Measurement and Variation in Estimation of Quality of Life Effects of Patients Undergoing Treatment for Papillary Thyroid Carcinoma

Carrie C. Lubitz,[1,2] Lucia De Gregorio,[1] Abbey L. Fingeret,[1] Konstantinos P. Economopoulos,[1,2]
Diana Termezawi,[1] Mursal Hassan,[1] Sareh Parangi,[1] Antonia E. Stephen,[1] Elkan F. Halpern,[2,3]
Karen Donelan,[2,4] and J. Shannon Swan[2,3]

Background: Thyroid cancer incidence is increasing. The effect of diagnosis and treatment on health-related quality of life (HRQoL) is an essential variable in the absence of a change in life span for the majority of patients. HRQoL instruments, with data useful for between-disease comparisons, are being increasingly used for health policy and outcomes evaluation. Variation exits among the instruments based on the impact of a specific disease. We assessed which of four well-validated, preference-based surveys detect changes in health and clinical intervention in patients diagnosed with papillary thyroid cancer (PTC).
Methods: Four commonly used HRQoL questionnaires (Short Form-12v2® [SF6D], EuroQol-5D [EQ5D], and Health Utilities Index Mark 2 and 3 [HUI2, HUI3]) were administered to patients with the diagnosis of PTC at three perioperative time points during the first year of treatment. Clinicopathological and treatment course data were assessed for HRQoL impact including complications from surgery, re-operation for persistence/early recurrence, and adjuvant radioactive iodine treatment. We compared standard metrics, including ceiling effect, intraclass correlation coefficient, effect sizes, and quality-adjusted life-years between the four instruments.
Results: Of 117 patients, 27% had a preoperative diagnosis of anxiety or depression, 41% had regional lymph node metastases, three had distant metastases and 49% underwent adjuvant radioactive iodine treatment. The ceiling effect (i.e., proportion with a perfect score) was greatest with EQ5D and least with SF6D. Index scores ranged from 0.77 (SF6D) to 0.90 (EQ5D). All scores declined at two weeks postoperatively and returned to pretreatment levels at six months. The SF6D was the only instrument to exceed the conventional minimally important difference between all three time points. Quality-adjusted life-years were as follows: SF6D, 0.79; EQ5D, 0.90; HUI2, 0.88; and HUI3, 0.86.
Conclusions: Our results reflect the general good health of PTC patients. The effect on quality of life is primarily related to emotional and social impacts of treatment. The results support the measurement of a similar underlying construct, although variation in detecting changes in health exists between the instruments. Of the instruments assessed, the SF6D is the most responsive to treatment effects and should be utilized in future economic analyses in this patient population.

Keywords: thyroid cancer, clinical research, surgery, quality of life

Introduction

THYROID CANCER INCIDENCE is increasing worldwide at a rate higher than any other cancer and constitutes four percent of all cancers in the United States, with over 62,000 patients diagnosed in 2015 (1,2). The majority of the increase in incidence is in small papillary thyroid carcinomas (PTC) with no effect on disease-specific survival (3,4). The prevalence of thyroid cancer survivors is likewise steadily increasing given the young age at diagnosis and 98% five-year survival (5).

Most people diagnosed with thyroid cancer will not become symptomatic or die from their disease—a state referred to as overdiagnosis (6). This has led to a discussion among

Departments of [1]Surgery and [3]Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.
[2]Institute for Technology Assessment and [4]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts.

experts about the presence and extent of overtreatment and a need for weighing the benefits of our interventions against undue harm. Recent guidelines reflect a shift toward less aggressive approaches, and a number of groups have instituted active surveillance protocols for patients with low-risk PTC (7,8–10). Likewise, efforts to stratify patients has led to the recent reclassification of noninvasive encapsulated follicular-variant PTC as noninvasive follicular thyroid neoplasm with papillary-like nuclear features (11). Assessment of the appropriateness and effectiveness of current and proposed changes to care is warranted, yet challenging. The extensive cost and follow-up time necessary for large-scale clinical trials is prohibitive. Assessment of comparative effectiveness through computer simulation modeling integrating health-related quality of life adjustments is a next best alternative.

Most thyroid cancer related research focuses on oncological outcomes. Given the longevity of patients with PTC, survival alone as an outcome measure is not sufficient. Nearly a fifth of all thyroid cancer patients have recurrences. While most patients experience recurrence within the first few years following diagnosis, it can occur decades after initial treatment (12). This means that survivors require lifelong surveillance and adjustments to administered thyroid hormone. The effect of diagnosis and treatment on quality of life over the prolonged course of survivorship is, therefore, increasingly relevant.

Thyroid cancer diagnoses and therapy have immediate and long-standing effects on patients (13). Quality of life research has primarily focused on the physical side effects, and prospective, longitudinal data during the course of treatment are lacking (14–16). Neuropsychometric surveys have reported a decrement to quality of life for up to five years following treatment, on par with other cancers carrying worse prognoses (17–19). Qualitative analysis reveals emotional and financial impact on thyroid cancer on survivors of all stages of disease (20). These reports have highlighted the importance of the patients' perspectives of well-being in thyroid survivorship care; however, these data are not translatable to values useful for economic analysis or comparative effectiveness research. In order to make comparisons between different diseases, and to potential inform resource allocation, generic survey instruments are necessary. Prior studies in the thyroid cancer population have required extended recall times from patients and record a single snapshot in the treatment course.

In this work, we aim to assess the health-related quality of life (HRQoL) effects of patients diagnosed with PTC from diagnosis through the initial treatment period. We survey patients diagnosed with PTC during evaluation and treatment using well-validated instruments, and calculate a preference-based utility (i.e., a patient's relative value assigned to health) that can be compared between various disease states. Additionally, we aimed to assess which commonly used and validated HRQoL instrument(s)—EuroQol-5D-5L [EQ5D] (21,22), Short Form-12v2/SF6D® (SF6D) (23), or Health Utility Indices Mark 2 and Mark 3 [HUI2 and HUI3] (24,25)—are useful for comparative effectiveness research (i.e., to obtain quality-adjusted life-year estimates). Based on the previously reported sensitivity to changes in health of the SF6D in healthier populations, similar to most patients diagnosed with PTC, we hypothesized that this instrument

would most accurately reflect the impact of PTC diagnosis and treatment.

