Cancer Management and Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Impact of the Extent of Thyroidectomy on Quality of Life in Differentiated Thyroid Cancer Survivors: A Propensity Score Matched Analysis

Xiongtao Yang[1]
Qing Yang[1]
Yuan Tang[1]
Jing Ma[2]
Huamei Ye[3]

[1]Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, 610041, People's Republic of China; [2]Business School, Sichuan University, Chengdu, 610064, People's Republic of China; [3]People's Hospital of Dujiangyan City, Chengdu, 611830, People's Republic of China

Purpose: As most thyroid cancer patients survive for more than ten years, it has become increasingly important to understand whether the different surgery types have any effect on the quality of life (QoL) of patients.

Patients and Methods: Using observational data from head and neck surgery at the Sichuan Cancer Hospital in China, three scoring methods – sum scoring, domain-based scoring and IRT-based scoring, were employed to measure the QoL in differentiated thyroid cancer (DTC) patients and a propensity score matched analysis performed to explore the impact of surgery type on QoL as measured by the Treatment of Cancer Quality of Life core Questionnaire version 3.0 (EORTC QLQ-C30) and a disease-specific health-related quality of life questionnaire (THYCA-QoL).

Results: No statistically significant patient QoL differences were found between the two surgery types regardless of which questionnaire was used and which scoring method was used (ATE $= -0.400$, $p = 0.834$ using the EORTC QLQ-C30 and the sum scoring; ATE $= -0.4491$, $p = 0.807$ using the EORTC QLQ-C30 and the domain-based scoring; and ATE $= -0.442$, $p = 0.114$ using the EORTC QLQ-C30 and the IRT-based scoring; ATE $= -0.827$, $p = 0.586$ using the THYCA-QoL and the sum scoring; ATE $= -1.692$, $p = 0.406$ using the THYCA-QoL and the domain-based scoring; and ATE $= -0.032$, $p = 0.908$ using the THYCA-QoL and the IRT-based scoring).

Conclusion: This study confirmed that the surgery type (hemithyroidectomy or total thyroidectomy) for DTC patients did not appear to influence their general QoL.

Keywords: thyroid cancer, quality of life, type of surgery, propensity score matching, bifactor model

Plain Language Summary

Nowadays, more than 95% of patients with differentiated thyroid cancer can survive for more than ten years, so the quality of life of these patients is of great concern. Recently, hemithyroidectomy and total thyroidectomy are the most important ways to treat differentiated thyroid cancer patients, so we employed a propensity score matching method to build a counterfactual inference framework to detect causality between the extent of thyroidectomy and quality of life in differentiated thyroid cancer patients. Besides, we used three scoring methods (sum scoring, domain-based scoring and IRT-based scoring) and two popular questionnaires (the Treatment of Cancer Quality of Life core Questionnaire version 3.0 and a disease-specific health-related quality of life questionnaire) to measure the quality of life of differentiated thyroid cancer patients. Our research conclusions provided evidence that the different surgery types (hemithyroidectomy and total thyroidectomy) did not influence the quality of life of the differentiated thyroid cancer patients.

Correspondence: Jing Ma
Business School, Sichuan University, Chengdu, 610064, People's Republic of China
Email jingma_mjjmm@foxmail.com

Received: 8 June 2021
Accepted: 16 August 2021
Published: 7 September 2021

Introduction

The incidence of differentiated thyroid cancer (DTC) has been rising over the past few decades and has now become one of the more common cancers.[1] At present, the first choice and standard DTC treatment is surgery, which is generally a total thyroidectomy (with or without neck dissection as required) or a -hemithyroidectomy.[2] Due to the excellent prognoses in recent years, 89% of patients can now expect to live for at least five years, with 85% of these living for an average of ten or more years.[3]

As having cancer can lead to a low mood, some patients may gradually develop symptoms that could increase the likelihood of depression.[4] Therefore, although thyroid cancer has a lower mortality than most other cancers, recent studies have found that the self-reported quality of life (QoL) of patients with thyroid cancer is similar to and often worse than that of patients diagnosed with cancers that have poorer prognoses.[5–7] Further, as thyroid cancer prognoses are more favorable, particular emphasis should be placed on QoL issues.[8] However, as patient QoL cannot be directly observed, it must be inferred (using a mathematical model) from other observable variables. Therefore, to ensure a more dependable conclusion, this paper used three scoring methods: sum scoring, domain-based scoring and IRT-based scoring: to measure DTC patient QoL.

It has been well documented that thyroid cancer surgery can have numerous potential complications, such as the risk of scarring, hypocalcemia, pain, infection, dysphonia, and dysphagia.[9] Therefore, it is important to study the impact the surgery type has on QoL in thyroid cancer survivors. However, this assessment is very complex as surgery type impacts can be confounded by other factors such as postoperative care. Therefore, a propensity score matching (PSM) method[10] that builds a counterfactual inference framework was used to detect whether surgery type affected QoL outcomes.

The remainder of this paper is structured as follows. Section 2 briefly reviews the determinants and impacts of surgery type on the QoL of thyroid cancer survivors; Section 3 presents the research methodology: study design, study sample, QoL patient scoring, and statistical analysis; Section 4 discusses the empirical results and findings; and Section 5 discusses the implications and gives the conclusion.

Literature Review

While thyroid cancer often has a good prognosis, it can adversely impact QoL.[8] Since the recognition of QoL in the 1970s, QoL assessments have received significant research attention, which has enriched clinical research, public health, and daily medical practice[3] with a great deal of QoL research having been focused on evaluating the associated QoL factors.

Patient QoL evaluation research has developed and verified several questionnaires, the most widely applied of which had been the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire version 3.0 (EORTC QLQ-C30)[11,12] and a disease-specific health-related quality of life questionnaire (THYCA-QoL).[13] Specifically, Nordin et al[11] investigated the known-group validity of a two-item global QoL scale and three alternative scoring algorithms for the EORTC QLQ-C30 based on: (1) the 15 EORTC QLQ-C30 scale means; (2) the sum of all individual EORTC QLQ-C30 items (except for the financial problems item); and (3) the sum of the scales assessing physical function, emotional function, QoL, fatigue, nausea/vomiting, pain, appetite, and diarrhea. Hinz et al[12] then used a total score derived from summing all 30 questionnaire items and two separate summary scores based on the sum of all items in the functioning and symptom domains. Husson et al[13] developed and pretested a thyroid cancer specific HRQoL questionnaire that can be used in addition to the more general EORTC QLQ-C30.

In recent years, factor analysis models have also become increasingly popular in clinical research to measure health behavior and QoL. For example, Efficace et al[14] estimated a single-factor model using confirmatory factor analysis (CFA) and verified that the single-factor model in the EORTC QLQ-C30 exhibited an adequate fit for hematologic malignancy patients, and Giesinger et al[15] found a robust single higher order factor model to be the best performing measurement model for the EORTC QLQ-C30. As the EORTC QLQ-C30 and THYCA-QoL items use a Likert scale, item response theory (IRT) was found to be more suitable for measuring the QoL as it allowed respondents to indicate a level of agreement (using a rating or Likert scale).[16] Compared with the higher-order IRT model, the bifactor model was found to provide more accurate project parameters, trait estimates and test reliability.[17,18] Therefore, this study measured the QoL of thyroid cancer patients using the bifactor model underlying the EORTC QLQ-C30 and the THYCA-QoL.

Research from socioeconomic and pathological perspectives has been conducted to identify possible associated QoL factors. Goswami et al[9] for example, found that patient age, postoperative hypocalcemia, dysphonia, dysphagia, scar appearance, and complications from radioactive iodine were all associated with QoL, and Haraj et al[28] found that radioiodine therapy and its doses, the metastasis, multifocality, lymph node dissection and the microcarcinoma were QoL influencing factors.

