



## Surgical Outcomes

Validation of a novel patient-reported outcome measure for parathyroid and thyroid disease (PROMPT)<sup>☆</sup>

Talia Burneikis, MD\*, Jennifer Colvin, MD, Judy Jin, MD, Eren Berber, MD, Vikram D Krishnamurthy, MD, Joyce Shin, MD, Allan Siperstein, MD

Cleveland Clinic Foundation



## ARTICLE INFO

## Article history:

Accepted 3 April 2018

Available online 3 November 2018

## ABSTRACT

**Background:** Patient-reported outcome measures are being used increasingly to assess disease severity and response to surgery. The purpose of this study was to create and validate a patient-reported outcome measure for symptoms of thyroid enlargement and hyperparathyroidism, 2 conditions where the presence of preoperative symptoms and response to surgery is often questioned.

**Methods:** A questionnaire reviewing common symptoms was developed from a literature review and expert opinion. Internal validity, reliability, and initial responsiveness to surgery were evaluated.

**Results:** Patient-Reported Outcome Measure for Parathyroid and Thyroid Disease (PROMPT) consists of 30 items: 10 compressive items and 20 hyperparathyroidism items; we evaluated 302 surveys collected over 10 months. PROMPT showed high internal consistency for compressive and hyperparathyroid constructs (Cronbach's  $\alpha$  0.84 and 0.95). Constructs were scored from 0–100, with greater scores corresponding to increased severity of symptoms. Preoperatively, patients with a goiter demonstrated greater compressive scores compared with other thyroid patients and hyperparathyroid patients (goiter, 47.5; nodule/other, 38.4; hyperparathyroid, 29.8;  $P < .0001$ ). PROMPT demonstrated high test–retest reliability with acceptable intraclass correlation coefficients for both compressive score and hyperparathyroid score (0.840 and 0.646). Hyperparathyroid scores improved 2 weeks after (48.6 postop, 44.0 preop,  $P = .0470$ ).

**Conclusion:** We validated a novel measure for symptoms of hyperparathyroidism and goiter. PROMPT demonstrates high internal consistency, test–retest reliability, and preliminary analysis, which suggests that it is sensitive to change after surgery.

© 2018 Elsevier Inc. All rights reserved.

## Background

Assessing the quality of patient care takes many forms, from objectively collected metrics on preventable complications to more patient-centered, albeit subjective, surveys. There is increasing emphasis being put on the impact of care as perceived by the patient and the use of patient-reported outcome measures (PROMs) to assess disease severity and response to treatment including surgery. PROMs provide a means of eliciting the patient's perspective and translating that information into objective data. Endocrine surgery represents a field where the opportunity for using PROMs is ample, and the impact potential is great because of the nuances of certain disease states.

Hyperparathyroidism is characterized by overproduction of parathyroid hormone from 1 or more autonomously functioning parathyroid glands. The severe downstream effects of excess parathyroid hormone are classically described as osteitis fibrosa cystica, nephrocalcinosis, peptic ulcer disease, and proximal myopathy; in addition, more subtle neuropsychiatric symptoms are well described, including bone pain, fatigue, depressed mood, irritability, anxiety, and poor concentration.<sup>1</sup> Patients presenting without clinically abnormal laboratory findings (ie, normocalcemic hyperparathyroidism) or mild symptoms are often underdiagnosed or dismissed as “asymptomatic” despite multiple studies highlighting the risk of disease progression without intervention and the symptom benefit of parathyroidectomy.<sup>2,3,4</sup>

Another common endocrine problem, goiter, encompasses a spectrum of diseases, including endemic, nodular, and immune-mediated disease. Thyroid enlargement results in cosmetic disfigurement and compressive symptoms, including esophageal and tracheal compression in severe cases.<sup>5</sup> Yet, thyroid enlargement is a continuum. Improvement of subtle symptoms, (ie, dysphagia), after thyroidectomy is often unclear.

<sup>☆</sup> Presented at the annual meeting of the American Association of Endocrine Surgeons in Durham, North Carolina, May 6–8, 2018.

\* Corresponding author: Cleveland Clinic Foundation, General Surgery, 9500 Euclid Avenue, Mail Code F20, Cleveland, Ohio 44195.  
E-mail address: [burneit@ccf.org](mailto:burneit@ccf.org) (T. Burneikis).

The use of PROMs is of interest to multiple stakeholders in surgical care, including patients, surgeons, and insurers. Practitioners can increase patient engagement by using PROMs. Evaluation of preoperative symptoms using PROMs objectively informs the discussion regarding expected improvement from surgery. Patients can review their current health status in comparison to the average outcomes achieved by prior patients. Leading the way of continuous quality improvement, the American College of Surgeons is incorporating PROMs as another pillar of its quality programs and advocating their use in measuring surgical outcomes alongside National Safety and Quality Improvement Program measures.<sup>6</sup>

