doi: 10.4183/aeb.2018.192

THE ROMANIAN VERSION OF THE THYROID-RELATED PATIENT-REPORTED OUTCOMES THYPRO AND THYPRO-39. TRANSLATION AND ASSESSMENT OF RELIABILITY AND CROSS-CULTURAL VALIDITY

A.E. Zahan[1,4,*], T. Watt[5], I. Pascanu[2,4], A.K. Rasmussen[5], L. Hegedüs[7], S.J. Bonnema[7], U. Feldt-Rasmussen[5], J.B. Bjorner[6], V. Nadasan[3], A. Boila[1], I. Merlan[4], A. Borda[1]

University of Medicine and Pharmacy of Tîrgu Mures, [1]Dept. of Histology, [2]Dept. of Endocrinology, [3]Dept. of Hygiene, [4]Emergency County Hospital, Tîrgu Mures, Dept. of Endocrinology, Tirgu Mures, Romania, [5]Copenhagen University Hospital Rigshospitalet, Dept. of Medical Endocrinology, [6]Institute of Public Health, University of Copenhagen, Copenhagen, [7]Odense Universitetshospital, Odense, Denmark

Abstract

Background. ThyPRO is a recently developed thyroid-specific quality of life (QoL) questionnaire applicable to patients with benign thyroid disorders(BTD). The aim of the present study was to translate ThyPRO and ThyPRO-39 into Romanian, and to evaluate reliability and cross-cultural validity.

Methods. Standard methodology for translation and linguistic validation of patient-reported outcomes (PRO) was applied. The questionnaire was completed by 130 patients with benign thyroid diseases seen at Department of Endocrinology in the Emergency County Hospital, Tîrgu Mureș, Romania, between October 2015 and March 2016. Internal reliability of the Romanian version of the ThyPRO (ThyPROro) scales was assessed for multi-item scales using Cronbach's alpha coefficient. An efficient method for testing cross-cultural validity is analysis of differential item functioning (DIF). Uniform DIF between the Romanian and the original Danish sample was investigated using ordinal logistic regression. The translation process proceeded without difficulties, and any disagreements were revised by one of the developers and the language coordinator.

Results. Internal reliability for ThyPRO was satisfactory. Cronbach`s alpha coefficients for the 13 scales ranged from 0.78 to 0.93 for the ThyPROro and 0.78 to 0.87 for the ThyPROro-39. In the 85-item ThyPRO, nine instances of DIF were found. Most were minor, explaining <3% of the variation in scale score, but DIF in positively worded items were larger, with explained variance (R2's) around 10-15%.

Conclusion. The ThyPROro questionnaire is ready for assessment of health-related quality of life in Romanian patients with benign thyroid diseases.

Key words: benign thyroid disease, quality of life, questionnaire, Romanian versions.

INTRODUCTION

Benign thyroid diseases are very common endocrine disorders which affect individuals of all ages, mainly women. For most patients a main goal is longevity in good health, because many thyroid disorders are associated with increased morbidity and excess mortality (1, 2). Therefore the treatment should be focused on optimizing the quality of life (QoL) (3, 4). According to the available literature, health-related quality of life (HRQoL) is impaired in patients with benign thyroid disorders, such as non-toxic goiter, hyperthyroidism (either as toxic nodular goiter or Graves' disease - with or without Graves' orbitopathy) and autoimmune or post-thyroidectomy hypothyroidism (5-9). In the past decades, the importance of involving measurements of HRQoL aspects in the evaluation of thyroid patients has been increasingly recognized (3-8). HRQoL concerns the patients' point of view in terms of how they feel and what they are able to do in everyday life. It has been defined as the subjective assessment of the impact of disease and its treatment on the physical, psychological, social, and somatic domains of functioning and well-being (4, 10). To an increasing extent, the broader, but also more neutral term 'patient-reported outcomes (PROs)' is replacing HRQoL.

QoL can be assessed by generic and/or disease specific QoL questionnaires (11). ThyPRO is the first comprehensive thyroid-specific patient-reported outcome questionnaire measuring QoL, recently developed by Watt *et al.*(12) in Denmark. This instrument has been validated for patients with any benign thyroid disease (4, 12-14) with a good cross-cultural validity (15), and is recommended for use in

*Correspondence to: Ancuta Elena Zahan MD, University of Medicine and Pharmacy of Tîrgu Mures, Dept. of Histology, Gh. Marinescu, 38, Tîrgu Mureș, Mureș, 540000, Romania, E-mail: ancutza_ct@yahoo.com

clinical studies (16) as well as in daily clinical practice (17). Recently, an abbreviated, 39-item version of ThyPRO was developed - ThyPRO-39 - with optimized cross-cultural validity (18).