There have been differing views as to whether surgery type affects the QoL of thyroid cancer survivors. For example, Nickel et al[19] used a multivariate regression model and found that hemithyroidectomies had fewer adverse treatment effects and better QoL outcomes than total thyroidectomies for DTC patients, and Hauch et al[20] suggested that as thyroidectomies may cause vocal cord paralysis and hypoparathyroidism, total thyroidectomy risks were higher than the hemithyroidectomy risks. However, several studies have found no significant QoL differences in patients treated with hemithyroidectomy and those treated with total thyroidectomy. For example, Bongers et al[21] employed univariate comparisons and a multivariate regression analysis and found that there were no significant QoL differences between low-risk DTC patients treated with total thyroidectomy and those treated with a hemithyroidectomy.

There are usually strict selection criteria for hemithyroidectomy surgeries for DTC patients: tumors less than 4cm; no history of radiation exposure; no distant transfer; no cervical lymph node metastasis; and no extracapsular invasion. Therefore, because of this treatment selection bias, there may be systematic differences in the baseline variable distributions in the two groups, which could bias the treatment effect estimates.[10,22] As this means that using traditional least squares methods to assess the QoL differences between the two surgical types could lead to biased results, this study sought to provide solid evidence for the QoL surgery impact in DTC survivors by employing propensity score matching (PSM) and conditioning the confounders to reduce the bias in the treatment effect estimations in an observational data set.

Materials and Methods

Study Design and Study Sample

A cross-sectional, self-administered survey of DTC patients was conducted to assess their QoL. The target population was a consecutive cohort of adults being treated for DTC between November and December, 2019, at the Head and Neck Surgery in the Sichuan Cancer Hospital in China. The inclusion criteria were: (1) 18 years or older at the time of the survey; (2) pathological diagnosis for DTC; (3) primary school education and above; (4) knowing the diagnostics results; (5) having clear thinking and being able to understand and cooperate with the surveys and barrier-free language communication; and, (6) volunteering for the study and signing the informed consent. The exclusion criteria were: (1) any combination with serious primary diseases of the liver, kidney, hematopoiesis, or endocrine system with severe organ failure; (2) a history of mental illness, personality disorders, cognitive impairments, or organic brain disease; and (3) participating in other clinical studies.

A unified instruction was used to explain the survey purpose, significance and questionnaire completion method to the research subjects. After obtaining the informed consent, the questionnaire was distributed online to avoid the need to enter data again and increase the workload. Questionnaire was completed by the research object, and at the same time, the researcher was on the side to assist. Patients who had difficulties in independently understanding the questions were assisted by the researchers, who avoided using any language that could have affected the patient's judgment. Finally, 186 questionnaires were distributed and 150 valid questionnaires received, an effective return rate of 80.6%.

Scoring the Patients' QoL

The QoL of the DTC survivors was the dependent variable of interest, which was measured using the classical questionnaire, the Treatment of Cancer Quality of Life core Questionnaire version 3.0 (EORTC QLQ-C30)[23] and a disease-specific health-related quality of life questionnaire (THYCA-QoL),[13] which are widely used and validated questionnaires to evaluate the QoL of oncology patients and thyroid cancer patients.

The questionnaire EORTC QLQ-C30 has a global quality-of-life subscale (GQ, 2 items), five functioning subscales – physical functioning (PF, 5 items); role functioning (RF, 2 items); cognitive function (CF, 2 items); emotional functioning (EF, 2 items); and social functioning (SF, 4 items) – nine symptom subscales – fatigue (FA, 3 items); pain (PA, 2 items); and nausea/ vomiting (NV, 3 items) – and six single items that assesses additional symptoms commonly reported by cancer patients – dyspnea (DY), lack of appetite (LA), insomnia (IN), constipation (CO), diarrhea (DI), and financial difficulties(FD). The time frame was the week before the survey with each item being scored on a 4-point response scale ranging from 1 = "not at all", to 4 = "very much", except for global QoL scale, which was scored on a seven-point modified linear analogue scale ranging from 1 = "very poor" to 7 = "excellent". Except for the global QoL scale, all items in the other scales were inverse items, that is, the higher the score, the lower the QoL.

The questionnaire THYCA-QoL consists of 24 items that assess seven subscales – neuromuscular (NM, 3 items), voice (VO, 2 items), concentration (CT, 2 items), sympathetic (ST, 2 items), throat/mouth problems (TM, 2 items), psychological (PC, 4 items), and sensory problems (SE, 2 items) – and six single items: scar (SC), chilly (CH), tingling hands/feet (THF), weight gain (GW), headache (HA), and interest of sex (SXI). Except the time frame of interest of sex item is the past 4 weeks, all other items are in the past 1 week. Each item is scored on a four-point response scale ranging from 1 = "not at all" to 4 = "very much". Except for the interest of sex item, other items were inverse items, that is, the higher the score, the lower the QoL.

To ensure a solid result, the questionnaire was scored using conventional sum scoring, domain-based scoring and IRT-based scoring. For the sum scoring, the QoL score was determined by summing all item scores and transforming the inverse items so that higher scores represented a greater QoL:

$$Y_i^{\mathrm{sum}(e)} = \sum_{n=1}^{28}\left(5 - Q_{ij}^{e}\right) + Q_{i29}^{e} + Q_{i30}^{e} \tag{1}$$

$$Y_i^{\mathrm{sum}(t)} = \sum_{n=1}^{23}\left(5 - Q_{ij}^{t}\right) + Q_{i24}^{t}, \tag{2}$$

where $Y_i^{\mathrm{sum}(e)}$ and $Y_i^{\mathrm{sum}(t)}$ represented the QoL scores measured by EORTC QLQ-C30 and THYCA-QoL respectively for patient i measured by summing all item scores, and Q_{ij}^{e} and Q_{ij}^{t} were the response of patient i on item$j$ on questionnaire EORTC QLQ-C30 and THYCA-QoL.

The domain-based scoring and the EORTC QLQ-C30 domain division and the THYCA-QoL domain division followed three steps (Tables 1 and 2): (1) The raw scores (RS) for each domain were calculated separately, with the RS being equal to the sum of the item scores in this domain divided by the number of items in this domain, that is, $\mathrm{RS} = (Q_1 + Q_2 + \ldots + Q_n)/n$

To compare the scores from the various domains, a linear transformation method was used to transform the RS into a standard score (SS) ranging from 0 to 100. And this transformation method ensured that a higher SS represented a greater QoL. The domain-based scored for EORTC QLQ-C30 (Equation (3)) and for THYCA-QoL (Equation (4)) were then determined by averaging the standard score (SS):

$$Y_i^{\mathrm{domain}(e)} = \frac{\mathrm{SS}_i^{\mathrm{PF}} + \mathrm{SS}_i^{\mathrm{EF}} + \cdots + \mathrm{SS}_i^{\mathrm{GQ}}}{15}, \tag{3}$$

$$Y_i^{\mathrm{domain}(t)} = \frac{\mathrm{SS}_i^{\mathrm{NM}} + \mathrm{SS}_i^{\mathrm{VO}} + \cdots + \mathrm{SS}_i^{\mathrm{SXI}}}{13}, \tag{4}$$

where $Y_i^{\mathrm{domain}(e)}$ and $Y_i^{\mathrm{domain}(t)}$ were the QoL measured by EORTC QLQ-C30 and THYCA-QoL respectively for patient i using the domain-based scoring and $\mathrm{SS}_i^{\mathrm{PF}}$, $\mathrm{SS}_i^{\mathrm{EF}}$, $\mathrm{SS}_i^{\mathrm{GQ}}$, $\mathrm{SS}_i^{\mathrm{NM}}$, $\mathrm{SS}_i^{\mathrm{VO}}$, and $\mathrm{SS}_i^{\mathrm{SXI}}$ were the standard scores for the PF, EF, GQ, NM, VO, and SXI domains of patient i.