Materials and Methods

Patient selection

From July 2012 to July 2015, we identified consecutive eligible patients from the outpatient clinic schedules and approached for participation in the study. To be eligible, patients had to be adult (>18 years old), English-speaking, and have a fine-needle aspiration biopsy positive for PTC. Prior to surgery, a questionnaire packet was sent to each patient by postal mail. Participants could either respond by phone with a trained interviewer, by mail, or, for subsequent surveys, online. Enrolled patients were surveyed at three relevant clinical time points: (1) prior to surgery (pre-op); (2) at the first postoperative visit, two weeks to one month following thyroidectomy (post-op), and (3) at the first surveillance visit with the surgeon or endocrinologist, 6–12 months after surgery (follow-up). Patients who failed to respond to the post-op survey were again contacted at the follow-up visit. Patient consent was implied by completing the questionnaire packet and remuneration in the amount of U.S. $20 was given to those participants who completed the last survey. The study protocol was approved by the Institutional Review Board at Massachusetts General Hospital.

Quality of life assessment

The survey instrument was nine pages long and consisted of quality of life assessment, clinical indicators, and patient demographics, taking approximately 15 minutes to complete. The questionnaire packet included four widely used preference-based generic utility instruments: EQ5D, SF6D, and HUI2 and HUI3 (21,23–25). All instruments are based on a multidimensional health state classification system from which a standardized scoring (or weighting) system is generated. Levels are self-reported by the respondent and a community preference-based scoring function translates the descriptive system into the index score. Study data were collected and catalogued with secure REDCap™ electronic data capture tools hosted at Partners Healthcare (26). Survey responses were later translated into interval scale utilities [0 (dead) to 1.0 (perfect health)] based on the above-established methods. A score closer to 1.0 reflects a "better" health status. The highest score in the range for each instrument is considered the "ceiling" and the lowest score in the range the "floor."

The EQ5D defines health in five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The time frame in the questions is "today." In an effort to minimize the EQ5D well-known ceiling effect, we used the latest five-level response version. The U.S. algorithm scores range from −0.11 to 1.0, and are based on a temporary cross-walk from the regression-based three-level version (22,27). The EQ5D also features a visual analogue rating scale for the respondent to assess overall health on the current day. The visual analogue rating scale rating, ranging from 0 to 100, can also be interpreted as a quantitative valuation of the patients' own health but is not on a dead to perfect health scale. It instead uses a scale of "best" to "worst imaginable health state."

The SF6D consists of six dimensions: physical functioning, role limitations (combined physical and emotional), social functioning, pain, mental health, and vitality. This index utility is derived from applying a regression-based scoring method to seven items from the Short Form-12v2 (23). Respondents are asked about their health in the past four weeks and two non-preference-based summary scores for physical and mental health status are generated. The SF6D has been found to be more responsive than EQ5D, HUI2, and HUI3 in healthy populations, given its relatively low ceiling effect (28–30). However, it has a known floor effect, since scores can range from 0.3 to 1.0.

The HUI2 health status classification system describes six attributes—sensation, mobility, emotion, cognition, self-care, and pain—while HUI3 describes eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Both scores are obtained from a 40-item questionnaire that assesses the respondents' recall of health in the last four weeks. These attributes consider physical and emotional dimensions but exclude social interactions (thus limiting the HUI to ''within the skin'' attributes) (31). Scores range from −0.03 to 1.0 and −0.36 to 1.0 for the HUI2 and HUI3 respectively.

We also considered including the Quality of Well-Being Scale (32) given the known lower ceiling effect. However, we chose not to include the Quality of Well-Being Scale due to its length and less optimal psychometrics (30,33).

Demographic and clinical variables

Patients self-reported demographic data including race/ethnicity, education level, marital status, and number of children, together with the open-ended question, ''What concerns you most about your diagnosis and treatment?'' Clinical variables recorded were: past history of neck radiation, family history of thyroid cancer, thyroid cancer presentation, extent of surgery, postoperative complications, pathological findings, postoperative thyrotropin (TSH), and thyroglobulin evels, radioactive iodine (RAI) therapy, and evidence of recurrent or persistent disease. In order to assess for nonresponse bias, data on age, sex, race/ethnicity, and educational level were obtained from the medical charts in nonrespondents and compared with respondents. Responses to the open-ended question were recorded verbatim coded into one of the following categories: anxiety/emotional well-being, cancer prognosis/death, financial burden to family members, risk of children getting cancer, surgical complications/morbidity, radioactive iodine treatment, and long-term consequences of hormone replacement.

Missing data

Twenty-two cases had missing data; 18 of the 22 values were missing due to a single systematic error (i.e., missed a question while turning the packet pages). We imputed the missing data using a modified hot deck procedure that stratified respondents by sex, age group (18–39 years old, 40–64 years old, and 65+ years old), and survey time point (pre-op, post-op, or follow-up) (34). Coherence among similar items from different utilities was revised so that the imputed value would make sense within the patient's responses. For example, question 1 on the EQ5D asks about mobility problems

and is similar in content to questions 2a and 2b on SF6D and question 16 on HUI2/3.

Assumptions

In order to assess comparative responsiveness of the four instruments, we made some general assumptions. First, we assumed that all patients experience a decrease in their HRQoL post-op, mainly attributable to physical discomfort. Secondly, with the exception of patients sustaining surgical complications (i.e., vocal cord palsy or hypoparathyroidism), persistence or early recurrence of disease, or those undergoing RAI treatment, all patients were expected to return to pretreatment health status by the time of the follow-up survey (6–12 months after surgery).

Statistical analyses

Response rate was calculated as follows: Respondents / (Respondents + Nonrespondents) × 100. Respondents were all patients who completed the pretreatment survey. Nonrespondents were patients approached by mail or phone to a verified address or phone number and who either expressed refusal to participate or from whom we did not get response prior to their surgery date. Age group, sex, race/ethnicity, and educational level among nonrespondents and respondents were compared by Fisher's exact test.

Distributions of overall scores and by domain were compared at the three different time points and between instruments. The proportion of respondents reporting perfect health (ceiling effect) was calculated for each index at all three time points. Additionally, we assessed the change over the three time points in those that reported perfect health.

To assess for reliability (i.e., measurements of the same health status in the same patients should be similar) among the instruments we calculated intra-class correlation coefficients (ICC) for absolute agreement with the same subjects as raters at the same time points. In this analysis we used a two-way mixed model with single measures. Interpretation of ICCs were as follows: poor, <0.40; moderate, 0.40–0.59; good, 0.60–0.74; and excellent, >0.75 (35). For a second assessment of reliability among indexes, we assessed Spearman correlation coefficients that were interpreted as follows: absent, <0.20; poor, 0.20–0.34; moderate, 0.35–0.50; and good, >0.50 (35).

