ORIGINAL RESEARCH

Psychometric evaluation of the hypoparathyroidism symptom diary

Theresa Coles[1]
Kristina Chen[2]
Lauren Nelson[1]
Nimanee Harris[1]
Montserrat Vera-Llonch[2]
Alan Krasner[2]
Susan Martin[3]

[1]RTI Health Solutions, Research Triangle Park, NC, USA; [2]Shire Human Genetic Therapies, Lexington, MA, USA; [3]RTI Health Solutions, Ann Arbor, MI, USA

Purpose: To conduct an initial psychometric evaluation of the reliability and validity of the Hypoparathyroidism Symptom Diary (HPT-SD).

Patients and methods: Data were collected during a cross-sectional, observational study. Participants with self-reported hypoparathyroidism (HPT) completed the HPT-SD, the Functional Assessment in Cancer Therapy–Cognitive Function (FACT-Cog), the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue), and the Hospital Anxiety and Depression Scale (HADS) measures. Item- and scale-level internal consistency reliability, known-groups validity, and construct validity were evaluated. Subscales were identified and preliminary scoring algorithms were developed.

Results: The study included 52 participants (mean age, 51 years). Overall, the measurement properties of the HPT-SD were very good. Item-level response frequency distributions showed evidence of possible floor effects for four muscle-related symptom items. Inter-item correlations revealed a pattern of relationships among symptom items (r=0.3–0.8) and among impact items (r=0.5–0.7) and provided evidence for two HPT-SD subscales: Symptoms and Impacts. Construct validity correlations supported a priori convergent validity hypotheses ($|r|\geq0.4$) between HPT-SD subscales and the FACT-Cog, FACIT-Fatigue, and HADS. Mean HPT-SD Symptom and Impact scores were in the expected direction and significantly different between subgroups of patients with high and low HPT disease severity.

Conclusion: Results indicate that the HPT-SD is an appropriate measure of HPT-related symptoms and impacts. Floor effects may be attributed to the observational study design: participants manage symptoms with calcium and active vitamin D supplements prior to an escalation in severity. Future studies should assess the HPT-SD measurement properties using longitudinal study designs.

Keywords: hypoparathyroidism, symptom, psychometric, validation, impacts, patient-reported outcome

Correspondence: Lauren Nelson
RTI Health Solutions, 200 Park Offices Drive, Research Triangle Park, NC 27709, USA
Tel +1 919 541 6590
Email lnelson@rti.org

Plain language summary

- Patients with hypoparathyroidism experience considerable symptom burden – including physical symptoms (eg, fatigue, pain, muscle spasms, paresthesia), cognitive symptoms (eg, "brain fog"), and emotional symptoms (eg, depression, anxiety) – resulting in impairments in health-related quality of life. The Hypoparathyroidism Symptom Diary (HPT-SD) questionnaire has been developed to assess the key symptoms and impacts of hypoparathyroidism from the patient perspective.
- The current analysis represents the first psychometric evaluation of the HPT-SD in patients with hypoparathyroidism. The items performed well in assessments of reliability and validity, and preliminary scoring rules were developed.

- The HPT-SD is an appropriate measure of hypoparathyroidism-related symptoms and impacts in individuals diagnosed with hypoparathyroidism, providing a new patient-centered framework for the evaluation of current and emerging treatments for hypoparathyroidism. Future longitudinal studies should evaluate the responsiveness of the HPT-SD and determine thresholds for meaningful change.

Introduction

Chronic hypoparathyroidism (HPT) is a rare disorder that causes persistent low calcium levels in the blood due to insufficient levels of parathyroid hormone, lasting at least 6 months.[1] The most common etiologies for HPT are neck surgery, autoimmune disorders, and genetic disorders.[2] Chronic HPT is three times more prevalent in women than in men and occurs most often in individuals aged 55 years and older.[3]

The symptom burden of HPT is considerable. Hypocalcemia due to HPT most commonly results in neuromuscular excitability (tetany),[2] and calcification in the kidney, brain, eye, or vasculature may also occur. Patients with HPT may experience physical symptoms (eg, fatigue, pain, muscle spasms, paresthesia), cognitive symptoms (eg, "brain fog"), and emotional symptoms (eg, depression, anxiety), which can impair health-related quality of life.[4]

To assess the key symptoms and impacts of HPT from the patient perspective, the Hypoparathyroidism Symptom Diary (HPT-SD) questionnaire was recently developed. Martin et al[5] describe development and content validation of the HPT-SD in detail. Briefly, the HPT-SD was developed according to recommendations in the US Food and Drug Administration (FDA) guidance for industry.[6] Development of the HPT-SD included a literature review, clinical expert input, and concept elicitation interviews with six individuals with HPT to inform the draft item set. Iterative cognitive debriefing interviews were then conducted with six additional individuals with HPT. The qualitative results supported a 13-item instrument (HPT-SD) assessing relevant symptoms (muscle cramping, tingling and muscle spasms/twitching, fatigue, and cognition, as well as anxiety, sadness, and depression) and impacts on sleep, ability to exercise, ability to work, and family relationships.

The objective of the current study was to conduct an initial psychometric evaluation of the reliability and validity of the HPT-SD symptoms in patients with HPT in the US. The goal is for the HPT-SD to be a brief, useful tool for assessing symptoms and impacts of HPT from the patient perspective, and to use the HPT-SD in future studies that

evaluate disease burden or treatment efficacy relating to symptom improvement.

Methods
Study design and population

This was a cross-sectional, observational study involving individuals with a self-reported diagnosis of HPT. Staff from the Hypoparathyroidism Association, a patient advocacy group, contacted individuals from the association's database to complete the questionnaire via mail or, for those individuals planning to attend, in person, at the 10th International Conference on Hypoparathyroidism in Jacksonville, Florida. Individuals interested in participating were provided with contact information to be screened for study eligibility.