Bifactor models were used to evaluate the QoL score for the IRT-based scoring (shown as a graphical representation in Figures 1 and 2). The structure of the final bifactor model for the EORTC QLQ-C30 included a general factor on which all the items were loaded, and three specific factors: Physical burden, Mental function, and Overall evaluation. The global quality-of-life subscale was loaded on the overall evaluation (F3, consisting of 2 items, as shown in Figure 1), and the remaining subscales were loaded only on the physical burden factor (F1, including PF, FA, NV, PA, DY, DI, CO, FD, and IN subscale, consisting of 17 items, as shown in Figure 1) or on the mental function factor (F2, including EF, CF, SF and RF subscale, consisting of 11 items, as shown in Figure 1), with the specific factors being orthogonal to the general factor. The structure of the bifactor model for the THYCA-QoL included a general factor on which all the items were loaded, and three specific factors: Physiological feeling, Psychological feeling, and Pain. Each item was loaded only on the Physiological feeling (F1, consisting of 10 items), on the Psychological feeling factor (F2, consisting of 9 items) or on the Pain factor (F3, consisting of 5 items, as shown in Figure 2), with the specific factors being orthogonal to the general factor. Based on bifactor model, the estimation of the overall QoL Y_i^{IRT} was expressed as follows, which was

Table 1 Domain Division and Domain-Based Scoring Method for EORTC QLQ-C30

Domain	Property	Item Numbers	R	RS	SS
Physical functioning (PF)	Functioning	5	3	$\frac{Q_1+Q_2+Q_3+Q_4+Q_5}{5}$	$(1-\frac{RS-1}{R})*100$
Emotional functioning (EF)	Functioning	4	3	$\frac{Q_{21}+Q_{22}+Q_{23}+Q_{24}}{4}$	$(1-\frac{RS-1}{R})*100$
Cognitive functioning (CF)	Functioning	2	3	$\frac{Q_{20}+Q_{25}}{2}$	$(1-\frac{RS-1}{R})*100$
Social functioning (SF)	Functioning	2	3	$\frac{Q_{26}+Q_{27}}{2}$	$(1-\frac{RS-1}{R})*100$
Role functioning (RF)	Functioning	2	3	$\frac{Q_6+Q_7}{2}$	$(1-\frac{RS-1}{R})*100$
Fatigue (FA)	Symptom	3	3	$\frac{Q_{10}+Q_{12}+Q_{18}}{3}$	$(1-\frac{RS-1}{R})*100$
Nausea & vomiting (NV)	Symptom	2	3	$\frac{Q_{14}+Q_{15}}{2}$	$(1-\frac{RS-1}{R})*100$
Pain (PA)	Symptom	2	3	$\frac{Q_9+Q_{19}}{2}$	$(1-\frac{RS-1}{R})*100$
Dyspnoea (DY)	Symptom	1	3	Q_8	$(1-\frac{RS-1}{R})*100$
Diarrhoea (DI)	Symptom	1	3	Q_{17}	$(-\frac{RS-1}{R})*100$
Constipation (CO)	Symptom	1	3	Q_{16}	$(1-\frac{RS-1}{R})*100$
Financial difficulties (FD)	Symptom	1	3	Q_{28}	$(1-\frac{RS-1}{R})*100$
Insomnia (IN)	Symptom	1	3	Q_{11}	$(1-\frac{RS-1}{R})*100$
Lack of appetite (LA)	Symptom	1	3	Q_{13}	$(1-\frac{RS-1}{R})*100$
Global quality of life (GQ)	Overall	2	6	$\frac{Q_{29}+Q_{30}}{2}$	$(\frac{RS-1}{R})*100$

Note: R=maximum score-minimum score.

Table 2 Domain Division and Domain-Based Scoring Method for THYCA-QoL

Domain	Item Numbers	R	RS	SS
Neuromuscular (NM)	3	3	$\frac{Q_{10}+Q_{12}+Q_{13}}{3}$	$(1-\frac{RS-1}{R})*100$
Voice (VO)	2	3	$\frac{Q_3+Q_4}{2}$	$(1-\frac{RS-1}{R})*100$
Concentration (CT)	2	3	$\frac{Q_{20}+Q_{21}}{2}$	$(1-\frac{RS-1}{R})*100$
Sympathetic (ST)	2	3	$\frac{Q_8+Q_9}{2}$	$(1-\frac{RS-1}{R})*100$
Throat/mouth problems (TM)	3	3	$\frac{Q_1+Q_2+Q_5}{3}$	$(1-\frac{RS-1}{R})*100$
Psychological (PC)	4	3	$\frac{Q_{17}+Q_{19}+Q_{22}+Q_{23}}{4}$	$(1-\frac{RS-1}{R})*100$
Sensory (SE)	2	3	$\frac{Q_{15}+Q_{16}}{2}$	$(1-\frac{RS-1}{R})*100$
Scar (SC)	1	3	Q_6	$(1-\frac{RS-1}{R})*100$
Chilly (CH)	1	3	Q_7	$(1-\frac{RS-1}{R})*100$
Tingling hands/feet (THF)	1	3	Q_{11}	$(1-\frac{RS-1}{R})*100$
Gained weight (GW)	1	3	Q_{14}	$(1-\frac{RS-1}{R})*100$
Headache (HA)	1	3	Q_{18}	$(1-\frac{RS-1}{R})*100$
Interest of sex (SXI)	1	3	Q_{24}	$(\frac{RS-1}{R})*100$

Note: R=maximum score-minimum score.

https://doi.org/10.2147/CMAR.S321988
DovePress

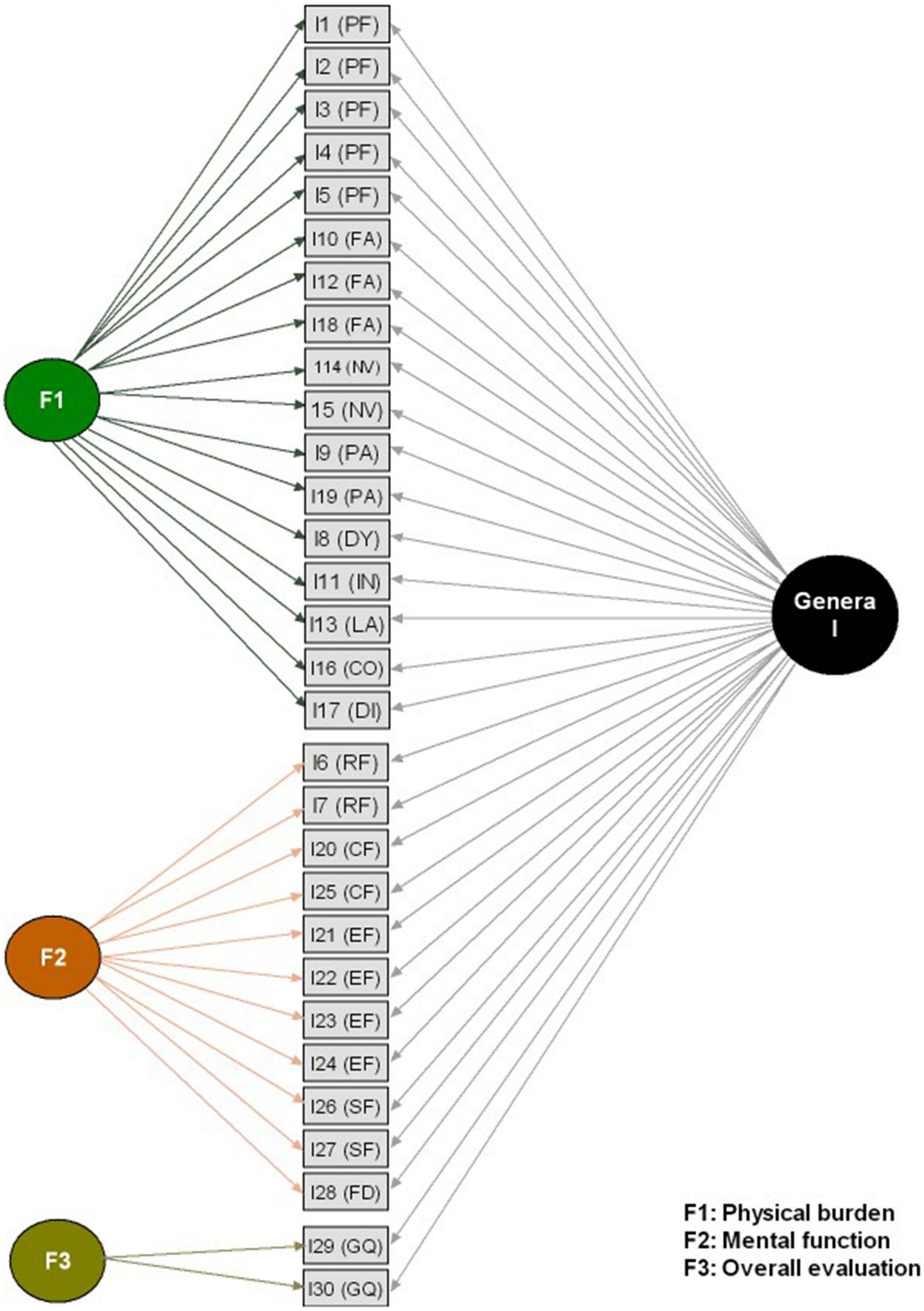

Figure 1 QoL bifactor model for the EORTC QLQ-C30.