In the field of endocrine surgery, few tools designed to assess benign thyroid disease exist. Two such tools include the Thyroid Patient-Reported Outcome and the quality of life (QOL) questionnaire specific to Grave's ophthalmopathy (GO-specific QOL survey).<sup>7,8,9</sup> Validated PROMs, however, are still lacking in the areas of thyroid enlargement and hyperparathyroidism. The effect of parathyroidectomy on physical and mental health has been investigated thoroughly using a mixture of ad hoc surveys and various combinations of previously validated, generic measures.<sup>10,11</sup> Questionnaire-based studies have even shown symptom improvement after parathyroidectomy in patients with physiologically normocalcemic hyperparathyroidism.<sup>12</sup> The Pasiëka Illness Questionnaire, the most robustly validated tool in the current literature, was designed to assess the impact of parathyroidectomy on preoperative symptoms of primary hyperparathyroidism.<sup>13,14</sup> Similar studies have correlated symptom improvement after parathyroidectomy with generic, QOL outcome measures including the Patient Health Questionnaire-9, the Generalized Anxiety Disorder 7-item scale, and Short Form 36 Health Survey.<sup>13,15,16</sup> The major disadvantage to using these measures in the assessment of hyperparathyroidism is that they are either generic QOL measures or they are specific to other unrelated diseases that happen to share the vague neuropsychiatric symptoms associated with hyperparathyroidism. Most recently, the Primary Hyperparathyroidism Quality of Life questionnaire was developed and validated in Spain in 2016 and consists of 16 questions.<sup>17</sup> To our knowledge, no other group has created a tool that can be applied to both goiter and hyperparathyroid patients. We posited multiple advantages of creating a PROM incorporating both disease states. Foremost, each group serves as an internal control for the other because both populations undergo neck surgery, thus controlling for the placebo effect from surgery itself. In addition, goiter symptoms generally include focal compressive symptoms, in contrast to hyperparathyroidism, which is a systemic disease. Finally, the Patient-Reported Outcome Measure for Parathyroid and Thyroid Disease (PROMPT) can be applied to patients who present with mixed disease states.

Quantifying subjective symptoms otherwise impossible to measure explicitly requires rigorous evaluation through established methods of validation. Methods of validation are well described by the Patient-Reported Outcomes Measurement Information System of the National Institutes of Health.<sup>18</sup> In addition, the Scientific Advisory Committee of the Medical Outcomes Trust for health-related outcome measures has established guidelines to describe the principal attributes of measurement model, reliability, validity, and responsiveness.<sup>19</sup> A construct of interest must be defined (ie, compressive goiter symptoms). An instrument must produce consistent scores (reliability). Internal consistency is calculated with Cronbach's  $\alpha$ . Reproducibility refers to the test-retest reliability and is evaluated using intraclass correlation coefficients (ICC). Validity refers to the extent to which an instrument measures what it was intended to measure. Finally, responsiveness refers to a sensitivity of the instrument to change over time.

The primary purpose of this study was to design and validate a novel PROM to assess the constructs of hyperparathyroidism and compressive symptoms attributable to thyroid enlargement using

modern validation methods. We address the principles of measurement model, reliability, and validity. Initial responsiveness was examined but does not represent a major objective of this study.

## Methods

### Survey design

A questionnaire reviewing common symptoms of hyperparathyroidism and general compressive symptoms was developed from review of the literature and expert opinion.<sup>20–23</sup> Patients rated the frequency of symptoms according to a 5-point Likert scale (Never, Rarely, Sometimes, Often, Always). Similar symptoms with alternative wording were intentionally included to increase the reliability of testing a given construct. Negatively worded items (similar questions worded oppositely) were included to minimize response bias from automatic cognitive processing (checking the same box the whole way down). Finally, question order was randomized between constructs. The questionnaire was easily completed in less than 10 minutes.

