

Patient-Reported Outcomes Following Total Thyroidectomy for Graves' Disease

Alexander H. Gunn,^{1,2} Nicholas Frisco,¹ Samantha M. Thomas,^{3,4} Michael T. Stang,¹
Randall P. Scheri,¹ and Hadiza S. Kazaure¹

Background: Graves' disease accounts for ~80% of all cases of hyperthyroidism and is associated with significant morbidity and decreased quality of life. Understanding the association of total thyroidectomy with patient-reported quality-of-life and thyroid-specific symptoms is critical to shared decision-making and high-quality care. We estimate the change in patient-reported outcomes (PROs) before and after surgery for patients with Graves' disease to inform the expectations of patients and their physicians.

Methods: PROs using the MD Anderson Symptom Inventory (MDASI) validated questionnaire were collected prospectively from adult patients with Graves' disease from January 1, 2015, to November 20, 2020, on a longitudinal basis. Survey responses were categorized as before surgery (≤ 120 days), short term after surgery (< 30 days; ST), and long term after surgery (≥ 30 days; LT). Negative binomial regression was used to estimate the association of select covariates with PROs.

Results: Eighty-five patients with Graves' disease were included. The majority were female (83.5%); 47.1% were non-Hispanic white and 35.3% were non-Hispanic black. The median thyrotropin (TSH) value before surgery was 0.05, which increased to 0.82 in ST and 1.57 in LT. In bivariate analysis, the Total Symptom Burden Score, a composite of all patient-reported burden, significantly reduced shortly after surgery (before surgery mean of 56.88 vs. ST 39.60, $p < 0.001$), demonstrating improvement in PROs. Furthermore, both the Thyroid Symptoms Score, including patient-reported thermoregulation, palpitations, and dysphagia, and the Quality-of-Life Symptom Score improved in ST and LT (thyroid symptoms, before surgery 13.88 vs. ST 8.62 and LT 7.29; quality of life, before surgery 16.16 vs. ST 9.14 and LT 10.04, all $p < 0.05$). After multivariate adjustment, the patient-reported burden in the Thyroid Symptom Score and the Quality-of-Life Symptom Score exhibited reduction in ST (thyroid symptoms, rate ratio [RR] 0.55, confidence interval [CI]: 0.42–0.72; quality of life, RR 0.57, CI: 0.40–0.81) and LT (thyroid symptoms, RR 0.59, CI: 0.44–0.79; quality of Life, RR 0.43, CI: 0.28–0.65).

Conclusions: Quality of life and thyroid-specific symptoms of Graves' patients improved significantly from their baseline before surgery to both shortly after and longer after surgery. This work can be used to guide clinicians and patients with Graves' disease on the expected outcomes following total thyroidectomy.

Keywords: Graves' disease, patient-reported outcomes, quality of life, thyroidectomy

Introduction

GRAVES' DISEASE ACCOUNTS for nearly 80% of all cases of hyperthyroidism (1), and causes significant morbidity and decreased quality of life (2–4). Treatment options include medications, radioactive iodine ablation (RAI), and surgical management; the treatment choice is based on patient preference as well as the clinical presentation (2,3,5).

Surveys indicate that RAI is the dominant treatment modality in the United States, accounting for more than half of the treatment, however, the use of antithyroid medication is increasing (6,7). Recent analysis showed that even though surgery is the most definitive treatment (99%), it is only used as first-line treatment for 6% of patients and subsequently in 9% of patients with first-line treatment failure, and 3% with second-line treatment failure (7,8). With the use of shared

¹Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA.

²Duke-Margolis Center for Health Policy, Durham, North Carolina, USA.

³Duke University Cancer Institute, Durham, North Carolina, USA.

⁴Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA.

decision-making in treatment decisions with Graves' patients, understanding the effect of surgical treatment on patient-reported quality of life, core symptoms, and thyroid-specific symptoms is critical to providing high-quality care.

Patient-reported outcomes (PROs) represent a powerful tool to help physicians and their patients understand the dynamic changes in symptoms and quality of life due to surgical treatment. For Graves' disease in particular, patients may present with multiple nonspecific symptoms, such as fatigue, mood dysregulation, or heat sensitivity, and would like to know what changes in these symptoms they can expect as they undergo treatment. Previous research on PROs in Graves' disease has been limited by several key factors. First, baseline values and short-term time points with relation to thyroidectomy are largely omitted, limiting the understanding of short-term changes following surgery (2,3,9–17). Second, many prior studies with Graves' patients rely on mailed questionnaires with varied response rates (2,3,9–17) or on retrospective PROs, which are at risk for recall bias (18). Last, peer-reviewed literature regarding PROs and Graves' disease is predominantly from studies based outside of North America. Published studies include patients from Europe (2,3,9,10,12,14,15,17,19), Brazil (13), and New Zealand (16), where patients have different social, cultural, and economic contexts. Consequently, current evidence that is applicable to the patients in the United States is not available. We aim to fill this gap in the literature by assessing the quality of life, core symptoms, and thyroid-specific symptoms of patients before, in the short term, and in the long term following thyroid surgery. We estimate the change in PROs in response to surgical treatment and characterize the dynamics of these responses to better inform the expectations of patients with Graves' disease and their physicians.

Methods

Patients were included if they were age ≥ 18 years, had a diagnosis of Graves' disease, and underwent total thyroidectomy. Patients who did not have data for at least one MD Anderson Symptom Inventory (MDASI) survey ≤ 120 days before surgery or any time after surgery were excluded.

Data were extracted from a novel thyroid data platform created by the Private Diagnostic Clinic Outcome Research Team (PORT-Thyroid database) developed in conjunction with the Duke Endocrine Neoplasia Laboratory. The database collected prospective PROs with the MDASI, which includes a thyroid symptom module that has previously been validated for use in thyroid patients, beginning from January 1, 2015, for all patients with thyroid disease seen at the Duke University Medical Center (Supplementary Appendix SA1) (20). In addition, patient characteristics, such as demographics, disease, and treatment information, were collected over the same period. The data were exported on November 20, 2020, and encompass patient visits that occurred between January 1, 2015, and November 19, 2020. Patient factors collected included age, race/ethnicity, sex, marital status, employment status, religion, and insurance. Disease and treatment factors included body mass index, medical history, and thyrotropin (TSH). Timing of survey responses was categorized as before surgery (≤ 120 days), short term after surgery (< 30 days; ST), and long term after surgery (≥ 30 days after; LT). Survey responses that occurred > 120 days before

surgery were excluded. If multiple responses were recorded during a specific time interval, the mean score was computed and utilized throughout this analysis. Therefore, patients could have at most three score entries for each survey item. TSH values were collected via chart review and categorized as before surgery, ST, or LT. If multiple TSH measures were recorded during a specific time interval, the mean value during the specific interval was estimated and utilized.

