Am J Transl Res 2026;18(11):9708-9723
www.ajtr.org /ISSN:1943-8141/AJTR0175346

Original Article

Retrospective cross-sectional survey of patient-reported outcomes after trans-breast approach endoscopic thyroidectomy: a multidimensional analysis of cosmetic satisfaction, voice function, and scar assessment

Lixue Xu, Caozhen Chen, Lingdiao Zeng, Dandan Cui, Haoyong Liu

Thyroid and Breast Surgery, The Second Affiliated Hospital of Guangdong Medical University, Zhanjiang 524000, Guangdong, China

Received July 13, 2026; Accepted August 28, 2026; Epub November 15, 2026; Published November 30, 2026

Abstract: Objective: To evaluate patient-reported outcomes (PROs) across postoperative time windows after trans-breast approach endoscopic thyroidectomy (TBAET), identify factors independently associated with cosmetic satisfaction, and explore the relationships among PRO dimensions and the interaction between surgical extent and time window. Methods: This single-center retrospective cross-sectional study included 298 patients with benign thyroid disease who underwent TBAET (2021-2023) and completed postoperative PRO assessment. Patients were classified into 6-month (3-9 months), 12-month (10-18 months), and long-term (≥19 months) time-window groups (n=102, 106, 90). The instruments used were BREAST-Q (≥80 = satisfied), VHI-10 (≥11 = voice handicap), and POSAS (PSAS, OSAS). Groups were compared using Kruskal-Wallis, Jonckheere-Terpstra, and Dunn-Bonferroni tests; multivariable logistic regression with AUC identified factors associated with cosmetic satisfaction. Results: The cohort was predominantly female (92.3%); 60.1% were aged <40 years and 55.7% had BMI <24 kg/m^2. Median BREAST-Q was 67, 75, and 82 across the groups (H=44.51, P<0.001); ≥80 rates were 29.4%, 44.3%, and 65.6% (overall 45.6%). VHI-10 ≥11 occurred in 13.8%, without between-group difference. PSAS and OSAS scores decreased across the groups (both P<0.001) and correlated strongly (r=0.70). The long-term window (OR=4.62), BMI <24 (OR=2.86), age <40 years (OR=2.08), and absence of complications (OR=2.20) were independently associated with BREAST-Q ≥80 (AUC=0.759); sensitivity analyses were consistent, and the extent-time window interaction was not significant. Conclusion: PRO scores after TBAET differed across time windows, with higher cosmetic satisfaction and better scar scores in the long-term window. BMI <24, age <40 years, absence of complications, and the long-term window independently predicted cosmetic satisfaction. Patients rated scars more critically than observers, supporting the value of PRO assessment.

Keywords: Trans-breast approach endoscopic thyroidectomy, patient-reported outcomes, cosmetic satisfaction, voice handicap index-10, patient and observer scar assessment scale, retrospective cross-sectional study

Introduction

Benign thyroid nodules, nodular goiter, and follicular adenoma are common indications for thyroid surgery. A systematic review and meta-analysis by Mu et al. [1] showed that thyroid nodules are prevalent in the general population and are more common in women. The 2023 European Thyroid Association guideline also emphasizes that diagnosis and treatment of thyroid nodules should integrate ultrasound risk stratification, clinical symptoms, and patient preferences to facilitate individualized decision-making [2]. Although conventional open thyroidectomy is effective and safe, the permanent anterior neck scar may affect body image, social confidence, and psychological well-being [3]. In recent years, with the development of minimally invasive surgical techniques and increasing patient expectations regarding postoperative cosmesis, various remote-access endoscopic thyroid procedures have emerged,

https://doi.org/10.62347/QAVW3536

including the trans-breast, transaxillary, transoral vestibular, and bilateral axillo-breast approaches [4]. The latest consensus on remote-access thyroid surgery states that the core objective of these techniques is to avoid or reduce visible cervical scarring while ensuring surgical safety and complete lesion management, and that their implementation should be based on strict patient selection, standardized training, and adequate preoperative communication [5]. Reviews by Zhou et al. [6] and Nguyen et al. [7] both suggest that remote-access or minimally invasive thyroid surgery provides clear cosmetic benefits overall, although heterogeneity persists across approaches, study populations, and outcome measures. The trans-breast approach endoscopic thyroidectomy (TBAET) has been widely used in Asia because it provides a relatively spacious working field, uses a familiar operative route, and can address bilateral thyroid lesions.

In recent years, the evaluation paradigm in surgical clinical research has expanded from physician-centered objective outcomes to patient-reported outcomes (PROs). The literature indicates that patient-reported outcome measures (PROMs) can complement traditional perioperative indicators by capturing symptoms, function, cosmetic satisfaction, and quality of life from the patient's perspective [8]. In thyroid surgery, patients are concerned not only with surgical safety but also with subjective experiences such as cervical or anterior chest scarring, voice function, swallowing, and neck sensation. A systematic review by Xiong et al. [9] suggested that many thyroid-related PROMs are available; however, the selection of scales, reporting methods, and cross-study comparability still require improvement. Previous studies have mainly focused on perioperative indicators such as operative time, intraoperative blood loss, and complication rates; PROs have often been treated as secondary or exploratory endpoints and are frequently limited to a single dimension or a single time point. Among the few small-sample studies specifically evaluating patient-reported outcomes after TBAET, a randomized trial demonstrated that transareolar single-site endoscopic thyroidectomy achieved significantly higher patient-rated cosmetic scores than minimally invasive video-assisted thyroidectomy [10], and a comparative study of 131 patients showed that voice and swallowing impairment after the transareolar approach were comparable to those after the transoral vestibular approach, with good cosmetic satisfaction for both [11]. However, these studies were limited by small sample sizes and short or single-point follow-up, and none have simultaneously examined the relationships among cosmetic satisfaction, voice function, and scar appearance. A review by Beka and Gimm et al. [12] noted that subjective voice changes may occur after thyroidectomy even without evident laryngeal nerve injury. A recent review of thyroid scar assessment also showed that instruments combining subjective and objective components, such as POSAS, have been used in post-operative scar research, although evaluation methods remain nonuniform [13]. Therefore, simultaneous assessment of cosmetic satisfaction, voice function, and scar appearance within the same cohort can provide a more comprehensive description of patients' experiences after TBAET. Of note, the BREAST-Q originates from breast plastic surgery [14] and was therefore used off-label in this study, and no reference thresholds have been established for patients undergoing thyroidectomy via areolar incisions, an inherent limitation warranting cautious interpretation of the cosmetic-satisfaction scores reported here.

Evidence regarding factors influencing PROs after TBAET remains limited. Existing studies have generally focused on satisfaction at a single time point and have insufficiently explored the internal relationships among PRO dimensions-for example, whether patients with higher cosmetic satisfaction also report better voice function and scar assessment, or whether trade-offs exist across dimensions. In addition, quantitative comparisons of PRO differences by surgical extent and postoperative time window remain inadequate, limiting evidence-based management of patients' cosmetic expectations during preoperative counseling. Accordingly, this study used retrospective PRO data from a relatively large single-center cohort to systematically evaluate cosmetic satisfaction, voice function, and scar appearance among TBAET patients across different postoperative follow-up time windows; analyze independent factors associated with cosmetic satisfaction; explore internal associations among PRO dimensions; and test the interaction between surgical extent and postoperative time window.

Methods

Sample size calculation

The primary binary outcome of this study was the proportion of patients achieving BREAST-Q ≥80. Because no direct prior data were available for the achievement rate of BREAST-Q ≥80 after TBAET, sample-size estimation was based on a reported treatment satisfaction rate of approximately 65% in a study of endoscopic thyroid lobectomy via the axillary-breast-shoulder approach [15]. With a two-sided α=0.05 and an allowable error d=0.06, the single-sample proportion method indicated that at least 243 patients were required. The present study included 298 patients, satisfying the sample-size requirement. We acknowledge that the satisfaction rate used for sample-size estimation was derived from a remote-access endoscopic approach other than TBAET, because no direct data on the achievement rate of BREAST-Q ≥80 after TBAET were available at the time of study design. To address the uncertainty introduced by this indirect reference, we recomputed the required sample size using the actually observed achievement rate in the present cohort (45.6%, 136/298). This yielded a requirement of approximately 265 patients, which was still exceeded by the 298 enrolled patients. Eight candidate variables were planned for multivariable logistic regression; with 136 events for BREAST-Q ≥80, the events-per-variable ratio was approximately 17, meeting the basic sample-size requirement for model analysis.

Study design and ethics

This was a single-center retrospective cross-sectional study. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Ethics Committee of The Second Affiliated Hospital of Guangdong Medical University (Approval No. PJKT2026-057). The requirement for informed consent was waived.

