. Baumgartner, W.P.J. den Elzen, M.A. Ikram, M.R. Blum, T.H. Collet, et al., Subclinical hypothyroidism and the risk of stroke events and fatal stroke: an individual participant data analysis, J Clin Endocrinol Metab 100 (2015) 2181–2191, https://doi.org/10.1210/jc.2015-1438.
10. S. Li, M. Yu, Z. Zhao, Y. Wei, L. Peng, Y. Li, Changes in thyroid function after thermal ablation of thyroid nodules, Front Endocrinol (Lausanne) 16 (2025) 1557725, https://doi.org/10.3389/fendo.2025.1557725.
11. Z. Zhao, S. Wang, J. Kuo, B. Çekiç, L. Liang, H.A. Ghazi, et al., 2024 international expert consensus on US-guided thermal ablation for T1N0M0 papillary thyroid cancer, Radiology 315 (2025) e240347, https://doi.org/10.1148/radiol.240347.
12. X. Zhu, G. Zhou, Y. Zhou, C. Chen, L. Sui, D. Ou, et al., Early efficacy of radiofrequency ablation for multifocal T1N0M0 papillary thyroid carcinoma: a multicenter study, Int J Hyperthermia 42 (2025) 2482716, https://doi.org/10.1080/02656736.2025.2482716.
13. M.L. LeFevre; U.S. Preventive Services Task Force, Screening for thyroid dysfunction: U.S. Preventive Services Task Force recommendation statement, Ann Intern Med 162 (2015) 641–650, https://doi.org/10.7326/M15-0483.
14. Z.L. Zhao, Y. Wei, C.H. Liu, L.L. Peng, Y. Li, N.C. Lu, et al., Changes in thyroid antibodies after microwave ablation of thyroid nodules, Int J Endocrinol 2022 (2022) 1–6, https://doi.org/10.1155/2022/7916327.
15. Y. Fei, Y. Qiu, D. Huang, Z. Xing, Z. Li, A. Su, et al., Effects of energy-based ablation on thyroid function in treating benign thyroid nodules: a systematic review and meta-analysis, Int J Hyperthermia 37 (2020) 1090–1102, https://doi.org/10.1080/02656736.2020.1806362.
16. J.E. Zhu, C.J. Sheng, H.L. Zhang, J.X. Li, X.W. Bo, J.J. Yin, et al., Ultrasound-guided percutaneous microwave ablation of primary hyperthyroidism: safety and efficacy analysis, Int J Hyperthermia 41 (2024) 2424903, https://doi.org/10.1080/02656736.2024.2424903.

17. I.A. Chang, Considerations for thermal injury analysis for RF ablation devices, Open Biomed Eng J 4 (2010) 3–12, https://doi.org/10.2174/1874120701004010003.
18. L. Qiu, Y. Huang, Y. Ge, X. Zhao, C. Su, Y. Yang, et al., Thyroid dysfunction following thermal ablation of large solid and solid-predominant thyroid nodules, Endocr Pract 31 (2025) 599–606, https://doi.org/10.1016/j.eprac.2025.01.004.
19. X. Luo, E. Kandil, Microwave ablation: a technical and clinical comparison to other thermal ablation modalities to treat benign and malignant thyroid nodules, Gland Surg 13 (2024) 1805–1813, https://doi.org/10.21037/gs-24-221.
20. M.S. Lui, K.N. Patel, Current guidelines for the application of radiofrequency ablation for thyroid nodules: a narrative review, Gland Surg 13 (2024) 59–69, https://doi.org/10.21037/gs-23-18.
21. G. Rossi, M.C. Petrone, G. Capurso, L. Albarello, S.G.G. Testoni, L. Archibugi, et al., Standardization of a radiofrequency ablation tool in an ex-vivo porcine liver model, Gastrointest Disord 2 (2020) 300–309, https://doi.org/10.3390/gidisord2030027.
22. R.G. Yoon, J.H. Baek, S.R. Chung, Y.J. Choi, J.H. Lee, Ex vivo comparison between thyroid-dedicated bipolar and monopolar radiofrequency electrodes, Int J Hyperthermia 34 (2018) 624–630, https://doi.org/10.1080/02656736.2018.1437283.
23. X. Liang, B. Jiang, Y. Ji, Y. Xu, Y. Lv, S. Qin, et al., Complications of ultrasound-guided thermal ablation of thyroid nodules and associated risk factors: experience from 9667 cases, Eur Radiol 35 (2024) 2307–2319, https://doi.org/10.1007/s00330-024-11023-9.
24. T. Huang, S. Wang, H. Tseng, G.W. Randolph, G. Dionigi, Y. Lin, et al., Thyroid radiofrequency ablation – thermal effects on recurrent laryngeal nerve using continuous intraoperative neuromonitoring animal model, Otolaryngol Head Neck Surg 172 (2025) 63–73, https://doi.org/10.1002/ohn.1017.
25. C.F. Sinclair, J.H. Baek, K.E. Hands, S.P. Hodak, T.C. Huber, I. Hussain, et al., General principles for the safe performance, training, and adoption of ablation techniques for benign thyroid nodules: an American Thyroid Association statement, Thyroid 33 (2023) 1150–1170, https://doi.org/10.1089/thy.2023.0281.