Currently, no thyroid-specific instrument for measuring thyroid-specific HRQoL exists in Romanian. When pursuing a measurement instrument in a new population, already existing validated measures in other languages should be translated, adapted and cross-culturally validated, to enhance cross-cultural comparability (19). The evaluation of cross-cultural validity implies the analysis of differential item functioning (DIF) using either item response theory methods or logistic regression (20-23). DIF is used when there is a difference in the response to a specific item for subjects from different countries, with the same overall score or, more precisely, the same level of a measured attribute (24-27). The previous cross-cultural validity assessment of the ThyPRO questionnaire, using DIF analysis, has had good results (15).

The aim of the present study was to translate ThyPRO and ThyPRO-39 into Romanian, and to evaluate reliability and cross-cultural validity of both versions.

MATERIALS AND METHODS

ThyPRO questionnaire – characteristics and translation steps

The ThyPRO consists of 85 items summarized in 13 scales covering physical and mental symptoms, function and well-being as well as impaired participation in important life activities, as shown in Table 1 (12, 13). Subsets of 39 of these 85 items are included in ThyPRO-39, the scales of which can thus be constructed from the responses to the 85-item version. The selected questions for ThyPRO-39 are presented in Table 1.

Approximately 15-20 minutes are needed to answer the 85-item version, which has a reference period of four weeks. Each item is rated by the patient on a five-point Likert scale: 0 — not at all; 1 — a little; 2 — some; 3 — quite a bit; 4 — very much/completely. The 13 scales are derived by averaging and linearly transforming these item scores into their respective 0–100 scale score, increasing scores indicating decreasing QoL (i.e., more symptoms or greater impact of disease) and lower scores indicating better QoL. The score is summarized according to the formula "Transformed score = (raw sum score/highest possible raw sum score)*100". For example, the Tiredness scale contains: TQ2A TQ2B TQ2C TQ2D TQ3A TQ3B TQ3C; if the patient`s response is "not at all" to all 7 questions, the scale score is 0. One of the hypothetical versions of answers could be: two responses -"a little", two responses -"some", two responses -"quite a bit" and one response - "very much", then the scale score will be (1 +1+ 2 + 2 + 3 + 3 + 4)/ 28 *100 = 16 /28 *100 = 57.

After the developers' approval, the questionnaire was translated into Romanian. Standard methodology for translation and linguistic validation of PRO was applied (28, 29): Initially, two independent translations of the English version of the ThyPRO questionnaire into Romanian were performed by two independent professional translators whose native language is Romanian. A reconciled version was constructed on the basis of these translations, in collaboration with a language coordinator. This "reconciled forward translation" was translated back into English by a professional translator whose native language is English. The backwards translation was compared to the original English version by a consultant expert native in English, one of the developers of ThyPRO (TW) and the language coordinator. Any issue arising during the process was documented and if revisions were made, the forward-backward translation was repeated, until a final version was reached. Finally, ThyPROro was tested by the cognitive interview techniques in five patients with thyroid disorders (hyperfunctioning nodule with subclinical hyperthyroidism, Hashimoto's thyroiditis in two patients, multinodular non-toxic goiter and Graves' disease with orbitopathy), with a median age of 56 (range 40 - 66 years) and different educational levels. The procedure included, but it was not limited to five questions: "1. Did you have any difficulty understanding this item?; 2. What does this item mean to you?; 3. Would you rephrase this item?; 4. Is this item relevant to your situation?; 5. If the response options are not clear and consistent with this item, how would you rephrase them?".