Study design and follow-up time-window definitions

This study was not a prospective longitudinal study; rather, it retrospectively extracted PRO assessments that were completed during different postoperative time windows. The study population consisted of patients with benign thyroid disease who underwent TBAET at our hospital from 2021 to 2023, with data collated through March 2025. According to the interval from surgery to PRO assessment, patients were divided into three follow-up time windows: the 6-month time-window group (3-9 months after surgery), the 12-month time-window group (10-18 months after surgery), and the long-term time-window group (≥19 months after surgery). The boundaries of the three time windows were not pre-specified before data analysis; they were determined after data collection, based on the observed distribution of the intervals between surgery and PRO assessment. Because PRO assessments in this retrospective cohort were completed at routine postoperative follow-up visits rather than at prospectively scheduled time points, the intervals clustered around approximately 6 and 12 months after surgery. Accordingly, patients were classified into the 6-month time-window group (3-9 months after surgery), the 12-month time-window group (10-18 months), and the long-term time-window group (≥19 months), with boundaries set to reflect these conventional follow-up points and to yield non-overlapping, gap-free windows with adequate group sizes (n= 102, 106, and 90, respectively). Given the retrospective cross-sectional design, the groups reflect between-patient differences at corresponding postoperative stages and should not be interpreted as longitudinal changes within the same individuals.

Inclusion and exclusion criteria

Inclusion criteria were as follows: (1) age ≥18 years; (2) postoperative pathology confirming benign thyroid disease, including nodular goiter, thyroid follicular adenoma, and other benign lesions; (3) receipt of TBAET, including unilateral lobectomy or total thyroidectomy; and (4) completion of at least one postoperative PRO assessment including BREAST-Q, VHI-10, and POSAS. Exclusion criteria were as follows: (1) postoperative pathological diagnosis of thyroid malignancy; (2) previous neck surgery or neck radiotherapy; (3) concomitant parathyroid disease requiring simultaneous parathyroid surgery; (4) more than 20% missing items in any PRO scale; and (5) missing key follow-up data in the electronic medical record. A total of 298 patients were included in the final analysis.

Surgical technique

All TBAET procedures were performed by the same surgical team, which had extensive clinical experience in endoscopic thyroid surgery. The basic concept and technical route of breast/anterior chest approach endoscopic thyroidectomy can be traced to early clinical series using the breast approach [16], and subsequent larger-sample studies have also suggested that this approach is feasible and associated with low complication rates for benign and low-risk thyroid disease [17]. All procedures in this study were performed under general anesthesia with endotracheal intubation. Patients were placed in the supine position with the shoulders elevated to achieve slight neck extension, and both upper limbs were abducted and fixed. After routine sterilization and draping, a 10-mm incision was made at the superomedial edge of each areola, and a 5-mm auxiliary incision was made between the medial side of the right areola and the parasternal line. Epinephrine saline (1:500,000) was infiltrated into the subcutaneous tissue layer of the anterior chest wall to reduce intraoperative bleeding. Through the areolar incision, the subcutaneous tissue was bluntly dissected, and a subcutaneous working tunnel was gradually established along the superficial plane of the pectoralis major fascia using an ultrasonic scalpel and dissecting forceps. The tunnel extended superiorly to the level of the thyroid cartilage and laterally to the anterior border of the sternocleidomastoid muscles. After tunnel creation, CO_2 was insufflated to maintain the working space at a pressure of 6-8 mmHg. A 5-mm endoscope (30° oblique lens) was inserted through the auxiliary incision, and the ultrasonic scalpel and grasping forceps were introduced through the bilateral areolar incisions. The linea alba cervicalis was first exposed and incised longitudinally, the strap muscles were retracted laterally, and the thyroid gland was exposed. The superior thyroid vessels, middle thyroid vein, and inferior pole vessels were divided sequentially using the ultrasonic scalpel, with attention to preserving the recurrent laryngeal nerve and parathyroid glands. For lobectomy, the affected thyroid lobe was completely removed after transection of the isthmus; for total thyroidectomy, the contralateral side was treated in the same manner. The specimen was placed in a retrieval bag and removed through the areolar incision. At the end of surgery, the operative field was irrigated with saline, hemostasis was confirmed, and a negative-pressure drain was placed through the areolar incision. Incisions were closed with intradermal sutures. Postoperatively, prophylactic antibiotics were routinely administered, and the drain was removed when output was less than 20 mL/24 h, usually 48-72 hours after surgery.

Data collection

Two trained researchers independently extracted the following data retrospectively from the hospital electronic medical record system; disagreements were resolved by discussion or adjudication by a third investigator: (1) demographic characteristics: age and sex; (2) anthropometric indicators: height and weight, with body mass index (BMI) calculated; (3) perioperative data: operation date, PRO assessment date, surgical extent (lobectomy or total thyroidectomy), operative time (from skin incision to completion of skin closure), and wet weight of the resected specimen (measured by electronic scale in the pathology department); (4) complications within 30 days after surgery: transient hypoparathyroidism (postoperative serum calcium <2.0 mmol/L accompanied by clinical manifestations of hypocalcemia such as perioral numbness or limb spasm), transient recurrent laryngeal nerve (RLN) injury (postoperative laryngoscopy confirming restricted vocal cord mobility, with recovery within 6 months after surgery), postoperative hematoma (requiring surgical evacuation or prolonged hospitalization), wound infection (incisional redness, swelling, heat, and pain with purulent discharge requiring antibiotic treatment), and subcutaneous fluid collection or seroma (requiring puncture aspiration or prolonged drainage); and (5) postoperative pathological diagnosis, as classified in the pathology report (nodular goiter, thyroid follicular adenoma, or other benign lesions).

PRO assessment instruments

BREAST-Q cosmetic satisfaction module: BREAST-Q is a patient-reported outcome instrument for breast surgery developed and validated by Pusic et al. [14]. Its standardized Q-Score ranges from 0 to 100, with higher scores indicating greater satisfaction. It should be noted that BREAST-Q was not originally developed for thyroid surgery. In this study, its cosmetic satisfaction module was therefore used off-label to quantify satisfaction related to the anterior

chest/areolar incisions after thyroid surgery-an application for which the scale was not originally designed-and a score ≥80 was predefined as the threshold for cosmetic satisfaction. Voice Handicap Index-10 (VHI-10): This 10-item scale ranges from 0 to 40, with higher scores indicating more severe voice handicap. The scale was developed and validated by Rosen et al. [18], and a score of ≥11 on the VHI-10 was defined as subjective voice handicap, based on normative data [19]. Patient and Observer Scar Assessment Scale (POSAS): This instrument includes the Patient Scar Assessment Scale (PSAS) and the Observer Scar Assessment Scale (OSAS), each with six items scored from 1 to 10 (total score 6-60), and allows simultaneous patient and observer assessment of scars [20]. It has been validated for assessment of linear surgical scars. Observer assessment was completed by two independent, trained evaluators who were blinded to the PRO results; the average score was used as the OSAS score. Administration of the PRO instruments. All instruments were administered using their validated Chinese versions. PRO assessments were collected through two channels: (1) paper-based self-administration in the outpatient clinic during routine follow-up visits, with the questionnaires checked on the spot for completeness; and (2) electronic questionnaires completed through a WeChat-based platform, in which the system prevented submission unless all items had been answered. When both the outpatient paper version and the electronic version were available for the same patient, the paper version was used. The BREAST-Q cosmetic satisfaction module, the VHI-10, and the PSAS were self-administered by the patients, whereas the OSAS was completed independently by a trained observer. Each patient was assessed once, at the follow-up visit corresponding to his or her postoperative time window. Questionnaires with >20% missing items in any scale were excluded, in accordance with the exclusion criteria.

Outcome definitions

The primary outcomes were the BREAST-Q cosmetic satisfaction score (continuous variable, 0-100) and the predefined achievement rate of BREAST-Q ≥80 (binary variable). Secondary outcomes included the VHI-10 score and the incidence of VHI-10 ≥11; PSAS score; and OSAS score.

Statistical analysis

Non-normally distributed data are expressed as median (IQR). Comparisons among the three groups were performed using the Kruskal-Wallis test and Dunn-Bonferroni pairwise comparisons, and trend testing was conducted using the Jonckheere-Terpstra test. Categorical variables were expressed as n (%), and between-group comparisons were performed using the chi-square test or Fisher's exact test. With BREAST-Q ≥80 as the dependent variable, univariate logistic regression was first performed. Subsequently, eight variables - postoperative time window (6-month window as the reference), age <40 years, BMI <24 kg/m^2, absence of complications, sex, surgical extent, operative time, and specimen weight - were entered into the multivariable model using the Enter method. Variance inflation factor (VIF) was used to assess collinearity, the Hosmer-Lemeshow test was used to evaluate goodness of fit, and area under the curve (AUC) was used to assess discrimination. Sensitivity analyses were performed using thresholds of ≥70 and ≥90. Subgroup analyses used combined stratification by BMI × complication status × age. These subgroup analyses were post hoc exploratory analyses and were not pre-specified endpoints. The interaction between surgical extent and time window was evaluated using a likelihood ratio test in an ordinal logistic regression model. Because both lobectomy and total thyroidectomy share the same areolar and anterior chest incisions and the same subcutaneous working tunnel, we hypothesized that the temporal pattern of cosmetic satisfaction would be comparable between the two surgical extents; however, total thyroidectomy involves more extensive bilateral tissue dissection and a potentially different postoperative recovery course. This hypothesis was therefore formally tested by comparing ordinal logistic regression models with and without the time window × surgical extent interaction term using a likelihood ratio test. The proportional odds assumption underlying the ordinal logistic regression was verified using a likelihood ratio test comparing the proportional odds model with a model allowing non-proportional (category-specific) effects for all predictors; the assumption was not violated (P=0.634). Correlations among PRO dimensions were assessed using Spearman rank correlation. Analyses were performed