suitable for both the EORTC QLQ-C30 and the THYCA-QoL:

$$\log it(Q_{ij}) = \mu_j + \lambda_j^g\theta_i^g + \lambda_j^1\theta_i^1 + \lambda_j^2\theta_i^2 + \lambda_j^1\theta_i^1 + \varepsilon_{ij}$$

$$Y_i^{IRT} = \omega_g\theta_i^g + \omega_1\theta_i^1 + \omega_2\theta_i^2 + \omega_3\theta_i^3$$

where Q_{ij} was the response of patient i on item$j$, μ_j was the intercept for item j, which was listed as a mean as this was typically what it became, θ was the latent factor scores, θ_i^g was the overall ability of the general factor for patient i, θ_i^1, θ_i^2 and θ_i^3 were the specific abilities for the specific factors for patient i, λ_j^g, λ_j^1, λ_j^2 and λ_j^3 were the standardized factor loadings associated with the general and specific factors for item j, with $\lambda_j^f = 0$ if item j load on factor f, and $\in_{ij}$ was the residual for patient i on item j. Y_i^{IRT} was the QoL score for

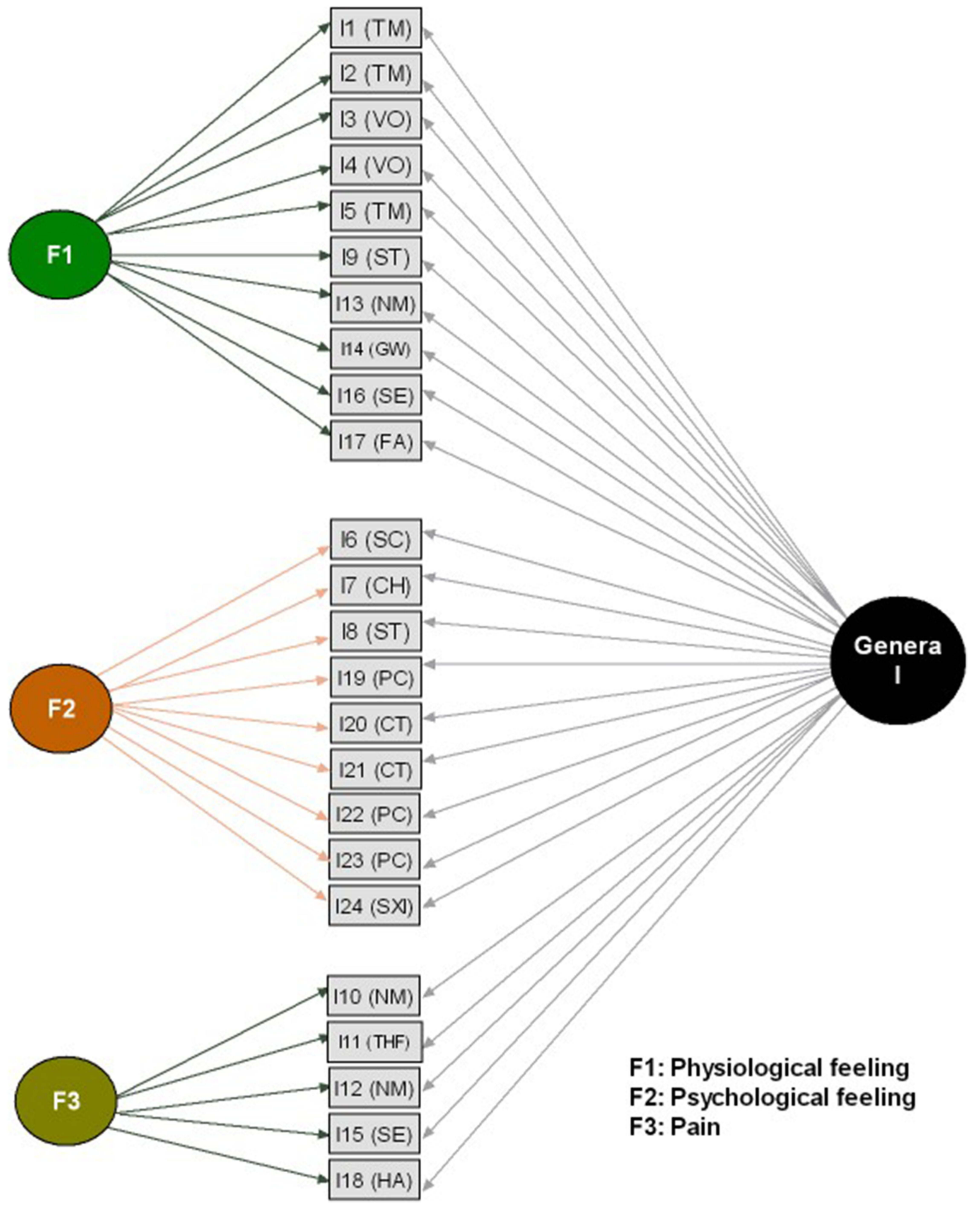

Figure 2 QoL bifactor model for the THYCA-QoL.

patient i measured using the IRT-based scoring, w_g was the weight of the general factor for the overall score, and w_1, w_2 and w_3 were the weights for the three specific factors for the overall score. A bifactor general model was employed using 1 and 0 as the weights: $w_g = 1, w_1 = 0, w_2 = 0$, and $w_3 = 0$.

The structural QoL model was evaluated using R software and the multidimensional item response theory (MIRT) package[24] using a confirmatory maximum likelihood bifactor model under an item response theory (IRT) paradigm. The IRT model was fitted using a dimensional reduction EM algorithm.

Statistical Analysis

Propensity score matching was performed to reduce any treatment selection bias and potential confounding and to adjust for any significant differences in patient characteristics or lesions.[25] The propensity scores were estimated using a logit regression model in both hemithyroidectomy and total thyroidectomy DTC patients. Generally, model fit or parsimony is not a concern when estimating propensity score models as the goal of the ps-logit is to find the model that results in the best covariate balance. The following variables were selected to calculate the propensity score and adjust the baseline or general characteristic discrepancies between the two groups: age, income level, employment status, marital status, and means of payment: with age being a continuous variable and the others being categorical variables: income level had three levels:≤3000, 3000~5000, and ≥5000; employed status had two levels: yes and no; marital status had two levels: yes and no; and means of payment had five levels: own expense, medical insurance (urban), medical insurance (rural), business insurance, and other. Absolute standardized differences were used to determine the balance after matching, with all absolute standardized differences after matching being less than 0.1. Using caliper matching with a caliper width of 0.01 standard deviations of the propensity score, propensity score matching was used to calculate the average treatment effect for surgery type (total thyroidectomy versus hemithyroidectomy) on QoL:

$$ATE = E[Y(T = 1) - Y(T = 0)]$$

$$ATT = E[Y(T = 1) - Y(T = 0)|T = 1]$$

$$ATU = E[Y(T = 1) - Y(T = 0)|T = 0]$$

where ATE was the average treatment effect that evaluated the expected effect on the outcome if patients were randomly assigned to be treated with either total thyroidectomy or hemithyroidectomy, ATT was the average treatment effect on the treated group that explicitly evaluated the effects on patients who were actually treated with total thyroidectomy, and ATU was the average treatment effect on the untreated group that explicitly evaluated the effects on those patients actually treated with hemithyroidectomy. $Y(1)$ was the QoL score when the patient was treated with a total thyroidectomy, $Y(0)$ was the QoL score when the patient was treated with a hemithyroidectomy, and T was the type of surgery: $T = 1$ when the patient was treated with a total thyroidectomy and $T = 0$ when the patient was treated with a hemithyroidectomy.