Data base and study design

Patients with benign thyroid disease, seen in the Department of Endocrinology at Emergency County Hospital, Tîrgu Mures, Romania, between October 2015 and March 2016, were included (Romanian sample). Inclusion criteria were: age above 18 years and benign thyroid disease. The benign thyroid diseases comprised: non-toxic goiter (diffuse non-toxic goiter, multinodular non-toxic goiter, uni-nodular non-toxic goiter and solitary cyst), toxic nodular goiter (multinodular toxic goiter and toxic adenoma), Graves'

Table 1. Scales and items of ThyPRO and ThyPRO - 39

Physical symptom scales

Goitre Symptoms	Hyperthyroid Symptoms	Hypothyroid Symptoms	Eye Symptoms
1a - Sense of fullness in neck	*1l -Trembling hands*	*1q - Sensitive to cold*	1u - Watery eyes
1b - Visible swelling in front of neck	*1m - Increased sweating*	*1cc - Swollen hands or feet*	1v - Bags under the eyes
1c - Pressure in throat	*1n - Palpitations*	*1dd - Dry skin*	*1w - Grittiness in eyes*
1d - Pain in front of neck	1o - Shortness of breath	*1ee - Itching skin*	*1x - Reduced sight*
1e - Throat pain felt in ears	1p - Sensitive to heat		1y -Pressure in eyes
1f - Lump in throat	1r - Increased appetite		1z - Double vision
1g - Clear throat often	1s - Loose stools		1aa - Pain in eyes
1h - Discomfort swallowing	*1t - Upset stomach*		*1bb - Sensitive to light*
1i - Difficulty swallowing			
1j - Sense of suffocating			
1k – Hoarseness			

Function and well-being scales

Tiredness	Cognitive Impairment	Emotional Susceptibility
2a - Been tired	*4a - Problems remembering*	7a - Difficulty coping
2b - Been exhausted	*4b - Slow or unclear thinking*	7b - Not like yourself
2c - Difficulty getting motivated	4c - Difficulty finding words	*7c - Easily stressed*
2d - Felt worn	4d - Been confused	*7d - Mood swings*
3a - Full of life	4e - Difficulty learning	7e - Irritable
3b - Energetic	*4f – Difficulty concentrating*	7f - Frustrated
3c - Able to cope with life		7g - Angry
		7h - Felt in control
		7i - Felt in balance

Psychological symptom scales

Anxiety	Depressivity
5a - Nervous	*6a - Sad*
5b - Afraid or anxious	6b - Depressed
5c - Felt tension	6c - Discouraged
5d - Concerned being seriously ill	6d - Crying easily
5e - Uneasy	*6e - Unhappy*
5f – Restless	6f - Happy
	6g - Self-confident

Participation scales

Impaired Social Life	Impaired Daily Life		Cosmetic Complaints
8a - Difficult being with other people	*9a - Difficulty managing daily life*	**Impaired Sex Life**	*11a - Disease affect appearance*
8b - A burden to other people	9b - Limit leisure activities	10a - Negative influence on sex life	11b - Unsatisfied with appearance
8c - Conflicts with other people	*9c - Difficulty participating in life*	10b - Decreased sexual desire	11c - Camouflage or mask visible signs
8d - People lack understanding	9d - Difficulty getting around		*11d - Bothered by other people looking*
	9e - Everything takes longer		*11e - Influence on clothes worn*
	9f - Difficulty managing job		11f - Felt too fat

Overall Quality of Life

12 - Has your thyroid disease had a negative effect on your quality of life?

disease with orbitopathy, Graves' disease without orbitopathy, Hashimoto's thyroiditis and post-operative hypothyroidism. Histopathology for all patients with post-operative hypothyroidism was benign. Exclusion criteria were major psychiatric disorders, malignancy or severe comorbidity.

Sociodemographic data were self-reported as part of the questionnaire and included: gender, age, education level, marital status and current employment status. The clinical characteristics were retrieved from medical records and contained: biochemical thyroid function tests (TSH, free T4, free T3), thyroid peroxidase antibodies (TPOAb), TSH receptor antibodies (TRAb) at the time of diagnosis, thyroid ultrasound (thyroid volume), type of thyroid disorder (if available), duration of the disease, previous and current treatment, and comorbidity.

The patients completed the ThyPROro questionnaire in the presence of a physician, who dealt with HRQoL assessments, so there were no missing data, nor reading or writing problems.

The study was approved by the Ethical Committee of the University of Medicine and Pharmacy of Tîrgu Mures, Romania, and written informed consent was obtained from all patients prior to their inclusion in the study.

As described elsewhere (12, 15), patients from Denmark comprised patients referred to the endocrine outpatient clinics at Copenhagen University Hospital, Rigshospitalet, and Odense University Hospital for treatment of benign thyroid diseases. All completed the ThyPRO questionnaire and had blood drawn a few days prior to or as part of their scheduled visit to the clinics.