Eligible patients were adults (aged 18 years or older) with a diagnosis of chronic HPT who were currently receiving oral calcium and vitamin D for the treatment of HPT; reported experiencing symptoms due to chronic HPT despite ongoing treatment; were willing and able to complete study questionnaires (either in person at the conference or at home); were living in the US; and were able to read, speak, and understand English. Those with a calcium-sensing receptor gene mutation or impaired responsiveness to parathyroid hormone or who were currently or previously treated with recombinant human parathyroid hormone (rhPTH [1-84]) therapy were ineligible. Interested and eligible individuals provided written informed consent before completing the survey. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki (2008) and was approved by the Office of Research Protection and Ethics at RTI International. A total of 52 adult participants with self-reported HPT completed the paper-based survey.

Patient-reported outcome (PRO) measures

Participants were administered the HPT-SD and three additional PRO questionnaires – specifically, the Functional Assessment in Cancer Therapy–Cognitive Function (FACT-Cog) Version 3, the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue) Version 4, and the Hospital Anxiety and Depression Scale (HADS) – to assess fatigue, cognition, anxiety, and depression. These questionnaires were chosen based on a literature and instrument search as being the most appropriate measures to capture the concepts most relevant to patients with HPT.[5] The FACT-Cog is a 37-item, self-reported questionnaire measuring cognitive function over the past 7 days.[7] The measure yields four

domain scores: perceived cognitive impairments, impact of perceived cognitive impairments on quality of life, comments from others, and perceived cognitive abilities. Higher scores indicate better cognitive functioning and the theoretical range for each scale is different because they are calculated by summing FACT-Cog item scores. The FACIT-Fatigue is a 13-item, patient-reported questionnaire measuring fatigue in patients diagnosed with chronic illnesses.[8] The measure uses a 7-day recall period and yields a single overall score ranging from 0 to 52, with higher scores indicating less fatigue. The HADS is a screening tool for clinically meaningful anxiety and depression in a general outpatient setting that utilizes a recall period of "the past week".[9] The HADS includes two subscales, anxiety and depression, with maximum scores of 21. Higher scores indicate worse anxiety or depression; scores of 11 or more are considered to indicate significant psychological morbidity.

The FACT-Cog, FACIT-Fatigue, and HADS questionnaires were included in the study to evaluate the validity of the HPT-SD. The survey also collected demographic information, along with self-reported HPT-related clinical and treatment information, as well as an item assessing patients' global impression of symptom severity (PGIS). A subset of the participants (those who completed the questionnaires via mail) also completed additional questions developed to gage the relevance of the supporting PRO measures (ie, FACT-Cog, FACIT-Fatigue, and HADS) to his or her experience with HPT.

Analyses

The psychometric evaluation of the HPT-SD was conducted in two stages. First, the measurement properties of the 13 HPT-SD items were evaluated. Based on key item-level results (eg, descriptive statistics, inter-item correlations), two subscales were identified for inclusion in the HPT-SD: a Symptom subscale and an Impact subscale. The measurement properties of the subscales were also evaluated. Specifically, the following psychometric properties were assessed in the first step for the item-level analyses.

Distributional characteristics

HPT-SD response distributions and other PRO measures were summarized using standard descriptive statistics (eg, means, SDs, missing) to assess distributional characteristics. Item-level response distributions for the HPT-SD were summarized to evaluate possible response biases, such as floor or ceiling effects.

Reliability of the HPT-SD

To evaluate the internal consistency of the HPT-SD, Cronbach's[10] coefficient alphas were computed. First, the coefficient alpha was computed using all 13 HPT-SD items. The coefficient alpha was then computed for the HPT-SD subscales. The approximate range of optimal alphas suggested by Streiner and Norman[11] is between 0.70 and 0.90; scores within this range indicate that a set of items is strongly related and capable of supporting a unidimensional scoring structure but is not redundant.

Structure of the HPT-SD

Inter-item correlations were computed and tabulated as a preliminary appraisal of the HPT-SD's structure. Patterns of correlations provided a basis for hypothesizing potential HPT-SD subscales. It was expected that all inter-item correlations would be positive in sign. Inter-subscale correlations were also computed for HPT-SD subscales.

Validity of the HPT-SD

Construct validity correlations were computed between patient-reported HPT-SD scores and supporting PRO measures administered in the study. The goal was to demonstrate stronger relationships among measures addressing similar constructs (convergent validity). Positive correlations were expected between HPT-SD items and the HADS, and negative correlations were expected between HPT-SD items and FACIT-Fatigue and FACT-Cog.

The hypothesized strength of correlations between HPT-SD items and other PRO measures was as follows: 1) cognitive issues: HPT-SD items 7 (brain fog) and 12 (impact work) were expected to be at least moderately correlated with FACT-Cog scales ($|r|>0.50$); 2) anxiety and depression: HPT-SD items 8 (anxiety) and 9 (depression) were hypothesized to be at least moderately correlated ($|r|>0.50$) with HADS anxiety and HADS depression scales, respectively; 3) fatigue: HPT-SD items 5 (heaviness), 6 (fatigue), and 10 (impact sleep) were expected to exhibit at least moderate correlations ($|r|>0.50$) with the FACIT-Fatigue scales; and 4) impacts: HPT-SD items 10 (impact sleep), 11 (impact exercise), 12 (impact work), and 13 (impact relationships) were expected to correlate moderately ($|r|>0.50$) or stronger with the FACT-Cog Impact of Perceived Cognitive Impairments on Quality of Life subscale.

Known-groups analyses comparing various subgroups of interest were conducted to provide support for the

discriminating ability of the HPT-SD. ANOVAs based on a priori hypotheses were conducted to examine mean differences in HPT-SD item scores between participants classified into subgroups of patient-reported HPT severity. It was hypothesized that participants who rated their HPT as "very mild" or "mild" would have lower HPT-SD item and subscale scores compared with participants who rated their HPT as "severe" to "very severe."