Table 1. Baseline characteristics by follow-up time window

Variable	Overall (n=298)	6 mo (n=102)	12 mo (n=106)	Long (n=90)	χ^2/H	P
Age <40, n (%)	179 (60.1)	58 (56.9)	66 (62.3)	55 (61.1)	χ^2=0.69	0.708
Female, n (%)	275 (92.3)	93 (91.2)	99 (93.4)	83 (92.2)	χ^2=0.36	0.835
BMI <24, n (%)	166 (55.7)	54 (52.9)	60 (56.6)	52 (57.8)	χ^2=0.51	0.776
Lobectomy, n (%)	186 (62.4)	65 (63.7)	66 (62.3)	55 (61.1)	-	-
Total thyroidectomy, n (%)	112 (37.6)	37 (36.3)	40 (37.7)	35 (38.9)	χ^2=0.14	0.932
Operative time, min	123.0±32.1	128.6±28.7	125.3±31.2	113.9±35.0	H=11.39	0.003
Specimen weight, g	58.0±24.6	57.4±24.1	57.0±23.9	59.7±26.0	H=0.38	0.828
Any complication, n (%)	49 (16.4)	19 (18.6)	16 (15.1)	14 (15.6)	χ^2=0.55	0.761

Note: TBAET, trans-breast approach endoscopic thyroidectomy; BMI, body mass index; SD, standard deviation. Continuous: mean ± SD; categorical: n (%). P from Kruskal-Wallis or chi-square/Fisher exact test; Between-group comparisons used the Kruskal-Wallis test (H) or the chi-square test (χ^2).

in R 4.3.2, and two-sided P<0.05 was considered statistically significant.

Results

Baseline characteristics

Between 2021 and 2023, 372 patients with benign thyroid disease who underwent TBAET at our hospital were assessed for eligibility. Of these, 74 were excluded (22 with malignant pathology on final pathology, 15 with previous neck surgery or neck radiotherapy, 8 with concomitant parathyroid disease requiring simultaneous parathyroid surgery, 18 with >20% missing items in any PRO scale, 11 with missing key follow-up data in the electronic medical record, and none were excluded for lacking a postoperative PRO assessment, leaving 298 patients for the final analysis (Figure S1); the loss to follow-up rate was 0%. A total of 298 patients were included, with 102, 106, and 90 patients in the three time-window groups, respectively.

In the overall cohort, 179 patients (60.1%) were younger than 40 years, 275 (92.3%) were female, and 166 (55.7%) had BMI <24 kg/m^2. Lobectomy was performed in 186 patients (62.4%), and total thyroidectomy in 112 (37.6%). Mean operative time was 123.0±32.1 min, and mean specimen weight was 58.0±24.6 g. The overall complication rate was 16.4% (49/298). Pathological diagnoses included nodular goiter in 194 patients (65.1%), adenoma in 66 (22.1%), and other benign lesions in 38 (12.8%). Baseline characteristics were generally comparable across the three groups, with the exception of operative time, which differed significantly (Kruskal-Wallis P=0.003; **Table 1**); operative time was therefore included as a covariate in the multivariable analyses.

BREAST-Q cosmetic satisfaction

Median BREAST-Q scores were 67 (interquartile range [IQR]: 56-80) in the 6-month window, 75 (IQR: 65-84) in the 12-month window, and 82 (IQR: 74-90) in the long-term window (H= 44.51, P<0.001; Jonckheere-Terpstra trend P<0.001, **Figure 1A**). Dunn-Bonferroni pairwise comparisons showed significant differences for long-term vs. 12-month (P<0.001), long-term vs. 6-month (P<0.001), and 12-month vs. 6-month (P=0.003). The overall achievement rate of BREAST-Q ≥80 was 45.6% (136/298), with rates of 29.4%, 44.3%, and 65.6% in the three groups, respectively (χ^2=25.29, P<0.001, **Figure 1B**).

VHI-10 voice function

Median VHI-10 scores were 6 (IQR: 4-9) in the 6-month window, 6 (IQR: 3-9) in the 12-month window, and 5 (IQR: 2-7) in the long-term window (H=6.13, P=0.047, **Figure 2A**). Because this overall between-group difference was borderline significant, pairwise comparisons were performed using Dunn's test with Bonferroni correction; after correction, only the long-term vs. 6-month comparison remained significant (P=0.021). The comparisons between the 12-month vs. 6-month (P=0.476) and the long-term vs. 12-month (P=0.193) were not significant. Therefore, the overall between-group difference should be interpreted with caution. Patients with VHI-10 ≥11 accounted for 13.8% (41/298), with no significant difference among the three groups (P=0.874, **Figure 2B**).

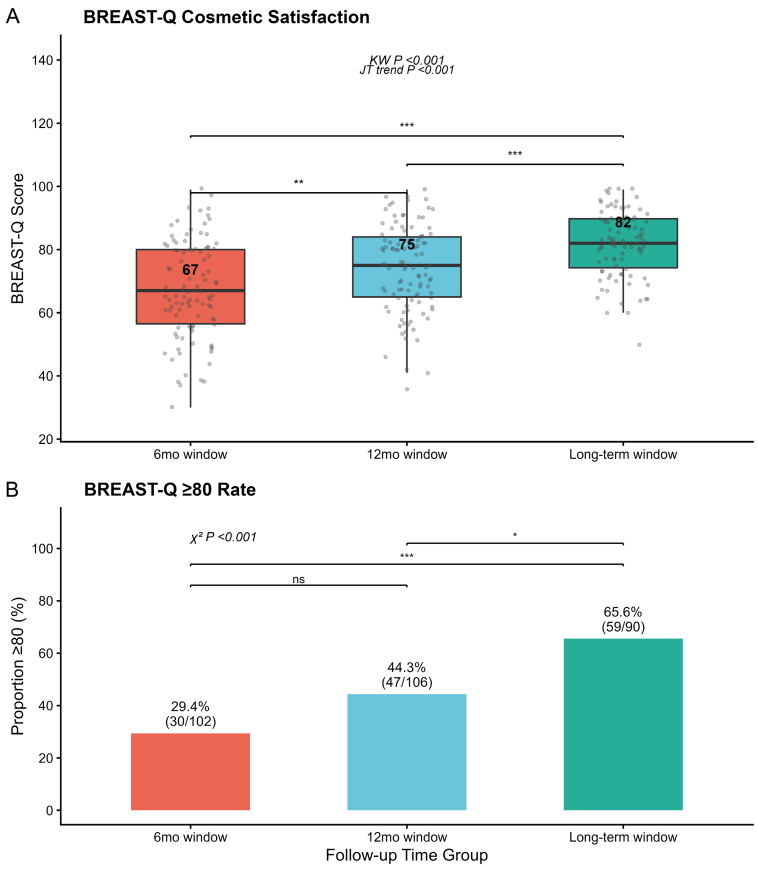


Figure 1. BREAST-Q cosmetic satisfaction scores across three follow-up time windows. A. Box-and-jitter plots of BREAST-Q scores by time window group. The horizontal line within each box denotes the median; box boundaries denote the interquartile range (IQR); whiskers extend to the most extreme values within 1.5 × IQR. Individual patient scores are overlaid as jittered points. The median value is displayed at the median line within each box. Kruskal-Wallis *P* value (KW P) and Jonckheere-Terpstra trend test *P* value (JT P) are annotated at the top of the panel. Dunn-Bonferroni-adjusted pairwise comparisons are indicated by horizontal brackets with asterisks in (A); Bonferroni-adjusted chi-square (or Fisher exact) pairwise comparisons are indicated by brackets in (B). B. Bar chart showing the proportion of patients achieving BREAST-Q ≥80 (cosmetic satisfaction threshold) in each time window group. The rate and corresponding count (events/total) are displayed above each bar. Chi-square test *P* value for overall between-group comparison is shown. Note: BREAST-Q, a patient-reported outcome measure originally developed for breast surgery (Q-Score standardized scoring, range 0-100, higher scores indicate greater satisfaction). The threshold of ≥80 was predefined in this study to indicate cosmetic satisfaction. ***P<0.001, **P<0.01, *P<0.05; ns, not significant (Dunn-Bonferroni corrected in A; Bonferroni-adjusted chi-square or Fisher exact test in B). The overall BREAST-Q ≥80 rate was 45.6% (136/298).

POSAS scar assessment

The median PSAS score in the overall cohort was 11 (IQR: 8-16), and the median OSAS score was 10 (IQR: 7-14), with Spearman r=0.70 (P<0.001, **Figure 3**). The Wilcoxon test showed that PSAS scores were higher than OSAS scores (median difference 1 point, P<0.001), indicating that patients rated scars more critically. PSAS scores decreased from 14 (IQR: 10-20) in the 6-month window to 12 (IQR: 10-15) in the 12-month window and 10 (IQR: 7-13) in the long-term window (H=35.00, P<0.001, **Figure 4A**, **4B**). OSAS scores decreased from 14 (IQR: 10-17) to 11 (IQR: 9-15) and 8 (IQR: 7-11), respectively (H=58.74, P<0.001, **Figure 5A**, **5B**). Dunn-Bonferroni comparisons were all P<0.01, and Jonckheere-Terpstra trend tests were all P<0.001.