### Survey administration and data collection

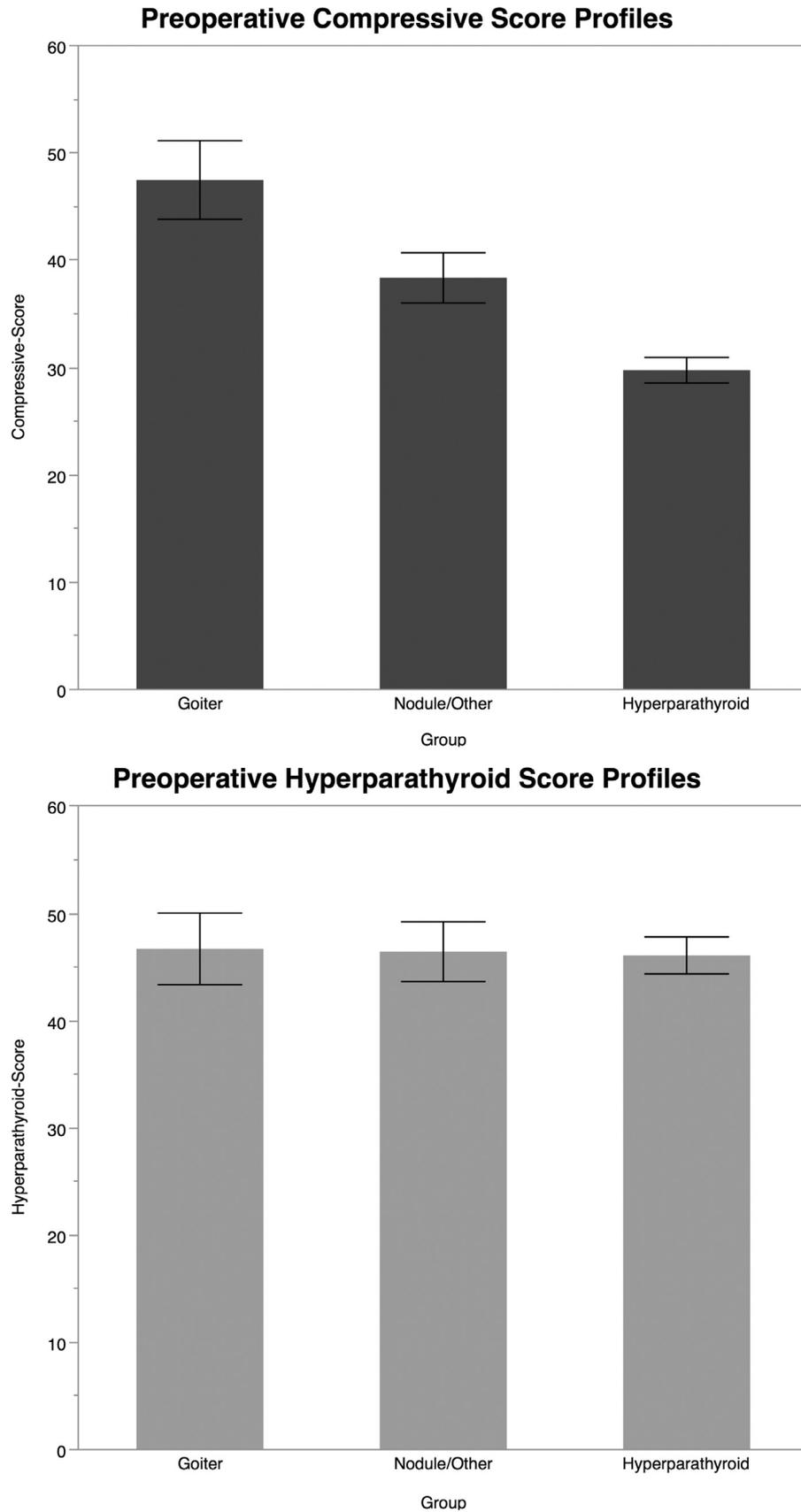
The Department of Endocrine Surgery at the Cleveland Clinic maintains a prospective database for all thyroid and parathyroid patients that is approved by an institutional review board and is compliant with the Health Insurance Portability and Accountability Act. After institutional review board approval, surveys were distributed to patients presenting with any thyroid or parathyroid complaints during their wait time. Results were collected from February to December 2017. Surveys were excluded for analysis based on the following criteria: patients with any prior neck surgery, patients presenting more than 6 months postoperatively, and surveys missing responses for greater than 50% of questions.

### Validation and scoring of the PROMPT

Internal reliability was evaluated. Question items were categorized according to construct. Cronbach's  $\alpha$  was calculated for questions to determine scale reliability and item redundancy. Individual items were examined for inter-item reliability and frequency of omission. There were 2 scores calculated based on PROMPT, a compressive score and a hyperparathyroid score. Each score is a weighted sum of the questions corresponding to the respective constructs. Scores range from 0 to 100, with greater scores corresponding to increased symptom severity. PROMPT includes twice as many hyperparathyroid questions as compressive symptoms questions because hyperparathyroid symptoms tend to be more nonspecific than compressive symptoms; however, each construct is weighted equally in the final scoring (both on a scale of 100). Preoperative scores were analyzed for discriminant validity using the Wilcoxon and Kruskal-Wallis (Rank Sums) test. Test-retest reliability was evaluated using ICC.

### Comparison of groups

Patients were grouped according to disease process (goiter, thyroid nodule or other thyroid, and hyperparathyroidism) and operative status (nonop, preop, and 2 weeks postop). Goiter was defined as obvious enlargement of the thyroid gland on physical exam or ultrasonography or any single nodule greater than 4 cm in size. The nodule or other thyroid group included patients having thyroid nodules without clinically relevant thyroid enlargement. For the purpose of initial validation of the PROMPT, the goiter and nodule or other thyroid groups did not separate hyperthyroid or



**Fig 1.** Preoperative compressive score and hyperparathyroid score profiles.

Preoperative compressive scores and hyperparathyroid scores were calculated and compared for all 3 patient groups (goiter, nodule or other thyroid, and hyperparathyroid). There was a difference in preoperative compressive scores among all groups (goiter,  $47.5 \pm 2.7$ ; nodule/other,  $38.4 \pm 2.1$ ; hyperparathyroid,  $29.8 \pm 1.5$ ;  $P < .0001$ ). There was no significant difference in preoperative hyperparathyroid scores among groups.