HPT-SD subscale evaluation

The HPT-SD Symptom and Impact subscales were evaluated using standard psychometric methods identical to the item-level evaluation, with the exception of the addition of the alpha-if-item-deleted method, which was used to determine the maximum tolerable number of missing item responses for the calculation of each HPT-SD subscale. Cronbach's[10] coefficient alpha was calculated sequentially as each item was deleted from the item set.[12] The order in which items were removed from the set was based on the size of the alpha with the item deleted, where the item with the smallest alpha-if-item-deleted (ie, the item with the largest contribution to the alpha of the item set) was removed at each step.[13] When Cronbach's alpha for the set of remaining items fell below 0.70 (an a priori set threshold), the number of items deleted from the scale minus one was identified as the maximum number of allowable missing item responses in the computation of a total score.

Analysis conventions

All statistical tests used in the psychometric evaluation were two-tailed, with a type I error rate of 5% (alpha =0.05) unless otherwise stated. Missing HPT-SD item-level responses were reported, but not imputed. The FACT-Cog, FACIT-Fatigue, and HADS were scored according to the instrument developers' guidelines (including missing rules).

Results

Sample characteristics

The 52 participants had a mean (SD) age of 51.3 (12.6) years (range, 24–78 years). Table 1 presents participants' demographic characteristics and HPT-related clinical information. Approximately two thirds of participants rated their HPT as moderate (34.6%) or severe (32.7%). On average, participants in this sample had been diagnosed with HPT for 10.5 years (SD, 11.1; range, 1–51).

Table S1 summarizes patients' treatment characteristics. Calcium carbonate (47.1%) and calcium citrate (46.2%) were the most commonly reported calcium treatments; some

Table 1 Demographic characteristics (N=52)

Characteristics	Statistics
Age (years), mean (SD), n	
Mean (SD), n	51.3 (12.6), 52
Median, min-max	51.5, 24–78
Sex, n (%)	
Male	8 (15.4)
Female	44 (84.6)
Educational level, n (%)	
High school graduate or equivalent	7 (13.5)
Some college but no degree	13 (25.0)
College degree (eg, BA, BS)	16 (30.8)
Some graduate school but no degree	3 (5.8)
Graduate degree (eg, MS, MD, PhD)	13 (25.0)
Employment status, n (%)	
Full-time	17 (32.7)
Part-time	9 (17.3)
Full-time homemaker	5 (9.6)
Unemployed	6 (11.5)
Disabled	9 (17.3)
Retired	6 (11.5)
Race/ethnicity, n (%) (check all that apply)	
American Indian or Alaskan Native	1 (1.9)
African American	1 (1.9)
White	50 (96.2)
Hispanic	1 (1.9)
HPT-related clinical information	
PGIS: How would you rate your hypoparathyroidism? n (%)	
Very mild	1 (1.9)
Mild	9 (17.3)
Moderate	18 (34.6)
Severe	17 (32.7)
Very severe	7 (13.5)
Was your hypoparathyroidism caused by surgery? n (%)	
Yes	45 (86.5)
No	6 (11.5)
Unknown	1 (1.9)
Years diagnosed with hypoparathyroidism, mean (SD)	10.5 (11.1), 52
Median, min-max	6.0, 1–51
Over the past year have you gone to urgent care or the emergency room due to your hypoparathyroidism? n (%)	
Yes	18 (34.6)
Number of times: mean (SD), median, min-max	2.6 (1.4), 2.0, 1–6
No	34 (65.4)
Over the past year have you been admitted to the hospital due to your hypoparathyroidism? n (%)	
Yes	9 (17.3)
Number of times: mean (SD), median, min-max	2.8 (2.4), 2.0, 1–8
No	43 (82.7)
Have you had kidney stones in the past 5 years? n (%)	
Yes	10 (19.2)
No	42 (80.8)

Note: Percent calculated from non-missing responses.
Abbreviations: BA, Bachelor of Arts; BS, Bachelor of Science; HPT, hypoparathyroidism; MD, Doctor of Medicine; MS, Master of Science; PGIS, patients' global impression of symptom severity; PhD, Doctor of Philosophy.

patients reported taking more than one form of calcium. Most participants (86.5%) took calcitriol. Most participants (80.8%) reported that they took additional calcium or vitamin D in the past 6 months due to HPT symptoms. Almost all participants (98.1%) had undergone a calcium level blood test in the past year, and 48% of participants had undergone a urine calcium test in the past year.

Study participants provided insight into which supporting PRO measures (eg, FACIT-Fatigue, FACT-Cog, HADS) were relevant to their experiences with HPT. Almost two thirds (64.1%) of participants endorsed the FACIT-Fatigue as the most relevant PRO measure, with 74.4% of participants indicating that most or all of the questions on the FACIT-Fatigue were relevant to their experiences with HPT.

Distributional characteristics

Table 2 presents the descriptive statistics for the HPT-SD, FACIT-Fatigue, FACT-Cog, and HADS scales. Among the HPT-SD Symptom items scored from 0 to 4, HPT-SD item 6 (physical fatigue) had the largest average (mean, 2.4; SD,

0.9). Among the impact-specific items, the average participant score was highest for the impact on exercise item, item 11 (mean,1.4; SD, 0.8). There were very few missing responses (n≤1). All patients reported experiencing at least two of the nine HPT symptoms, with the average number of reported symptoms per patient being 7.5. The average FACIT-Fatigue score was 20.3 (SD, 12.9) and average HADS scores were lower (depression: mean, 8.2 [SD, 4.9]; anxiety: mean, 9.8 [SD, 3.9]) than the 11-point threshold for significant psychological morbidity.[9]