Results and Discussion

Scoring and Model Checking

The patient QoLs were put directly into Equations (1) and (2) for the sum scoring and into Equation (3) and (4) for the domain-based scoring; however, a fit check was required for the IRT-based scoring method (bifactor models).

Therefore, the following fit indices were used to evaluate the bifactor model: root mean square error of approximation (RMSEA); standardized root mean square (SRMSR); comparative fit index (CFI); and the Tucker-Lewis index (TLI). The RMSEA provides an estimate of model fit that is unaffected by mode size, with an RMSEA below 0.06 indicating a good fit,[26] and an RMSEA between 0.06 and 0.08 indicating a mediocre fit.[27] The SRMSR fit index is a global fit measure that reflects the discrepancies between the observed and predicted model covariances, with an SRMSR below 0.08 indicating a good fit[26] and an SRMSR between 0.08 and 0.10 indicating an acceptable fit. A CFI and TLI above 0.95 indicates good model fit[26] and a CFI and TLI above 0.90 indicates acceptable fit.[27]

Two other competing QoL structure models that were derived from theory or previous research were also tested: (1) single factor models in which all items were loaded on one underlying QoL factor (Equation (5)); and (2) a three-factor model with three dimensions: physical burden, mental function, and overall evaluation for EORTC QLQ-C30, and a three-factor model with three dimensions: Physiological feeling, Psychological feeling, and Pain for THYCA-QoL (Equation (6)).

$$\mathrm{logit}(Q_{ij}) = \mu_j + \lambda_j \theta_i + \in_{ij} \tag{5}$$

where Q_{ij} was the response of patient i on item j, μ_j was the intercept for item j, which was listed as the mean as this was typically what it became, λ_j was the factor loading of item j on the only factor, θ_i was the latent factor score for patient i, and $\in_{ij}$ was the residual for patient i for item j.

$$\mathrm{logit}(Q_{ij}) = \mu_j + \lambda_j^p \theta_i^p + \lambda_j^m \theta_i^m + \lambda_j^o \theta_i^o + \in_{ij} \tag{6}$$

where Q_{ij} was the response of patient i on item j, μ_j was the intercept for item j, which was listed as the mean as this was typically what it became, λ_j^p, λ_j^m and λ_j^o were the respective factor loading for item j on the Physical burden, Mental function and Overall evaluation factors when using EORTC QLO-C30, or the respective factor loading for item j on the Physiological feeling, Psychological feeling, and Pain when using THYCA-QoL. $\lambda_j^f = 0$ if item j did not load on factor f. θ_i^f is the latent factor score for patient i for factor f, and $\in_{ij}$ was the residual for patient i for item j.

The single-factor model demonstrated a poor fit when using both EORTC QLO-C30 and THYCA-QoL ($\mathrm{RMSEA} = 0.0869(0.0902)$, $\mathrm{SRMSR} = 0.1058(0.0982)$, $\mathrm{TLI} = 0.7976(0.5854)$, $\mathrm{CFI} = 0.8135(0.6287)$), indicating that there was more than one factor for both EORTC QLQ-

C30 and THYCA-QoL, but the three-factor model demonstrated a worse fit (RMSEA = 0.0903(0.0949), SRMSR = 0.2062(0.1834), TLI = 0.7812(0.5412), CFI = 0.0.7985(0.8358)). The refined bifactor models were an improvement over the single- and three-factor models, as indicated by the lower RMSEA (RMSEA = 0.0578(0.0568)), lower SRMSR (SRMSR = 0.0913(0.0772)), higher TLI (TLI = 0.914(0.8358)), and higher CFI (CFI = 0.9245(0.8700)). The RMSEA value indicated a good fit, with the other fit indices indicating an acceptable fit for EORTC QLO-C30. The RMSEA and SRMSR values indicated a good fit for THYCA-QoL (shown in Table 3).

Study Profile and Covariate Balance

Before the matching, the cohort was 150 DTC patients, 59 (39.3%) who had been treated with a total thyroidectomy, and 91 (60.7%) who had been treated with a hemithyroidectomy. When the 59 total thyroidectomy (treated group) patients were propensity score matched with the 91 hemithyroidectomy patients with a radius matching of 0.01 calipers, 133 patients were matched (54 in the treated group and 79 in the untreated group). These 133 participants were therefore considered in the propensity score matched analysis and the other 17 participants (five in the treated group and twelve in the untreated group) were excluded because they lacked good propensity score matches.

As most observations were on support, this meant that the common support or overlap conditions were fulfilled (Figure 3), with the individual level characteristic differences being smaller after the propensity score matching (Table 4). Prior to the propensity score matching, the two groups were significantly different in terms of marital status ($p = 0.046$), with the proportion of married participants being higher in the treated group than in the control group (98.3% vs 86.8%). After the propensity score matching, there were no statistically significant differences in terms of age, income level, employment status, means of payment, and marital status in the matched cohort between the patients treated with total thyroidectomy and the patients treated with a hemithyroidectomy.

The standardized mean deviations (SMDs) for age, income, employment, and payment group were all less than 0.10 and for the marital status were zero, which signified that the matching covariates for the treated and control groups were well balanced (Table 5). The standard bias across the matched cohort covariates were closer to 0 than that of the unmatched cohort, which indicated that the SMDs for all covariates had been obviously reduced after the matching (Figure 4). Further, compared with the unmatched group, the Pseudo R^2, LR χ^2 MeanBias, and the MedBias in the matched group were all significantly reduced, with the Prob>χ^2 being very close to 1 (0.992, as shown in Table 6). All these results illustrated that a good balance had been achieved in the propensity score matched cohort.

Impact of Surgery Type on QoL

Overall QoL Differences Between the Two Surgery Types

When a non PSM cohort was used and the patient QoL was measured by the sum scoring or domain-based scoring methods regardless of using questionnaire EORTC QLO-C30 or THYCA-QoL, the surgery type was found to have no significant impact on the QoL of the DTC survivors (Table 7). Specifically, before being adjusted for the covariates, there were no statistically significant differences between the QoL of the DTC patients treated with total thyroidectomy and QoL of the DTC patients treated with hemithyroidectomy ($p = 0.766$ for the sum scoring using EORTC QLO-C30,

Table 3 Fit Indices for the Bifactor QoL Model

Model	df	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
EORTC QLQ-C30							
Single-factor model	351	0.0869	0.0780	0.0952	0.1058	0.7976	0.8135
Three-factor model	351	0.0903	0.0815	0.0985	0.2062	0.7812	0.7985
Bifactor model	321	0.0559	0.0445	0.0662	0.0913	0.9163	0.9295
THYCA QoL							
Single-factor model	206	0.0902	0.0788	0.1011	0.0982	0.5854	0.6287
Three-factor model	206	0.0949	0.0837	0.1056	0.1834	0.5412	0.5890
Bifactor model	182	0.0568	0.0417	0.0704	0.0772	0.8358	0.8700