Figures

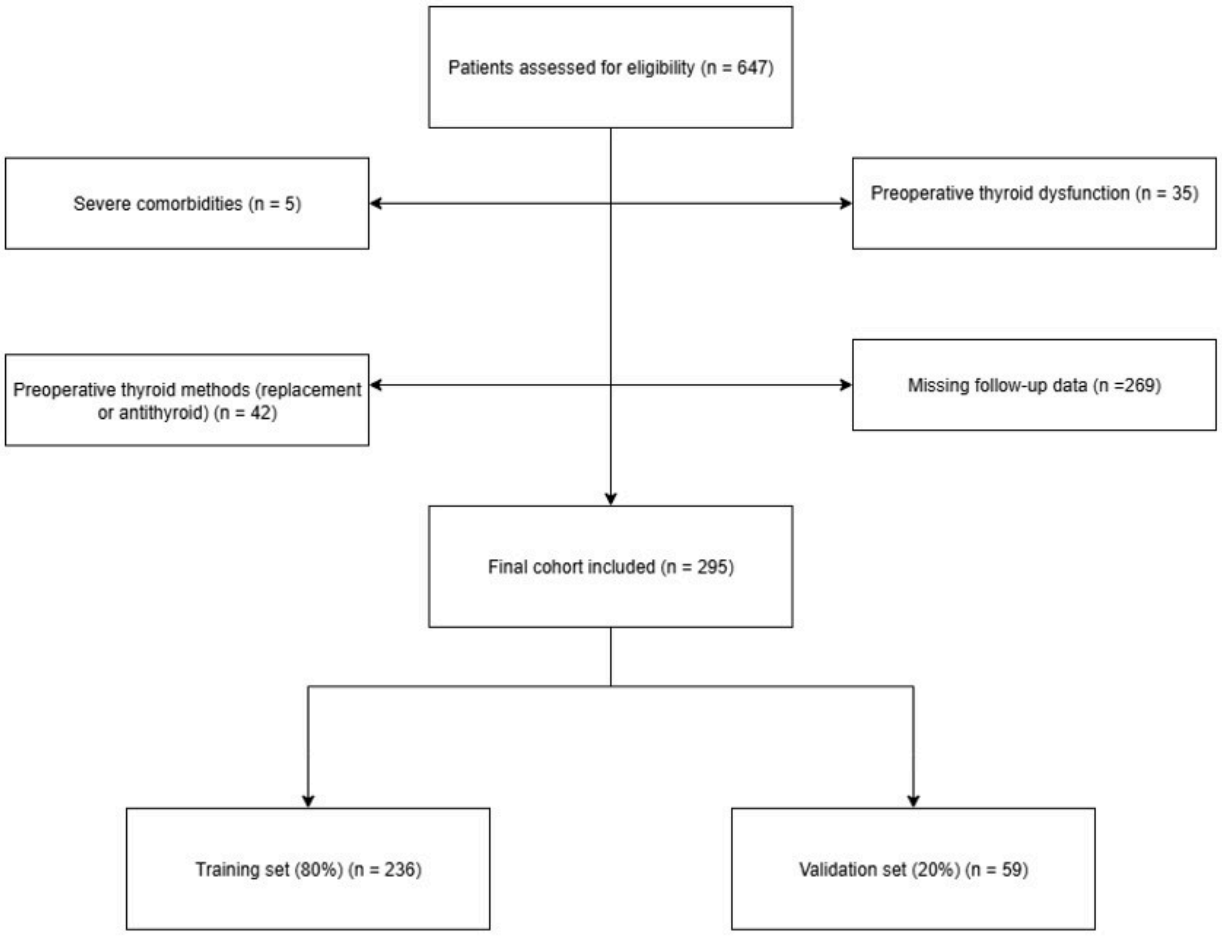


Figure 1

Flowchart of Patient Enrollment, Exclusion, and Dataset Allocation.