**Table 1**  
PROMPT questions and item-specific Cronbach's  $\alpha$ .

| Item | During the past month, how often have you:   | Domain      | Cronbach's $\alpha$ if Item Deleted |
|------|--|-------------|-------------------------------------|
| 1    | Had a feeling of fullness in your neck?  | compressive | 0.8                                 |
| 2    | Felt fatigued?   | fatigue     | 0.94                                |
| 3    | Been satisfied with your sleep?  | sleep       | 0.95                                |
| 4    | Felt depressed?  | mood        | 0.94                                |
| 5    | Noticed swelling or "a lump" in the front of your neck?  | compressive | 0.84                                |
| 6    | Felt exhausted?  | fatigue     | 0.94                                |
| 7    | Had trouble concentrating?   | thinking    | 0.94                                |
| 8    | Had muscle aches?  | aches       | 0.94                                |
| 9    | Had pain in the front of your neck?  | compressive | 0.83                                |
| 10   | Had trouble starting and finishing tasks because you were too tired?                             | fatigue     | 0.94                                |
| 11   | Experienced snoring at night?  | compressive | <b>0.87</b>                         |
| 12   | Had slow or unclear thinking?  | thinking    | 0.94                                |
| 13   | Felt like you had a lump in your throat?   | compressive | 0.8                                 |
| 14   | Had difficulty falling asleep?   | sleep       | 0.95                                |
| 15   | Had trouble staying asleep   | sleep       | 0.95                                |
| 16   | Felt unhappy?  | mood        | 0.95                                |
| 17   | Felt pressure in your neck while lying flat?   | compressive | 0.82                                |
| 18   | Had difficulty learning new things?  | thinking    | 0.95                                |
| 19   | Had bone and/or joint aches?   | aches       | 0.95                                |
| 20   | Had difficulty swallowing?   | compressive | 0.82                                |
| 21   | Felt that nothing was interesting?   | mood        | 0.95                                |
| 22   | Had hoarseness?  | compressive | 0.84                                |
| 23   | Felt weak?   | fatigue     | 0.95                                |
| 24   | Had to limit your activities because of weakness?  | fatigue     | 0.95                                |
| 25   | Been self-conscious about the appearance of your neck (due to swelling, mass/lump, and/or scar)? | compressive | 0.83                                |
| 26   | Wanted to sleep all the time?  | fatigue     | 0.95                                |
| 27   | Had difficulty remembering things?   | thinking    | 0.94                                |
| 28   | Had muscle or joint aches that affected your daily activities?                                   | aches       | 0.95                                |
| 29   | Had mood swings or irritability?   | mood        | 0.94                                |
| 30   | Tried to cover the appearance of your neck? (for example, wearing turtlenecks or scarves)        | compressive | <b>0.85</b>                         |

PROMPT is a Patient-reported outcome measure (PROM) designed to assess 2 constructs: compressive symptoms attributable to goiter and hyperparathyroid symptoms. Compressive symptom items appear in gray. Hyperparathyroid items appear in white. Cronbach's  $\alpha$ , a measure of internal consistency, was 0.84 for compressive symptom items and 0.95 for hyperparathyroid symptom items. Item-specific Cronbach's  $\alpha$  values appear in bold if removing the corresponding item from the measure would increase the internal consistency of PROMPT overall.

hypothyroid states (Graves' disease, toxic nodules, or Hashimoto's thyroiditis). Similarly, the hyperparathyroid group included all hyperparathyroid patients regardless of etiology (primary hyperparathyroidism, or renal-related hyperparathyroidism). Preoperative profiles were examined for each group. Then, PROMPT was used to assess responsiveness to surgery by comparing preoperative scores with the 2-week postoperative scores within each group. Paired observations were compared using Student's *t*-test.

## Results

### Survey overview

PROMPT consists of 30 items to avoid survey fatigue: 10 questions assessed the construct of compressive symptoms owing to

thyroid enlargement; 20 questions assessed the construct of hyperparathyroidism, including the domains of fatigue, sleep, mood, mental clarity, and body aches (Fig 1). The measure was field-tested over 10 months. After exclusions, 302 surveys were analyzed representing 235 unique patients (the number of surveys reflects additional surveys for the same visit obtained at either preoperative, postoperative, or nonoperative follow-up visits). A total of 10 patients presented with both hyperparathyroidism and thyroid disease (either goiter or nodule or other thyroid). Patient characteristics, including age, sex, and body mass index (BMI), were analyzed for newly presenting patients. For comparisons,  $\chi^2$  test was used for categorical variables and 1-way analysis of variance for continuous variables. There was no statistically significant difference in distribution of sex or BMI between groups (Table 1). Patients with hyperparathyroidism tended to be older than patients presenting

**Table 2**  
Baseline demographics.

|                          | Goiter <i>n</i> mean (SE) / % | Nodule/other thyroid <i>n</i> mean (SE) / % | Hyperparathyroidism <i>n</i> mean (SE) / % | <i>P</i> value |    |             |      |
|--------------------------|-------------------------------|---|--|----------------|----|-------------|------|
| Mean age (SE) (y)        | 33                            | 57 (2)                                      | 96   | 54 (1)         | 77 | 62 (2)      | <.01 |
| Sex                      |                               |   |  |                |    |             | <.2  |
| Male                     | 3                             | 9%  | 12   | 13%            | 16 | 21%         |      |
| Female                   | 30                            | 91%   | 84   | 88%            | 61 | 79%         |      |
| BMI (kg/m <sup>2</sup> ) |                               | 29.2 (1.2)                                  |  | 29.2 (0.7)     |    | 29.7 (0.80) | .91  |

SE, standard error

Note: Baseline demographics of our patient populations were examined. Patients with hyperparathyroidism were older than goiter or nodule or other thyroid patients. There was no significant difference in distribution based on sex or BMI.