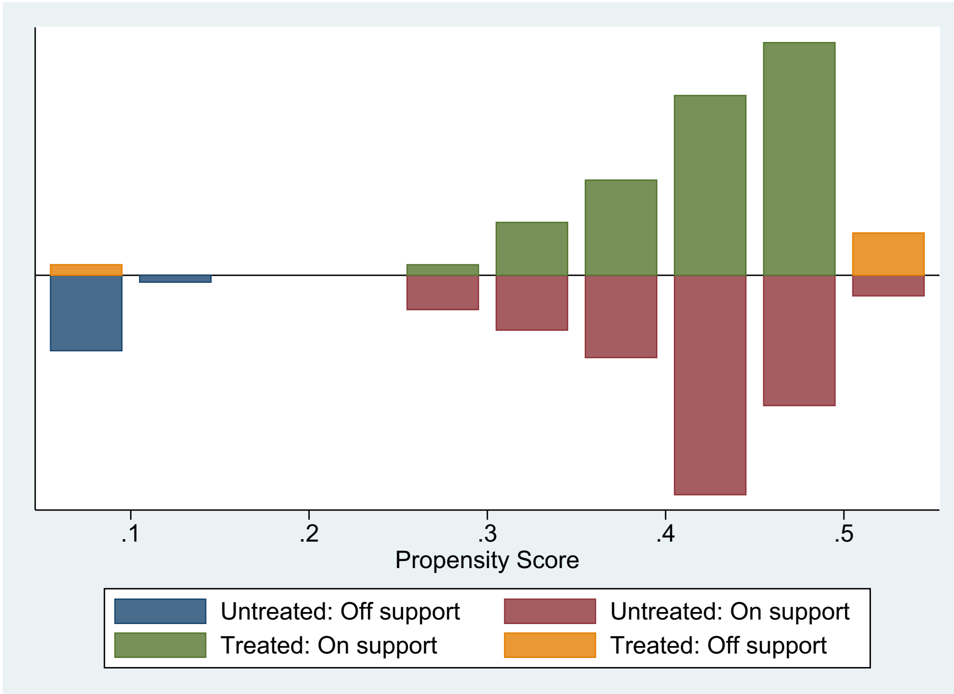

Figure 3 Common support region of propensity scores.

$p = 0.734$ for the sum scoring using THYCA-QoL, $p = 0.602$ for the domain-based scoring using EORTC QLO-C30 and $p = 0.523$ for the domain-based scoring using THYCA-QoL). After being adjusted for the covariates in the multivariate analysis, the same conclusion as in the univariate analysis was obtained ($p = 0.569$ for the sum scoring

Table 4 Participant Characteristics Before and After the Propensity Score Matching

	Before PS Matching				After PS Matching			
	Untreated No. (%)	**Treated No. (%)**	**Total N**	***p* value**	**Untreated No. (%)**	**Treated No. (%)**	**Total N**	***p* value**
Age/years	40 (10.2, 23, 67)	41 (10.2, 24, 67)	150	0.844	41 (9.42, 27, 67)	40 (8.5, 24, 61)	133	0.821
Income level								
≤3000	19 (20.9)	10 (16.9)	29	0.446	17 (21.5)	9 (16.7)	26	0.521
3000~5000	17 (18.7)	11 (18.6)	28		15 (19)	11 (20)	26	
≥5000	55 (60.4)	38 (64.4)	93		47 (59.5)	34 (63)	81	
Employed status								
No	36 (39.6)	26 (44.1)	62	0.700	33 (42)	21 (39)	54	0.915
Yes	55 (60.4)	33 (55.9)	88		46 (58)	33 (61)	79	
Marital status								
No	12 (13.2)	1 (1.7)	13	0.046	0 (0)	0 (0)	0	–
Yes	79 (86.8)	58 (98.3)	137		79 (100)	54 (100)	133	
Means of payment								
Own expense	4 (4.4)	5 (8.5)	9	0.256	4 (5.1)	4 (7.4)	8	0.231
Medical Insurance (Urban)	67 (73.6)	41 (69.5)	108		57 (72)	38 (70)	95	
Medical Insurance (Rural)	4 (4.4)	7 (11.9)	11		3 (3.8)	7 (13)	10	
Business Insurance	0 (0)	1 (1.7)	1		0 (0)	0 (0)	0	
Other	16 (17.6)	5 (8.5)	21		15 (19)	5 (9.3)	20	
Total	91	59	150		79	54	133	

Note: Continuous data are presented as mean (standard deviations, minimum, maximum), and categorical data are presented as frequencies with percentages.

Table 5 Covariate Balance Before and After Propensity Score Matching

Covariates	Before PS Matching			After PS Matching		
	Mean in Treated Group	Mean in Untreated Group	SMD (%)	Mean in Treated Group	Mean in Untreated Group	SMD (%)
Age/years	41.407	39.846	15.3	40.463	40.958	−4.9
Income	2.4746	2.3956	9.9	2.463	2.4523	1.3
Employed	0.5593	0.6044	−9.1	0.611	0.5664	9.0
Marital	0.9831	0.8681	44.6	1	1	0.0
Payment	2.322	2.5275	−18.9	2.333	2.358	−2.3

using EORTC QLO-C30, $p = 0.830$ for the sum scoring using THYCA-QoL, $p = 0.499$ for the domain-based scoring using EORTC QLO-C30 and $p = 0.632$ for the domain-based scoring using THYCA-QoL). When a non-PSM cohort was used and the QoL of the DTC patients was measured using the IRT-based scoring and using questionnaire THYCA-QoL, the surgery type was also found to have no significant impact on the QoL of the DTC survivors (Table 7). Specifically, no matter before being adjusted for the covariates or after being adjusted for the covariates in the multivariate analysis, there were no statistically significant differences between the QoL of the DTC patients treated with total thyroidectomy and QoL of the DTC patients treated with hemithyroidectomy ($p = 0.852$ before being adjusted for covariates, $p = 0.713$ after being adjusted for covariates in the multivariate analysis).

When a non-PSM cohort was used and the QoL of the DTC patients was measured using the IRT-based scoring and using questionnaire EORTC QLO-C30, however, there was a statistically significant difference between the QoL of the DTC patients treated with a total thyroidectomy and those treated with a hemithyroidectomy ($p = 0.076$ before being adjusted for covariates, $p = 0.027$ after being adjusted for covariates in the multivariate analysis, as shown in Table 7), which suggested that the QoL of the DTC patients treated with hemithyroidectomy was higher than for the DRC patients treated with a total thyroidectomy, which was in line with the results in Nickel et al[19] and Hauch et al.[20]

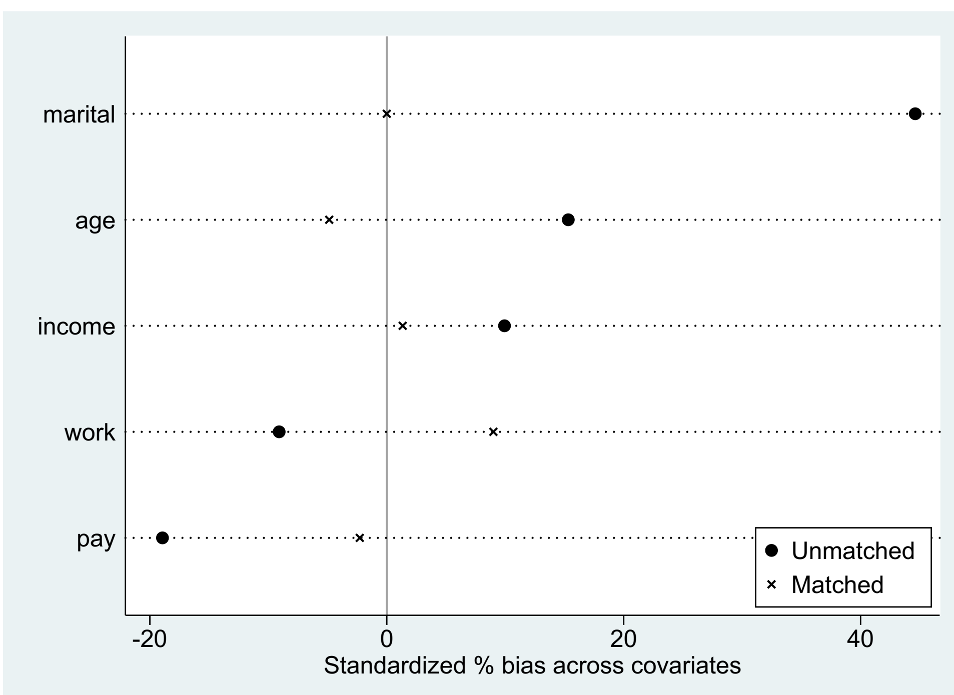

Figure 4 Standard mean deviation of covariates.