The tool assesses common hyperthyroid patient symptoms, such as heat sensitivity, palpitations, dysphagia, and diarrhea, among others. The primary outcome of this study is the Total Symptom Burden Score, the composite score of the MDASI tool (20,21), which is a synthesis of the following variables: "pain," "patient-reported fatigue," "nausea," "sleep," "distress," "shortness of breath," "memory," "drowsiness," "appetite," "dry mouth," "sadness," "vomiting," "numbness," "hoarseness," "feeling hot," "racing heartbeat," "feeling cold," "difficulty swallowing," "diarrhea or loose stools," "activity," "ability to work," "relationships with others," "enjoyment of life," "walking," and "mood." To precisely describe the quality-of-life changes in relation to surgical management, analyses assessed each of these symptoms as individual outcomes and as composite categories: (i) Core Symptom Score, (ii) Quality-of-Life Score, and (iii) Thyroid Symptom Score. The Core Symptom Score includes patient-reported "pain," "fatigue," "nausea," "sleep," "distress," "shortness of breath," "memory," "drowsiness," "appetite," "dry mouth," "sadness," "vomiting," and "numbness." The Quality-of-Life Score includes patient-reported "activity," "ability to work," "relationships with others," "enjoyment of life," "walking," and "mood." The Thyroid Symptom Score includes patient-reported "hoarseness," "feeling hot," "racing heartbeat," "feeling cold," "difficulty swallowing," and "diarrhea or loose stools." Notably, higher MDASI scores correspond with greater symptom burden and lower MDASI scores correspond with less symptom burden, and therefore, an improved symptom score is a reduction in score. Due to a data collection error, one component of the Quality-of-Life Score, "work (including work around the house)," was not reported universally for all patients, however, this score was collected and incorporated into the Quality-of-Life Score and Total Score. Therefore, a score for this field was estimated for all patients as the total Quality-of-Life score minus the sum of the remaining five fields used to calculate the composite Quality-of-Life Score. This estimated score is reported throughout.

Patient characteristics were summarized with N (%) for categorical variables and median (interquartile range, IQR) for continuous variables (Table 1). MDASI survey responses were summarized as mean with standard deviation by timing of the MDASI survey responses (Table 2). Differences between time groups were tested using pairwise nonparametric paired Wilcoxon signed-rank tests.

Negative binomial regression was used to estimate the association of time point with the subscale and total scores, after adjustment for TSH, age, sex, race/ethnicity, obesity, hypertension, marital status, and insurance type. TSH was included in all models as a repeated measure. The distributions of the subscale scores and total score were examined and several regression methods were tested to determine which method would be the best fit to the data. Negative

TABLE 1. BASELINE DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF PATIENTS WITH GRAVES' DISEASE

Characteristic	Patients (N = 85)
Age at baseline—median (IQR)	42 (29–54)
Sex	
Female	71 (83.5%)
Male	14 (16.5%)
Race/ethnicity	
Non-Hispanic white	40 (47.1%)
Non-Hispanic black	30 (35.3%)
Non-Hispanic other or multiple races	4 (4.7%)
Hispanic	6 (7.1%)
Marital status	
Married or life partner	42 (49.4%)
Not married and no life partner	40 (47.1%)
Primary payer	
Private	22 (25.9%)
Government	36 (42.4%)
Other	5 (5.9%)
Unreported	22 (25.9%)
BMI	
Underweight (<18.5)	2 (2.4%)
Normal weight (18.5–24.9)	10 (11.8%)
Overweight (25.0–29.9)	14 (16.5%)
Obese (30.0–39.9)	17 (20%)
Morbidly obese (40.0+)	16 (18.8%)
Medical history	
Hypertension	15 (17.6%)
Anemia	3 (3.5%)
Diabetes	2 (2.4%)
Kidney disease	2 (2.4%)
Congestive heart failure	1 (1.2%)
Atrial fibrillation	1 (1.2%)
Dyspnea	1 (1.2%)
Peripheral vascular disease	1 (1.2%)
Surgery type	
Total thyroidectomy	82 (96.5%)
Total with limited neck dissection	1 (1.2%)
Substernal thyroidectomy (transcervical)	2 (2.4%)
Thyroid profile (TSH)—median (IQR)	
Before surgery	0.05 (0.01–1.76)
Short term after surgery	0.82 (0.14–4.82)
Long term after surgery	1.57 (0.51–10.08)
Days between measurement and surgery—median (IQR)	
Before surgery	38 (29–53)
Short term after surgery	13 (10–15)
Long term after surgery	205 (94–548)

Data presented as *N* (%) unless otherwise specified. Percentages may not add up to 100 due to rounding or missing values.

BMI, body mass index; IQR, interquartile range; TSH, thyrotropin.

binomial regression was selected based on the non-normal distributions of the outcomes and the lowest quasi-information criteria score, which measures the goodness of fit of a model to the data, with lower values indicating a better fit (22). All regression models were built in the generalized estimating equation framework and included an autoregressive covariance structure to account for the correlation of

the repeated measures for each patient. Only observations with complete data for all covariates were included for each regression.

To quantify potential differences between patients with differing data availability, an additional unadjusted analysis was conducted to compare those with data for all three time points versus those with data for the presurgery and one postsurgery time point versus those with data for the short-term and long-term time points versus those with data for only one time point. These results are summarized in Supplementary Table S1.

All statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC). No adjustments were made for multiple comparisons. This study has been approved by the Duke University Institutional Review Board.

Results

Baseline demographic and clinical characteristics

There were 85 patients with Graves' disease who met the study inclusion criteria (Fig. 1). Median age before surgery was 42 years. The majority were female (83.5%) and were either overweight (16.5%) or obese (38.8%). A total of 47.1% of patients identified as non-Hispanic white, 49.4% of patients had a life partner or spouse, and 42.4% of patients used government insurance (Table 1). A total of 81 (95.3%) were on antithyroid medication before surgery, 2 (2.4%) were not, and 2 (2.4%) had incomplete data. One patient (1.2%) had previously been treated with radioactive iodine. A total of 54 patients (63.5%) had a normal free thyroxine (T4) level (as defined by our institutional range of 0.52–1.21) at the before surgery time point. Median TSH before surgery was 0.05 and increased to 0.82 in ST and to 1.57 in LT (Table 1). The median number of days from presurgery PRO measurement to surgery was 38 (IQR 29–53), the median number of days from surgery to short term after surgery PRO measurement was 13 days (IQR 10–15), and the median number of days from surgery to LT PRO measurement was 205 (94–575).