Factors associated with cosmetic satisfaction and complication analysis

The results of the univariate and multivariable logistic regression analyses are shown in **Table 2**. Using the 6-month time-window group as the reference, the 12-month time-window group had odds ratio (OR)=1.83 (95% confidence interval [CI]: 1.00-3.41, P=0.053), and the long-term time-window group had OR=4.62 (95% CI: 2.39-9.14, P<0.001). BMI <24 (OR=2.86, P<0.001), age <40 years (OR=2.08, P=0.008), and absence of complications (OR=2.20, P=0.038) were also independently associated with cosmetic satisfaction. Sex, surgical extent, operative time, and specimen weight did not reach statistical significance. VIF values were <2.00, Hosmer-Lemeshow P=0.632, and AUC=0.759 (95% CI: 0.705-0.813). Among the five types of postoperative complications, transient RLN injury was significantly associated with higher VHI-10 score (P<0.001), whereas the other complications were not significantly associated with any PRO score (**Table 3**).

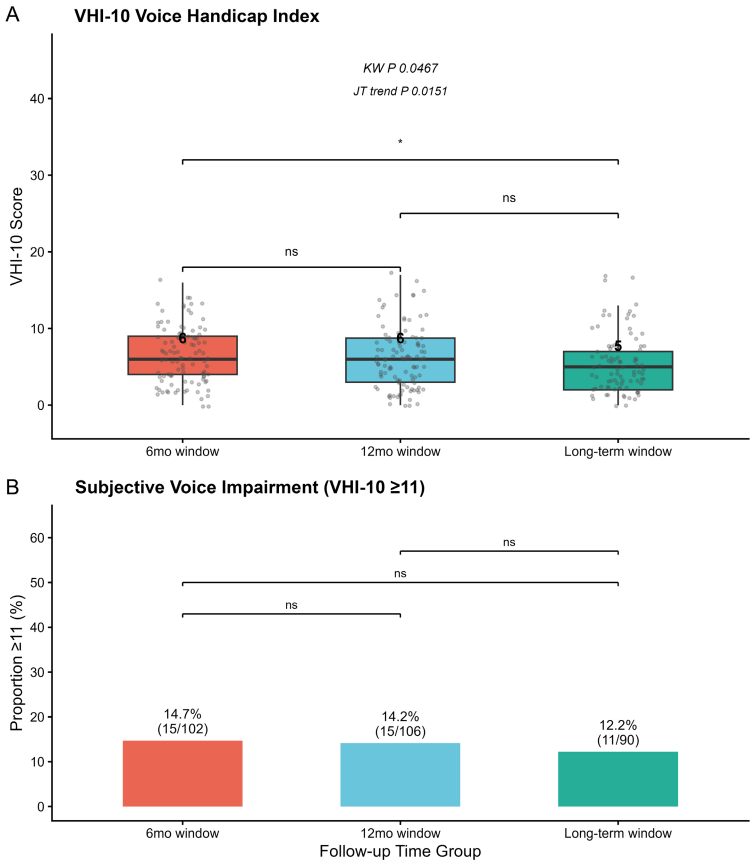


Figure 2. VHI-10 voice handicap scores across three follow-up time windows. A. Box-and-jitter plots of VHI-10 scores by time window group. Box-and-whisker conventions, median labels, and significance annotations follow the format described in **Figure 1**. B. Bar chart showing the proportion of patients with VHI-10 ≥11 (threshold for subjective voice impairment) in each time window group. The rate and corresponding count (events/total) are displayed above each bar; pairwise comparisons (Bonferroni-adjusted chi-square test) are indicated by horizontal brackets. Note: VHI-10, Voice Handicap Index-10 (range 0-40, higher scores indicate greater voice handicap). The threshold of ≥11 was used to define subjective voice impairment according to published normative values. The overall VHI-10 ≥11 rate was 13.8% (41/298). The proportion of patients with VHI-10 ≥11 did not differ significantly across the three time windows (chi-square test, P=0.874), and all Bonferroni-adjusted pairwise comparisons were non-significant. *P<0.05; ns, not significant.

Sensitivity analysis, subgroup stratification, and interaction effect

Sensitivity analyses (**Table 4**) showed AUC values of 0.829, 0.759, and 0.814 at thresholds of ≥70, ≥80, and ≥90, respectively. Long-term time window, BMI <24, and age <40 years remained significant and directionally consistent across all three thresholds. Absence of complications did not reach significance at the ≥70 threshold (OR=1.35, P=0.448) or the ≥90 threshold (OR=2.48, P=0.129), suggesting that its effect was mainly reflected at a moderate satisfaction level. The number of events and the events-per-variable (EPV) ratio for each cutoff are reported in Table S1; the ≥70 and ≥80 models satisfied the EPV≥10 criterion, whereas the ≥90 model (EPV=5.6) was underpowered and is interpreted as exploratory, although a reduced four-variable model (EPV=11.25) confirmed the core findings. Subgroup analysis showed that patients with BMI <24, no complications, and age <40 years had the highest achievement rate (65.3%, 64/98), exceeding the overall cohort rate of 45.6% (**Figure 6**). These subgroup analyses were post hoc exploratory analyses and were not pre-specified endpoints; any apparent gradient across subgroups should therefore be interpreted with caution. In analyses stratified by surgical extent, BREAST-Q median scores did not differ significantly between lobectomy and total thyroidectomy within any time window (6 months: 68 vs. 65, P=0.181; 12 months: 77 vs. 73, P=0.481; long-term: 83 vs. 81, P=0.360, **Figure 7**). An ordinal logistic regression likelihood ratio test showed no significant time window × surgical extent interaction (P=0.952); the proportional odds assumption was satisfied (LRT P=0.634). This finding is consistent with our hypothesis that the temporal pattern of cosmetic satisfaction was independent of surgical extent.

Correlations among PRO dimensions and predictors of VHI-10

BREAST-Q was moderately negatively correlated with PSAS (r=-0.38) and OSAS (r=-0.33), and weakly negatively correlated with VHI-10 (r=-0.19). PSAS and OSAS were strongly positively correlated (r=0.70). VHI-10 was weakly correlated with PSAS/OSAS (|r|<0.20), suggesting that voice and scar appearance repre-

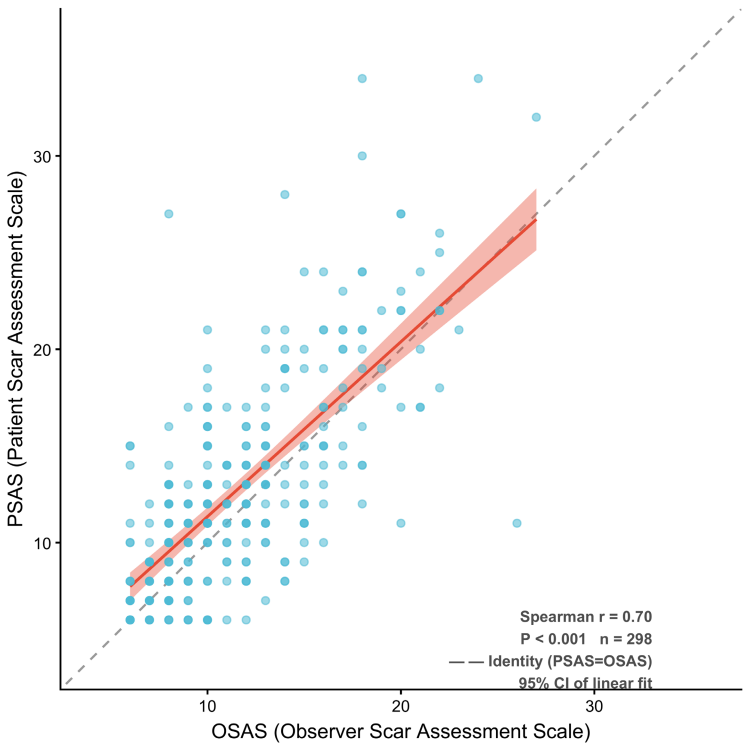


Figure 3. Agreement between patient-reported and observer-reported scar assessment. Scatter plot of PSAS (Patient Scar Assessment Scale) versus OSAS (Observer Scar Assessment Scale) total scores. The dashed diagonal line represents the identity line (PSAS=OSAS), where points above the line indicate patients rating their scar as worse than the observer. The solid red line represents the linear regression fit, with the shaded area denoting the 95% confidence interval. The Spearman rank correlation coefficient (r), *P* value, and sample size are annotated in the lower right corner. Note: PSAS, Patient Scar Assessment Scale (range 6-60); OSAS, Observer Scar Assessment Scale (range 6-60). Lower scores indicate better scar appearance. The Wilcoxon signed-rank test confirmed that PSAS scores were significantly higher than OSAS scores (median difference 1 point, P<0.001), indicating that patients systematically rated their scars more critically than trained observers.

sent relatively independent dimensions (**Figure 8**). In univariate logistic regression for VHI-10 ≥11, only one variable had P<0.1; therefore, no multivariable model was constructed (**Table 5**).