Table 6 Summary of the Distribution of the Bias

Sample	Pseudo R^2	LR X^2	Prob>χ^2	MeanBias (SMD)	MedBias
Unmatched	0.045	9.11	0.105	19.6	15.3
Matched	0.002	0.27	0.992	3.5	2.3

Table 7 Impact of Surgery Type on the QoL in DTC Patients Before the Propensity Matched Analysis

		EORTC QLQ-C30			THYCA-QoL		
		Coef.	Std. Err.	p value	Coef.	Std. Err.	p value
Sum scoring	Univariate analysis	−0.448	1.506	0.776	−0.416	1.226	0.734
	Multivariate analysis	−0.904	1.587	0.569	−0.272	1.267	0.830
Domain-based scoring	Univariate analysis	−0.824	1.575	0.602	−1.083	1.693	0.523
	Multivariate analysis	−1.135	1.676	0.499	−0.841	1.752	0.632
IRT-based scoring	Univariate analysis	−0.401	0.225	0.076	0.044	0.233	0.852
	Multivariate analysis	−0.512	0.229	0.027	0.091	0.247	0.713

When QoL of patients was measured by using the questionnaire EORTC QLO-C30 and the PSM cohort was used, there was no statistically significant differences between the QoL of patients treated with a total thyroidectomy and random patients treated with a hemithyroidectomy regardless of the scoring method (Table 8). The expected impact on the QoL of DTC patients who were randomly assigned to be treated with either a total thyroidectomy or a hemithyroidectomy were -0.4 (ATE $= -0.4$, $p = 0.834$), -0.491 (ATE $= -0.491$, $p = 0.807$) and -0.442 (ATE $= -0.442$, $p = 0.114$) when the QoL was respectively calculated using sum scoring, domain-based scoring, and IRT-based scoring. The differences between the QoL of the patients treated with a total thyroidectomy and those not treated with a total thyroidectomy were -0.655 (ATT $= -0.655$, $p = 0.765$), -0.703 (ATT $= -0.703$, $p = 0.766$) and -0.480 (ATT $= -0.480$, $p = 0.119$) when the QoL was respectively calculated using the sum scoring, domain-based scoring, and IRT-based scoring. The differences between the QoL of patients treated with hemithyroidectomy and those not treated with hemithyroidectomy were -0.226 (ATU $= -0.226$, $P = 0.914$), -0.347 (ATU $= -0.347$, $P = 0.874$) and -0.416 (ATU $= -0.416$, $P = 0.174$) when the QoL was respectively calculated using the sum scoring, domain-based scoring, and IRT-based scoring. Besides, the same conclusion was detected when QoL of patients was measured by using the questionnaire THYCA-QoL and the PSM cohort was used. There was no statistically significant differences between the QoL of patients treated

Table 8 Impact of Surgery Type on the QoL in DTC Patients After the Propensity Matched Analysis

		EORTC QLQ-C30			THYCA-QoL		
		Coef.	Std. Err.	p value	Coef.	Std. Err.	p value
Sum scoring	PSM analysis (ATE)	−0.400	1.906	0.834	−0.827	1.516	0.586
	PSM analysis (ATT)	−0.655	2.191	0.765	−0.817	1.739	0.639
	PSM analysis (ATU)	−0.226	2.107	0.914	−0.834	1.643	0.612
Domain-based scoring	PSM analysis (ATE)	−0.491	2.015	0.807	−1.692	2.036	0.406
	PSM analysis (ATT)	−0.703	2.358	0.766	−1.788	2.369	0.450
	PSM analysis (ATU)	−0.347	2.193	0.874	−1.627	2.312	0.482
IRT-based scoring	PSM analysis (ATE)	−0.442	0.279	0.114	−0.032	0.278	0.908
	PSM analysis (ATT)	−0.480	0.308	0.119	−0.062	0.319	0.845
	PSM analysis (ATU)	−0.416	0.306	0.174	−0.011	0.307	0.971

with a total thyroidectomy and random patients treated with a hemithyroidectomy regardless of the scoring method with all *p* values for ATE, ATT or ATU over than 0.1 (Table 8).

Domains of QoL Differences Between the Two Surgery Types

To make a solid conclusion and learn more about the QoL of DTC survivors, this paper analyzed domains of QoL differences between patients treated with a hemithyroidectomy and those treated with a total thyroidectomy. The domain divisions were shown in Tables 1 and 2. Differences on 15 domains of EORTC QLO-C30 and 13 domains of THYCA-QoL were explored after the thyroidectomy patients were propensity score matched with the hemithyroidectomy patients. When QoL of patients was evaluated by EORTC QLO-C30, results showed that no matter which surgery was chosen by patients, they had similar function, symptom and global QoL. Specifically, for all domains of the EORTC QLO-C30 questionnaire, there were no statistically significant difference between patients treated with a hemithyroidectomy and those treated with a total thyroidectomy ($p>0.1$, Table 9). Meanwhile, when QoL of patients was evaluated by THYCA-QoL, there was no significant difference between every domain of QoL of patients treated with hemithyroidectomy and those treated with total thyroidectomy ($p>0.1$, Table 10). Therefore, based on the results from the PSM estimations, a conclusion was drawn that the QoL of DTC patients treated with total thyroidectomy and those treated with hemithyroidectomy were basically the same.

Conclusion

Because of concerns as to whether surgery type has an impact on the QoL of DTC patients, this study used a propensity score matching method to study the association relationships to reduce the selection bias common in traditional methods such as one-way analysis of variance and covariance analysis and reveal the causality between surgery type and QoL in DTC patients. Based on the EORTC QLQ-C30 and the THYCA-QoL, three scoring methods (sum scoring, domain-based scoring and IRT-based scoring) were developed to

Table 9 Impact of Surgery Type on the Domain of EORTC QLQ-C30 in DTC Patients After the Propensity Matched Analysis

	Domain	Coef.	Std. Err.	*p* value	Domain	Coef.	Std. Err.	*p* value
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	FA	3.613 3.873 3.436	3.838 4.293 4.186	0.346 0.367 0.412	LA	−0.351 −0.079 −0.538	1.268 1.368 1.436	0.782 0.954 0.708
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	NV	−0.500 −0.518 −0.488	2.046 2.539 2.116	0.807 0.939 0.817	PF	−0.118 −0.147 −0.098	1.725 1.917 1.909	0.946 0.939 0.959
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	PA	1.450 1.885 1.154	3.167 3.741 3.519	0.647 0.614 0.743	EF	2.821 2.556 3.001	3.238 3.593 3.634	0.384 0.477 0.409
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	DY	3.849 3.462 4.114	4.169 4.669 4.684	0.356 0.458 0.38	CF	0.962 0.599 1.21	3.712 4.466 3.882	0.796 0.893 0.755
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	DI	−1.027 −1.519 −0.691	3.289 3.494 3.685	0.755 0.664 0.851	SF	−1.384 −3.361 −0.033	3.73 4.193 4.06	0.711 0.423 0.994
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	CO	−4.420 −3.193 −5.259	3.889 4.554 4.136	0.256 0.483 0.204	RF	−4.621 −4.433 −4.750	3.231 3.523 3.593	0.153 0.208 0.186
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	FD	1.318 1.313 1.321	2.929 3.223 3.34	0.653 0.684 0.693	GQ	−1.245 −0.681 0.053	3.939 4.184 4.565	0.95 0.871 0.991
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	IN	−0.054 0.687 0.379	2.703 2.957 3.108	0.984 0.816 0.9903				

Table 10 Impact of Surgery Type on the Domain of THYCA-QoL in DTC Patients After the Propensity Matched Analysis

	Domain	Coef.	Std. Err.	*p* value	Domain	Coef.	Std. Err.	*p* value
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	NM	1.67 0.874 2.214	2.912 3.174 3.295	0.566 0.783 0.502	SC	7.536 8.5 6.876	4.97 5.288 5.69	0.129 0.108 0.227
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	VO	6.078 6.013 6.123	4.444 5.085 4.741	0.171 0.237 0.197	CH	3.062 4.511 2.071	5.33 5.563 6.42	0.566 0.417 0.747
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	CT	−1.654 −2.341 −1.185	3.936 4.802 4.098	0.674 0.626 0.772	THF	4.928 5.359 4.633	4.228 4.778 4.715	0.244 0.262 0.326
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	ST	1.478 1.254 1.632	4.133 4.34 4.626	0.721 0.778 0.778	GW	−2.366 −2.491 −2.280	5.574 6.106 6.308	0.671 0.683 0.718
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	TM	−2.002 −1.779 −2.154	3.674 4.141 3.993	0.586 0.667 0.59	HA	1.217 0.977 1.381	4.181 4.836 4.414	0.771 0.84 0.754
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	PC	0.545 0.551 0.54	2.967 3.214 3.383	0.854 0.864 0.873	SXI	−3.377 −3.558 −3.253	4.168 4.608 4.53	0.418 0.44 0.473
PSM analysis (ATE) PSM analysis (ATT) PSM analysis (ATU)	SE	−2.425 −2.689 −2.245	2.866 3.449 3.038	0.397 0.436 0.46				

measure the latent variable, the QoL of DTC patients, with the IRT-based scoring method (the bifactor model) showing an adequate model-data fit for the DTC patients.