Mean differences and standard deviation (SDΔ) between time points were estimated for each instrument. Two-way (time by utility) repeated measures ANOVA was used to test the equivalence of means and variances. The standardized response mean (SRM), used to detect the magnitude of change, was calculated for each index from pretreatment to post-op and post-op to follow-up as a measure of the magnitude of change. The following formula for SRM was utilized, where U is utility score and T_1 and T_2 are time points:

$$\frac{\bar{x}_{(U_{T2} - U_{T1})}}{s_{(U_{T2} - U_{T1})}}$$

SRMs were interpreted as follows: small, <0.2; medium, 0.2–0.5; and large, >0.8 (i.e., Cohen's effect size criteria) (36). The minimally important difference (MID) is an accepted reference for the smallest yet clinically meaningful difference

in HRQoL scores (37). For each health-state transition (pre-op to post-op, post-op to follow-up, and pre-op to follow-up) the percentage of patients with a mean change greater than the MID was calculated. Estimates of the MID for SF6D, HUI2, and HUI3 were taken from a previous publication (38). In the case of the five-level version of the EQ5D, one-half of the standard deviation at pretreatment ($0.5 \times SD_{T1}$) was considered as threshold for MID (39).

Those patients experiencing complications from surgery, re-operation for persistence/early recurrence, or adjuvant radioactive iodine treatment were expected to have a decrease in HRQoL throughout the study period. The remaining patients were expected to have their health status back to pretreatment at the follow-up survey. In these patients, coefficients of stability (intraclass correlation coefficient) between pre-op and follow-up were calculated for each index, as an indicator of the index's test–retest reliability. These data serve to assess how much a score could vary when no change in their health status is expected. A two-way mixed effects model provided the variance components. Also in this subgroup of patients, the standard error of measurement (SEM) of each index was calculated as $(1 - \text{coefficient of stability}) \times SD_{T1}$ (33).

Quality adjusted life-years were calculated using the data from the four instruments, assuming a linear change between time points. Specifically, we summed the average utility score for each time period $[(|U_{\text{pre-op}} + U_{\text{post-op}}|/2) \times 30/365] + [(|U_{\text{post-op}} + U_{\text{follow-up}}|/2) \times 153/365] + [(|U_{\text{pre-op}} + U_{\text{follow-up}}|/2) \times 183/365]$. Differences were considered statistically significant if p values were lower than 0.05. The statistical software used for analysis was STATA® for Windows version 13.1 (Copyright ©1985–2013 StataCorp LP) and MedCalc for Windows, version 16.2 (MedCalc Software).

Results

Survey response

Of the 275 total outpatient clinic patients approached, 244 were eligible for the study. Reasons for being ineligible included: survey packet returned undeliverable and address and phone numbers of record not reachable, patient not fluent in English, patient was a minor, or the operation was for another type of tumor or lymph node recurrence. Of the 244 eligible patients, 9 patients deliberately refused. One hundred seventeen patients had complete data from the preoperative survey packet for a response rate of 48% (Fig. 1). Of the 117 patients with complete pretreatment data, 95 completed the postoperative survey packets, and 92 completed the follow-up survey. There was no significant difference in age, sex, race/ethnicity, or education level between respondents and nonrespondents (Table 1).

Study population

Seventy-two percent of the responders were female and mean age was 50 (±15), which corresponds with reported sex and age distributions in the thyroid cancer population (5). Consistent with the population at our institution, the majority of patients were white (92%), married/living with a partner (74%), and had at least a college education (71%), with 72% employed at least part-time. Of the respondents, 15% had a family history of thyroid cancer and 27% had a prior diagnosis of anxiety or depression.
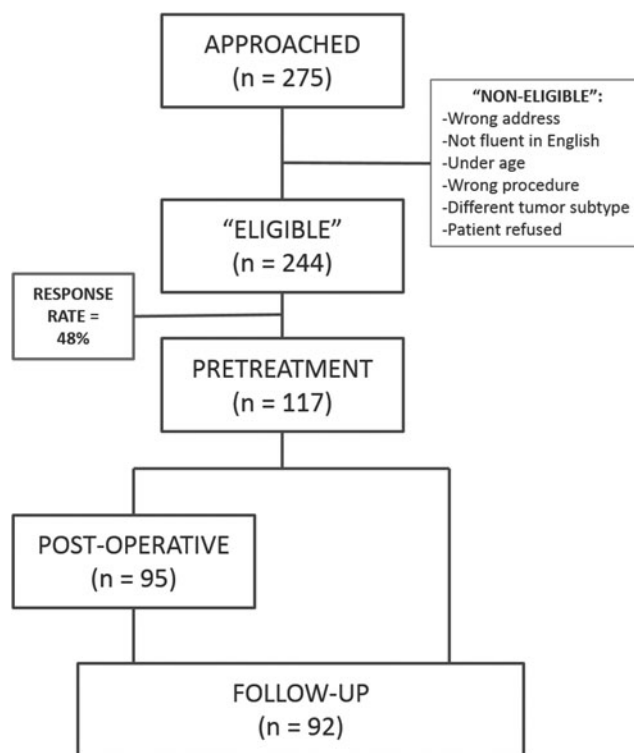

FIG. 1. Study flow diagram.

The majority of patients (57%) presented with an incidental finding on radiological imaging for an unrelated problem. Ninety-six percent of patients (112/117) had at least a total thyroidectomy with 49 undergoing concomitant lymph node dissection; 11% underwent a simultaneous modified radical neck dissection. The average size of the tumor was 1.8 (±0.9) cm. Consistent with known distributions of PTC subtypes, 61% had classical PTC, 16% had follicular variant of PTC, and 13% had aggressive variants (tall cell and diffuse sclerosing). Forty-one percent of the respondents had metastatic thyroid cancer to the lymph nodes on final pathology and 49% underwent postoperative radioactive iodine therapy (treatment given prior to the final survey). Three patients had distant metastatic disease at the time of diagnosis.