Discussion

Based on retrospective cross-sectional PRO data from 298 patients after TBAET, this study systematically described PRO characteristics across three core dimensions-cosmetic satisfaction, voice function, and scar appearance-across different postoperative follow-up time windows. Unlike many previous studies that focused on a single PRO dimension or a single time point, this study simultaneously incorporated three complementary PRO dimensions within the same cohort and compared PRO differences across three postoperative time windows, thereby providing a more complete profile of patient experience after TBAET. The concept that PROMs can complement traditional complication and perioperative indicators in surgical quality assessment (PMID: 38922601) is consistent with the multidimensional PRO design of this study. The main findings were as follows. First, BREAST-Q scores differed significantly across time windows, with the long-term group having the highest achievement rate (65.6%) and an adjusted OR of 4.62 compared with the 6-month group. Second, BMI <24, age <40 years, absence of complications, and the long-term time window were independently associated with cosmetic satisfaction; the AUC was 0.759, and core factors showed consistent directions across three BREAST-Q thresholds. Third, voice function was generally preserved, and RLN injury was significantly associated with higher VHI-10, but preoperative macroscopic variables had limited ability to predict voice handicap. Fourth, PSAS and OSAS were moderately positively correlated, and patient self-assessment was systematically higher than observer assessment. Fifth, no interaction was observed between surgical extent and time window.

The long-term time-window group in this study had a BREAST-Q ≥80 achievement rate of 65.6%, suggesting that patients assessed at longer postoperative intervals reported greater satisfaction with the cosmesis of their anterior chest/areolar incisions than those assessed at shorter intervals. A network meta-analysis by Nguyen et al. [7] indicated that remote-access and minimally invasive thyroid surgery generally achieve better cosmetic outcomes than conventional cervical approaches. Zhou et al. [6] also noted clear cosmetic benefits of endoscopic thyroid surgery, although het-

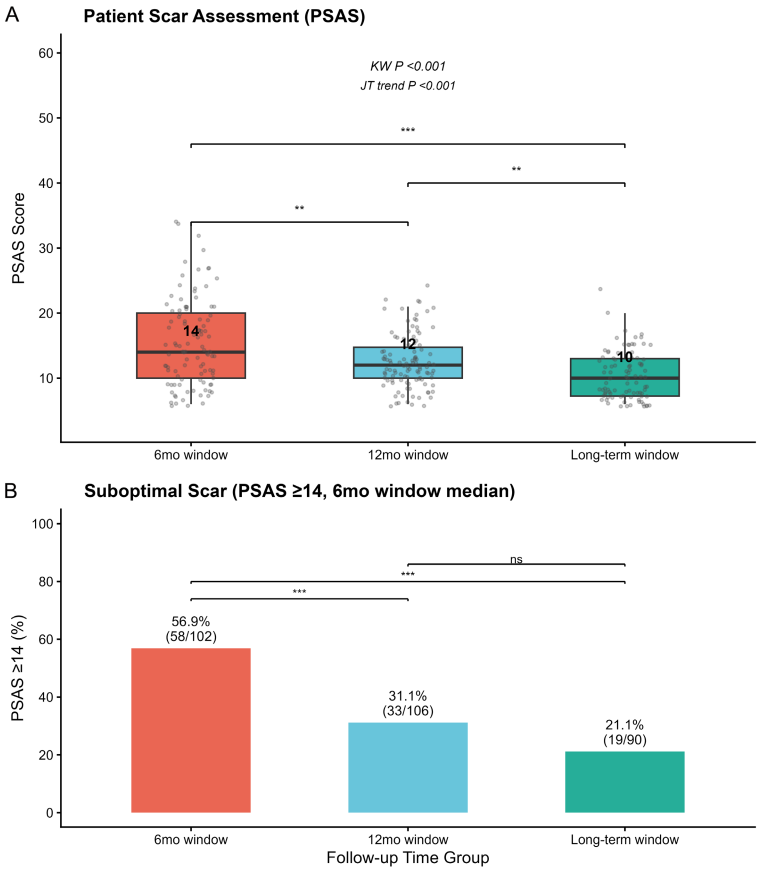


Figure 4. PSAS scores across three follow-up time windows. A. Box-and-jitter plots of PSAS scores by time window group, with conventions and annotations as described in **Figure 1**. B. Bar chart showing the proportion of patients with a suboptimal scar, defined as a PSAS score ≥14 (the median PSAS score of the 6-month window group), in each time window group. The rate and corresponding count (events/total) are displayed above each bar; Bonferroni-adjusted pairwise comparisons (chi-square or Fisher exact test) are indicated by horizontal brackets with asterisks. Note: PSAS, Patient Scar Assessment Scale. The suboptimal scar threshold was defined as the median PSAS score of the 6-month window group. Jonckheere-Terpstra trend test P<0.001 across the three time windows. ***P<0.001, **P<0.01; ns, not significant.

erogeneity remains in sensory alteration, quality of life, and different PROMs. A recent propensity-score matched study showed that endoscopic thyroid surgery was superior to open surgery in cosmetic outcomes and overall satisfaction without a marked increase in major complications [21]. Another systematic review and meta-analysis suggested that overall quality-of-life outcomes may be similar between endoscopic and conventional open thyroidectomy, indicating that cosmetic benefit does not necessarily translate into improvement in all quality-of-life dimensions [22]. It must be emphasized that the cross-sectional design of the present study means that each time-window group represents a different set of patients assessed once; the between-group differences reported here therefore reflect comparisons among different patients and must not be interpreted as longitudinal improvement within the same individuals. Nevertheless, the baseline characteristics were balanced across the three groups, and BREAST-Q scores showed a clear ordinal gradient. This finding has important clinical implications for expectation management. During preoperative counseling, patients may be informed that satisfaction with the areolar and anterior chest incisions was higher among patients assessed at longer postoperative intervals than among those assessed soon after surgery; whether such a difference reflects within-patient improvement over time would require prospective longitudinal confirmation. The latest remote-access consensus also emphasizes that such procedures should be performed on the basis of patient selection and expectation management [5].

BMI <24 kg/m^2 was one of the strongest independent factors associated with cosmetic satisfaction in this study (OR= 2.86, P<0.001). Its effect size was slightly higher than that reported in some literature, possibly because patients with a BMI ≥24 accounted for a relatively large proportion of this cohort (44.3%). Trans-breast endoscopic surgery requires creation of a working tunnel in the subcutaneous fat layer of the chest wall. In patients with higher BMI, the thicker subcutaneous fat layer may necessitate a wider range of tissue dissection during tunnel creation, complicate local contour recovery, and reduce the visibility of the aesthetic cervicothoracic angle. Recent studies have shown that higher BMI is associated with prolonged operative time in thyroid surgery [23]. A systematic review and meta-

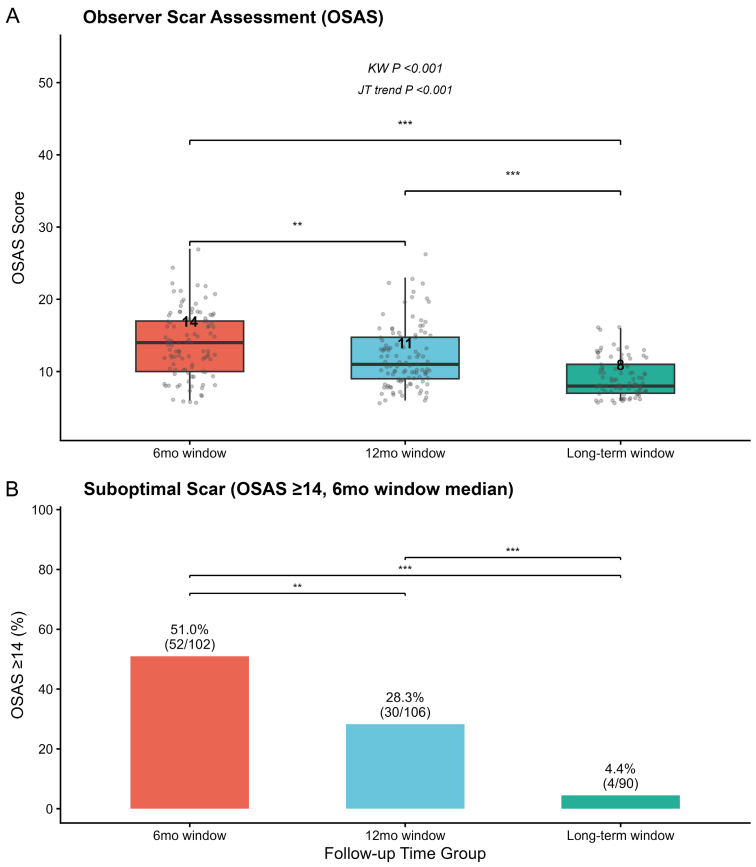


Figure 5. OSAS scores across three follow-up time windows. A. Box-and-jitter plots of OSAS scores by time window group, with conventions and annotations as described in **Figure 1**. B. Bar chart showing the proportion of patients with a suboptimal scar, defined as an OSAS score ≥14 (the median OSAS score of the 6-month window group), in each time window group. The rate and corresponding count (events/total) are displayed above each bar; Bonferroni-adjusted pairwise comparisons (chi-square or Fisher exact test) are indicated by horizontal brackets with asterisks. Note: OSAS, Observer Scar Assessment Scale. The suboptimal scar threshold was defined as the median OSAS score of the 6-month window group. Jonckheere-Terpstra trend test P<0.001 across the three time windows. ***P<0.001, **P<0.01; ns, not significant.

analysis by Ghani et al. [24] also suggested that obesity may influence certain outcomes after thyroidectomy. Unlike those studies, however, the primary outcome in the present study was not a safety event but patient-reported cosmetic satisfaction; therefore, the influence of BMI may more strongly reflect tissue thickness in the operative region, local recovery, and differences in appearance perception. Age <40 years was also an independent associated factor (OR=2.08), possibly because younger patients have better skin elasticity and scar-repair capacity, higher expectations for body image, and more active engagement in scar management. Recent findings regarding the relationship between age and scar-related quality of life are not entirely consistent. Noguera et al. [25] suggested that age may be associated with the impact of scars on quality of life, indicating that preoperative expectation management should be individualized according to age, aesthetic needs, and psychological concerns. Sensitivity analysis further showed that BMI <24 and age <40 years remained significant at the extreme thresholds of ≥70 and ≥90, whereas the 12-month time window was significant only at the ≥70 threshold. This suggests that the 12-month window may primarily influence achievement of a "basic satisfaction" level, whereas the association with the long-term time window was more prominent for achieving higher satisfaction.