Thyroid-specific symptoms

The Thyroid Symptoms Score improved from a mean of 13.88 before surgery to 8.62 in the ST ($p < 0.001$), and to 7.29 in the LT ($p = 0.02$), indicating less symptom burden in ST and LT (Table 2 and Fig. 2). In bivariate analysis, specific components of the Thyroid Symptoms Score improved significantly from before surgery to ST, including “feeling hot” (4.00–1.92, $p = 0.001$), “racing heartbeat” (3.42–1.35, $p < 0.001$), “diarrhea” (1.36–0.59, $p = 0.04$), and “difficulty swallowing” (2.28–1.65, $p = 0.005$) (Fig. 3). There were no significant changes in hoarseness or feeling cold from before surgery to ST. Patient-reported “racing heartbeat” (3.42–1.39, $p = 0.02$) and “difficulty swallowing” (2.28–0.77, $p = 0.008$) significantly improved from before surgery to LT. “Hoarseness” reported by patients significantly improved from ST to LT (1.49–0.80, $p = 0.001$) (Table 2).

Quality-of-life symptoms

The Quality-of-Life Symptoms Score improved from a mean of 16.16 before surgery to 9.14 in the ST ($p < 0.001$), and to 10.04 in the LT ($p = 0.02$) (Table 2 and Fig. 2).

TABLE 2. PATIENT-REPORTED OUTCOMES OF GRAVES' PATIENTS BEFORE, SHORT TERM AFTER, AND LONG TERM AFTER SURGERY

	Timing			P		
	Before surgery (N=50)	Short term after surgery (N=59)	Long term after surgery (N=32)	Before vs. short term (N=33)	Before vs. long term (N=15)	Short vs. long term (N=19)
Part I: Rate how severe the following symptoms have been in the last 24 hours from 0 (symptom has not been present) to 10 (symptom was as bad as imaginable).						
#1: Your pain at its worst?	2.02 (3.05)	2.33 (2.61)	1.14 (2.20)	0.74	0.70	0.001
#2: Your fatigue (tiredness) at its worst?	4.64 (3.70)	3.34 (2.87)	3.45 (2.87)	0.03	0.83	0.74
#3: Your nausea at its worst?	1.60 (2.77)	0.86 (2.07)	0.50 (1.39)	0.04	0.25	0.94
#4: Your disturbed sleep at its worst?	3.96 (3.74)	2.96 (3.26)	2.57 (2.69)	0.04	0.20	0.81
#5: Your feeling of being distressed (upset) at its worst?	2.43 (3.26)	1.23 (2.39)	2.76 (2.98)	0.04	0.69	0.35
#6: Your shortness of breath at its worst?	2.06 (2.84)	1.27 (2.51)	1.39 (2.54)	0.004	0.31	0.84
#7: Your problem remembering things at its worst?	2.00 (2.99)	1.59 (2.73)	1.91 (2.73)	0.14	0.61	0.20
#8: Your problem with lack of appetite at its worst?	1.41 (2.42)	1.72 (2.75)	1.73 (2.48)	0.36	0.59	0.64
#9: Your feeling drowsy (sleepy) at its worst?	3.04 (3.72)	2.41 (2.77)	2.41 (2.60)	0.47	0.91	0.45
#10: Your having a dry mouth at its worst?	1.23 (2.40)	1.25 (2.29)	1.48 (2.55)	0.37	0.94	0.58
#11: Your feeling sad at its worst?	1.69 (3.02)	0.97 (2.10)	2.03 (2.82)	0.20	1.00	0.37
#12: Your vomiting at its worst?	0.25 (0.81)	0.31 (1.47)	0.16 (0.80)	0.38	0.50	0.25
#13: Your numbness or tingling at its worst?	1.45 (2.48)	1.61 (2.78)	2.03 (3.13)	0.37	0.63	0.56
#14: Your hoarseness at its worst?	1.42 (2.44)	1.49 (2.30)	0.80 (1.86)	0.17	0.16	0.001
#15: Your problem with feeling hot at its worst?	4.00 (3.90)	1.92 (2.95)	2.01 (2.52)	0.001	0.28	0.97
#16: Your problem with racing heartbeat at its worst?	3.42 (3.78)	1.35 (2.39)	1.39 (2.19)	<0.001	0.02	0.31
#17: Your problem with feeling cold at its worst?	1.38 (2.14)	1.61 (2.72)	1.23 (2.48)	0.46	0.57	0.92
#18: Your difficulty swallowing at its worst?	2.28 (3.02)	1.65 (2.18)	0.77 (1.77)	0.005	0.008	0.11
#19: Your diarrhea or loose stools at its worst?	1.36 (2.61)	0.59 (1.61)	1.09 (2.43)	0.04	0.72	0.008
Part II: How much have your symptoms interfered with the following items in the last 24 hours from 0 (did not interfere) to 10 (interfered completely)?						
#20: General activity?	2.96 (3.59)	2.10 (2.71)	1.78 (2.60)	0.01	0.47	0.17
#21: Mood?	2.99 (3.55)	1.46 (2.38)	2.30 (3.21)	0.001	0.05	0.49
#22: Work (including work around the house)?	3.12 (3.60)	1.98 (2.69)	1.80 (2.81)	0.003	0.31	0.33
#23: Relationships with other people?	2.41 (3.46)	1.10 (2.31)	1.58 (2.64)	<0.001	0.25	0.75
#24: Walking?	2.09 (3.38)	1.26 (2.29)	0.73 (1.93)	0.009	0.50	0.44
#25: Enjoyment of life?	2.73 (3.50)	1.25 (2.29)	1.85 (2.99)	0.003	0.31	0.52
Summary Scores						
Core Symptoms Score ^a	27.78 (27.00)	21.84 (23.84)	23.56 (22.60)	0.04	0.45	0.68
QOL Symptoms Score ^b	16.16 (18.80)	9.14 (13.14)	10.04 (14.38)	<0.001	0.03	0.94
Thyroid Symptoms Score ^c	13.88 (12.46)	8.62 (10.70)	7.29 (9.17)	<0.001	0.02	0.63
Total Symptom Burden Score ^d	56.88 (54.53)	39.60 (44.55)	40.90 (42.88)	<0.001	0.08	1.00

Data presented as mean (standard deviation). The presented *p*-values are the Wilcoxon signed-rank test paired *p*-values. Higher scores indicate higher symptom burden and lower scores indicate lower symptom burden.