All response choices were reported for nine out of 13 HPT-SD items, indicating no severe item-level floor or floor effects for these nine items (Table 3). The four remaining HPT-SD items exhibited possible evidence of floor effects: HPT-SD item 1 (muscle cramps), item 2 (tingling), item 3 (numbness), and item 4 (muscle spasms or twitching) (Table 3). For HPT-SD item 1 (muscle cramps), participants did not endorse two of the three highest response choices, 8 and 10 (worst possible muscle cramps). Similarly, no participants chose the "very severe" response choice for items 2, 3,

Table 2 Descriptive statistics for HPT-SD and other PRO measures (N=52)

Measure and items	n	Mean (SD)	Median	Min-Max	Missing (%)
HPT-SD					
Item 1. Muscle cramps	52	4.0 (2.6)	4.5	0–9	0 (0)
Item 2. Tingling	52	1.4 (0.8)	1.0	0–3	0 (0)
Item 3. Numbness	52	1.2 (1.0)	1.0	0–3	0 (0)
Item 4. Muscle spasms or twitching	52	1.5 (1.0)	1.0	0–3	0 (0)
Item 5. Feelings of heaviness in arms or legs	52	1.3 (1.0)	1.0	0–4	0 (0)
Item 6. Physical fatigue	52	2.4 (0.9)	2.0	0–4	0 (0)
Item 7. Slowed or confused thinking (brain fog)	52	1.8 (1.1)	2.0	0–4	0 (0)
Item 8. Anxiety	52	1.7 (1.2)	1.5	0–4	0 (0)
Item 9. Sadness or depression	52	1.3 (1.1)	1.0	0–4	0 (0)
Item 10. Impact sleep	52	1.0 (0.7)	1.0	0–2	0 (0)
Item 11. Impact exercise	52	1.4 (0.8)	2.0	0–2	0 (0)
Item 12. Impact work	51	0.9 (0.7)	1.0	0–2	1 (1.9)
Item 13. Impact relationships	52	0.8 (0.7)	1.0	0–2	0 (0)
HPT-SD Symptom subscale score					
(1: Muscle cramps[a]; 2: Tingling; 3: Numbness; 4: Muscle spasms or twitching; 5: Feelings of heaviness in arms; 6: Physical fatigue; 7: Slowed or confused thinking)	52	1.6 (0.8)	1.5	0.3–3.1	0 (0)
HPT-SD Impact subscale score					
(10: Sleep; 11: Exercise; 12: Work; 13: Relationships)	51	1.0 (0.6)	1.0	0.0–2.0	1 (1.9)
FACIT-Fatigue	52	20.3 (12.9)	17.0	0–48	0 (0)
FACT-Cog					
Perceived cognitive impairments	52	37.9 (18.5)	39.5	3–69	0 (0)
Impact of perceived cognitive impairments on quality of life	52	5.6 (5.3)	4.5	0–16	0 (0)
Comments from others	52	12.5 (4.0)	14.0	0–16	0 (0)
Perceived cognitive abilities	52	13.9 (6.5)	12.8	0–28	0 (0)
HADS					
Depression	52	8.2 (4.9)	9.0	1–19	0 (0)
Anxiety	52	9.8 (3.9)	9.5	0–17	0 (0)

Note: [a]To calculate the HPT-SD Symptom subscale, item 1 was rescaled from an 11-point response scale to a 5-point response scale ranging from 0 to 4.
Abbreviations: FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; FACT-Cog, Functional Assessment in Cancer Therapy–Cognitive Function; HADS, Hospital Anxiety and Depression Scale; HPT-SD, Hypoparathyroidism Symptom Diary; PRO, patient-reported outcome.

Table 3 HPT-SD item-level response distributions (N=52)

Symptom items[a]	Response option, n (%)					
	None	Mild	Moderate	Severe	Very severe	Missing
Item 2. Tingling, n (%)	7 (13.5)	23 (44.2)	18 (34.6)	4 (7.7)	0 (0)	0 (0)
Item 3. Numbness, n (%)	15 (28.8)	20 (38.5)	11 (21.2)	6 (11.5)	0 (0)	0 (0)
Item 4. Muscle spasms or twitching, n (%)	7 (13.5)	23 (44.2)	11 (21.2)	11 (21.2)	0 (0)	0 (0)
Item 5. Feelings of heaviness in arms or legs, n (%)	13 (25.0)	16 (30.8)	16 (30.8)	6 (11.5)	1 (1.9)	0 (0)
Item 6. Physical fatigue, n (%)	2 (3.8)	3 (5.8)	26 (50.0)	15 (28.8)	6 (11.5)	0 (0)
Item 7. Slowed or confused thinking (brain fog), n (%)	7 (13.5)	15 (28.8)	15 (28.8)	12 (23.1)	3 (5.8)	0 (0)
Item 8. Anxiety, n (%)	7 (13.5)	19 (36.5)	12 (23.1)	10 (19.2)	4 (7.7)	0 (0)
Item 9. Sadness or depression, n (%)	15 (28.8)	16 (30.8)	14 (26.9)	5 (9.6)	2 (3.8)	0 (0)
Impact items	**Response option, n (%)**					
	Not at all		Somewhat		Very much	Missing
Item 10. Impact sleep, n (%)	14 (26.9)		24 (46.2)		14 (26.9)	0 (0)
Item 11. Impact exercise, n (%)	9 (17.3)		15 (28.8)		28 (53.8)	0 (0)
Item 12. Impact work, n (%)	16 (31.4)		24 (47.1)		11 (21.6)	1 (1.9)
Item 13. Impact relationships, n (%)	22 (42.3)		21 (40.4)		9 (17.3)	0 (0)

Notes: Percent calculated from non-missing responses. [a]Item 1 was initially scored on an 11-point NRS (0 [no muscle cramps] to 10 [worst possible muscle cramps]), with distributions for the response options as follows: 0: seven patients (13.5%); 1: four patients (7.7%); 2: four patients (7.7%); 3: nine patients (17.3%); 4: two patients (3.8%); 5: seven patients (13.5%); 6: ten patients (19.2%); 7: seven patients (13.5%); 8: 0 patients; 9: two patients (3.8%); 10: 0 patients.
Abbreviations: HPT-SD, Hypoparathyroidism Symptom Diary; NRS, numeric rating scale.