The absolute differences in VHI-10 scores among the three groups were small (median scores of 5-6), and approximately 14% of patients had subjective voice handicap. Although the overall between-group difference in VHI-10 scores was borderline significant (P=0.047), only the long-term vs. 6-month comparison remained significant after Bonferroni correction, and the absolute difference was small (median 6 vs. 5); therefore, this finding should be interpreted with caution. An important negative finding of this study is that, aside from RLN injury, preoperative macroscopic variables such as age, BMI, surgical extent, and operative time did not predict the occurrence of VHI-10 ≥11. Beka and Gimm et al. [12] noted that voice changes after thyroidectomy are not fully explained by laryngeal nerve injury and may also be related to strap muscle manipulation, endotracheal intubation, local inflammation, and functional factors. Therefore, the present results suggest that the mechanisms of postoperative voice handicap may be more closely related to

Table 2. Univariate and multivariate logistic regression for BREAST-Q ≥80

Variable	Univariate OR (95% CI)	Univariate P	Multivariate OR (95% CI)	Multivariate P
12-month window (vs. 6-month)	1.91 (1.08-3.41)	0.027	1.83 (1.00-3.41)	0.053
Long-term window (vs. 6-month)	4.57 (2.51-8.50)	<0.001	4.62 (2.39-9.14)	<0.001
Age <40 years	2.43 (1.51-3.96)	<0.001	2.08 (1.21-3.60)	0.008
BMI <24 kg/m^2	3.15 (1.96-5.15)	<0.001	2.86 (1.70-4.88)	<0.001
No complications	2.14 (1.13-4.24)	0.023	2.20 (1.06-4.75)	0.038
Female sex	1.10 (0.47-2.66)	0.829	0.99 (0.38-2.65)	0.988
Total thyroidectomy	0.70 (0.44-1.13)	0.143	0.85 (0.48-1.52)	0.587
Operative time (per 10-min)	0.90 (0.83-0.97)	0.004	0.93 (0.85-1.02)	0.127
Specimen weight (per 10-g)	1.02 (0.93-1.12)	0.737	1.03 (0.92-1.15)	0.644

Note: OR, odds ratio; CI, confidence interval. AUC=0.759 (95% CI: 0.705-0.813); VIF all <2.00; HL P=0.632. Multivariate model includes all 8 variables with 6-month window as reference.

Table 3. Complication types and PRO scores (median, Yes vs. No)

Complication (n)	BREAST-Q	VHI-10	PSAS	OSAS
Hypoparathyroidism (28)	72 vs. 74	6 vs. 6	12 vs. 11	10 vs. 9
RLN injury (12)	70 vs. 74	8* vs. 6	13 vs. 11	11 vs. 9
Hematoma (5)	71 vs. 74	6 vs. 6	12 vs. 11	10 vs. 9
Wound infection (4)	70 vs. 74	6 vs. 6	14 vs. 11	11 vs. 10
Seroma (5)	73 vs. 74	6 vs. 6	12 vs. 11	10 vs. 9

Note: RLN, recurrent laryngeal nerve. *P* values are from the Mann-Whitney U test comparing complication-present vs. complication-absent groups. *P < 0.001 for the comparison of VHI-10 scores between patients with and without transient RLN injury; all other complication-PRO comparisons were non-significant (P > 0.05).

detailed intraoperative factors, such as anatomical variation of the RLN, the duration and force of traction, and the safe distance between energy devices and the nerve, rather than being fully captured by preoperative macroscopic variables. This finding has implications for intraoperative neural protection strategies and refined operative training.

POSAS analysis provided meaningful findings at two levels. First, both PSAS and OSAS scores differed significantly across time-window groups, with patients assessed at longer postoperative intervals reporting lower (i.e., better) scar scores. This supports an association between longer postoperative intervals and better scar appearance assessment at the group level. Choi et al. [13] noted that POSAS is one of the commonly used instruments in studies of post-thyroidectomy scars. Dwivedi et al. [26] also showed that cosmetic scar evaluation after thyroid surgery is related to quality-of-life dimensions. A cross-sectional survey by Shen et al. [27] further suggested that scar perception after open thyroidectomy can change dynamically over time and affect quality of life; although their study population and surgical approach differed from the present study, their findings support the interpretation that postoperative time window may influence subjective scar perception. Second, patient self-assessment was systematically higher than observer assessment, indicating a non-negligible discrepancy between patient and observer scar evaluations in the specific context of TBAET, where incisions are located in the areolar and anterior chest regions rather than the conventional neck region. Draaijers et al. [20] already showed in the development of POSAS that patients and observers do not focus on exactly the same scar dimensions. POSAS also has a measurement basis for use in linear surgical scars [28]. If TBAET scar outcomes are judged only by observer assessment, the subjective burden of scarring perceived by patients may be systematically underestimated, further supporting routine inclusion of PROs in postoperative outcome evaluation.

The multivariable model in this study had an AUC of 0.759, and the core factors remained directionally consistent across three thresholds in sensitivity analyses, indicating that the model possesses only moderate discriminative power, albeit with a degree of robustness. Furthermore, the AUC of 0.759 implies that a large proportion of the variation in cosmetic satisfaction remains unexplained by the includ-

Table 4. Sensitivity analysis at three BREAST-Q cutoffs

Variable	≥70: OR (95% CI)	P	≥80: OR (95% CI)	P	≥90: OR (95% CI)	P
12-month window	2.35 (1.25-4.51)	0.009	1.83 (1.00-3.41)	0.053	2.01 (0.76-5.74)	0.169
Long-term window	12.29 (5.47-29.81)	<0.001	4.62 (2.39-9.14)	<0.001	4.52 (1.77-12.78)	0.003
Age <40 years	2.79 (1.55-5.09)	<0.001	2.08 (1.21-3.60)	0.008	4.50 (1.86-12.74)	0.002
BMI <24 kg/m^2	5.91 (3.30-10.88)	<0.001	2.86 (1.70-4.88)	<0.001	5.31 (2.23-14.82)	<0.001
No complications	1.35 (0.62-2.93)	0.448	2.20 (1.06-4.75)	0.038	2.48 (0.84-9.23)	0.129

Note: AUC, area under the curve; OR, odds ratio; CI, confidence interval. AUC: ≥70=0.829, ≥80=0.759, ≥90=0.814. All 8 variables included in each model. Core factors (long-term window, BMI<24, age <40) remained significant across all cutoffs.

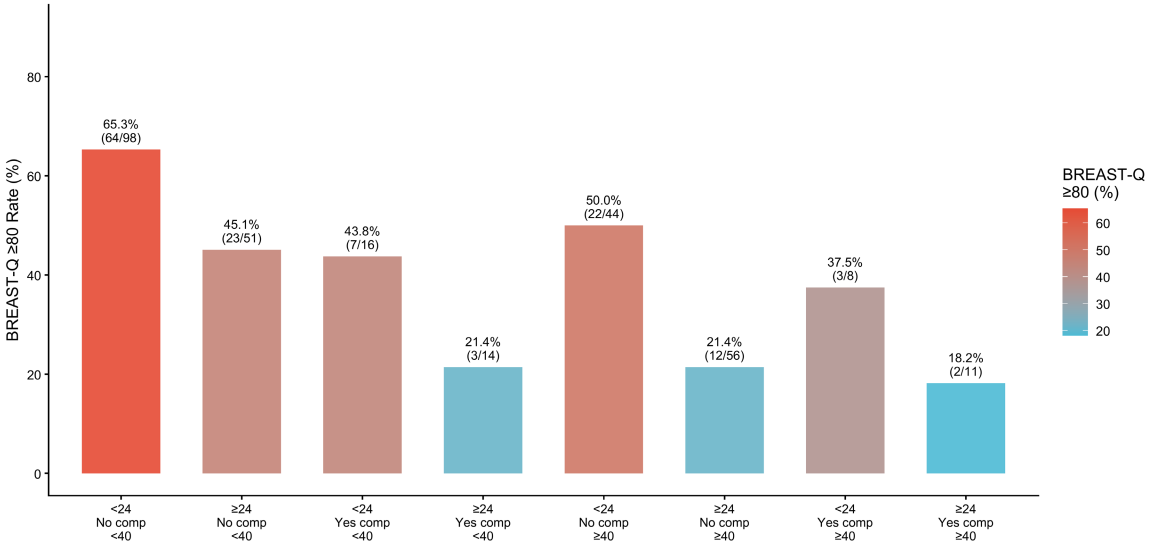


Figure 6. Subgroup analysis: BREAST-Q ≥80 rate stratified by BMI (<24 vs. ≥24 kg/m^2), postoperative complication (No vs. Yes), and age (<40 vs. ≥40 years). The bar height represents the proportion of patients achieving BREAST-Q ≥80 in each of the eight subgroups. The rate and corresponding count (events/total) are displayed above each bar. Color gradient from blue (lower rate) to red (higher rate) reflects the magnitude of the satisfaction rate. Note: BMI, body mass index. The optimal subgroup (BMI <24 kg/m^2, no complication, age <40 years) achieved a BREAST-Q ≥80 rate of 65.3% (64/98), compared with the overall cohort rate of 45.6% (136/298). Subgroups were defined post hoc based on clinically meaningful cutoffs derived from the logistic regression model. These subgroup analyses are exploratory in nature and were not pre-specified endpoints; any apparent gradient across subgroups should therefore be interpreted with caution.