**Table 3**  
Internal consistency of PROMPT by construct.

| Psychometric construct    | Number of items | Cronbach's $\alpha$ |
|---------------------------|-----------------|---------------------|
| Compressive symptoms      | 10              | 0.84                |
| Hyperparathyroid Symptoms | 20              | 0.95                |
| fatigue                   | 6               | 0.91                |
| sleep                     | 3               | 0.86                |
| mood                      | 4               | 0.87                |
| thinking                  | 4               | 0.90                |
| aches                     | 3               | 0.88                |

PROMPT consists of 2 psychometric constructs: compressive symptoms and hyperparathyroid symptoms. Hyperparathyroid symptoms are further categorized by subdomains (fatigue, sleep, mood, thinking, and aches). The number of items corresponding to each domain or subdomain are outlined here, and the respective Cronbach's  $\alpha$  was calculated. PROMPT demonstrated high levels of internal consistency with Cronbach's  $\alpha$  >0.8 in all domains or subdomains.

with thyroid disease (62 hyperparathyroidism, 57 goiter, 54 thyroid nodule or other thyroid;  $P=.0006$ ). [Table 2](#).

### Initial validation

Internal reliability was evaluated by calculating Cronbach's  $\alpha$ . When separated by constructs, each set of questions showed high internal consistency. Cronbach's  $\alpha$  was 0.84 for compressive symptom items and 0.95 for hyperparathyroid symptom items. Because PROMPT assesses 2 constructs in parallel (compressive symptoms and hyperparathyroid symptoms), the inter-item reliability was calculated for each construct. Hyperparathyroid subdomains (fatigue, sleep, mood, thinking, and body aches) were further analyzed for internal consistency and were all similarly high ([Table 3](#)).

### Individual item review

After demonstrating validity of PROMPT, additional evaluation of the relative contribution of individual items to the overall validity of the measure was assessed. To assess the impact of an individual item on the overall validity of the measure, Cronbach's  $\alpha$  was recalculated as if the given item were deleted. If Cronbach's  $\alpha$  increased, removal of that item should be considered. Recalculated Cronbach's  $\alpha$  coefficients after individual item deletion are shown in [Fig 1](#). There were 2 items in the compressive-score that showed poor inter-item reliability: 11 "experienced snoring at night" and 30 "tried to cover the appearance of your neck." Overall, hyperparathyroid symptom items showed high internal consistency without any single item warranting removal. The most frequently omitted questions included c11 ("experienced snoring at night"; 6.0% omitted), followed by c1 ("had a feeling of fullness in your neck"; 3.2% omitted), and c5 ("noticed swelling, or a 'lump' in the front of your neck"; 2.8% omitted).

### Discriminant validity and reliability

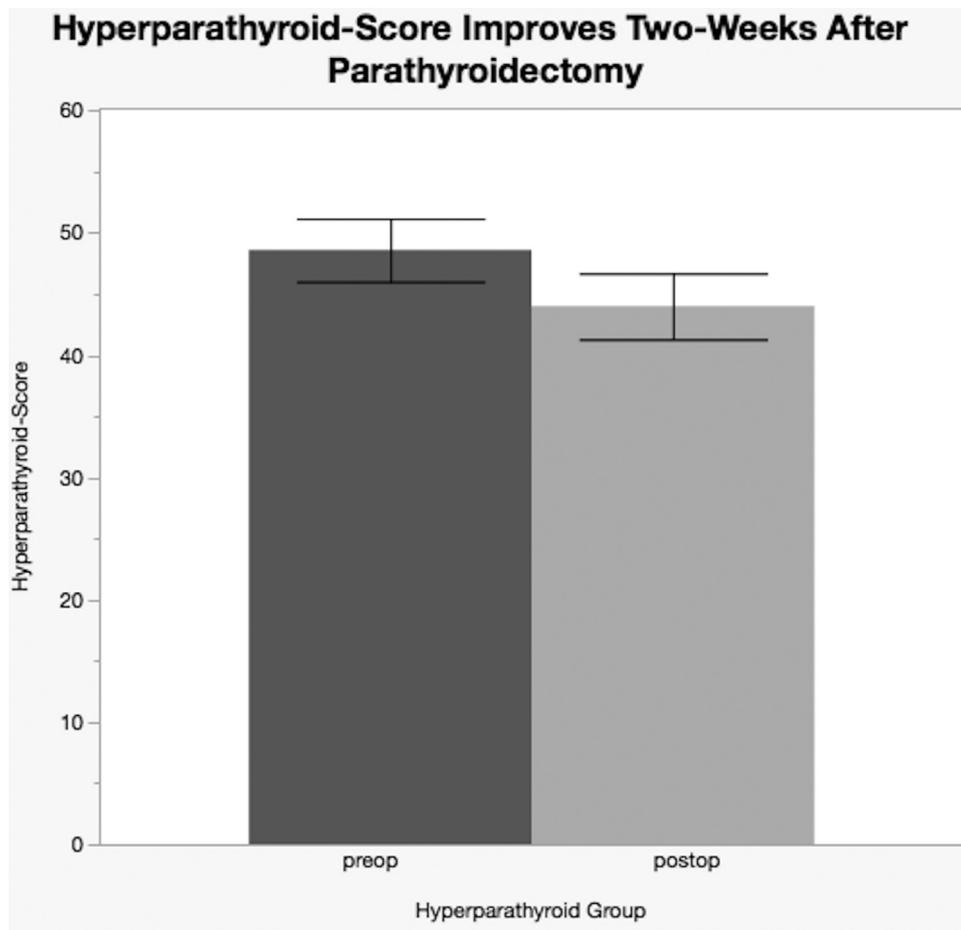
To assess the discriminant validity of whether PROMPT distinguishes between groups, a Wilcoxon and Kruskal-Wallis (Rank Sums) test was used to compare compressive scores and hyperparathyroid scores for surveys obtained at initial visit. Preoperatively, goiter patients demonstrated greater compressive scores when compared with nodule or other thyroid patients and hyperparathyroid patients (goiter,  $47.5 \pm 2.7$ ; nodule or other,  $38.4 \pm 2.1$ ; hyperparathyroid,  $29.8 \pm 1.5$ ;  $P < .0001$ ). There was no significant difference in preoperative hyperparathyroid scores between groups (goiter,  $46.7 \pm 3.2$ ; nodule/other,  $46.5 \pm 2.6$ ; hpt,  $46.09 \pm 1.83$ ;  $P=.984$ ; [Fig 2](#)).