Table 4 HPT-SD inter-item correlations (n=51–52)

HPT-SD items	Correlation coefficient, r												
	HPT-SD item												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Item 1. Muscle cramps	–												
Item 2. Tingling	0.55	–											
Item 3. Numbness	0.60	0.74	–										
Item 4. Muscle spasms or twitching	0.75	0.72	0.71	–									
Item 5. Feelings of heaviness in arms or legs	0.55	0.61	0.72	0.64	–								
Item 6. Physical fatigue	0.50	0.66	0.69	0.57	0.56	–							
Item 7. Slowed or confused thinking (brain fog)	0.57	0.43	0.53	0.55	0.68	0.70	–						
Item 8. Anxiety	0.30	0.42	0.50	0.48	0.55	0.51	0.63	–					
Item 9. Sadness or depression	0.34	0.47	0.49	0.49	0.53	0.57	0.63	0.81	–				
Item 10. Impact sleep	0.45	0.50	0.35	0.62	0.32	0.29	0.39	0.50	0.59	–			
Item 11. Impact exercise	0.57	0.53	0.59	0.69	0.52	0.63	0.63	0.51	0.69	0.49	–		
Item 12. Impact work	0.49	0.62	0.63	0.46	0.66	0.70	0.60	0.61	0.59	0.48	0.67	–	
Item 13. Impact relationships	0.24	0.57	0.58	0.53	0.52	0.52	0.56	0.57	0.73	0.45	0.66	0.62	–

Abbreviation: HPT-SD, Hypoparathyroidism Symptom Diary.

and 4 (tingling, numbness and muscle spasms, or twitching, respectively).

Reliability and structure of the HPT-SD items

The estimate of internal consistency reliability, as measured by Cronbach's[10] coefficient alpha for all 13 HPT-SD items, was 0.92. The structure of the HPT-SD (Table 4) shows the HPT-SD inter-item correlations. As expected, all inter-item correlations were positive. The strongest inter-item correlation (r=0.81) was between HPT-SD item 8 (anxiety) and

item 9 (sadness or depression). A cluster of strong inter-item correlations was present among HPT symptoms related to muscle or nerve sensations. Specifically, HPT-SD item 4 (muscle spasms or twitching) was strongly correlated with items 1 (muscle cramps), 2 (tingling), and 3 (numbness) ($0.71 \leq r \leq 0.75$). In addition, HPT-SD item 5 (feelings of heaviness in arms or legs) was strongly correlated with item 3 (numbness) (r=0.72).

Other strong correlations were observed between pairs of symptom- and impact-specific items. HPT-SD item 6 (physical fatigue) was strongly correlated with item 12 (impact

work) (*r*=0.70), and HPT-SD item 9 (sadness or depression) was strongly correlated with item 13 (impact relationships) (*r*=0.73). All correlations were above 0.30 in magnitude except for HPT-SD item 1 (muscle cramps) and item 13 (impact relationships) (*r*=0.24).

Validity of the HPT-SD items

As expected, in the construct validity analyses, positive correlations were present between HPT-SD items and the HADS, and correlation coefficients between HPT-SD items and FACIT-Fatigue and FACT-Cog were negative (Table 5). The hypotheses related to HPT-SD items related to cognitive issues were supported. Specifically, correlations between HPT-SD items 7 (brain fog) and 12 (impact work) were strongly correlated with FACT-Cog scales (0.53≤ |*r*|≤0.75). It also was anticipated that HPT-SD items 7 (brain fog) and 12 (impact work) would exhibit stronger correlations with the FACT-Cog scales than with FACIT-Fatigue and HADS scales. The correlations between HPT-SD item 7 (slowed or confused thinking) and supporting measures also followed this pattern.

The expected relationships between the HPT-SD items addressing anxiety and depression and supporting measures addressing similar concepts were also met. HPT-SD items

8 (anxiety) and 9 (sadness or depression) yielded strong correlations with the HADS anxiety (*r*=0.63) and HADS depression (*r*=0.63), respectively. In addition, correlations between FACIT-Fatigue and HPT-SD items 5 (heaviness) 6 (fatigue), and 10 (impact sleep) were moderate to strong (0.40≤ |*r*|≤0.78). As expected, the HPT-SD items 10 (impact sleep), 11 (impact exercise), 12 (impact work), and 13 (impact relationships) were strongly correlated with the FACT-Cog Impact of Perceived Cognitive Impairments on Quality of Life subscale (0.52≤ |*r*|≤0.75).

ANOVAs supported the discriminating ability of the HPT-SD score based on a priori-defined known groups derived from patient-reported HPT severity on the PGIS (Table S2). As anticipated, the HPT-SD item-level scores were higher for the subgroup of patients with more severe HPT compared to the "very mild or mild" group. All item-level group differences were statistically significant at different severity levels except for HPT-SD item 2 (tingling).