Operative complications were similar to reported incidences, with 5% of the population experiencing symptoms of hypocalcemia (i.e., perioral and finger numbness), 1%

TABLE 1. COHORT DEMOGRAPHICS

	Responders	Nonresponders	p Value
n (%)	117 (47.5)	127 (52.5)	
Age (mean ± SD)	49.8 ± 14.6	48.2 ± 16.8	0.42
18–39 years	27 (23.1)	43 (33.9)	0.12
40—64 years	74 (63.2)	64 (50.4)	
>65 years	16 (13.7)	20 (15.8)	
Males (n, %)	36 (28.4)	33 (28.2)	>0.99
White/caucasian	108 (92.3)	104 (88.1)	0.38
Latino	4 (3.4)	4 (3.4)	>0.99
College graduate/ post-graduate	83 (70.9)	66(68)	0.66

TABLE 2. PRETREATMENT SCORES FOR THE HEALTH-RELATED QUALITY OF LIFE INDEXES ($N=117$)

	Mean in U.S. normal population[a]	Mean	SD	Median	IQR
EQ5D-5L	0.87	0.895	0.103	0.876	[0.82–1]
SF6D	0.79	0.773	0.125	0.793	[0.66–0.86]
HUI2	0.84	0.875	0.133	0.917	[0.83–0.95]
HUI3	0.82	0.859	0.185	0.919	[0.79–0.97]

[a]Norms for females aged 45–54 years (53).
IQR, interquartile ratio; SD, standard deviation.

requiring reoperation for hematoma, 8% of patients experiencing hoarseness from recurrent laryngeal neuropraxia (two with continued symptoms at follow-up), and one patient with Horner's syndrome. Seven patients (6%) included in the study had structural disease persistence and/or early recurrence of disease.

Mean pretreatment overall scores ranged from 0.77 (SF6D) to 0.90 (EQ5D) (Table 2), reflective of the relatively young age and health of the thyroid cancer population. The HUI3 had the greatest range (0.10–1) and SD (±0.18). In contrast to the normal distribution of the SF6D results, the other indexes are skewed toward higher scores (Fig. 2). The EQ5D had 48 patients (41%) with a perfect index score, with >15% representing a significant ceiling effect, compared with only 3 (2.6%) of SF6D scores (40). Of the patients with a perfect pretreatment EQ5D score, 52% and 62% maintained this score at post-op and follow-up time points, respectively. The 3 patients with a perfect SF6D score showed a decline in scores on follow-up. There were no patients that reported the worst level of health (floor effect).

There was varied correlation with each instrument when assessing pairwise comparisons between indexes at each time point. Spearman correlation coefficients (ρ) were all highly significant ($p < 0.0001$) and good between each of the index comparisons at pre-op ($\rho=0.58$–0.91), post-op ($\rho=0.52$–0.77), and at follow-up evaluation ($\rho=0.53$—0.85). ICCs for absolute agreement for pretreatment and follow-up comparisons were poor to excellent (0.38–0.84), with the least cor-

FIG. 2. Mean scores for each of the four indexes at the three study time points (boxes represent standard deviation and whiskers represent 95% confidence interval).

relation between the SF6D and the other indexes and highest for the HUI2 versus HUI3. At the post-op time point, the same trends were seen with poor to good ICCs.

Table 3 shows the percentage of responses in each domain by level. The greatest impairments were in the emotional and mental health domains in each of the instruments. This spread to higher (i.e., worse) scores is noted in all of the indexes. In Table 4, the mean changes at key time points are reported. At all three time points, the SF6D had the highest proportion of responses with greater than reported minimal important differences. The standardized response means (SRM), reflecting the responsiveness (i.e., magnitude of change) of the index, were greatest at all three time points for the SF6D. Average post-op SF6D utility scores declined immediately following surgery ($p=0.007$) and improved from early post-op to follow-up ($p=0.001$). No significant differences were seen with any instrument when comparing pretreatment to follow-up scores.

The ICC agreement comparing preoperative with follow-up scores in in the group of patients with no expected change on quality of life (test–retest) was 0.43 (EQ5D), 0.51 (SF6D), 0.53 (HUI2), and 0.46 (HUI3) (Table 5), representing moderate to good correlation.

There was a significant effect of the instrument used on the scores ($p < 0.001$) by repeated measures ANOVA, but no effect of time point or the interaction of instrument and time point on scores. When comparing all patients expected to have decreasing scores compared with pretreatment (i.e., those with surgical complications, receiving RAI, or experiencing disease recurrence/persistence), there were no significant differences in mean change between key time points with any of the instruments. When comparing patients with RAI versus without RAI, the HUI2 showed a significant difference between postop to follow-up ($p=0.03$). When comparing patients with recurrence/persistence against those with localized disease, the SF6D and HUI3 showed significant differences between pre-op to post-op ($p=0.03$ and $p=0.03$).

Quality adjusted life-year estimates for patients with papillary thyroid carcinoma undergoing initial treatment (+/−) adjuvant RAI based on clinical guidelines were 0.786 (SF6D), 0.904 (EQ5D), 0.858 (HUI2), and 0.877 (HUI3).

Patients were asked the open-ended question, ''What concerns you most about your diagnosis and treatment?'' At the pre-op time point, most concerns related to complications from surgery, changes in lifestyle secondary to treatment, need for lifelong thyroid hormone replacement, and disease prognosis. The proportion of patients reporting concerns decreased but the similar worries about disease recurrence
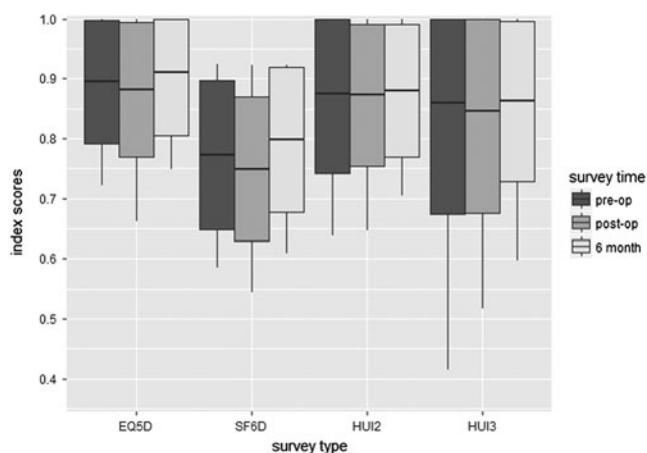
TABLE 3. PERCENTAGE OF RESPONSES BY INDEX AND DOMAIN AT THREE TIME POINTS