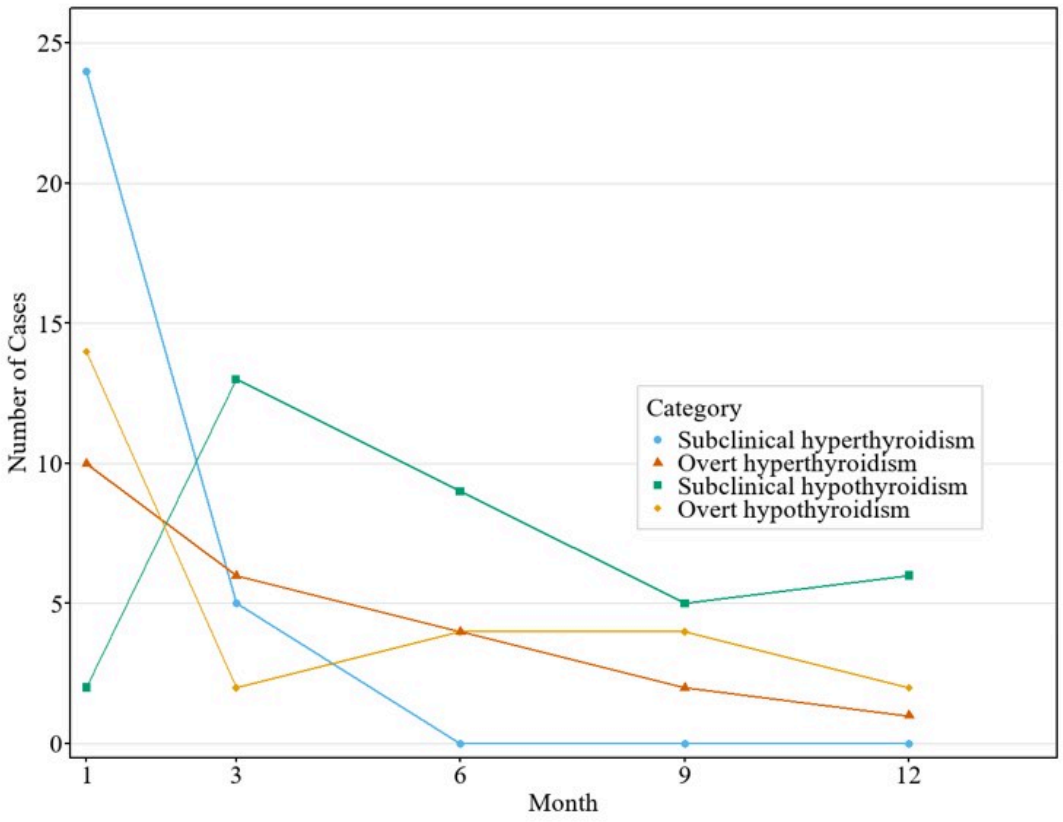


Figure 2

Line chart showing the temporal trends of TD at different postoperative time points.