HPT-SD subscale-level results
HPT-SD subscale scoring algorithms
Item-level response distributions show that some patients did not experience all of the HPT-related symptoms

Table 5 Construct validity correlations (n=51–52)

HPT-SD score	Correlation coefficient (r)						
	FACIT-Fatigue	FACT-Cog impairments	FACT-Cog impact	FACT-Cog comments	FACT-Cog abilities	HADS depression	HADS anxiety
Item 1. Muscle cramps	–0.53	–0.41	–0.59	–0.37	–0.45	0.33	0.33
Item 2. Tingling	–0.51	–0.34	–0.46	–0.29	–0.31	0.42	0.36
Item 3. Numbness	–0.50	–0.40	–0.54	–0.42	–0.48	0.47	0.49
Item 4. Muscle spasms or twitching	–0.54	–0.43	–0.61	–0.44	–0.37	0.40	0.37
Item 5. Feelings of heaviness in arms or legs	**–0.61**	–0.57	–0.54	–0.52	–0.64	0.50	0.47
Item 6. Physical fatigue	**–0.78**	–0.48	–0.59	–0.43	–0.46	0.49	0.34
Item 7. Slowed or confused thinking (brain fog)	–0.57	**–0.68**	**–0.58**	**–0.68**	**–0.62**	0.49	0.30
Item 8. Anxiety	–0.53	–0.57	–0.49	–0.61	–0.57	0.64	**0.63**
Item 9. Sadness or depression	–0.59	–0.49	–0.54	–0.54	–0.52	**0.63**	0.48
Item 10. Impact sleep	**–0.40**	–0.44	**–0.52**	–0.31	–0.35	0.28	0.23
Item 11. Impact exercise	–0.67	–0.59	**–0.59**	–0.62	–0.49	0.62	0.51
Item 12. Impact work	–0.67	**–0.59**	**–0.75**	**–0.53**	**–0.57**	0.54	0.58
Item 13. Impact relationships	–0.51	–0.45	**–0.59**	–0.53	–0.47	0.52	0.52
HPT-SD Symptom subscale score	–0.70	–0.57	–0.67	–0.55	–0.58	0.54	0.46
(1: Muscle cramps[a]; 2: Tingling; 3: Numbness; 4: Muscle spasms or twitching; 5: Feelings of heaviness in arms; 6: Physical fatigue; 7: Slowed or confused thinking)							
HPT-SD Impact subscale score	–0.66	–0.60	–0.71	–0.55	–0.54	0.57	0.54
(10: Sleep; 11: Exercise; 12: Work; 13: Relationships)							

Notes: Correlations in bold indicate a priori hypotheses. [a]To calculate the HPT-SD Symptom subscale, item 1 was rescaled from an 11-point response scale to a 5-point response scale ranging from 0 to 4.
Abbreviations: FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; FACT-Cog, Functional Assessment in Cancer Therapy–Cognitive Function; HADS, Hospital Anxiety and Depression Scale; HPT-SD, Hypoparathyroidism Symptom Diary.

(13.5%–28.8%) or impacts (17.3%–42.3%) covered in the HPT-SD questionnaire (Table 3); therefore, subscale scoring algorithms were calculated using average item scores. In addition, HPT-SD item 1 (muscle cramps) was rescaled from an 11-point response scale to a 5-point response scale (ranging from 0 to 4), making all symptom-related response choices consistent across the symptom items. Using the rescaled item 1, the Symptom subscale was calculated. While relevant to patients, HPT-SD items 8 (anxiety) and 9 (sadness or depression) cover less specific symptoms of HPT that could be attributed to other conditions; therefore, these items were not included in the Symptom subscale.

HPT-SD Symptom subscale (seven items)

Using the rescaled HPT-SD item 1, patient responses for HPT-SD items 1 through 7 were averaged to compute the HPT-SD Symptom subscale. The allowable range of the 7-item HPT-SD Symptom subscale is 0 to 4, with higher scores indicating worse HPT symptoms.

HPT-SD Impact subscale (four items)

Patient responses for items 10 through 13 were averaged to compute the HPT-SD Impact subscale. The allowable range of the HPT-SD Impact subscale is 0–2, with higher scores indicating worse HPT impacts.

Summary of the HPT-SD subscale psychometric evaluation results

Cronbach's[10] coefficient alphas were satisfactory for the HPT-SD Symptom (0.89) and Impact (0.77) subscales. The results of the alpha-if-item-deleted analyses assessing the maximum tolerable number of missing responses for the HPT-SD Symptom subscale suggested that, in order to calculate a subscore, responses for up to four out of seven items can be missing from the 7-item HPT-SD Symptom subscale and that no missing item responses could be present among the 4-item HPT-SD impact items (after one item was removed from the impact scale, alpha fell below 0.70). However, the authors adopted a more conservative approach with the Symptom score calculation, allowing up to three missing symptom items for calculating the Symptom subscale score. This decision was informed by recommendations of Fairclough and Cella,[14] to calculate a score where at least half of the item responses were non-missing.

The correlation between HPT-SD subscales was 0.73, indicating that the subscales are measuring related, but distinct, concepts. The mean (SD) HPT-SD Symptom and Impact subscale scores were 1.6 (0.8) and 1.0 (0.6), respectively (Table 2).

Construct validity correlations between HPT-SD subscale scores and all supporting PRO measures were strong ($|r|>0.53$) except for the correlation between the HPT-SD Symptom subscale score and HADS anxiety ($r=0.46$) (Table 5). Among construct validity correlations with the HPT-SD Symptom subscale score, the strongest correlation was with the FACIT-Fatigue ($r=-0.67$). Among the correlations with the HPT-SD Impact subscale scores, the strongest construct validity correlation was with the FACT-Cog impact subscale ($r = -0.71$).

Finally, known-groups results showed that the average HPT-SD subscale scores were higher for patients who rated their HPT as "severe" or "very severe" compared with patients who rated their HPT as "very mild" or "mild" (Table S2): HPT-SD Symptom subscale score mean =1.9, SD =0.8 vs mean =0.8, SD =0.3; HPT-SD Impact subscale score mean =1.3, SD =0.5 vs mean =0.4, SD =0.5. The differences between known groups were statistically significant.

Discussion

This study examined the psychometric properties of the HPT-SD, a patient-reported questionnaire that assesses the symptoms and impacts of HPT. The HPT-SD was developed according to qualitative methodology recommended in the FDA's PRO guidance,[6] and the content validity of the measure has been described previously.[5] The current analysis represents the first psychometric evaluation of the HPT-SD in patients with HPT and includes assessments of reliability and validity, as well as preliminary scoring rules. Future studies should assess responsiveness and explore preliminary responder definitions of the HPT-SD using longitudinal study designs.