	Percentage of responses by level[a]																			
Instrument and domain	Pretreatment						Postoperative						Follow-up					
EQ5D	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5	
Mobility	93.2	6.0	0.9	–	–		93.7	5.3	–	1.1	–		92.4	4.3	2.2	–	1.1	
Self-care	100	–	–	–	–		95.8	2.1	1.1	–	1.1		99.0	1.1	–	–	–	
Usual activities	88.0	6.8	4.3	0.9	–		75.8	17.9	5.3	1.1	–		82.6	15.2	2.2	–	–	
Pain/discomfort	68.4	23.9	6.8	0.9	–		52.6	36.8	9.5	1.1	–		68.5	25.0	6.5	–	–	
Anxiety/Depression	47.0	37.6	12.8	1.7	0.9		65.3	26.3	4.2	4.2	–		65.2	28.3	5.4	1.1	–	
SF6D	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Physical functioning	82.9	12.0	5.1	–	–	–	72.6	20.0	7.4	–	–	–	79.3	17.4	3.3	–	–	–
Role limitations	42.7	11.1	19.7	26.5			20.0	31.6	12.6	35.6			47.8	15.2	13.0	23.4		
Pain	68.4	26.5	3.4	0.9	0.9	–	50.1	34.5	6.3	6.3	2.1	–	69.6	19.6	9.8	1.1	–	–
Vitality	6.8	49.6	23.1	14.5	6.0		5.3	42.1	31.6	15.8	5.3		5.4	47.8	30.4	13.0	3.3	
Mental health	32.5	35.9	22.2	7.7	1.7		40.0	40.0	14.7	5.3	–		45.6	37.0	14.1	2.2	1.1	
Social functioning	59.0	21.4	12.8	5.1	1.7		51.6	31.6	9.5	6.3	1.1		70.7	16.3	7.6	3.3	2.2	
HUI2	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Sensation	41.0	53.9	5.1	–		–	41.1	54.7	4.2	–		–	37.0	51.1	11.9	–		–
Mobility	93.2	6.8	–	–	–	–	89.5	10.5	–	–		–	91.3	8.7	–	–	–	–
Emotion	58.1	31.6	6.8	0.9	2.6		62.1	34.7	2.1	–	1.1		66.3	29.3	3.3	–	1.1	
Cognition	75.2	22.2	2.6	–		–	70.5	27.4	2.1	–		–	73.9	26.1	–	–		–
Self-care	99.2	–	–	0.9			97.9	1.1	–	1.1			97.8	1.1	1.1	–		
Pain	62.4	25.6	12.0	–	–		53.7	25.3	20.0	1.1	–		53.3	33.7	13.0	–	–	
HUI3	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Vision	41.9	55.6	0.9	1.7	–	–	41.0	59.0	–	–	–	–	38.0	54.3	2.2	4.3	1.1	–
Hearing	94.9	3.4	0.9	0.9	–	–	96.8	1.1	2.1	–	–	–	93.5	4.3	2.2	–	–	–
Speech	97.4	2.6	–	–	–		95.8	3.2	–	1.1	–		95.7	4.3	–	–	–	–
Cognition	75.2	7.7	12.0	2.6	2.6	–	70.5	12.6	9.5	4.2	1.1	–	73.9	6.5	13.3	6.5	–	–
Emotion	66.7	23.1	7.7	1.7	0.9		69.5	22.1	7.4	1.1	–		70.7	23.9	4.3	1.1	–	
Pain	62.4	25.6	9.4	2.6	–		53.7	25.3	16.8	3.2	1.1		53.3	33.7	8.7	4.3	–	
Ambulation	94.9	5.1	–	–	–	–	96.8	3.2	–	–	–	–	95.7	4.3	–	–	–	–
Dexterity	94.9	5.1	–	–	–	–	96.8	3.2	–	–	–	–	96.7	3.3	–	–	–	–

[a]Level 1 represents the ''best'' and the highest number level represents the ''worst.''
EQ5D, EuroQol-5D-5L; HUI2, Health Utility Indices Mark 2; HUI3, Health Utility Indices Mark 3; SF6D, Short Form-12v2/SF6D health survey.

and long-term medication remained at both the post-op and follow-up time points (Fig. 3).

Discussion

Our study assessed the HRQoL impact of the initial diagnosis and treatment in a group of patients with papillary thyroid carcinoma. To our knowledge, this is the first study to follow patients longitudinally; capturing patient-reported health related quality of life from diagnosis during the first year of treatment. Additionally, we performed a comparative analysis of four commonly used generic (i.e., across diseases) HRQoL instruments. As we hypothesized, the SF6D was the most responsive index with a decline in scores at the post-op time point and return to pretreatment score at follow-up and maintained the highest proportion of score greater than the MID. However, by Cohen's criteria (0.2 to 0.5 regarded as small, 0.5 to 0.8 as moderate, and >0.8 as large), the SRM for the SF6D in this application is in the small range (41).

The SRM is a frequently used measure of responsiveness and therefore it is presented here. The ICC is often used for assessing relative test–retest reliability, as is SEM, which is

TABLE 4. HEALTH-RELATED QUALITY OF LIFE SCORE CHANGES AT KEY TIME POINTS IN THE INITIAL TREATMENT

	MID	Mean Δ $T_1 \rightarrow T_2$ (SD)	Patients mean Δ > MID	SRM $T_1 \rightarrow T_2$	Mean Δ $T_2 \rightarrow T_3$ (SD)	Patients mean Δ > MID	SRM $T_2 \rightarrow T_3$	Mean Δ $T_2 \rightarrow T_3$ (SD)	Patients mean Δ > MID	SRM $T_1 \rightarrow T_3$
EQ5D[a]	0.052	0.022 (0.12)	53%	0.185	(−0.027) (0.12)	48%	−0.229	(−0.003) (0.11)	41%	−0.031
SF6D	0.027	0.038 (0.13)	81%	0.286	(−0.049) (0.13)	80%	−0.372	(−0.005) (0.12)	74%	−0.046
HUI2	0.045	0.008 (0.14)	59%	0.055	(−0.006) (0.09)	61%	−0.059	0.012 (0.12)	58%	0.097
HUI3	0.032	0.021 (0.19)	63%	0.112	(−0.014) (0.15)	67%	−0.091	0.019 (0.16)	60%	0.121

[a]For the EQ5D-5L (0.5 × SD pre-op) considered threshold for MID.
MID, minimal important difference; SRM, standardized response mean; T_1, pretreatment; T_2, postoperative; T_3, follow-up.