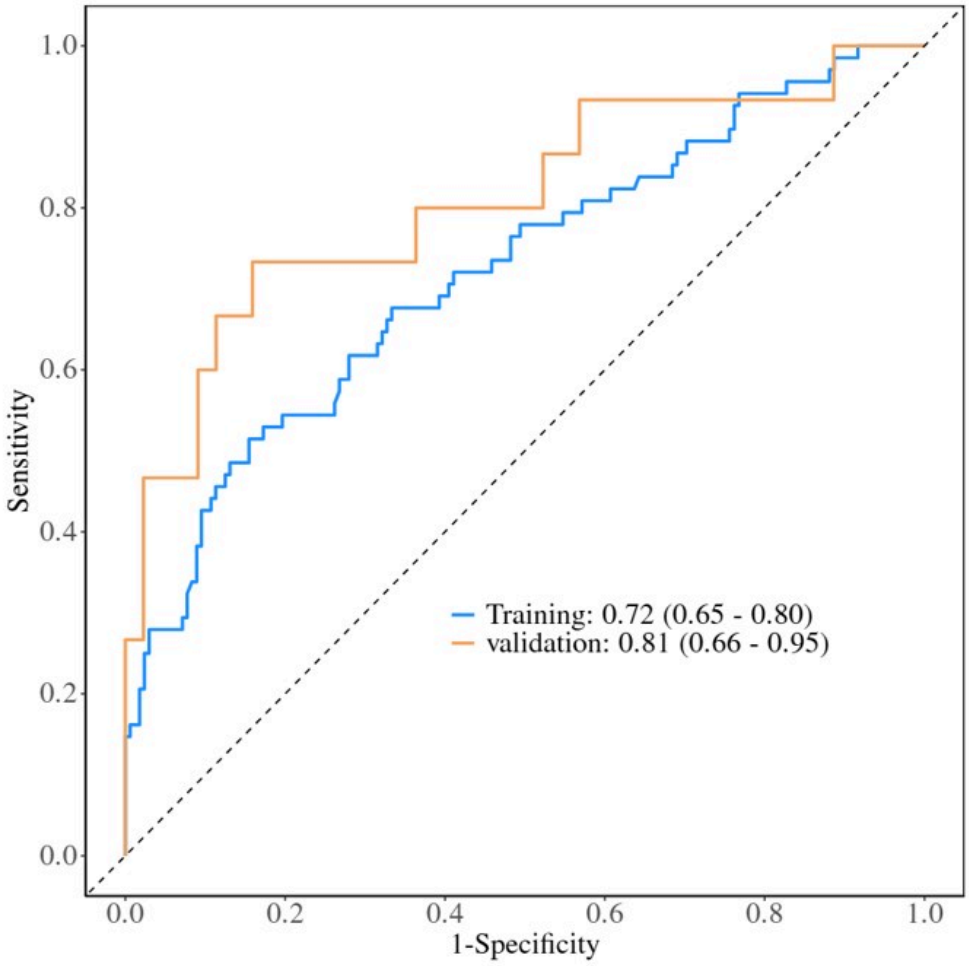


Figure 3

ROC curves for the prediction model in the training set and validation set. The x-axis is 1 – specificity (false positive rate) and the y-axis is sensitivity (true positive rate). The area under the curve (AUC) quantifies the model's discrimination; higher AUC indicates better discriminative performance. The AUC is 0.72 (95% CI: 0.65–0.80) in the training set and 0.81 (95% CI: 0.66–0.95) in the validation set, indicating the model has good predictive ability in both sets.