Although 46% of patients in the study rated their HPT as "severe" or "very severe," item-level response frequency distributions and descriptive statistics showed evidence of floor effects for neuromuscular HPT-SD items (HPT-SD item 1 [muscle cramps], item 2 [tingling], item 3 [numbness], and item 4 [muscle spasms or twitching]). One possible explanation for the floor effects is that patients may experience severe symptoms less frequently than milder symptoms. Almost all study participants (81%) reported that they took extra calcium or vitamin D to control their symptoms. The onset of milder neuromuscular symptoms may be a signal for patients to take additional supplements, thereby reducing the probability of experiencing a severe symptom and resulting in floor effects.

For each HPT symptom captured with the HPT-SD, between 13.5% and 28.8% of patients indicated that they did not experience it; similarly, for each impact evaluated, between 17.3% and 42.3% of patients did not experience it. However, previous qualitative research on the HPT-SD supports the relevance and inclusion of each HPT-SD item.[5] These data support the notion that patients with HPT can experience a variety of combinations of HPT-related symptoms and impacts. Treatments for HPT were not controlled in this observational study and could play a role in symptom and impact severity. Future research should investigate whether symptoms and impacts experienced by patients vary by HPT severity level and whether modifications to the scoring algorithm should be made.

A preliminary HPT-SD scoring algorithm was based on descriptive statistics and inter-item correlations, as well as careful consideration of the qualitative research conducted during the development of the HPT-SD.[5] The 13 items of the HPT-SD were scored as the HPT-SD Symptom subscale (first seven items) and HPT-SD Impact subscale (last four items). Two HPT-SD items (anxiety and depression) were not included in either subscale because anxiety and depression could be attributed to other conditions. The results of the internal consistency reliabilities and missing data analyses provided further evidence for the scoring algorithm.

Internal consistency reliabilities of the HPT-SD scores were satisfactory, and inter-item correlations provided evidence that HPT-related symptoms and impacts were indeed related but separate concepts. For example, strong correlations were observed between HPT-SD item 6 (physical fatigue) and item 12 (impact work) (r=0.70), and HPT-SD item 9 (sadness or depression) and item 13 (impact relationships) (r=0.73), suggesting that there may be different symptom drivers for different HPT-related impacts.

Mean FACIT-Fatigue scores (mean, 20.3) suggest that the analysis population experienced severe fatigue: more severe, on average, than patients with iron-deficiency anemia (mean, 24.1 for patients on Ferumoxytol; mean, 24.7 for patients on placebo)[15] and more severe than anemic cancer patients (mean, 23.9).[16] Fatigue may play an important role in patients' ability to work or socialize with friends and family, and it may contribute to mood disturbance and other observed impacts on quality of life.[17]

With respect to validity, construct validity correlations provided support for convergent validity. Construct validity hypotheses were always in the anticipated direction and of the approximate size expected. Known-groups analyses demonstrated the discriminant validity of the HPT-SD to distinguish between known groups of patients classified according to HPT severity.

Some limitations of this study must be considered. The survey sample was relatively small (n=52), and respondents constituted a convenience sample. Therefore, generalizability of the study findings should be carefully evaluated. Future studies may be warranted to further validate the current results and explore the potential impact of demographic factors (eg, sex) or clinical factors (eg, menopausal syndrome). Moreover, clinical data, including respondents' HPT diagnosis, severity of disease, and HPT-related clinical information, were self-reported and were not independently confirmed by a clinician; thus, this information may be limited to patient recall bias and subjective assessment (eg, severity).

Conclusion

The present study provides important results regarding the psychometric properties of the HPT-SD in a sample of individuals with HPT. The 13 HPT-SD items generally met accepted psychometric criteria and corroborate the findings of the qualitative research undertaken in the development of the HPT-SD.[5] Collectively, the results of the psychometric analyses indicate that the HPT-SD is an appropriate measure of HPT-related symptoms and impacts.

Acknowledgments

Disclosure

are employees of RTI Health Solutions, a unit of RTI International, a not-for-profit research organization that provides consulting and research services to governmental organizations, businesses, and pharmaceutical companies. Kristina Chen and Montserrat Vera-Llonch are employees of Shire Human Genetic Therapies. Alan Krasner was an employee of Shire Human Genetic Therapies when this research was conducted. The authors report no other conflicts of interests in this work.