Test-retest reliability was assessed by calculating ICC for compressive and hyperparathyroid scores among nonoperative patients followed longitudinally ( $n = 11$ ). ICCs were acceptable for compressive and hyperparathyroid scores (0.840 and 0.646).

### Discussion

The present study describes the development and validation of a novel Patient-Reported Outcome Measure for Parathyroid and Thyroid Disease (PROMPT), which can assess the symptoms of both hyperparathyroidism and goiter-related compressive symptoms. Multiple methods were used to evaluate key psychometric properties, including concept/measurement model, reliability, and validity. The results indicate that PROMPT has high internal reliability for both compressive and hyperparathyroidism constructs. Methodologists recommend a minimum Cronbach's  $\alpha$  coefficient of 0.7 to 0.8; thus PROMPT demonstrates good to excellent internal reliability.<sup>24</sup> Appropriately, PROMPT also has high discriminant validity. Using the compressive score, PROMPT differentiates between goiter, thyroid nodule, and hyperparathyroid patients. Importantly, PROMPT is reliable with acceptable ICCs among the few nonoperative patients who were followed longitudinally. Although not an explicit goal at the outset of this project, PROMPT was evaluated for initial responsiveness to surgery. In a matched-pairs analysis at 2 weeks postoperatively, hyperparathyroid patients showed improvement in hyperparathyroid symptoms ( $n = 27$ , preop hyperparathyroid score 48.6, postop hyperparathyroid score 44.0, mean difference  $-4.59$ , standard error 2.20,  $P = .0470$ ; [Table 4](#), [Fig 2](#)). At 2 weeks postoperatively, hyperparathyroid patients demonstrated statistically significant improvement in hyperparathyroid scores. In contrast, a similar improvement in compressive scores among goiter patients was not observed at 2 weeks postoperatively, likely owing to the shared cervical symptoms among patients with compressive goiters and patients having just recently undergone a cervical neck surgery. Greater follow-up is needed.

PROMPT represents a novel tool in the management of hyperparathyroidism and goiter. Furthermore, PROMPT was developed using specific methods of evaluation. Although Thyroid Patient-Reported Outcome has been rigorously validated for the assess-



**Fig 2.** Hyperparathyroid score improves 2 weeks after parathyroidectomy. Change in compressive scores and hyperparathyroid scores from baseline or preoperative evaluation and at 2 weeks postoperatively are shown. Statistically significant improvement in hyperparathyroid scores among hyperparathyroid patients was seen at 2 weeks postoperatively.

**Table 4**  
Change in scores after thyroidectomy or parathyroidectomy.

| Group               | n  | Score | Preop (mean) | Postop, 2wk (mean) | Change in score (mean ± SD) | P value |
|---------------------|----|-------|--------------|--------------------|-----------------------------|---------|
| Goiter              | 12 | c     | 43           | 38                 | -4.8 ± 27.5                 | .56     |
|                     | 12 | h     | 47           | 38                 | -8.8 ± 15.3                 | .08     |
| Nodule/other        | 9  | c     | 36           | 42                 | 5.6 ± 9.2                   | .11     |
|                     | 9  | h     | 46           | 47                 | 1.7 ± 6.2                   | .45     |
| Hyperparathyroidism | 27 | c     | 29           | 36                 | 7.0 ± 12.3                  | .01     |
|                     | 27 | h     | 49           | 44                 | -4.6 ± 11.5                 | .047    |

SD, standard deviation; c, compressive; h, hyperparathyroidism  
 Note: Change in compressive scores and hyperparathyroid scores from baseline/preoperative evaluation and at 2 weeks postoperatively are shown. Statistically significant improvement in hyperparathyroid scores among hyperparathyroid patients was seen at 2 weeks postoperatively. Compressive symptoms and hyperparathyroid symptoms showed minimal or no improvement at 2 weeks postoperatively.

ment of benign thyroid diseases, it is limited in its application because it requires an exhaustive, 84-question assessment. In contrast, PROMPT is feasible to administer quickly, and the compressive-symptom construct is specific to goiter patients. Another advantage of PROMPT is that it can be administered to patients with a range of types of endocrine disease and results can be compared with different disease states serving as control groups for each other. Specifically, PROMPT was distributed to 3 patient populations: goiter, nodules, and hyperparathyroid patients. A priori, we did not expect symptoms attributed to hyperparathyroidism to improve after thyroidectomy for goiter or small thyroid

nodules. In contrast, we did expect hyperparathyroid patients to have symptoms of neck discomfort during the immediate postoperative period because of the simple fact that they underwent a cervical neck operation.