ed variables, pointing to the likely influence of unmeasured patient-centered factors. These may include: (1) preoperative cosmetic expectations, which shape how patients judge surgical outcomes relative to their prior hopes; (2) psychological traits, such as anxiety, depressive symptoms, neuroticism, and perfectionism, which can modulate body-image perception and satisfaction; (3) social and family factors, such as marital status, social support, and family members' attitudes toward visible scarring; and (4) postoperative scar-management behaviors, including adherence to scar-care regimens, sun protection, and the use of silicone-based scar products. None of these factors were available in the present retrospective dataset, and they merit systematic evaluation in future prospective studies. It is important to note that BREAST-Q was developed in the field of breast surgery [16] and it is not a thyroid surgery-specific PRO instrument. In this study its cosmetic satisfaction module was therefore used off-label to quantify satisfaction related to the anterior chest/areolar incisions after TBAET. BREAST-Q ≥80 was a predefined threshold for satisfaction, and its measurement suitability and threshold interpretation in the TBAET population require further validation. Limitations of this study include its single-center retrospective design, the fact that cross-

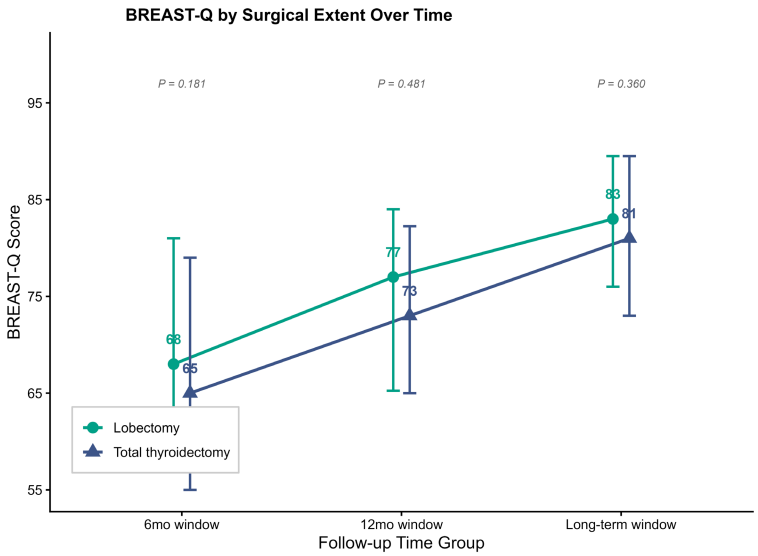


Figure 7. BREAST-Q median scores stratified by surgical extent (lobectomy vs. total thyroidectomy) across three follow-up time windows. Points represent median BREAST-Q scores; error bars represent the interquartile range (IQR). Mann-Whitney U test *P* values for the comparison between lobectomy and total thyroidectomy are annotated at each time point. The two curves display parallel trajectories across time windows. Note: Within each time window, no statistically significant difference in BREAST-Q scores was observed between lobectomy and total thyroidectomy (all P>0.05). Ordinal logistic regression likelihood ratio test confirmed no significant time window × surgical extent interaction (P=0.952), indicating that the BREAST-Q score distribution patterns across time windows were similar for both surgical extents. BREAST-Q, patient-reported cosmetic satisfaction score.

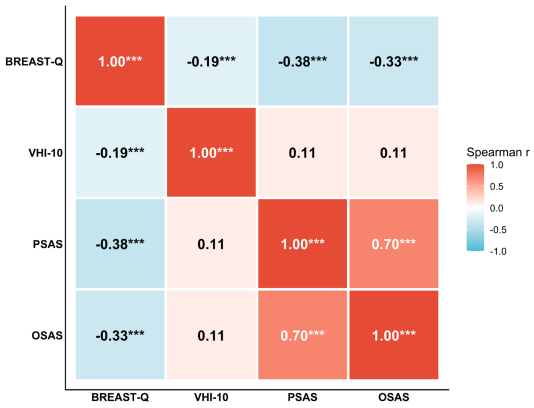


Figure 8. Spearman rank correlation matrix among four PRO dimensions: BREAST-Q cosmetic satisfaction, VHI-10 voice handicap, PSAS patient-reported scar assessment, and OSAS observer-reported scar assessment. The matrix displays Spearman correlation coefficients (r) with significance asterisks. Tile color intensity and direction reflect the strength and sign of the correlation: red indicates positive correlations, blue indicates negative correlations, and white indicates weak or absent correlations. Note: BREAST-Q (higher = better) is inversely correlated with PSAS, OSAS, and VHI-10 (higher = worse), hence the negative correlation coefficients. The strong positive correlation between PSAS and OSAS (r=0.70, P<0.001) supports the concurrent validity of patient- and observer-based scar assessment. The weak, non-significant correlations between VHI-10 and both PSAS (r=0.11, P=0.052) and OSAS (r=0.11, P=0.054) suggest that voice function and scar appearance represent relatively independent PRO dimensions. ***P<0.001 (off-diagonal cells; diagonal cells, r=1.00, are self-correlations shown for reference).

sectional data cannot be equated with longitudinal change, possible follow-up selection bias, absence of preoperative PRO baselines and psychological assessment, lack of physician evaluation, and limited sample size in the long-term window. Future multicenter prospective longitudinal studies should incorporate psychological assessment and physician evaluation and compare long-term PRO differences across different endoscopic approaches.

Conclusion

PRO scores in patients after TBAET differed significantly across postoperative follow-up time windows, and patients in the long-term time window had better cosmetic satisfaction and scar scores. BMI <24 kg/m^2, age <40 years, absence of postoperative complications, and the long-term follow-up time window were independently associated with higher cosmetic satisfaction. Patients rated scars more critically than observers, and PRO assessment has important clinical value in evaluating outcomes after TBAET.

Disclosure of conflict of interest

None.

Address correspondence to: Haoyong Liu, Thyroid and Breast Surgery, The Second Affiliated Hospital of Guangdong Medical University, No.12 Minyou Road, Xiashan District, Zhanjiang 524000, Guangdong, China. E-mail: 18320492292@163.com

Table 5. Univariate logistic regression for VHI-10 ≥11 (no multivariable model constructed)

Variable	OR (95% CI)	P
Long-term window (vs. 6-month)	0.80 (0.36-1.70)	0.568
Age <40 years	0.92 (0.46-1.86)	0.809
BMI <24 kg/m^2	0.85 (0.43-1.68)	0.638
No complications	1.10 (0.47-2.73)	0.831
Female sex	1.25 (0.35-5.82)	0.746
Total thyroidectomy	1.05 (0.52-2.10)	0.884
Operative time (per 10-min)	0.96 (0.87-1.05)	0.377
Specimen weight (per 10-g)	1.01 (0.89-1.13)	0.926

Note: VHI-10, Voice Handicap Index-10; OR, odds ratio; CI, confidence interval; BMI, body mass index.

References

[1] Mu C, Ming X, Tian Y, Liu Y, Yao M, Ni Y, Liu Y and Li Z. Mapping global epidemiology of thyroid nodules among general population: a systematic review and meta-analysis. Front Oncol 2022; 12: 1029926.

[2] Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, Soares P, Solymosi T and Papini E. 2023 European thyroid association clinical practice guidelines for thyroid nodule management. Eur Thyroid J 2023; 12: e230067.

[3] Kim DH, Kim SW, Kim GJ, Basurrah MA and Hwang SH. Efficacy and safety of minimally invasive thyroid surgery: a network meta-analysis. Laryngoscope 2023; 133: 2470-2479.

[4] Mirghani HO and Albalawi IA. A comparison between transoral endoscopic thyroidectomy vestibular approach and transareolar thyroidectomy regarding postoperative pain and cosmetic outcomes: a systematic review and meta-analysis. Pak J Med Sci 2024; 40: 2726-2731.

[5] Russell JO, Tae K, Roth MY, Hou R, Suh I, Tufano RP, Banucchi V, Kandil E, Grogan RH, Berber E, Anuwong A, Chung WY, Kim HY, Yang TL, Lee KE, Koh YW, Raffaelli M, Makay Ö, Barczynski M, Lira RB, Sanabria A and Kowalski LP. Remote-access thyroidectomy and parathyroidectomy: a 2025 consensus statement from the Asia-pacific society of thyroid surgery, American Head and neck society, American association of endocrine surgeons, American thyroid association, European society of endocrine surgeons, and Latin American thyroid society. Thyroid 2025; 35: 1285-1296.