References

1. Bilezikian JP, Khan A, Potts JT, et al. Hypoparathyroidism in the adult: epidemiology, diagnosis, pathophysiology, target-organ involvement, treatment, and challenges for future research. *J Bone Miner Res*. 2011;26(10):2317–2337.
2. Clarke BL, Brown EM, Collins MT, et al. Epidemiology and diagnosis of hypoparathyroidism. *J Clin Endocrinol Metab*. 2016;101(6): 2284–2299.
3. Powers J, Joy K, Ruscio A, Lagast H. Prevalence and incidence of hypoparathyroidism in the United States using a large claims database. *J Bone Miner Res*. 2013;28(12):2570–2576.
4. Shoback DM, Bilezikian JP, Costa AG, et al. Presentation of hypoparathyroidism: etiologies and clinical features. *J Clin Endocrinol Metab*. 2016;101(6):2300–2312.
5. Martin SA, Chen K, Harris NI. Development of a patient-reported outcome measure for chronic hypoparathyroidism. Presented at the ISPOR 19th Annual European Congress; November 1, 2016; Vienna, Austria. *Value Health.* 2016;19(7):A679.
6. FDA. Guidance for industry. *Patient-reported outcome measures: use in medical product development to support labeling claims*. FDA; 2009. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed July 13, 2017.
7. Jacobs SR, Jacobsen PB, Booth-Jones M, Wagner LI, Anasetti C. Evaluation of the functional assessment of cancer therapy cognitive scale with hematopoietic stem cell transplant patients. *J Pain Symptom Manage*. 2007;33(1):13–23.
8. Webster K, Cella D, Yost K. The Functional Assessment of chronic illness therapy (FACIT) measurement system: properties, applications, and interpretation. *Health Qual Life Outcomes*. 2003;1:79.
9. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361–370.
10. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334.
11. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 2nd ed. New York: Oxford University Press; 1995.
12. Mitchell J, Bradley C. Psychometric evaluation of the 12-item Well-being Questionnaire for use with people with macular disease. *Qual Life Res*. 2001;10(5):465–473.
13. Coon CD, Williams VSL, Nelson LM, et al. Determining missing data rules for patient-reported outcomes: alpha-if-item-deleted. Presented at the ISPOR 15th Annual Meeting; May 17, 2010; Atlanta, Georgia.
14. Fairclough DL, Cella DF. Functional assessment of cancer therapy (FACT-G): non-response to individual questions. *Qual Life Res*. 1996;5(3):321–329.
15. Acaster S, Dickerhoof R, Debusk K, Bernard K, Strauss W, Allen LF. Qualitative and quantitative validation of the FACIT-fatigue scale in iron deficiency anemia. *Health Qual Life Outcomes*. 2015;13:60.
16. Cella D, Lai JS, Chang CH, Peterman A, Slavin M. Fatigue in cancer patients compared with fatigue in the general United States population. *Cancer*. 2002;94(2):528–538.
17. Bianchi GP, Zaccheroni V, Solaroli E, et al. Health-related quality of life in patients with thyroid disorders. *Qual Life Res*. 2004;13(1):45–54.

Supplementary materials

Table S1 Treatment characteristics (N=52)

	Statistics
Calcium type, n (%)[a]	
Calcium carbonate	24 (47.1)
Calcium citrate	24 (46.2)
Calcium gluconate	0 (0)
Other	12 (23.5)
Do not know	0 (0)
Vitamin D type, n (%)[a]	
Over the counter	10 (19.2)
Ergocalciferol	3 (5.9)
Calcitriol	45 (86.5)
Other	6 (11.8)
Do not know	0 (0)
Over the past 6 months have you taken additional calcium/vitamin D due to symptoms? n (%)	
Yes	42 (80.8)
No	10 (19.2)
Have you had a blood test of your calcium level in the past year? n (%)	
Yes	51 (98.1)
Number of blood tests: mean (SD), median, min-max	8.6 (13.0), 4.0, 1–60
No	1 (1.9)
Was your most recent blood calcium level within normal range, n (%)[b]	
Yes	29 (55.8)
No	23 (44.2)
Have you had a urine test of your calcium level in the past year? n (%)	
Yes	25 (48.1)
Number of urine tests: mean (SD), median, min-max	1.7 (1.5), 1.0, 0–6
No	27 (51.9)
Was your most recent urine calcium level within normal range, n (%)[c]	
Yes	18 (34.6)
No	16 (30.8)
Not applicable	18 (34.6)

Notes: Percent calculated from non-missing responses. [a]Percentages may not total 100 because patients checked all options that applied. [b]The most recent blood calcium level may or may not have been measured in the past year. [c]The most recent urine calcium level may or may not have been measured in the past year.

Table S2 Known-groups results

HPT-SD score	Patient-reported HPT severity[a] = very mild or mild		Patient-reported HPT severity = severe[a] or very severe		_P_-value[b]
	N	Mean (SD)	N	Mean (SD)	
Item 1. Muscle cramps	10	1.0 (1.1)	24	4.9 (2.2)	<0.0001
Item 2. Tingling	10	1.0 (0.5)	24	1.6 (0.9)	0.0698
Item 3. Numbness	10	0.4 (0.5)	24	1.5 (1.1)	0.0039
Item 4. Muscle spasms or twitching	10	0.7 (0.5)	24	1.9 (1.1)	0.0024
Item 5. Feelings of heaviness in arms or legs	10	0.6 (0.7)	24	1.5 (1.1)	0.0277
Item 6. Physical fatigue	10	1.7 (0.7)	24	2.7 (1.0)	0.0084
Item 7. Slowed or confused thinking (brain fog)	10	0.7 (0.8)	24	2.0 (1.1)	0.0018
Item 8. Anxiety	10	0.9 (0.9)	24	2.0 (1.3)	0.0223
Item 9. Sadness or depression	10	0.5 (0.7)	24	1.6 (1.2)	0.0153
Item 10. Impact sleep	10	0.6 (0.7)	24	1.3 (0.7)	0.0236
Item 11. Impact exercise	10	0.5 (0.7)	24	1.7 (0.5)	<0.0001
Item 12. Impact work	10	0.3 (0.5)	23	1.2 (0.7)	0.0005
Item 13. Impact relationships	10	0.2 (0.4)	24	1.0 (0.8)	0.0073
HPT-SD symptom subscale score (1: Muscle cramps[c]; 2: Tingling; 3: Numbness; 4: Muscle spasms or twitching; 5: Feelings of heaviness in arms; 6: Physical fatigue; 7: Slowed or confused thinking)	10	0.8 (0.3)	24	1.9 (0.8)	0.0002
HPT-SD impact subscale score (10: Sleep; 11: Exercise; 12: Work; 13: Relationships)	10	0.4 (0.5)	23	1.3 (0.5)	<0.0001

Notes: [a]Groups based on responses to PGIS. [b]ANOVA tests "very mild to mild" subgroup with "severe to very severe" subgroup. [c]To calculate the HPT-SD Symptom subscale, item 1 was rescaled from an 11-point response scale to a 5-point response scale ranging from 0 to 4.
Abbreviations: HPT, hypoparathyroidism; HPT-SD, Hypoparathyroidism Symptom Diary; PGIS, patients' global impression of symptom severity.