When a non PSM cohort was used and the IRT-based scoring was applied for the EORTC QLQ-C30, the results indicated that hemithyroidectomies offered better QoL outcomes than total thyroidectomies for DTC patients; however, when the sum scoring and the domain-based scoring were applied, no differences were found. When the PSM cohort was used, regardless of the scoring method for EORTC QLQ-C30, no QoL differences were found for the two surgery types. Because of the confounder conditioning and observation selection bias control, the PSM cohort results made more sense. Besides, regardless of non PSM cohort or PSM cohort was used and no matter which scoring method was applied for THYCA-QoL, results showed that the QoL of DTC survivors treated with the two surgery types are similar.

This study, which was conducted in November and December 2019, sought to develop a reference framework for similar effect evaluations; however, the original plan to continue this investigation in January 2020 was suspended due to the coronavirus pandemic. Therefore, as the sample may not have been rich enough, further studies are planned to consider more covariates to detect the QoL causalities for DTC patients in other areas such as postoperative care and psychotherapy.

Ethical Approval

This study was approved by the Institutional Review Board of Sichuan Cancer Hospital and Institutional Ethics Committee and performed according to the ICH GCP principle. And the study design complies with the Declaration of Helsinki ethical standards.

Informed Consent

We obtained written informed consent from all of the individual participants included in the study.

Acknowledgments

This study was supported by a grant from the Sichuan Provincial Health Commission [No. 19PJ274].

Disclosure

The authors report no conflicts of interest in this work.

References

1. Davies L, Morris L, Hankey B. Increases in thyroid cancer incidence and mortality. *JAMA*. 2017;318(4):389–390. doi:10.1001/jama.2017.7906
2. Callender GG, Carling T, Christison-Lagay E, Udelsman R. Surgery for thyroid cancer. *Endocrinol Metab Clinics*. 2014;43(2):443–458. doi:10.1016/j.ecl.2014.02.011
3. McIntyre C, Jacques T, Palazzo F, Farnell K, Tolley N. Quality of life in differentiated thyroid cancer. *Int J Surg*. 2018;50:133–136. doi:10.1016/j.ijsu.2017.12.014
4. Sotelo JL, Musselman D, Nemeroff C. The biology of depression in cancer and the relationship between depression and cancer progression. *Int Rev Psychiatry*. 2014;26(1):16–30. doi:10.3109/09540261.2013.875891
5. Applewhite MK, James BC, Kaplan SP, et al. Quality of life in thyroid cancer is similar to that of other cancers with worse survival. *World J Surg*. 2016;40(3):551–561. doi:10.1007/s00268-015-3300-5
6. Aschebrook-Kilfoy B, James B, Nagar S, et al. Risk factors for decreased quality of life in thyroid cancer survivors: initial findings from the North American thyroid cancer survivorship study. *Thyroid*. 2015;25(12):1313–1321. doi:10.1089/thy.2015.0098
7. Goswami S, Mongelli M, Peipert BJ, Helenowski I, Yount SE, Sturgeon C. Benchmarking health-related quality of life in thyroid cancer versus other cancers and United States normative data. *Surgery*. 2018;164(5):986–992. doi:10.1016/j.surg.2018.06.042
8. Hoftijzer HC, Heemstra KA, Corssmit EP, van der Klaauw AA, Romijn JA, Smit JW. Quality of life in cured patients with differentiated thyroid carcinoma. *J Clin Endocrinol Metab*. 2008;93(1):200–203. doi:10.1210/jc.2007-1203
9. Goswami S, Peipert BJ, Mongelli MN, et al. Clinical factors associated with worse quality-of-life scores in United States thyroid cancer survivors. *Surgery*. 2019;166(1):69–74. doi:10.1016/j.surg.2019.01.034
10. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39(1):33–38.
11. Nordin K, Steel J, Hoffman K, Glimelius B. Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. *Br J Cancer*. 2001;85(9):1265–1272. doi:10.1054/bjoc.2001.2046
12. Hinz A, Einenkel J, Briest S, Stolzenburg JU, Papsdorf K, Singer S. Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? *Eur J Cancer Care (Engl)*. 2012;21(5):677–683. doi:10.1111/j.1365-2354.2012.01367.x
13. Husson O, Haak HR, Mols F, et al. Development of a disease-specific health-related quality of life questionnaire (THYCA-QoL) for thyroid cancer survivors. *Acta Oncol*. 2013;52(2):447–454. doi:10.3109/0284186X.2012.718445
14. Efficace F, Cottone F, Sommer K, et al. Validation of the European organisation for research and treatment of cancer quality of life questionnaire Core 30 summary score in patients with hematologic malignancies. *Value Health*. 2019;22(11):1303–1310. doi:10.1016/j.jval.2019.06.004
15. Giesinger JM, Kieffer JM, Fayers PM, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. *J Clin Epidemiol*. 2016;69:79–88. doi:10.1016/j.jclinepi.2015.08.007
16. Embretson SE, Reise SP. *Item Response Theory*. Psychology Press; 2013.
17. DeMars CE. Application of the bi-factor multidimensional item response theory model to testlet-based tests. *J Educ Measure*. 2006;43(2):145–168. doi:10.1111/j.1745-3984.2006.00010.x
18. Cai L, Yang JS, Hansen M. Generalized full-information item bifactor analysis. *Psychol Methods*. 2011;16(3):221. doi:10.1037/a0023350
19. Nickel B, Tan T, Cvejic E, et al. Health-related quality of life after diagnosis and treatment of differentiated thyroid cancer and association with type of surgical treatment. *JAMA Otolaryngol Head Neck Surg*. 2019;145(3):231–238. doi:10.1001/jamaoto.2018.3870
20. Hauch A, Al-Qurayshi Z, Randolph G, Kandil E. Total thyroidectomy is associated with increased risk of complications for low-and high-volume surgeons. *Ann Surg Oncol*. 2014;21(12):3844–3852. doi:10.1245/s10434-014-3846-8
21. Bongers PJ, Greenberg CA, Hsiao R, et al. Differences in long-term quality of life between hemithyroidectomy and total thyroidectomy in patients treated for low-risk differentiated thyroid carcinoma. *Surgery*. 2020;167(1):94–101. doi:10.1016/j.surg.2019.04.060
22. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv*. 2008;22(1):31–72. doi:10.1111/j.1467-6419.2007.00527.x
23. Aaronson NK, Ahmedzai S, Bergman B, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Nat Cancer Instit*. 1993;85(5):365–376. doi:10.1093/jnci/85.5.365
24. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw*. 2012;48(1):1–29. doi:10.18637/jss.v048.i06
25. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265–2281. doi:10.1002/(SICI)1097-0258-(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B
26. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1–55. doi:10.1080/10705519909540118
27. Brown TA. *Confirmatory Factor Analysis for Applied Research*. Guilford publications; 2015.
28. Haraj NE, Bouri H, El Aziz S, Nani S, Habti N. Evaluation of the quality of life in patients followed for differentiated cancer of the thyroid. *Annales D'endocrinologie*. 2018;80(1):26–31. doi:10.1016/j.ando.2018.01.003.