Validation of the MOST Instrument for Graves' Disease and Goiter

Lucy Xu MD[1], Amr Ahmed MD MMSc[1], Marika Russell MD FACS[1], Gregory Randolph MD[1], Jennifer Shin MD SM[1,2]

[1]*Department of Otolaryngology-Head and Neck Surgery, Mass General Brigham, Harvard Medical School, Boston, MA*

[2]*Center for Surgery and Public Health, Mass General Brigham, Boston, MA*

Background

Multiple treatment options can be considered for patients with Graves' disease or goiter including observation, medical therapy, radioactive iodine, selective ablation, and surgery.1-3 Treatment decisions depend on not only disease characteristics but also on patient concerns, values, risk tolerance, and preferences. Patient-reported outcome measures (PROMs) are a useful tool for quantifying and tracking patient preferences and concerns. Existing PROMs for benign thyroid disease such as the ThyPRO[7], ThyPRO-39[8] and PROMPT[9] explore disease-related symptoms over 30-85 items but do not specifically elicit patient concerns around treatment itself.

The Measure for Observation or Surgery for Thyroid lesions (MOST) instrument was developed to offer a concise means for evaluating and tracking both disease- and treatment-related concerns regarding their thyroid lesions to supporting shared decision making around surgery.

Study aims

To assess the psychometric validity of the Measure for Observation or Surgery for Thyroid lesions (MOST) in patients with Graves' disease or goiter undergoing surgical evaluation.

Methods: survey development and data collection

Survey item content originated from qualitative patient comments, assessments of face validity, and direct utility in routine endocrine surgical practice for Graves and goiter patients. Each item response was assigned an ordinal score of 0 (best, no problem) to 4 (worst, extreme problem). Four and five item variants were proposed for further testing (Table 1). Overall scores were calculated as the sum of item responses.

The prospective validation study included patients ≥13 years old who presented for surgical evaluation of Graves' disease or thyroid goiter and agreed to complete the prospectively designed questionnaires in English. An established electronic platform was utilized for prospective data collection at a single academic medical center in the US. Gender was self-reported. The study was approved by the Mass General Brigham Institutional Review Board.

Table 1: MOST 4- and 5-item variants

4-item	5-item	Item Description
✓	✓	Worry about thyroid cancer
✓	✓	Worry about thyroid surgery and its effects
✓	✓	Dissatisfaction with thyroid-related problems
✓	✓	Bulge concern
	✓	Scar concern

Methods: analysis

Psychometric evaluation included assessment of dimensionality, internal consistency, discriminant and convergent validity, intra-rater reliability, and responsiveness to change. Exploratory factor analysis assessed dimensionality and item loadings, while Cronbach's α with one-sided 95% confidence intervals s evaluated internal consistency (α ≥0.70 considered acceptable). Discriminant validity was assessed by comparing scores before and after endocrine surgeon consultation, and convergent validity by correlation with the global Patient-Reported Outcomes Measurement Information System (PROMIS) short form[10]. Intra-rater reliability was evaluated by repeat administration within 3 weeks without intervening treatment. Responsiveness was assessed among patients reporting ≥1-point change in overall health on a 5-point scale, with effect sizes and 95% confidence intervals calculated.

Results

Among 239 consecutive patient responses (mean age 54 years (range 24-88), 83% female, 88% goiter, 12% Graves' disease; Table 2), exploratory factor analysis indicated two factors (Table 3). Internal consistency was measured across both 4-item and 5-item variants (Cronbach alpha 0.7, 95% CI > 0.6). Sum scores discriminated between patients who had and had not received counseling from an endocrine surgeon (difference in means (1.16-1.74, Cohen d 0.35-0.52, p<0.01)). Convergent validity was supported by inverse correlations between MOST sum scores and PROMIS health measures, strongest for physical health (Spearman ρ=−0.36 to −0.39, p<0.001; Table 4). Intra-rater reliability was high (Spearman 0.7-1.0). Responsiveness to change was demonstrated (Cohen d 0.38-0.51, p<0.01).

Table 2: Patient Characteristics

Patient Characteristics	All Responses (n=239)
Age (mean, SD, range)	54 (15, 25-89)
Gender (%)	Female 83% Male 17%
Goiter (%)	Goiter 88%
Graves (%)	Graves 12%
Instrument Items (mean, range)	
1. Worry about thyroid cancer	0.7 (0-4)
2. Worry about thyroid surgery and its effects	1.4 (0-4)
3. Dissatisfaction with thyroid-related problems	1.3 (0-4)
4. Bulge concern	1.3 (0-4)
5. Scar concern	0.7 (0-4)
Sum score items 1,2,3,4	4.8 (0-16)
Sum score items 1,2,3,4,5	5.5 (0-20)