Statistical analysis

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS, version 20, Chicago, IL, USA) and SAS V9.3. Quantitative variables were tested for normality of distribution using the Kolmogorov-Smirnov test and were described by mean ± standard deviation, or median and percentiles (25; 75%), whenever appropriate.

Reliability analysis

Reliability is defined as the ability of a test to "yield the same results on repeat trials under the same conditions"(30). Internal reliability of ThyPROro scales was assessed for multi-item scales using Cronbach's α coefficient. An alpha coefficient above 0.90 is considered "excellent", above 0.80 as "very good", while above 0.70 is "adequate" (30, 31).

Cross-cultural validity analysis

The Romanian versions were compared pairwise to the original Danish version by examining uniform and non-uniform DIF (12). For the DIF, ordinal logistic regression was used (32), with a level of significance alpha=0.05 with Bonferroni correction (33, 34). The uniform and non-uniform DIF were examined by comparing pairwise the Danish and Romanian versions. Apart from the statistical significance, a DIF must induce a change in the determination coefficient (r2) of at least 0.02 (32, 35), i.e. a change of more than 2% in the variance of the dependent variable which is explained by the independent variable (32, 36, 37). For the estimation of the score level, the sum score was used, taking into account the item being tested and excluding the other items with DIF, thus purifying the sum score (32, 38).

RESULTS

Translation of the ThyPRO

During the forward translation step, a few translator disagreements were observed. A reconciled version was constructed on the basis of two independent translations of the English version (A, B), in collaboration with an expert native in English. The translations were identical for 27 items. Other 6 items were chosen combining translation A and B. For 29 items, translation A was closer to the intended meaning, more adequate, explicit and colloquial than translation B, and vice versa for 23 other items. After the backwards translation had been compared to the original English version by a consultant expert native in English, the developer and the language coordinator, 21 questions had to be revised. For example, the original English: "Page 4, item 4d - been confused?"; Reconciled Version Romanian: "Pagina 4, punctul 4d - ati avut momente de confuzie?"; Back Translation: "Page 4, item 4d - have you had moments of confusion?"; Evaluation: "Developer: Please explain why it is necessary to include the term "moments" in this item. Consultant: In Romanian this term leads to a better and a clearer understanding of the question. Developer decision: Accepted."

During the 5 cognitive interviews, 8 problems in the questionnaire response process were identified. Two problems involved comprehension: "During the last 4 weeks, have you had the sensation of fullness in the neck?" and "During the last 4 weeks, have you had difficulty managing your job (for example, finding it hard to cope or calling in sick)". Four of 5 patients had

difficulty in understanding this item. This situation was solved using and keeping the colloquial meaning of the words, and it was accepted by the developer and the language coordinator.

Internal reliability and cross-cultural validity of the ThyPROro and ThyPROro-39

One hundred and thirty Romanian patients (89% females), with a mean age of 53 years, were included. All completed the ThyPROro questionnaire, which on average took 20 minutes. Sociodemographic and clinical characteristics of the patients with thyroid disease are shown in Table 2, along with a description of the 902 patients from the original Danish validation sample.

The samples had comparable age, gender, education level, employment status, disease duration and previous treatment. A higher proportion of patients with Hashimoto's disease were included in the Romanian sample. In the Danish sample the majority of patients had non-toxic goiter. Mean ThyPRO scores of the 13 scales, according to language and type of thyroid disease, are presented in Table 3.

Cronbach`s alpha coefficients, for each scale of both versions, are summarised in Table 4. The reliability estimates for these 13 scales ranged from 0.78 to 0.93 for the ThyPROro (85-item) and from 0.78 to 0.87 for the ThyPROro (39-item).

Table 5 illustrates the results of DIF analysis, with nine instances of DIF for the 85-item ThyPRO. The majority were minor, with less than 3% influence on the scales scores' variance, but the DIF for positively worded items had R2 around 10-15%, for which there was a consistent pattern. In comparison with the Danish subjects, the Romanians reported in a more exaggerated manner. I.e. for similar overall levels of health, they reported more positive effects at the best end of the scale and mirrored negative effects at the worst end of the scale. Similar results were obtained for the ThyPRO-39, with all three instances of DIF in the three positively worded items (Table 6).