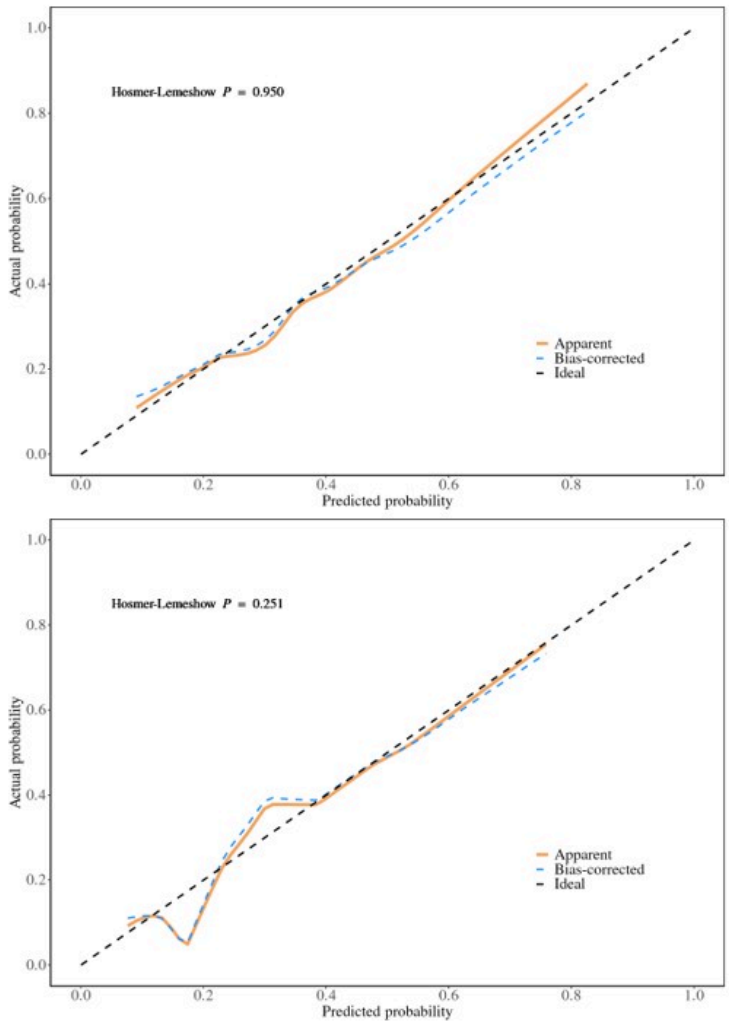


Figure 4

Calibration curves of the model in the training set (upper panel) and validation set (lower panel). The x-axis is the predicted probability of TD, and the y-axis is the observed probability. The diagonal line ("Ideal") represents perfect calibration. The solid line represents the actual calibration curve, and the dashed line is the bias-corrected curve. The Hosmer–Lemeshow test P values are 0.950 in the training set and 0.251 in the validation set, both > 0.05, indicating that the predicted probabilities align well with observed outcomes (good calibration).

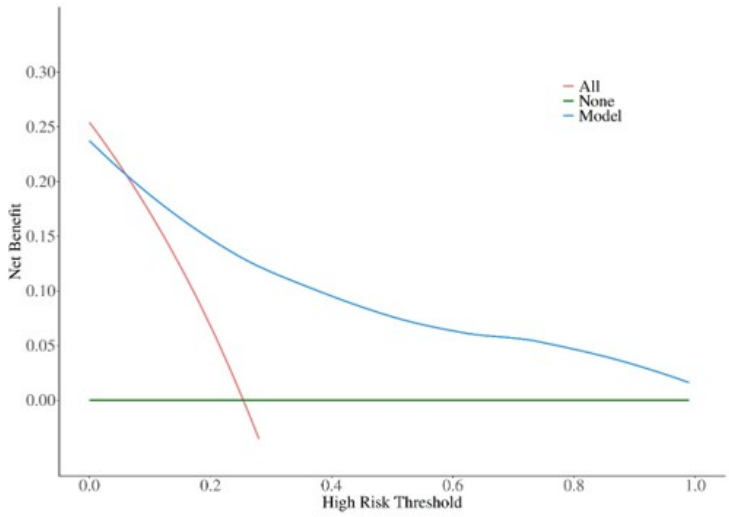


Figure 5

Decision curve analysis for the validation set (lower curve). The x-axis represents the threshold probability (the risk probability at which one would opt to intervene), and the y-axis represents net benefit. The blue curve is the net benefit of the prediction model; the gray diagonal line assumes all patients are treated (treat-all strategy); the black horizontal line assumes no patients are treated (treat-none strategy). Across most threshold probabilities, the model's net benefit is higher than that of the treat-all or treat-none strategies, indicating that the model has good clinical decision value.

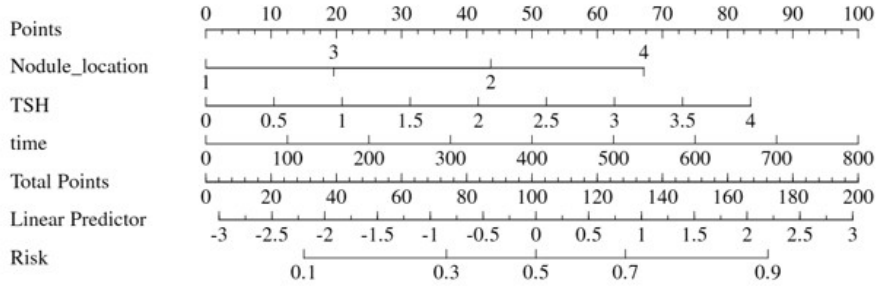


Figure 6

Nomogram for predicting the probability of post-ablation TD. **Usage**: For each predictor, draw a vertical line upward to the "Points" axis to determine the score for that value. Sum the scores for all predictors to get the "Total Points." Then draw a vertical line downward from the "Total Points" to the "Risk" axis at the bottom to read the predicted risk percentage of TD.