^aCore Symptoms Score: Sum of MDASI Core questions 1–13, range 0–130.

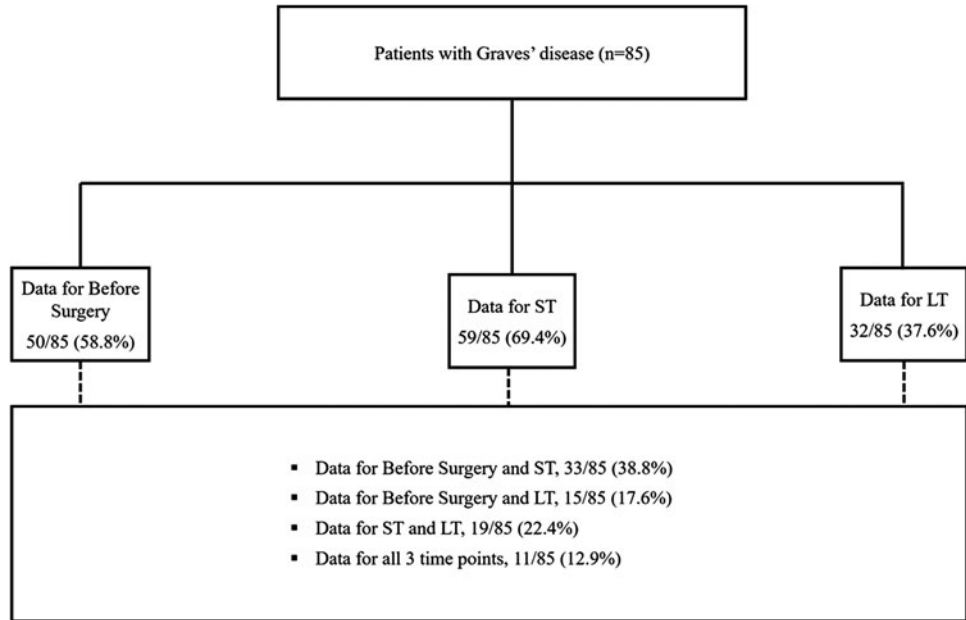
^bQOL Symptoms Score: Sum of MDASI Interference questions 20–25, range 0–60.

^cThyroid Symptoms Score: Sum of MDASI Thyroid-specific questions 14–19, range 0–60.

^dTotal Symptom Burden Score: Sum of all 25 MDASI questions, range 0–250.

MDASI, MD Anderson Symptom Inventory; QOL, quality of life.

FIG. 1. Summary of patient cohort and data collection. LT, Long Term After Surgery; ST, Short Term After Surgery.



In bivariate analysis, all the reported subcomponents of the Quality-of-Life Symptoms Score improved significantly from before surgery to ST, including “general activity” (2.96–2.10, $p=0.01$), “mood” (2.99–1.46, $p=0.001$), “relationship with other people” (2.41–1.10, $p<0.001$), “walking” (2.09–1.26, $p=0.009$), “enjoyment of life” (2.73–1.25, $p=0.003$), and “work (including work around the house)” (3.12–2.98, $p=0.003$) (Fig. 3). Patient-reported “mood” improved significantly from before surgery to LT (2.99–2.30, $p=0.05$). No significant change was observed in “general activity,” “relationship with others,” “walking,” and “enjoyment of life” from before surgery to LT.

Core symptoms

The Core Symptoms Score improved from a mean of 27.78 before surgery to 21.84 in the ST ($p=0.04$), but no improvement was seen from before surgery to LT (23.56, $p=0.45$) (Table 2 and Fig. 2). In bivariate analysis, specific components of the Core Symptom Score improved significantly from before surgery to ST, including “fatigue” (4.64–3.34, $p=0.03$), “nausea” (1.60–0.86, $p=0.04$), “disturbed sleep” (3.96–2.96, $p=0.04$), “distress” (2.43–1.23, $p=0.04$), and “shortness of breath” (2.06–1.27, $p=0.004$). There were no significant changes in “pain,”

FIG. 2. Summary of patient-reported outcomes with relation to surgery over time. Predicted means of the summary scores for the MDASI patient-reported outcomes at three time points: before surgery, ST, and LT. Higher scores indicate higher symptom burden and lower scores indicate lower symptom burden. MDASI, MD Anderson Symptom Inventory.