Table 2. Sociodemographic and clinical characteristics of Romanian and Danish thyroid patients

Characteristics	Romanian sample	Danish sample
N	**130**	**902**
Gender n (%)		
Female	115 (89)	784 (87)
Male	15 (12)	118 (13)
Age (mean ± SD, years)	53±11	51±15
Environment n (%)		
Urban	85 (65)	902 (100)
Rural	45 (35)	0
Marital status, n (%)		
Married/unmarried Couple	97 (75)	482 (56)
Single	9 (7)	192 (22)
Widow	17 (13)	61 (7)
Divorced	7 (5)	127 (15)
Education level		
Primary school (0-10 years)	8 (6)	151 (17)
Gymnasium school (11-14 years)	29 (22)	110 (12)
High school (15- 18 years)	64 (49)	261 (29)
University school (>18 year)	29 (23)	346 (38)
Current employment status, n (%)		
Unemployed	21 (16)	33 (4)
Pensioner	4 (3)	229 (30)
Unqualified worker	64 (49)	39 (5)
Qualified worker	23 (18)	85 (11)
Highly-qualified worker	18 (14)	370 (49)
Diagnosis		
Non-toxic goiter	41 (32)	259 (29)
Toxic nodular goiter	9 (7)	145 (16)
Graves` disease with/without orbitopathy	19 (15)	260 (28)
Hashimoto's thyroiditis	46 (35)	198 (22)
Other thyroid diagnosis	15 (12)	40 (4)
Disease duration (mean ± SD, years)	6±6.5	5±7.5
Current treatment, n (%)		
No treatment	64 (49)	448 (50)
L-Thyroxine	40 (31)	288 (32)
Antithyroid drugs	23 (18)	162 (18)
Other[a]	3 (2)	4 (0)
Current thyroid function[b] n (%)		
Euthyroid	95 (73)	530 (58)
Subclinical hypothyroid	18 (14)	124 (14)
Subclinical hyperthyroid	3 (2)	122 (13)
Hypothyroid	9 (7)	15 (2)
Hyperthyroid	5 (4)	98 (11)

Other thyroid diagnosis: Post-operative hypothyroidism, subacute thyroiditis. [a]Other: Radioiodine, thyroid surgery. [b]Thyroid function: Euthyroid: normal TSH and fT4. Subclinical hypothyroid: elevated TSH and normal fT4. Hypothyroid: elevated TSH and decreased fT4. Subclinical hyperthyroid: decreased TSH and normal fT4 and fT3. Hyperthyroid: decreased TSH and elevated fT4 or fT3.

DISCUSSION

The purpose of the present study was to translate the ThyPRO questionnaires into Romanian, and to establish a reliable and cross-culturally valid scale structure of the ThyPROro in Romanian patients with various benign thyroid diseases. We decided to translate the ThyPRO questionnaire because it is the only thyroid specific QoL-questionnaire for patients with benign thyroid diseases. It has appropriate reliability

Table 3. Mean ThyPRO scale scores (higher scores with worse symptoms/problems) according to language sample and thyroid disease

Scales	Goiter sympt	Hyper sympt	Hypo sympt	Eye sympt	Tiredness	Cognition	Anxiety	Depressivity	Emotion suscept	Social impact	Daily Impact	Sex life	Cosmetic complaints
Language													
Danish (n=902)	16 (17)	22 (18)	22 (21)	16 (17)	49 (26)	22 (22)	22 (23)	29 (22)	33 (23)	11 (17)	18 (24)	22 (31)	18 (21)
Romanian (n=130)	27 (20)	36 (19)	39 (22)	28 (22)	47 (10)	28 (22)	39 (25)	43 (15)	42 (14)	24 (22)	28 (24)	23 (25)	24 (21)
Thyroid disease													
Non-toxic goiter (n=41)	27 (19)	32 (19)	38 (22)	28 (21)	46 (10)	28 (20)	38 (19)	42 (13)	41 (14)	24 (19)	27 (23)	25 (25)	21 (17)
Toxic nodular goiter (n=9)	30 (16)	41 (17)	39 (18)	29 (26)	51 (13)	34 (24)	36 (28)	42 (14)	40 (14)	27 (20)	36 (21)	29 (33)	32 (29)
Graves` disease (n=2)	26 (8)	38 (13)	34 (4.4)	25 (8.8)	57 (15)	23 (21)	33 (6)	36 (5)	39 (8)	31 (27)	31 (29)	13 (17)	6 (3)
Graves` orbitopathy (n=17)	34 (23)	38 (20)	37 (16)	24 (20)	49 (9)	27 (22)	38 (25)	41 (12)	45 (14)	24 (21)	30 (26)	24 (27)	24 (16)
Hashimoto's thyroiditis (n=46)	30 (21)	27 (20)	31 (25)	31 (23)	46 (9)	29 (25)	41 (29)	44 (17)	43 (16)	24 (25)	29 (26)	24 (23)	25 (24)
Other thyroid disease (n=15)	24 (21)	37 (16)	39 (21)	26 (23)	46 (12)	24 (19)	40 (23)	44 (17)	42 (12)	20 (23)	17 (21)	17 (24)	24 (19)