Table 3: Factor Loading for 4- and 5-item MOST variants

Item Description	Factor 1 Loading	Factor 2 Loading	Uniqueness
Worry about thyroid cancer	0.504		0.732
Worry about thyroid surgery and its effects	0.462	0.390	0.635
Dissatisfaction with thyroid-related problems	0.446	0.432	0.591
Bulge concern	0.418	0.350	0.667
Scar concern		0.430	0.801

Results

Table 4: Convergent validity testing for MOST items and sum scores

Item Description	PROMIS Comparison Measure	Spearman rho (95% CI)	P-value
Worry about having thyroid cancer	Physical Health	-0.32 (-0.46 to -0.16)	-0.3203
	Mental Health	-0.18 (-0.34 to -0.02)	0.0334
	Global Health	-0.25 (-0.41 to -0.09)	0.0031
	Global Overview	-0.31 (-0.46 to -0.15)	0.0002
Worry about thyroid surgery and its effects	Physical Health	-0.12 (-0.28 to 0.06)	0.1851
	Mental Health	-0.24 (-0.39 to -0.07)	0.0054
	Global Health	-0.16 (-0.32 to 0.01)	0.0605
	Global Overview	-0.10 (-0.27 to 0.07)	0.2474
Dissatisfaction with thyroid-related problems	Physical Health	-0.36 (-0.50 to -0.20)	<0.0001
	Mental Health	-0.25 (-0.40 to -0.09)	0.0034
	Global Health	-0.20 (-0.36 to -0.03)	0.0208
	Global Overview	-0.23 (-0.39 to -0.07)	0.0066
Bulge concern	Physical Health	-0.26 (-0.41 to -0.09)	0.0027
	Mental Health	-0.21 (-0.37 to -0.04)	0.0144
	Global Health	-0.17 (-0.33 to 0.004)	0.0554
	Global Overview	-0.16 (-0.32 to 0.01)	0.0583
Scar concern	Physical Health	-0.13 (-0.29 to 0.04)	0.1399
	Mental Health	-0.12 (-0.29 to 0.05)	0.1577
	Global Health	-0.12 (-0.29 to 0.05)	0.1529
	Global Overview	0.03 (-0.14 to 0.20)	0.7052
Sum score items 1,2,3,4	Physical Health	-0.38 (-0.52 to -0.23)	<0.0001
	Mental Health	-0.32 (-0.46 to -0.16)	0.0002
	Global Health	-0.28 (-0.43 to -0.11)	0.0012
	Global Overview	-0.29 (-0.44 to -0.13)	0.0006
Sum score items 1,2,3,4,5	Physical Health	-0.39 (-0.52 to -0.24)	<0.0001
	Mental Health	-0.32 (-0.46 to -0.16)	0.0002
	Global Health	-0.28 (-0.43 to -0.11)	0.0011
	Global Overview	-0.27 (-0.42 to -0.10)	0.0017

Conclusions

Assessments of 4 and 5 item variants of a MOST instrument in patients with Graves' disease or goiter yield novel, practical functionality with demonstrated psychometric validity for use in high volume endocrine surgical settings.

References

1 Lee SY, Pearce EN. Hyperthyroidism: A Review. JAMA. 2023;330(15):1472-1483. doi:10.1001/jama.2023.19052
2 Burch HB, Cooper DS. Management of Graves Disease: A Review. JAMA. 2015;314(23):2544-2554. doi:10.1001/jama.2015.16535
3 Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. Thyroid. 2016;26(10):1343-1421. doi:10.1089/thy.2016.0229
4 Chen AY, Bernet VJ, Carty SE, et al. American Thyroid Association statement on optimal surgical management of goiter. Thyroid. 2014;24(2):181-189. doi:10.1089/thy.2013.0291
5 Bartsch DK, Luster M, Buhr HJ, et al. Indications for the Surgical Management of Benign Goiter in Adults. Dtsch Arztebl Int. 2018;115(1-02):1-7. doi:10.3238/arztebl.2018.0001
6 Jasim S, Patel KN, Randolph G, et al. American Association of Clinical Endocrinology Disease State Clinical Review: The Clinical Utility of Minimally Invasive Interventional Procedures in the Management of Benign and Malignant Thyroid Lesions. Endocr Pract. 2022;28(4):433-448. doi:10.1016/j.eprac.2022.02.011
7 Watt T, Hegedüs L, Groenvold M, et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *Eur J Endocrinol*. 2010;162(1):161-167. doi:10.1530/EJE-09-0521
8 Watt T, Bjorner JB, Groenvold M, et al. Development of a Short Version of the Thyroid-Related Patient-Reported Outcome ThyPRO. *Thyroid*. 2015;25(10):1069-1079. doi:10.1089/thy.2015.0209
9 Burneikis T, Colvin J, Jin J, et al. Validation of a novel patient-reported outcome measure for parathyroid and thyroid disease (PROMPT). *Surgery*. 2019;165(1):232-239. doi:10.1016/j.surg.2018.04.091
10 Cella D, Choi SW, Condon DM, et al. PROMIS® Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value Health*. 2019;22(5):537-544. doi:10.1016/j.jval.2019.02.004