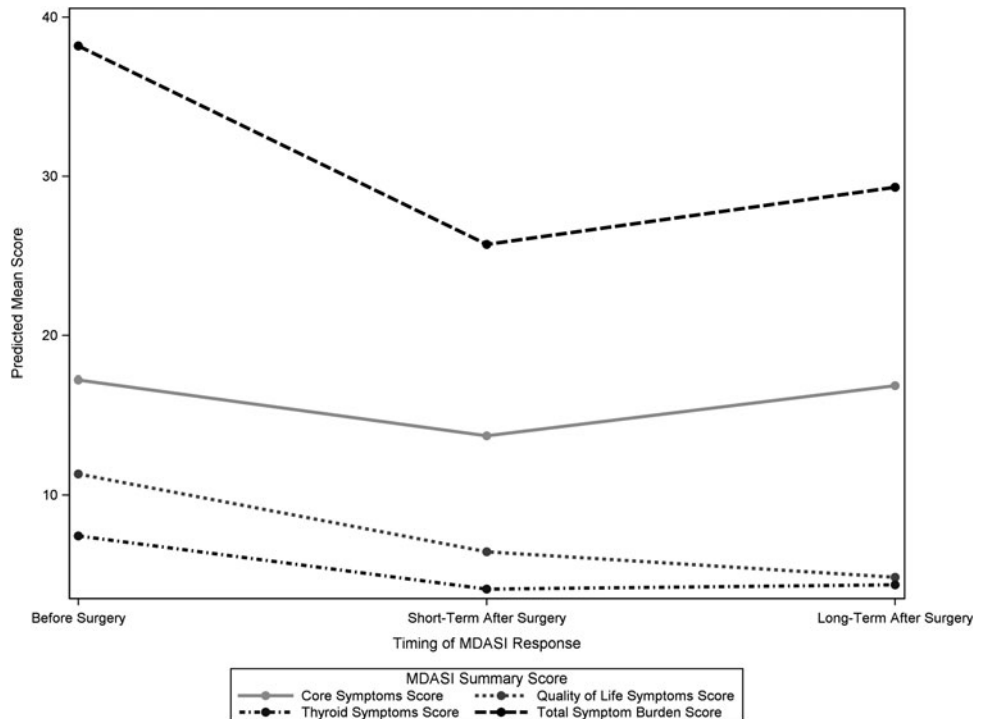




FIG. 3. Selected patient-reported outcomes before and after surgery. Specific patient-reported outcomes displayed by the mean value from before surgery and in the ST time point. All changes in the figure were statistically significant, $p < 0.05$. Higher scores indicate higher symptom burden, and lower scores indicate lower symptom burden.

“memory,” “appetite,” “drowsiness,” “dry mouth,” “sad feelings,” “vomiting,” and “numbness” from before surgery to ST.

Total Symptom Burden Score

The Total Symptom Burden Score significantly improved from a mean of 56.88 before surgery to 39.60 in the ST ($p < 0.001$) (Table 2 and Fig. 2). While not a statistically significant change from before surgery, the Total Symptom Burden Score was 40.90 for patients in the LT ($p = 0.08$). No significant change was observed in the score between ST and LT ($p = 1.00$). Subcategories and specific components of the Total Symptom Burden Score are reported in Table 2.

Multivariate analysis

After multivariate adjustment for patient and clinical characteristics and time related to surgery, the patient-reported burden in both the Thyroid Symptom Score and Quality-of-Life Symptom Score significantly decreased from before surgery to ST (thyroid symptoms, rate ratio [RR] 0.55, confidence interval [CI]: 0.42–0.72; quality of life, RR 0.57, CI: 0.40–0.81) and LT (thyroid symptoms, RR 0.59, CI: 0.44–0.79; quality of life, RR 0.43, CI: 0.28–0.65), demonstrating an improvement in PROs. The Total Symptom Burden Score improved from before surgery to ST (RR 0.67, CI: 0.53–

0.85), but there was no change from before surgery to LT (RR 0.77, CI: 0.54–1.08). The Core Symptom Score did not significantly change from before surgery to ST (RR 0.80, CI: 0.61–1.04) nor to LT (RR 0.98, CI: 0.62–1.54) (Table 3).

After multivariate adjustment, patient characteristics were associated with all the four composites scores. Older age (Total Symptoms, RR for 10-year increase 0.84, CI: 0.72–0.97; Core Symptoms, RR for 10-year increase 0.83, CI: 0.72–0.97) and male sex (Total Symptoms, RR 0.44, CI: 0.25–0.77; Core Symptoms, RR 0.49, CI: 0.28–0.86) were significantly associated with improved Total Symptom Burden Score and the Core Symptom Score. Male sex was also associated with improved Quality-of-Life Score (RR 0.36, CI: 0.18–0.71), while older age was also associated with improved Thyroid Symptom Score (RR 0.98, CI: 0.97–1.00) (Fig. 4). Non-Hispanic other or multiple races was associated with improved Thyroid Symptom Score compared with non-Hispanic white (RR 0.22, CI: 0.10–0.50), and patients with other insurance (RR 0.57, CI 0.34–0.94) were associated with improved Core Symptom Score compared with private insurance (Table 3). Finally, non-Hispanic black was associated with compromised PROs, including the Core Symptom Score (RR 1.80, CI 1.18–2.73), Quality-of-Life Score (RR 2.77, CI 1.50–5.12), and Total Symptom Burden Score (RR 1.66, CI 1.05–2.64) compared with non-Hispanic white.

TABLE 3. MULTIVARIATE NEGATIVE BINOMIAL REGRESSION FOR TOTAL SYMPTOM BURDEN, QUALITY-OF-LIFE SYMPTOM, AND CORE SYMPTOM SCORES

Independent variables	Total Symptom Burden Score			Quality-of-Life Symptom Score			Core Symptom Score		
	Rate ratio [CI]	p	Overall p-value	Rate ratio [CI]	p	Overall p-value	Rate ratio [CI]	p	Overall p-value
Time point			0.002			<0.001			0.09
Before surgery	Ref.			Ref.			Ref.		
Short term after surgery	0.67 [0.53-0.85]	<0.001		0.57 [0.4-0.81]	0.002		0.8 [0.61-1.04]	0.09	
Long term after surgery	0.77 [0.54-1.08]	0.13		0.43 [0.28-0.65]	<0.001		0.98 [0.62-1.54]	0.93	
Age (per 10 years)	0.84 [0.72-0.97]	0.02	0.02	0.88 [0.75-1.03]	0.12	0.12	0.83 [0.72-0.97]	0.02	0.02
Sex			0.003			0.003			0.01
Female	Ref.			Ref.			Ref.		
Male	0.44 [0.25-0.77]	0.004		0.36 [0.18-0.71]	0.003		0.49 [0.28-0.86]	0.01	0.02
Race/ethnicity			0.10			0.01			
Non-Hispanic white	Ref.			Ref.			Ref.		
Non-Hispanic black	1.66 [1.05-2.64]	0.03		2.77 [1.5-5.12]	0.001		1.8 [1.18-2.73]	0.006	
Non-Hispanic other or multiple races	0.76 [0.29-2.02]	0.58		2.15 [0.69-6.74]	0.19		0.72 [0.26-2.05]	0.54	
Hispanic	1.33 [0.68-2.61]	0.40		1.79 [0.84-3.86]	0.13		1.45 [0.83-2.55]	0.19	
Marital status			0.94			0.76			0.40
Married or life partner	Ref.			Ref.			Ref.		
Not married and no life partner	1.18 [0.81-1.73]	0.39		0.92 [0.55-1.54]	0.76		1.2 [0.79-1.83]	0.40	
Insurance			0.05			0.03			0.006
Private	Ref.			Ref.			Ref.		
Government	0.71 [0.43-1.15]	0.17		0.61 [0.34-1.12]	0.11		0.66 [0.42-1.05]	0.08	
Other	0.64 [0.33-1.22]	0.18		0.21 [0.04-1.07]	0.06		0.57 [0.34-0.94]	0.03	
Unreported	1.28 [0.75-2.19]	0.37		1.12 [0.58-2.15]	0.74		1.24 [0.73-2.11]	0.43	
Obese			0.40			0.09			0.39
Yes	Ref.			Ref.			Ref.		
No	1.16 [0.72-1.86]	0.55		0.84 [0.44-1.62]	0.61		1.25 [0.81-1.95]	0.32	
Unreported	1.47 [0.84-2.58]	0.18		1.63 [0.82-3.24]	0.16		1.43 [0.82-2.47]	0.21	
Hypertension			0.71			0.21			0.86
No	Ref.			Ref.			Ref.		
Yes	1.1 [0.68-1.77]	0.71		1.48 [0.8-2.76]	0.21		0.96 [0.6-1.53]	0.86	
TSH level	1.0 [0.99-1.01]	0.50	0.50	1.02 [1-1.04]	0.07	0.07	1.0 [0.99-1.02]	0.62	0.62