Table 4. Internal consistency reliability coefficients (Cronbach's alpha) from ThyPROro (85-item) and ThyPROro-39

Scales	Cronbach's α ThyPROro	Cronbach's α ThyPROro-39
Goitre symptoms	0.90	0.81
Hyperthyroid Symptoms	0.78	0.78
Hypothyroid Symptoms	0.87	0.82
Eye Symptoms	0.86	0.79
Tiredness	0.88	0.83
Cognitive Impairment	0.92	0.87
Anxiety	0.91	0.87
Depressivity	0.93	0.82
Emotional Susceptibility	0.88	0.82
Impaired Social Life	0.86	0.81
Impaired Daily Life	0.85	0.84
Impaired Sex Life	0.87	
Cosmetic Complaints	0.83	0.86

(equivalent to the "precision" in medical biochemistry) (12), clinical validity (equivalent to accuracy) (13), and cross-cultural validity as investigated in several languages (15).

Applying the recommendations of the methodology for translation and linguistic validation PRO translation guidelines (29) only two problems, involving comprehension, were identified. One of the reasons for these might be the structural difference between Romanian and English. Sometimes the word in English may not have an equivalent in Romanian or some words may have many different meanings where used. Finally, the developer and the consultant gave their consent, and ThyPROro and ThyPROro - 39 were approved.

Romanian patients had higher mean scores on the Goitre Symptoms, Hyperthyroid Symptoms, Hypothyroid Symptoms, Eye Symptoms, Anxiety, Depressivity, Emotional Susceptibility, Social life and Daily life scales, compared to Danish patients. Also, the standard deviation was lower in Romania for several scales: Tiredness, Depressivity, and Emotional susceptibility and Sex Life. A similar tendency has previously been found. Compared to English patients, Danish patients scored lower on Tiredness, Depressivity, Emotional Susceptibility and Cosmetic Complaints scales (15).

Precision, or reliability, can be evaluated by several techniques, but the most widely used approach is 'internal consistency reliability' using Cronbach's alpha (17, 18). Previous analyses have shown satisfactory internal consistency and reliability of ThyPRO as estimated by Cronbach`s alpha (10).

The alpha coefficients suggest that the ThyPROro (85-item) and ThyPROro-39 are reliable instruments for evaluating QoL of patients with benign thyroid diseases. The values of Cronbach`s alpha in our study, being in the range 0.78-0.93 for the ThyPROro (85-item) and 0.78-0.87 for the ThyPROro-39, were comparable to previous findings (12, 39). Watt *et al.* reported a Cronbach`s alpha above 0.9, a very high reliability estimate, from 904 respondents (12). Similarly, Branka *et al.* showed reliability coefficients for Serbian patients in the range 0.83-0.95 (39). As usual

Table 5. Differential item functioning of the Romanian ThyPRO (85-item) questionnaire

Abbreviated items wording	Significance of DIF	Variance explained by DIF	Direction of DIF
Goitre Symptoms scale			
Visible swelling in front of neck	<0.0001	2.9%	Romanian patients report more visible swelling
Sense of suffocating	<0.0001	2.9%	Romanian patients report less sense of suffocating
Hyperthyroid Symptoms scale			
Increased appetite*	0.0004	2.1%	Romanian patients report more appetite with few hyperthyroid symptoms but less increased appetite with more hyperthyroid symptoms
Anxiety scale			
Nervous*	0.0006	2.0%	Romanian patients report more nervousness, mostly at low levels of anxiety
Depressivity scale			
Happy*	<0.0001	11.3%	Romanian patients report being more happy at low level of depressivity and less happy at high levels of depressivity; i.e. they have a more exaggerated response
Self confident*	<0.0001	9.9%	Romanian patients report being more self-confident at low levels of Depressivity and less self-confidence at high levels of Depressivity; i.e. they have a more exaggerated response
Emotional Susceptibility			
In control*	<0.0001	13.1%	Romanian patients report being more in control at low levels of emotional susceptibility and less in control at high levels of emotional susceptibility; i.e. they have a more exaggerated response
In balance*	<0.0001	14.5%	Romanian patients report being more in balance at low levels of emotional susceptibility and less in balance at high levels of emotional susceptibility; i.e. they have a more exaggerated response
Impaired Daily Life			
Difficult getting around	<0.0001	2.4%	Romanian patients report greater difficulty getting around

*Non-uniform DIF.