[6] Zhou D, Zhang Z, Dou X, Xia F and Li X. Advances in the assessment of cosmetic outcomes, sensory alteration in surgical areas, and health-related quality of life of endoscopic thyroidectomy. World J Surg Oncol 2024; 22: 52.

[7] Nguyen VC, Song CM, Ji YB, Jeong JH, Russell JO, Chiang FY, Randolph GW and Tae K. Postoperative cosmetic outcomes and quality of life after thyroidectomy: a systematic review and network meta-analysis. Head Neck 2025; 47: 3443-3454.

[8] Temple LKF, Pusic AL, Liu JB, Melucci AD, Collins CE, Kazaure HS, Brajcich BC, Fordham MJ, Lapsley JC and Ko CY. Patient-reported outcome measures within a national multispecialty surgical quality improvement program. JAMA Surg 2024; 159: 1030-1039.

[9] Xiong L, Liu L, Xiang Y, Li C, Dai W, Hu J, Shui C, Zhou Y, Wang X and Cai Y. Patient-reported outcome measures for assessing health-related quality of life in patients with differentiated thyroid cancer: a systematic review. Int J Surg 2025; 111: 1311-1329.

[10] Shan YZ, Zhou LM, Yu ZF, Wang SG, Gao GL, Shen Y and Zhang XL. Comparison between transareola singlesite endoscopic thyroidectomy and minimally invasive video-assisted thyroidectomy. J Int Med Res 2012; 40: 2213-2219.

[11] Shen S, Hu X, Qu R, Guo Y, Luo L and Chen X. Comparing quality of life between patients undergoing trans-areola endoscopic thyroid surgery and trans-oral endoscopic thyroid surgery. BMC Surg 2021; 21: 277.

[12] Beka E and Gimm O. Voice changes without laryngeal nerve alterations after thyroidectomy: the need for prospective trials - a review study. J Voice 2024; 38: 231-238.

[13] Choi WK, Shin HY, Park YJ, Lee SH, Lee AY and Hong JS. Analysis of trends and status of evaluation methods in thyroid scar. Heliyon 2024; 10: e29301.

[14] Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG and Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. Plast Reconstr Surg 2009; 124: 345-353.

[15] Reunmarkkaew D and Sittitrai PJSS. Endoscopic thyroid lobectomy via axillary-breast-shoulder approach versus open thyroid lobectomy. 2018; 09: 8-16.

[16] Park YL, Han WK and Bae WG. 100 cases of endoscopic thyroidectomy: breast approach. Surg Laparosc Endosc Percutan Tech 2003; 13: 20-25.

[17] Kim YS, Joo KH, Park SC, Kim KH, Ahn CH and Kim JS. Endoscopic thyroid surgery via a breast

approach: a single institution's experiences. BMC Surg 2014; 14: 49.
[18] Rosen CA, Lee AS, Osborne J, Zullo T and Murry T. Development and validation of the voice handicap index-10. Laryngoscope 2004; 114: 1549-1556.
[19] Arffa RE, Krishna P, Gartner-Schmidt J and Rosen CA. Normative values for the voice handicap index-10. J Voice 2012; 26: 462-465.
[20] Draaijers LJ, Tempelman FR, Botman YA, Tuinebreijer WE, Middelkoop E, Kreis RW and van Zuijlen PP. The patient and observer scar assessment scale: a reliable and feasible tool for scar evaluation. Plast Reconstr Surg 2004; 113: 1960-1965; discussion 1966-1967.
[21] Wang M and Jin G. Quality of life and surgical outcomes of endoscopic thyroidectomy versus open thyroidectomy in papillary thyroid carcinoma. Medicine (Baltimore) 2025; 104: e46114.
[22] Agrawal V, Yadav SK, Agarwal P, Sharma D and Mishra SK. Quality of life after endoscopic vs. conventional open thyroidectomy: a systematic review and meta-analysis. Langenbecks Arch Surg 2025; 410: 144.
[23] Eliachar B and Ronen O. Patients' obesity is linked to prolonged thyroidectomy operative time. Langenbecks Arch Surg 2025; 410: 209.
[24] Ghani A, Al Darwashi LM, Hajeir MY, Ayesh AR, Witkowiak MM, Alogakos M, Hage K, Than CA, Nakanishi H, Shin J and Romero-Velez G. Clinical impact of obesity on thyroidectomy outcomes: a systematic review and meta-analysis. Surgery 2025; 186: 109598.
[25] Sánchez-Cabezudo Noguera FM, Muñoz de Nova JL, Garcia Virosta M, Valdés de Anca Á, Vaquero Pérez MA, Martin Perez E and Picardo Nieto AL. Influence of cosmetic results of thyroidectomy scars on patient quality of life. Cir Esp (Engl Ed) 2025; 103: 209-216.
[26] Dwivedi S, Kumar S, Saxena V, Dutta A, Dinker Y, Dwivedi S, Majumdar SS, Varshaney A and Sharma M. Evaluation of scar cosmesis and quality of life after thyroid surgery: a retrospective study. Indian J Otolaryngol Head Neck Surg 2025; 77: 1505-1512.
[27] Shen K, Liu Y, Tan M, Yan S, Yang S and Xu X. Long-term scar perception and quality of life after open thyroidectomy in Chinese patients: a cross-sectional survey study. Thyroid 2026; 36: 89-98.
[28] van de Kar AL, Corion LU, Smeulders MJ, Draaijers LJ, van der Horst CM and van Zuijlen PP. Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale. Plast Reconstr Surg 2005; 116: 514-522.

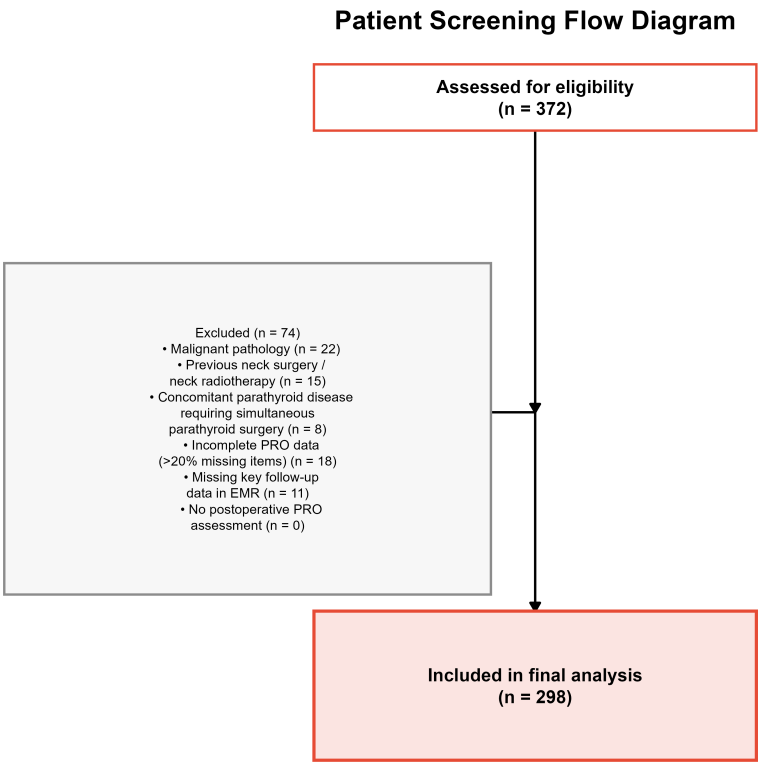


Figure S1. Patient screening flow diagram. Note: TBAET, trans-breast approach endoscopic thyroidectomy; PRO, patient-reported outcome; EMR, electronic medical record. A total of 372 patients with benign thyroid disease who underwent TBAET between 2021 and 2023 were assessed for eligibility; 74 were excluded (22 with malignant pathology, 15 with previous neck surgery or neck radiotherapy, 8 with concomitant parathyroid disease requiring simultaneous parathyroid surgery, 18 with >20% missing items in any PRO scale, 11 with missing key follow-up data in the EMR, and none without a postoperative PRO assessment), and 298 patients were included in the final analysis. Because completion of a full postoperative PRO assessment was a prerequisite for inclusion in this retrospective cross-sectional study, the loss to follow-up rate was 0%.

Table S1. Event numbers and events-per-variable (EPV) ratios for the three BREAST-Q satisfaction cutoffs in sensitivity analyses

Cutoff	Events, n (%)	Non-events, n (%)	Total	EPV (8 variables)	EPV ≥10 satisfied
≥70	188 (63.1)	110 (36.9)	298	23.5	Yes
≥80	136 (45.6)	162 (54.4)	298	17	Yes
≥90	45 (15.1)	253 (84.9)	298	5.6	No

Note: EPV, events per variable. EPV = number of events ÷ number of predictor variables (8 in the full multivariable models); a threshold of EPV ≥10 is generally recommended. The full eight-variable model for the ≥90 cutoff did not satisfy this criterion. As a robustness check, a reduced four-variable model (time window, age <40 years, BMI<24 kg/m^2, and absence of complications) was fitted for the ≥90 cutoff, yielding EPV=11.25, in which the long-term time window (OR=4.90, 95% CI 1.97-13.60), age <40 years (OR=4.46, 1.86-12.49), and BMI<24 kg/m^2 (OR=5.31, 2.25-14.73) remained statistically significant, and absence of complications remained non-significant.