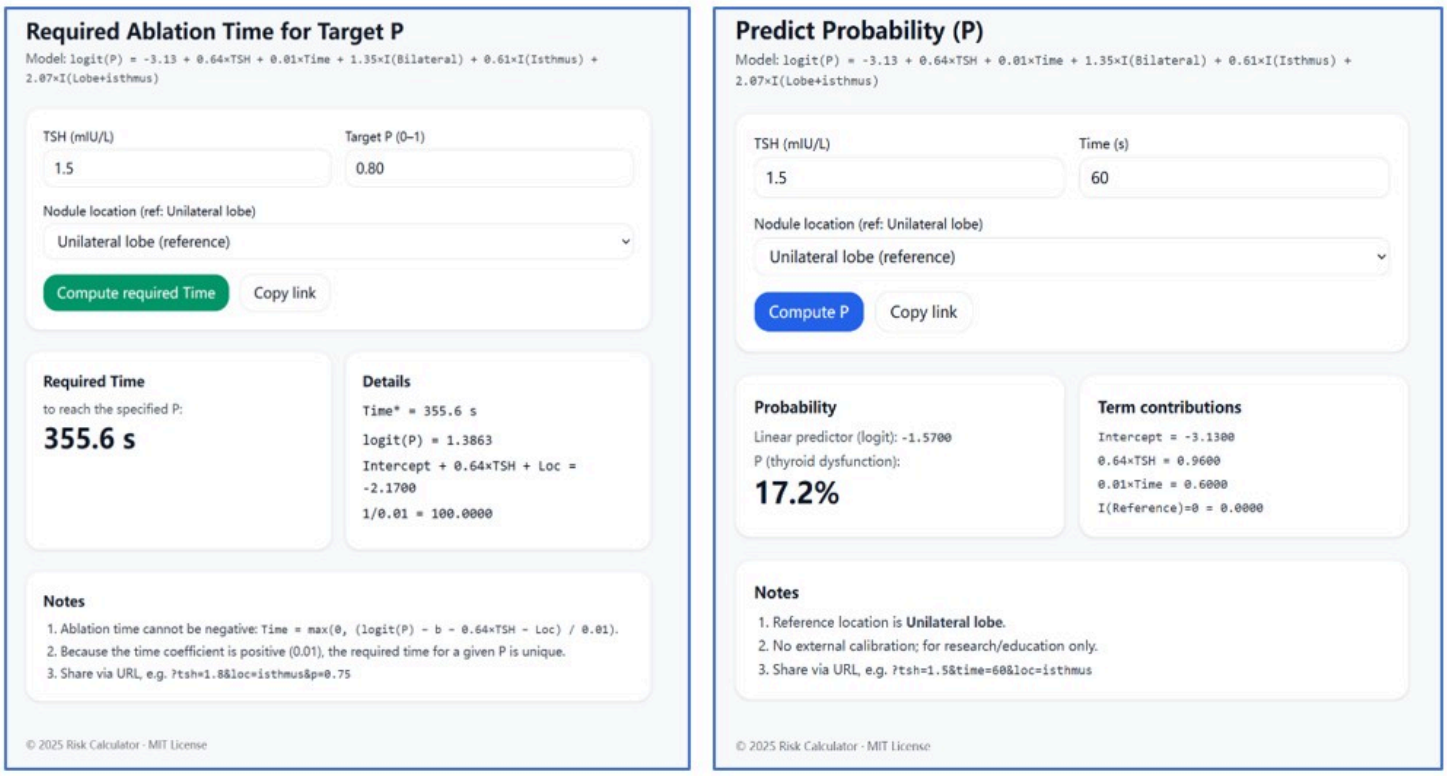


Figure 7

A view of the web-based tools for risk prediction and for determining acceptable ablation time given a target risk threshold.