Table 6. Differential item functioning of the Romanian ThyPRO-39 questionnaire

Abbreviated items wording	Significance of DIF	Variance explained by DIF	Direction of DIF
Tiredness scale			
Energetic*	<0.0001	10.9%	Romanian patients report being more energetic at low levels of tiredness and less energetic at high levels of tiredness; i.e. they have a more exaggerated response
Depressivity scale			
Self confident*	<0.0001	3.4%	Romanian patients report being less self-confident at higher levels of expressivity
Emotional Susceptibility scale			
In control	<0.0001	3.3%	Romanian patients report being less in control

*Non-uniform DIF.

with multi-item scales, the larger number of items in the 85-item version of ThyPRO led to increased reliability. The DIF evaluation showed that the Romanian version has acceptable cross-cultural validity, compared with the Danish original version with the exception of the positively worded items.

Of the negatively worded items, 'visible swelling in front of neck' has previously been identified as problematic (15, 40, 41). 'Sense of suffocating' also displayed DIF in a previous Serbian sample (15). The remaining three DIFs have not previously been identified as being prone to DIF, but their magnitude was very small, explaining only about 2% of the variation.

DIF in the positively worded items have also been identified previously, albeit with smaller magnitudes, as judged by the variance explained (R2) by the DIF. Although a minor violation of the unidimensionality assumption of the scale might be the explanation for this, the questionnaire's developers

have accepted this, first of all because it minimizes the ceiling effect, a desirable feature of the instrument (the ability to discriminate the less affected individuals) (18), second of all, because factor analysis confirmed an overall one factor solution as dominant (41).

Future studies should focus on the positively worded items to validate these further and possibly revise the wording, before applying the ThyPRO-39ro in cross-cultural multicentre studies. Future studies attempting to pursue this should test revised wordings in mixed methods set-ups. Until this has been settled, we recommend accounting for this in statistical analyses, e.g. by testing robustness of findings by evaluating effect of omitting the positively worded items from scale scoring. Dimensionality should also be evaluated further, e.g. by confirmatory factor or multi-trait analyses. In case differences in mean scale score levels between similar groups of patients from different countries, as found in this study, are also found in larger samples, wording of response options also of the negatively worded items should be critically reviewed.

In conclusion, good reliability of the ThyPROro (85 item) and ThyPROro-39 was found. Cross-cultural validity was good, but attention should be given to the three negatively worded items. The Romanian version of the ThyPRO questionnaire is ready for use in clinical studies assessing HRQoL in Romanian patients with benign thyroid diseases.

Conflict of interest

The authors declare that they have no conflicts of interest.

Researchers interested in using the Romanian ThyPROro or the ThyPRO-39ro may contact the corresponding author. For other language versions, contact Torquil Watt.