TABLE 5. TEST–RETEST RELIABILITY IN PATIENTS WITH SCORES EXPECTED TO REMAIN UNCHANGED OVER TIME

| | Pretreatment (n = 117) | | Test–retest[a] (n = 44) | | |
	Pre-op mean (SD)	Percent at ceiling	ICC	SEM	Spearman ρ
EQ5D	0.895 (0.103)	41.0	0.43 (0.15–0.64)	0.059	0.47
SF6D	0.773 (0.125)	2.6	0.51 (0.25–0.70)	0.050	0.43
HUI2	0.875 (0.133)	21.4	0.53 (0.28–0.71)	0.048	0.70
HUI3	0.859 (0.185)	24.8	0.46 (0.19–0.67)	0.056	0.66

[a]Assessed from pre-op to follow-up.
ICC, intraclass correlation coefficient; SEM, standard error of measurement.

an absolute reliability measure. The SEM is also compared with the MID. When evaluating the usefulness of a preference-index, experts often advocate presenting and comparing both the MID and SEM in assessing the clinical relevance of a score (33,42). In our case, the MID was the lowest for the SF6D and the SEM for the SF6D was second to lowest, being minimally (0.002) higher than the HUI2.

This work reflects HRQoL of patients undergoing treatment for PTC within the first year of treatment. While our data suggest that HRQoL returns to baseline at the follow-up visit, other studies suggest that the detriment may be prolonged. Qualitative and neuropsychometric analyses support the detriment to HRQoL in this population despite the misnomer ''the good cancer'' (13,43–46). In well-powered, cross-sectional, case-control studies of disease-free survivors of thyroid cancer, survivors in Korea (median follow-up 2.7 years) and the Netherlands (median follow-up 6.3 years) reported decreased scores compared with matched controls using validated neuropsychometric questionnaires (45,46). In another cross-sectional survey of 318 survivors over a mean 11-year follow-up, over half of patients continued to experience health deficits attributable to their cancer ten years following diagnosis and double the number of patients complained of health effects in comparison to patients with other types of cancers (18). Another cross-sectional study from Austria reported that the ''vitality'' and ''role-emotional'' domains remain reduced years after treatment in patients without evidence of disease (44).

Our study did not show a progressive change in HRQoL for those patients with expected decrements in QoL as a whole (i.e., RAI treatment, major surgical complications, and per-

sistent or recurrent disease). When looking at the effect of receiving RAI alone compared with those that did not, we did not see a significant difference between pretreatment and follow-up with any index, nor did we see a difference from post-op to follow-up except for the HUI2 scores. Others have reported a dose-related effect of RAI (18,43). This may be due to the fact that the majority of patients at our institution are prepared for RAI with recombinant TSH (rhTSH, Thyrogen®). Other trials assessing comparative outcomes for RAI preparation (i.e. thyroid hormone withdrawal versus rhTSH) report SF6D data (derived from SF36) indicating improved scores for those receiving rhTSH as moderate and stable following RAI treatment (0.71) (47,48). As expected, when comparing patients with persistent disease or an early recurrence with those with localized disease, the SF6D and HUI3 showed significant differences between pre-op to post-op time points. Additionally, we did not find a difference in scores for those patients having surgical complications including vocal cord paralysis or hypoparathyroidism. In our study, we had very few major complications documented (n = 10), and only 1% were permanent. To assess whether we lost patients with worse health to follow-up, we performed a comparison of the proportion of patients with and without expected decrease in HRQoL. No differences in completion of follow-up surveys were found.

Gallop and colleagues performed interviews to identify concepts affecting HRQoL in patients carrying a PTC diagnosis. These included treatment side effects, negative impact on personal relationships, mobility, daily activities, personal finances, work, and overall outlook. The effect seen was greater in patients with more advanced disease, however, all
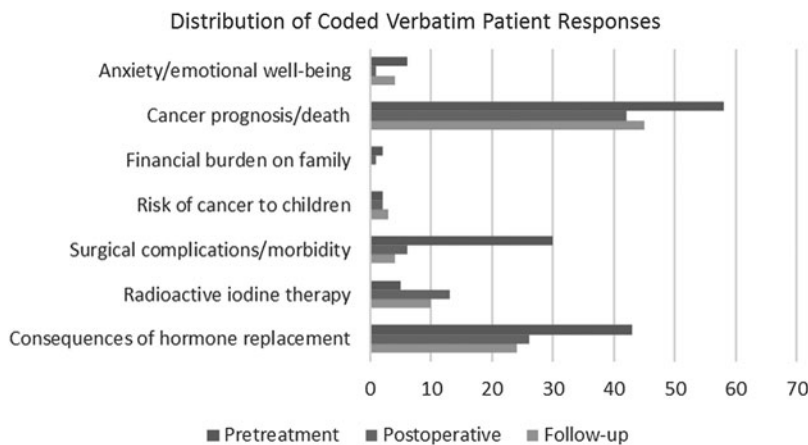

FIG. 3. Patient responses to open-ended question regarding their chief concerns for their diagnosis and treatment.

patients were affected (20). In the most recent cross-sectional survey of over 1000 thyroid cancer survivors, Grogan and colleagues reported a lower average overall score (5.6/10) compared with more aggressive cancers using a thyroid-specific neuropsychometric survey (City of Hope–QoL) (19). Additionally, patients reported a much higher prevalence of surgical complications and treatment related side effects than standard references.

Preference-based instruments assess the societally weighted (in the usual case) patient impressions of the effects of an intervention on quality of life. Each construct combines multidimensional health states to inform a summary index score on a scale of zero to one. Instruments differ in their conceptual frameworks. For instance, the EQ5D and SF6D are considered to reflect the influence of health on activities and mental health, while the HUI measures are focused on ''within skin'' effects. Further, the index algorithms for the EQ5D and SF6D are based on multiple regression, while the HUI instruments derive from multiattribute utility theory. Brazier *et al.* compared the intraclass correlation coefficient between the EQ5D and SF6D in seven different diseases. While there was a moderate correlation between the measures, there was more of a floor effect for SF6D and more ceiling effect with EQ5D (49). Others have noted the lack of interchangeability between instruments (50). When comparing six utility-based indexes, Hanmer and colleagues report significant variation between the instruments based on the impact of disease, concluding that scores using different instruments are not directly transferable (51).

Our results support the use of SF6D as the preferred generic preference-based tool for assessment of HRQoL in the thyroid cancer population. Given that this population, in general, is more physically healthy and young, we would expect there to be a skew toward higher scores. This is underscored by the high ceiling effect of the EQ5D. Likewise, based on our clinical experience, there is relatively less physical discomfort associated with PTC treatment, which may explain the relatively poor responsiveness of the HUI instruments. Gallop and colleagues performed a qualitative analysis of a group of PTC patients to assess the validity of the EQ5D and SF6D. By mapping responses to index domains, they concluded that the SF6D was more sensitive to a broad range of HRQoL impacts in the thyroid cancer population (20).