Higher scores indicate higher symptom burden, and lower scores indicate lower symptom burden. CI, confidence interval.

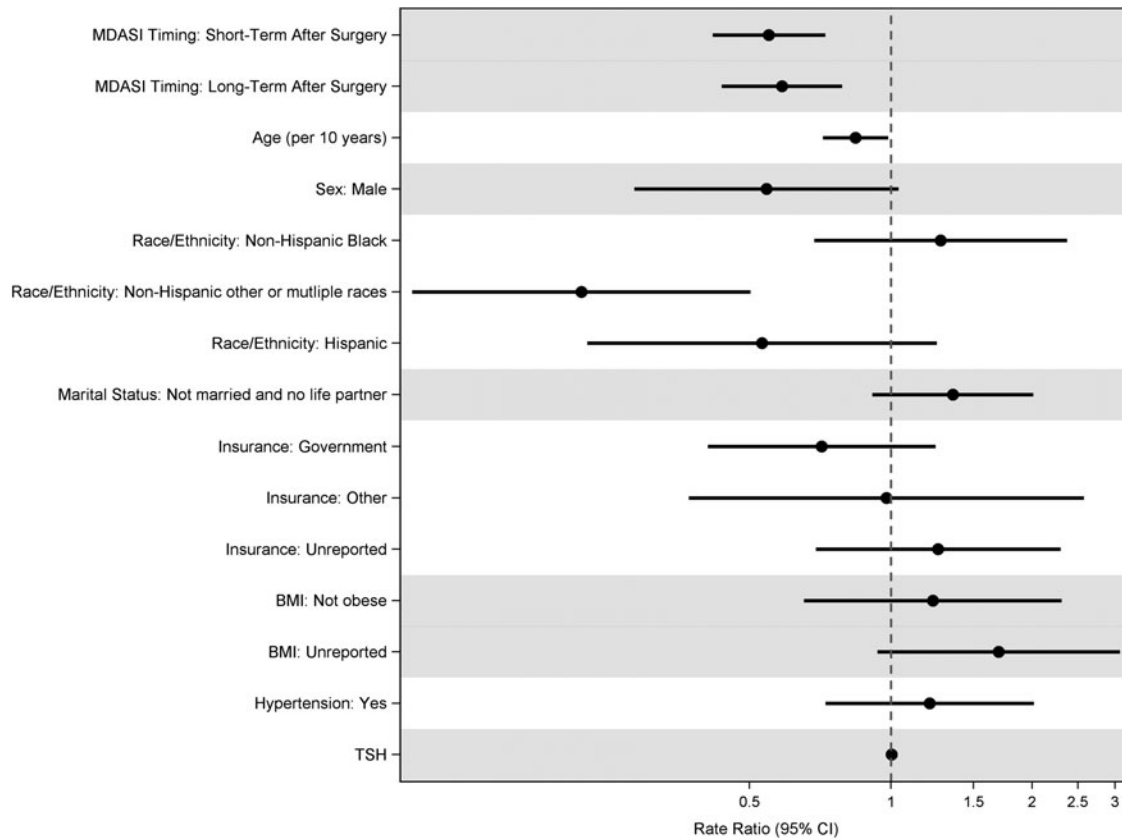


FIG. 4. Multivariate negative binomial regression for MDASI Thyroid Symptom Burden Score. Reference levels are: before surgery (MDASI timing), female (sex), non-Hispanic white (race/ethnicity), not married and no life partner (marital status), private (insurance), obese (BMI), and no hypertension. Higher scores indicate higher symptom burden, and lower scores indicate lower symptom burden. BMI, body mass index.

Discussion

To our knowledge, this is the first study to examine PROs in Graves’ disease patients undergoing thyroidectomy in North America and the first to analyze patient characteristics associated with PROs before and after thyroid surgery. After addressing the key limitations of published literature through our inclusion of data collection shortly after surgery, measurement of clinical characteristics (TSH levels), and analysis of the relationship between PROs and patient characteristics, we found that patient-reported symptoms, including quality-of-life and thyroid-specific symptoms, of Graves’ patients improved significantly from before surgery to shortly after, within 30 days of surgery, and to longer after the surgery. Specifically, patient-reported “fatigue,” “nausea,” “disturbed sleep,” “distress,” “shortness of breath,” “feeling hot,” “racing heartbeat,” “diarrhea,” “difficulty swallowing,” “general activity,” “mood,” “relationship with other people,” “walking,” and “enjoyment of life” significantly improved from before surgery to ST.

Previous research has demonstrated that patients with Graves’ disease have a poorer quality of life compared with the general population (3,9,13,14). Cramon *et al.* reported that thyroid-specific symptoms improved in patients with Graves 6 months after treatment (9). Despite this, compared with the general population (of Sweden), significant differ-

ences in health-related quality of life remained at 6 months after treatment (9). Torring *et al.* report congruent findings that PROs of patients with Graves’ disease remained worse compared with the general population (of Sweden), even 6–10 years after diagnosis (3). Furthermore, this group showed that patients treated with RAI had worse thyroid-related and general quality of life compared with patients treated with medications or thyroidectomy (3). The findings of long-term impairment of PROs in Graves’ patients have been corroborated in several other studies (10,12,14,15). Our study found that shortly after thyroid surgery, core symptoms, quality of life, and thyroid-specific symptoms all improve. Consistent with previous reports, quality of life and thyroid-specific symptoms maintained improvement in the LT time point after surgery. Despite power limitations from the small sample size, the Total Symptom Burden, which is a summary of all PROs collected, significantly improved before surgery to ST. In the context of the significant improvement in both quality-of-life symptoms and thyroid symptoms from before surgery to LT, the absence of significant change in Total Symptom Burden from before surgery to LT is likely due to the Core Symptom Score, which includes symptoms such as “nausea” and “sadness,” which are not specific to thyroid disease.