The key purpose of this study was the creation and validation of PROMPT. At this time, we are unable to make conclusions regarding the long-term responsiveness to surgery because of the small sample of paired surveys received and the short follow-up period. Early analysis at 2 weeks after surgery shows moderate improvement, which is promising for future studies. Eventually, with a larger dataset for comparison, in-office review of a patient's PROMPT scores can be used to review expected improvement from

surgery, thereby informing the indications for surgery and shared decision-making.

With increasing recognition of asymptomatic hyperparathyroidism in recent decades, clinical guidelines recommend a more thorough evaluation of the skeletal and renal systems than was previously recommended, but still, they lack clarity regarding nonclassic, neurocognitive symptoms as potential indications for surgery.<sup>25</sup> With additional confirmation of symptom resolution after parathyroidectomy with validated surveys, such as PROMPT, future guidelines hopefully will expand on the indications for parathyroidectomy in these so-called asymptomatic patients.

In conclusion, PROMPT appears to be a validated PROM that can complement the workup and treatment of hyperparathyroidism and goiter. In its current state, PROMPT is now validated and can be used clinically but could be further optimized through the deletion of 2 items (c11 and c30). Despite the relatively small number of patients in our preliminary analysis, PROMPT shows high test-retest reliability and promising ability to demonstrate improvement after surgery.

### Conflicts of interest

The authors indicate that they have no conflicts of interest regarding the content of this article.

### References

- Sneider MS, Solorzano CC, Lew JI. *McGraw-Hill Manual: Endocrine Surgery*. Chapter 8 Primary hyperparathyroidism. New York, NY: The McGraw-Hill Companies; 2010. Available from: <http://accesssurgery.mhmedical.com/content.aspx?aid=6162792>. Accessed: xxxx.
- Press DM, Siperstein AE, Berber E, et al. The prevalence of undiagnosed and unrecognized primary hyperparathyroidism: a population-based analysis from the electronic medical record. *Surgery*. 2013;154:1232–1237.
- Balentine CJ, Xie R, Kirklín JK, Chen H. Failure to Diagnose Hyperparathyroidism in 10,432 patients with hypercalcemia. *Ann Surg*. 2017;266:632–640.
- Niederle B, Wémeau J-L. Is surgery necessary for “mild” or “asymptomatic” hyperparathyroidism? *Eur J Endocrinol*. 2015;173:D13–D20.
- Mekel M, Hodin RA. *McGraw-Hill Manual: Endocrine Surgery*. Chapter 2 Goiter. New York, NY: The McGraw-Hill Companies; 2010. <http://accesssurgery.mhmedical.com/content.aspx?aid=6162090>. Accessed: xxxx.
- Liu J, Pusic A, Temple L, Ko C. Patient-reported outcomes in surgery: listening to patients improves quality of care | The Bulletin. *Bull Am Coll Surg*. 2017;102:19–23.
- Delfino LC, Zunino A, Sapia V, Croome M del CS, Ilera V, Gauna AT. Related quality of life questionnaire specific to dysthyroid ophthalmopathy evaluated in a population of patients with Graves' disease. *Arch Endocrinol Metab*. 2017;61:374–381.
- Hamdan A-L, Dowli A, Jabbour J, Sabri A, Azar ST. Pharyngeal symptoms and impact on quality of life in female patients with goiter. *Ear Nose Throat J*. 2016;95:E5–10.
- Watt T, Björner JB, Groenvold M, et al. Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. *Qual Life Res*. 2009;18:483–496.
- La J, Wang TS, Hammad AY, et al. Parathyroidectomy for primary hyperparathyroidism improves sleep quality: a prospective study. *Surgery*. 2017;161:25–34.
- Santos RSS, Coelho FMS, Silva BC da, et al. Parathyroidectomy improves restless leg syndrome in patients on hemodialysis. *PLoS One*. 2016;11.
- Siperstein AE, Shen W, Chan AK, Duh QY, Clark OH. Normocalcemic hyperparathyroidism. Biochemical and symptom profiles before and after surgery. *Arch Surg*. 1992;127:1157–1166.
- Greutelaers B, Kullen K, Kollías J, et al. Pasiëka Illness Questionnaire: its value in primary hyperparathyroidism. *ANZ J Surg*. 2004;74:112–115.
- Pasiëka JL, Parsons LL. Prospective surgical outcome study of relief of symptoms following surgery in patients with primary hyperparathyroidism. *World J Surg*. 1998;22:513–519.
- Liu JY, Saunders ND, Chen A, Weber CJ, Sharma J. Neuropsychological changes in primary hyperparathyroidism after parathyroidectomy. *Am Surg*. 2016;82:839–845.
- Zanocco K, Butt Z, Kaltman D, et al. Improvement in patient-reported physical and mental health after parathyroidectomy for primary hyperparathyroidism. *Surgery*. 2015;158:837–845.
- Webb SM, Puig-Domingo M, Villabona C, et al. Validation of PHPQoL, a disease-specific quality-of-life questionnaire for patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*. 2016;101:1571–1578.
- (revised May 2013) . [http://www.healthmeasures.net/images/PROMIS/PROMISStandards\\_Vers2.0\\_Final.pdf](http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf). Accessed January 22, 2018.
- Assessing health status and quality-of-life instruments: attributes and review criteria. [https://journals-ohiolink-edu.ccmain.ohionet.org/pg\\_99?112303803726577::NO::P99\\_ENTITY\\_ID,P99\\_ENTITY\\_TYPE:27322045,MAIN\\_FILE&cs=3mUVJISgUPJeUlbbvFd1uaK3F9pZUH9mRRwGKQk5UbzkaqYc2vghVm9Si93Dj1rSUV8BVIVT5BDcNXMIJLDiew](https://journals-ohiolink-edu.ccmain.ohionet.org/pg_99?112303803726577::NO::P99_ENTITY_ID,P99_ENTITY_TYPE:27322045,MAIN_FILE&cs=3mUVJISgUPJeUlbbvFd1uaK3F9pZUH9mRRwGKQk5UbzkaqYc2vghVm9Si93Dj1rSUV8BVIVT5BDcNXMIJLDiew). Accessed 29 January 2018.
- Mercieca-Bebber RL, Perreca A, King M, et al. Patient-reported outcomes in head and neck and thyroid cancer randomised controlled trials: a systematic review of completeness of reporting and impact on interpretation. *Eur J Cancer*. 2016;56:144–161.
- Wong CKH, Lang BHH, Lam CLK. A systematic review of quality of thyroid-specific health-related quality-of-life instruments recommends ThyPRO for patients with benign thyroid diseases. *J Clin Epidemiol*. 2016;78:63–72.
- Coker LH, Rorie K, Cantley L, et al. Primary hyperparathyroidism, cognition, and health-related quality of life. *Ann Surg*. 2005;242:642–650.
- Bratucu M.N., Garofil N.D., Radu P.A., et al. Measurement of quality of life after total parathyroidectomy in patients with secondary hyperparathyroidism and end stage renal disease. 2015;:511–17.
- Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997;314:572.
- Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*. 2014;99:3561–3569.