It is not surprising that the SF6D, although the most sensitive instrument in detecting change in HRQoL, still shows a relatively small SRM (reflective of the magnitude). This result might also suggest that generic instruments are missing morbidity that a disease-specific preference-based index might capture. The development and implementation of thyroid-specific neuropsychometric questionnaires, such as thyroid-specific patient report outcome measure (ThyPRO) and a disease-specific health-related quality of life questionnaire (THYCA-QoL) have been important in identifying relevant quality of life issues in this population (13,52). These reports have highlighted the importance of the patients' perspectives of well-being in thyroid survivorship care; however, these data are not translatable to values useful for comparative effectiveness research. On the other hand, a disease-specific instrument employable for comparative effectiveness research (i.e., translatable to quality-adjusted life-years) that covered a good balance of thyroid symptoms and

HRQoL concepts would be quite useful in thyroid treatment trials and, if developed with the intent of wider user access (like the EQ5D), would provide a substantial cost savings to investigators compared with the proprietary SF-suite of instruments (SF-36v2, SF-12v2/SF6D).

We acknowledge that there are limitations to our study. The cohort of study participants may not entirely reflect the U.S. population, in that our sample lacks the diversity of race/ethnicity and socioeconomic status of the general U.S. population. However, the high proportion of women and age are, indeed, reflective of the population of thyroid cancer survivors in the United States. (5). Our efforts to identify patients and survey them prior to surgery was successful in only 48% of eligible cases. While respondents and nonrespondents did not differ in our demographic analyses, it is possible that the prospect of answering a lengthy survey in the time window of accrual was a burden to some patients. It may also be that our survey did not reach patients who had near dates for surgery in a timely way. Future studies should take care to administer the pretreatment survey as close to the initial consultation as possible.

In conclusion, our results reflect the general good health of thyroid cancer patients and effects on quality of life seem primarily related to emotional and social impacts of treatment in the thyroid cancer population. Of the instruments assessed, the SF6D is the most responsive to treatment effects and should be utilized in future economic analyses in the thyroid cancer population. Sample sizes should reflect the expected effect size in such analyses. Lastly, development of a thyroid cancer–specific instrument that covered a good balance of thyroid symptoms and HRQoL concepts may be warranted to more accurately measure quality-of-life effects in this population.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Surveillance Epidemiology and End Results. http://seer.cancer.gov (accessed January 26, 2016).
2. Pellegriti G, Frasca F, Regalbuto C, Squatrito S, Vigneri R 2013 Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. J Cancer Epidemiol **2013:**965212.
3. Chen AY, Jemal A, Ward EM 2009 Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. Cancer **115:**3801–3807.
4. Hughes DT, Haymart MR, Miller BS, Gauger PG, Doherty GM 2011 The most commonly occurring papillary thyroid cancer in the United States is now a microcarcinoma in a patient older than 45 years. Thyroid **21:**231–236.
5. National Cancer Institute at the National Institutes of Health: Thyroid Cancer Statistics. http://seer.cancer.gov/statfacts/html/thyro.html (accessed January 27, 2016).

6. Welch HG, Black WC 2010 Overdiagnosis in cancer. J Natl Cancer Inst **102:**605–613.

7. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L 2016 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid **26:**1–133.

8. Ito Y, Miyauchi A, Inoue H, Fukushima M, Kihara M, Higashiyama T, Tomoda C, Takamura Y, Kobayashi K, Miya A 2010 An observational trial for papillary thyroid microcarcinoma in Japanese patients. World J Surg **34:**28–35.

9. Oda H, Miyauchi A, Ito Y, Yoshioka K, Nakayama A, Sasai H, Masuoka H, Yabuta T, Fukushima M, Higashiyama T, Kihara M, Kobayashi K, Miya A 2016 Incidences of unfavorable events in the management of low-risk papillary microcarcinoma of the thyroid by active surveillance versus immediate surgery. Thyroid **26:**150–155.

10. Brito JP, Ito Y, Miyauchi A, Tuttle RM 2016 A clinical framework to facilitate risk stratification when considering an active surveillance alternative to immediate biopsy and surgery in papillary microcarcinoma. Thyroid **26:**144–149.

11. Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, Barletta JA, Wenig BM, Al Ghuzlan A, Kakudo K, Giordano TJ, Alves VA, Khanafshar E, Asa SL, El-Naggar AK, Gooding WE, Hodak SP, Lloyd RV, Maytal G, Mete O, Nikiforova MN, Nose V, Papotti M, Poller DN, Sadow PM, Tischler AS, Tuttle RM, Wall KB, LiVolsi VA, Randolph GW, Ghossein RA 2016 Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. JAMA Oncol 1;2:1023–1029.

12. Grogan RH, Kaplan SP, Cao H, Weiss RE, Degroot LJ, Simon CA, Embia OM, Angelos P, Kaplan EL, Schechter RB 2013 A study of recurrence and death from papillary thyroid cancer with 27 years of median follow-up. Surgery **154:**1436–1446; discussion 1446–1437.

13. Husson O, Haak HR, Oranje WA, Mols F, Reemst PH, van de Poll-Franse LV 2011 Health-related quality of life among thyroid cancer survivors: a systematic review. Clin Endocrinol **75:**544–554.

14. Dow KH, Ferrell BR, Anello C 1997 Quality-of-life changes in patients with thyroid cancer after withdrawal of thyroid hormone therapy. Thyroid **7:**613–619.

15. Mendoza A, Shaffer B, Karakla D, Mason ME, Elkins D, Goffman TE 2004 Quality of life with well-differentiated thyroid cancer: treatment toxicities and their reduction. Thyroid **14:**133–140.

16. Sawka AM, Goldstein DP, Brierley JD, Tsang RW, Rotstein L, Ezzat S, Straus S, George SR, Abbey S, Rodin G, O'Brien MA, Gafni A, Thabane L, Goguen J, Naeem A, Magalhaes L 2009 The impact of thyroid cancer and postsurgical radioactive iodine treatment on the lives of thyroid cancer survivors: a qualitative study. PloS One **4:**e4191.

17. Bresner L, Banach R, Rodin G, Thabane L, Ezzat S, Sawka AM 2015 Cancer-related worry in Canadian thyroid cancer survivors. J Clin Endocrinol Metab **100:**977–985.