Our analysis shows that older age was associated with improved Total Symptom Burden Score, Core Symptoms

Score, and Thyroid Symptom Score. Male sex was also associated with improved Total Symptom Burden Score, Core Symptoms Score, and Quality-of-Life Symptom Score. This relationship is consistent with previous published work; however, the mechanisms remain unclear. In their cross-sectional analysis of patients with Graves' disease, Riguetto *et al.* found that male sex was associated with a higher quality of life (13). Our findings regarding age appear to disagree with traditional clinical outcomes, which suggest that older patients with Graves' disease suffer from greater morbidity compared with younger patients (23). However, these differences may be because traditional definitions of morbidity do not include patient-reported values. More research is needed to understand the relationship between age, male sex, and their clinical and PROs.

This study also showed that additional demographic factors, such as race and insurance type, were associated with PROs. Patients who identified as non-Hispanic black are significantly associated with a worse Total Symptom Burden Score, Core Symptoms Score, and Quality-of-Life Score compared with those identifying as non-Hispanic white. On the contrary, patients who identified as non-Hispanic other or multiple races were associated with improved Thyroid Symptom Score compared with non-Hispanic white. Patients with other insurance, which includes patients who reported "other" or "commercial," were significantly associated with improved Core Symptoms Score compared with private insurance. The observed differences in PROs by race and insurance type are difficult to contextualize as this is the first study based on the social and economic environment of the United States. In their analysis of older adults with and without cancer, Rincon *et al.* also report racial and ethnic disparities in health-related quality-of-life outcomes and that the disparities may be increasing over time (24). Even with the uncertainty of the observed differences, the prospective collection of PROs may be a useful tool to identify potential areas of improvement for reducing racial disparities in health care (25).

Several specific PROs displayed different patterns of change in relation to thyroid surgery. Hoarseness and pain significantly improved after 30 days from surgery (LT), but not prior (ST). This pattern of pain and hoarseness is likely the result of well-known operative complications of thyroid surgery or from the intubation procedure, and may not have yet resolved at the ST postsurgical time point (26–28). These findings are important in the shared decision-making between surgeons and patients. By demonstrating the dynamic changes of symptoms, patients with Graves' disease can be more effectively counseled on what to expect in the short-term period after surgery.

The primary limitations of this study are related to the use of single-institution data. While patients were recruited from a high-volume tertiary surgical center and underwent surgery by high-volume endocrine surgeons, the number of Graves' patients undergoing thyroidectomy remains relatively small; nonetheless, the fact that our analyses revealed statistically significant results indicates data robustness. In addition, because our data were collected prospectively in real time rather than by a mailed questionnaire or recall, participants did not necessarily have data for every time point. Despite this, our analysis showed that no meaningful differences in patient characteristics

existed between time points. Although the MDASI questionnaire was originally designed for thyroid cancer, the content is often associated with benign hyperthyroid disease, which provides insight into Graves' disease. Furthermore, we are unable to provide PROs from comparison populations treated with antithyroid medications or radioactive iodine; this could be a focus of future research. Despite these limitations, this study has key strengths, including the use of baseline data from before surgery, the inclusion of thyroid function, and the analysis of the relationship between patient characteristics and PROs, thus serving as the first report of PROs in patients with Graves' disease before and after thyroidectomy from North America.

In conclusion, our analysis showed that patient-reported symptoms, including quality of life and thyroid-specific symptoms, of Graves' patients improved significantly from their baseline before surgery to both shortly after and to long after the surgery. This comprehensive analysis can be used to guide clinicians and their patients with Graves' disease on what outcomes to expect as they undergo thyroidectomy. Furthermore, this is evidence that the surgical management of Graves' disease is associated with improved quality of life and symptom burden of patients.

Authors' Confirmation Statement

All authors meet the requirements for authorship, as defined by the International Committee of Medical Journal Editors, including the following:

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work.
- Drafting the work or revising it critically for important intellectual content.
- Final approval of the version to be published.
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgments

Data for this study were provided by the Private Diagnostic Clinic's Outcomes Research Team, and were specifically designed by Sarah Ahmadi, MD, Julie Ann Sosa, MD, and Susan Spratt, MD.

Authors' Contributions

A.H.G.: Conceptualization, methodology, data analysis and interpretation, writing—original draft, and writing—review and editing. N.F.: Conceptualization, methodology, data interpretation, writing—original draft, and writing—review and editing. S.M.T.: Collection and assembly of data, data analysis and interpretation, writing—original draft, and writing—review and editing. M.T.S.: Conceptualization, methodology, data interpretation, and writing—review and editing. R.P.S.: Conceptualization, supervision, methodology, data analysis and interpretation, writing—original draft, and writing—review and editing. H.S.K.:

Conceptualization, supervision, methodology, data analysis and interpretation, writing—original draft, and writing—review and editing.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

None for this project. S.M.T. (statistician) is, in part, supported by the Duke Cancer Institute through NIH grant P30CA014236 (Principal Investigator: Kastan) for the Biostatistics Core.