## Discussion

**Dr Sarah Oltmann** (Dallas, TX): It's great to see someone working on developing a tool for us to be able to see the patient's perspective of how our treatments are improving things or making a difference for them.

Moving forward, do you plan to continue to keep the parathyroid and thyroid questions together? Are you going to continue to ask your parathyroid patients about compressive symptoms and vice versa, your thyroid patients about hyperparathyroidism-specific symptoms? Will you break this apart into 2 different surveys for your patients so that they are getting questions specific to their issues?

**Dr Talia Burneikis**: Thank you very much for that question. Essentially, we have 2 measures in 1. The practical application of that is for the patient who walks into your clinic for a thyroid or maybe parathyroid complaint and half the time the patients don't even

know exactly what they are there for. So this is very easy to just distribute a single measure to every single patient you evaluate. It's not exhaustive. It doesn't seem like we are overly burdening them by asking them too many questions. We do want to ultimately shorten those to as few as possible to get the information that we are looking for. Ultimately, when you do the analysis, you have to break it down into separate groups. But practically speaking, there's a good reason to keep them together initially.

**Dr Sally E. Carty** (Pittsburgh, PA): Nice job. This is potentially quite valuable. Two questions:

There is no agreed upon or objective definition of goiter. How did you define it?

Also, what about Dr Janice Pasiëka's PAS form which is widely used in this manner and which has been validated?



**Dr Talia Burneikis:** For the purposes of this study, we chose obvious goiters or any visible enlargement that was noted in the visit note by the clinician, or obviously noted on ultrasound. And then for patients who presented with very large or multiple thyroid nodules, any single nodule greater than 4 centimeters was categorized into the goiter group for the purpose of this study.

Of course, I would do a disservice if I didn't mention Dr Pasioka and the Pasioka Illness Questionnaire. She did validate a tremendous tool previously, and we do ask some similar questions.

In terms of development of Patient-Reported Outcome Measures, when you are administering a given PROM, you have to administer it exactly as it was designed to be used. You can't just pick and choose a few questions and expect to interpret the same results. We had to go through the process because we were doing a combined measure of goiter and hyperparathyroidism; we had to be able to validate the study on its own.

**Dr Peter Mazzaglia** (Providence, RI): Thank you for a potentially very useful study. It appears that you didn't focus on patients that have smaller thyroid nodules. I am curious whether you think there's any applicability of this study to the patient that we all see who presents with a 1- to 2-centimeter nodule and swears that

they are having tremendous difficulty swallowing. Is there any applicability of your study to that patient?

**Dr Talia Burneikis:** Thank you for that important question. Having reviewed all of these patients' charts, I myself was surprised to find patients who come in with a large substernal goiter and 1 to 2 centimeters of tracheal deviation and they didn't complain of any compressive symptoms. And then in contrast, you had someone with a small nodule that did have a significant compressive symptom component.

In terms of how you could use this in your practice, we do distribute this to all of our patients. So those smaller thyroid nodule patients do still get a compressive score and a hyperparathyroid score. In its ideal form, as an electronic measure, we would be able to compare that patient's individual scores against the average of all of the patients that we have seen. Then perhaps we could try to say if their C score is a certain level preoperatively, then we expect it to drop a certain amount after, for that size of nodule.

Ideally, we will continue to accumulate more numbers and be able to stratify our patients according to other size criteria, or other subgroups of multi-nodular goiter versus Graves' versus Hashimoto's and so forth.