18. Schultz PN, Stava C, Vassilopoulou-Sellin R 2003 Health profiles and quality of life of 518 survivors of thyroid cancer. Head Neck **25:**349–356.

19. Aschebrook-Kilfoy B, James B, Nagar S, Kaplan S, Seng V, Ahsan H, Angelos P, Kaplan EL, Guerrero MA, Kuo JH, Lee JA, Mitmaker EJ, Moalem J, Ruan DT, Shen WT, Grogan RH 2015 Risk factors for decreased quality of life in thyroid cancer survivors: initial findings from the North American Thyroid Cancer Survivorship Study. Thyroid **25:**1313–1321.

20. Gallop K, Kerr C, Simmons S, McIver B, Cohen EE 2015 A qualitative evaluation of the validity of published health utilities and generic health utility measures for capturing health-related quality of life (HRQL) impact of differentiated thyroid cancer (DTC) at different treatment phases. Qual Life Res **24:**325–338.

21. EuroQol Group 1990 EuroQol–a new facility for the measurement of health-related quality of life. Health Policy **16:**199–208.

22. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X 2011 Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res **20:**1727–1736.

23. Brazier JE, Roberts J 2004 The estimation of a preference-based measure of health from the SF-12. Med Care **42:**851–859.

24. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M 2002 Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care **40:**113–128.

25. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q 1996 Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Med Care **34:**702–722.

26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG 2009 Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform **42:**377–381.

27. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS 2012 Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health **15:**708–715.

28. Hawthorne G, Richardson J, Day NA 2001 A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Ann Med **33:**358–370.

29. Johnson JA, Coons SJ 1998 Comparison of the EQ-5D and SF-12 in an adult US sample. Qual Life Res **7:**155–166.

30. Kopec JA, Willison KD 2003 A comparative review of four preference-weighted measures of health-related quality of life. J Clin Epidemiol **56:**317–325.

31. O'Brien BJ, Spath M, Blackhouse G, Severens JL, Dorian P, Brazier J 2003 A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Econ **12:**975–981.

32. Kaplan RM, Groessl EJ, Sengupta N, Sieber WJ, Ganiats TG 2005 Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. Med Care **43:**79–87.

33. Palta M, Chen HY, Kaplan RM, Feeny D, Cherepanov D, Fryback DG 2011 Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. Med Decis Making **31:**260–269.

34. Andridge RR, Little RJ 2010 A Review of Hot Deck Imputation for Survey Non-response. Int Stat Rev **78:**40–64.

35. McNamee P, Seymour J 2005 Comparing generic preference-based health-related quality-of-life measures: advancing the research agenda. Expert Rev Pharmacoecon Outcomes Res **5:**567–581.

36. Pickard AS, Johnson JA, Feeny DH 2005 Responsiveness of generic health-related quality of life measures in stroke. Qual Life Res **14:**207–219.

37. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR 2002 Methods to explain the clinical significance of health status measures. Mayo Clinic Proc **77:**371–383.

38. Luo N, Johnson J, Coons SJ 2010 Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. Med Care **48:**365–371.

39. Norman GR, Sloan JA, Wyrwich KW 2003 Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care **41:**582–592.

40. McHorney CA, Tarlov AR 1995 Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res **4:**293–307.

41. Cohen J 1988 Statistical Power Analysis for the Behavioural Sciences, second edition. Lawrence Earlbaum Associates, Hillsdale, NJ.

42. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC 2009 Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol **62:**1062–1067.

43. Almeida JP, Vartanian JG, Kowalski LP 2009 Clinical predictors of quality of life in patients with initial differentiated thyroid cancers. Arch Otolaryngol Head Neck Surg **135:**342–346.

44. Crevenna R, Zettinig G, Keilani M, Posch M, Schmidinger M, Pirich C, Nuhr M, Wolzt M, Quittan M, Fialka-Moser V, Dudczak R 2003 Quality of life in patients with non-metastatic differentiated thyroid cancer under thyroxine supplementation therapy. Support Care Cancer **11:**597–603.

45. Hoftijzer HC, Heemstra KA, Corssmit EP, van der Klaauw AA, Romijn JA, Smit JW 2008 Quality of life in cured patients with differentiated thyroid carcinoma. Journal Clin Endocrinol Metab **93:**200–203.

46. Lee JI, Kim SH, Tan AH, Kim HK, Jang HW, Hur KY, Kim JH, Kim KW, Chung JH, Kim SW 2010 Decreased health-related quality of life in disease-free survivors of differentiated thyroid cancer in Korea. Health Qual Life Outcomes **8:**101.

47. Mernagh P, Campbell S, Dietlein M, Luster M, Mazzaferri E, Weston AR 2006 Cost-effectiveness of using recombinant human TSH prior to radioiodine ablation for thyroid cancer, compared with treating patients in a hypothyroid state: the German perspective. European journal of endocrinology / European Federation of Endocrine Societies **155:**405–414.

48. Pacini F, Ladenson PW, Schlumberger M, Driedger A, Luster M, Kloos RT, Sherman S, Haugen B, Corone C, Molinaro E, Elisei R, Ceccarelli C, Pinchera A, Wahl RL, Leboulleux S, Ricard M, Yoo J, Busaidy NL, Delpassand E, Hanscheid H, Felbinger R, Lassmann M, Reiners C 2006 Radioiodine ablation of thyroid remnants after preparation with recombinant human thyrotropin in differentiated thyroid carcinoma: results of an international, randomized, controlled study. J Clin Endocrinol Metab **91:**926–932.

49. Brazier J, Roberts J, Tsuchiya A, Busschbach J 2004 A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ **13:**873–884.

50. Conner-Spady B, Suarez-Almazor ME 2003 Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med Care **41:**791–801.

51. Hanmer J, Cherepanov D, Palta M, Kaplan RM, Feeny D, Fryback DG 2016 Health Condition Impacts in a Nationally Representative Cross-Sectional Survey Vary Substantially by Preference-Based Health Index. Med Decis Making **36:**264–274.

52. Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, Feldt-Rasmussen U 2009 Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. Qual Life Res **18:**483–496.

53. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, Herrington SA, Hays RD, Kaplan RM, Ganiats TG, Feeny D, Kind P 2007 US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Med Care **45:**1162–1170.

Address correspondence to:
Carrie C. Lubitz, MD, MPH
Department of Surgery
Massachusetts General Hospital
55 Fruit Street
Yawkey 7B
Boston, MA 02114

E-mail: clubitz@partners.org