Supplementary Material

Supplementary Appendix SA1
Supplementary Table S1

References

- Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM, Okosieme OE 2018 Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol* **14**:301–316.
- Torrington O, Tallstedt L, Wallin G, Lundell G, Ljunggren JG, Taube A, Saaf M, Hamberger B 1996 Graves' hyperthyroidism: treatment with antithyroid drugs, surgery, or radioiodine—a prospective, randomized study. *Thyroid Study Group. J Clin Endocrinol Metab* **81**:2986–2993.
- Torrington O, Watt T, Sjölin G, Bystrom K, Abraham-Nordling M, Calissendorff J, Cramon PK, Filipsson Nystrom H, Hallengren B, Holmberg M, Khamisi S, Lantz M, Wallin G 2019 Impaired quality of life after radioiodine therapy compared to antithyroid drugs or surgical treatment for Graves' hyperthyroidism: a long-term follow-up with the thyroid-related patient-reported outcome questionnaire and 36-item Short Form Health Status Survey. *Thyroid* **29**:322–331.
- Watt T, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, Bjorner JB, Feldt-Rasmussen U 2006 Quality of life in patients with benign thyroid disorders. A review. *Eur J Endocrinol* **154**:501–510.
- Oltmann SC, Nwariaku FE 2019 Are PROMs ideally suited for most common endocrine surgical patients and procedures? *Surgery* **165**:240–241.
- Burch HB, Burman KD, Cooper DS 2012 A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab* **97**:4549–4558.
- Brito JP, Payne S, Singh Ospina N, Rodriguez-Gutierrez R, Maraka S, Sangaralingham LR, Iniguez-Ariza NM, Montori VM, Stan MN 2020 Patterns of use, efficacy, and safety of treatment options for patients with Graves' disease: a nationwide population-based study. *Thyroid* **30**:357–364.
- Hurtado C, Yeh MW, Livhits MJ 2020 Antithyroid medications are the most common treatment for Graves' disease in the United States despite high rates of treatment failure. *Clin Thyroidol* **32**:162–165.
- Cramon P, Winther KH, Watt T, Bonnema SJ, Bjorner JB, Ekholm O, Groenvold M, Hegedus L, Feldt-Rasmussen U, Rasmussen AK 2016 Quality-of-life impairments persist six months after treatment of Graves' hyperthyroidism and toxic nodular goiter: a prospective cohort study. *Thyroid* **26**:1010–1018.
- Abraham-Nordling M, Torring O, Hamberger B, Lundell G, Tallstedt L, Calissendorff J, Wallin G 2005 Graves' disease: a long-term quality-of-life follow up of patients randomized to treatment with antithyroid drugs, radioiodine, or surgery. *Thyroid* **15**:1279–1286.
- Taieb D, Bournaud C, Eberle MC, Catargi B, Schvartz C, Cavarec MB, Faugeron I, Toubert ME, Benisvy D, Archange C, Mundler O, Caron P, Abdullah AE, Baumstarck K 2016 Quality of life, clinical outcomes and safety of early prophylactic levothyroxine administration in patients with Graves' hyperthyroidism undergoing radioiodine therapy: a randomized controlled study. *Eur J Endocrinol* **174**:491–502.
- Al-Adhami A, Craig W, Krukowski ZH 2012 Quality of life after surgery for Graves' disease: comparison of those having surgery intended to preserve thyroid function with those having ablative surgery. *Thyroid* **22**:494–500.
- Riguetto CM, Neto AM, Tambascia MA, Zantut-Wittmann DE 2019 The relationship between quality of life, cognition, and thyroid status in Graves' disease. *Endocrine* **63**:87–93.
- Abraham-Nordling M, Wallin G, Lundell G, Torring O 2007 Thyroid hormone state and quality of life at long-term follow-up after randomized treatment of Graves' disease. *Eur J Endocrinol* **156**:173–179.
- Elberling TV, Rasmussen AK, Feldt-Rasmussen U, Hording M, Perrild H, Waldemar G 2004 Impaired health-related quality of life in Graves' disease. A prospective study. *Eur J Endocrinol* **151**:549–555.
- Conaglen HM, Tamatea JAU, Conaglen JV, Elston MS 2018 Treatment choice, satisfaction and quality of life in patients with Graves' disease. *Clin Endocrinol (Oxf)* **88**:977–984.
- Grove-Laugesen D, Cramon PK, Malmstroem S, Ebbehoj E, Watt T, Hansen KW, Rejnmark L 2020 Effects of supplemental vitamin D on muscle performance and quality of life in Graves' disease: a randomized clinical trial. *Thyroid* **30**:661–671.
- Boesen VB, Feldt-Rasmussen U, Bjorner JB, Cramon P, Groenvold M, Nygaard B, Rasmussen AK, Vilsboll T, Watt T 2018 How should thyroid-related quality of life be assessed? Recalled patient-reported outcomes compared to here-and-now measures. *Thyroid* **28**:1561–1570.
- Ljunggren JG, Torring O, Wallin G, Taube A, Tallstedt L, Hamberger B, Lundell G 1998 Quality of life aspects and costs in treatment of Graves' hyperthyroidism with antithyroid drugs, surgery, or radioiodine: results from a prospective, randomized study. *Thyroid* **8**:653–659.
- Gning I, Trask PC, Mendoza TR, Harle MT, Gutierrez KA, Kitaka SA, Sherman SI, Cleeland CS 2009 Development and initial validation of the thyroid cancer module of the M. D. Anderson Symptom Inventory. *Oncology* **76**:59–68.
- Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, Engstrom MC 2000 Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. *Cancer* **89**:1634–1646.
- Pan W 2001 Akaike's information criterion in generalized estimating equations. *Biometrics* **57**:120–125.

23. Samuels MH 2000 Hyperthyroidism in Aging. In: Feingold KR, Anawalt B, Boyce A, et al. (eds) Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK278986/> (accessed August 9, 2021).
24. Rincon MA, Smith AW, Yu M, Kent EE 2020 Trends in racial/ethnic disparity of health-related quality of life in older adults with and without cancer (1998–2012). *Cancer Epidemiol Biomarkers Prev* **29**:1188–1195.
25. Hostetter M, Klein S 2018 In Focus: Reducing Racial Disparities in Health Care by Confronting Racism. The Commonwealth Fund. Available at: <https://www.commonwealthfund.org/publications/2018/sep/focus-reducing-racial-disparities-health-care-confronting-racism> (accessed September 27, 2018).
26. Gunn AH, Oyekunle T, Stang M, Kazaure H, Scheri R 2020 Recurrent laryngeal nerve injury after thyroid surgery: an analysis of 11,370 patients. *J Surg Res* **255**: 42–49.
27. Mencke T, Echternach M, Kleinschmidt S, Lux P, Barth V, Plinkert PK, Fuchs-Buder T 2003 Laryngeal morbidity and quality of tracheal intubation: a randomized controlled trial. *Anesthesiology* **98**:1049–1056.
28. Tharakan T, Jiang S, Fastenberg J, Ow TJ, Schiff B, Smith RV, Mehta V 2019 Postoperative pain control and opioid usage patterns among patients undergoing thyroidectomy and parathyroidectomy. *Otolaryngol Head Neck Surg* **160**: 394–401.

Address correspondence to:

Hadiza S. Kazaure, MD

Department of Surgery

Duke University Medical Center

466G Seeley Mudd Building

Durham, NC 27710

USA

E-mail: hadiza.kazaure@duke.edu