References

1. Brandt F, Almind D, Christensen K, Green A, Brix TH, Hegedus L. Excess mortality in hyperthyroidism: the influence of preexisting comorbidity and genetic confounding: a Danish nationwide register-based cohort study of twins and singletons. J Clin Endocrinol Metab 2012;97(11):4123-4129.
2. Thvilum M, Brandt F, Almind D, Christensen K, Hegedus L, Brix TH. Excess mortality in patients diagnosed with hypothyroidism: a nationwide cohort study of singletons and twins. J Clin Endocrinol Metab 2013;98(3):1069-1075.
3. G.P. Bianchi VZ, E. Solaroli, F. Vescini, R. Cerutti, M. Zoli & G. Marchesini. Health-related quality of life in patients with thyroid disorders .A study based on Short-Form 36 and Nottingham Health Profile Questionnaires. Quality of Life Research 2004;13(9):45-54.
4. Watt T, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, Bjorner JB, Feldt-Rasmussen U. Quality of life in patients with benign thyroid disorders. A review. Eur J Endocrinol 2006;154(4):501-510.
5. Cramon P, Bonnema SJ, Bjorner JB, Ekholm O, Feldt-Rasmussen U, Frendl DM, Groenvold M, Hegedus L, Rasmussen AK, Watt T. Quality of life in patients with benign nontoxic goiter: impact of disease and treatment response, and comparison with the general population. Thyroid 2015;25(3):284-291.
6. Cramon P, Winther KH, Watt T, Bonnema SJ, Bjorner JB, Ekholm O, Groenvold M, Hegedus L, Feldt-Rasmussen U, Rasmussen AK. Quality-of-Life Impairments Persist Six Months After Treatment of Graves' Hyperthyroidism and Toxic Nodular Goiter: A Prospective Cohort Study. Thyroid 2016;26(8):1010-1018.
7. Winther KH, Cramon P, Watt T, Bjorner JB, Ekholm O, Feldt-Rasmussen U, Groenvold M, Rasmussen AK, Hegedus L, Bonnema SJ. Disease-Specific as Well as Generic Quality of Life Is Widely Impacted in Autoimmune Hypothyroidism and Improves during the First Six Months of Levothyroxine Therapy. PloS One 2016;11(6):e0156925.
8. Watt T, Hegedus L, Bjorner JB, Groenvold M, Bonnema SJ, Rasmussen AK, Feldt-Rasmussen U. Is Thyroid Autoimmunity per se a Determinant of Quality of Life in Patients with Autoimmune Hypothyroidism? Eur Thyroid J 2012;1(3):186-192.
9.Neagoe, RM., Cvasciuc IT, Muresan M, Sala DT Incidental Parathyroidectomy During Thyroid Surgery - Risk, Prevention and Controversies; an Evidence-Based Review. Acta Endocrinologica (Bucharest) 2017;13(4):467-475.
10. Ware JE. Conceptualization and measurement of health-related quality of life: comments on an evolving field. Arch Phys Med Rehabil 2003;84(4):S43-51.
11. Gasparik AI, Mihai G, Beaudart C, Bruyere O, Pop RM, Reginster JY, Pascanu IM. Psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. Archives of Osteoporosis 2017: 12:103.
12. Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, Feldt-Rasmussen U. Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. Qual Life Res 2009;18:483-496.
13. Watt T, Hegedus L, Groenvold M, Bjorner JB, Rasmussen AK, Bonnema SJ, Feldt-Rasmussen U. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. Eur J Endocrinol 2010;162(1):161-167.
14. Watt T, Cramon P, Frendl DM, Ware JE, Jr. ThyQo LG: Assessing health-related quality of life in patients with benign non-toxic goitre. Best Pract Res Clin Endocrinol Metab 2014;28(4):559-575.
15. Watt T, Barbesino G, Bjorner JB, Bonnema SJ, Bukvic B, Drummond R, Groenvold M, Hegedus L, Kantzer V, Lasch KE, Marcocci C, Mishra A, Netea-Maier R, Ekker M, Paunovic I, Quinn TJ, Rasmussen AK, Russell A, Sabaretnam M, Smit J, Torring O, Zivaljevic V, Feldt-Rasmussen U. Cross-cultural validity of the thyroid-specific quality-of-life patient-reported outcome measure, ThyPRO. Qual Life Res 2015;24(3):769-780.
16. Wong CK, Lang BH, Lam CL. 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.
17. Watt T, Cramon P, Hegedus L, Bjorner JB, Bonnema SJ, Rasmussen AK, Feldt-Rasmussen U, Groenvold M. The thyroid-related quality of life measure ThyPRO has good responsiveness and ability to detect relevant treatment effects. J Clin Endocrinol Metab 2014;99(10):3708-3717.
18. Watt T, Bjorner JB, Groenvold M, Cramon P, Winther KH, Hegedus L, Bonnema SJ, Rasmussen AK, Ware JE, Jr., Feldt-Rasmussen U. Development of a Short Version of the Thyroid-Related Patient-Reported Outcome ThyPRO. Thyroid 2015;25(10):1069-1079.

19. Fayers P, HaysR. Assessing Quality of Life in Clinical Trials - Methods and Practice, 2 edn. Oxford: Oxford University Press 2004.
20. Petersen MA, Groenvold M, Bjorner JB, Aaronson N, Conroy T, Cull A, Fayers P, Hjermstad M, Sprangers M, Sullivan M. European Organisation for R, Treatment of Cancer Quality of Life G: Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. Qual Life Res 2003;12(4):373-385..
21. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, Koller M, Petersen MA, Sprangers MA, Eortc, the Quality of Life Cross-Cultural Meta-Analysis G: The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. Qual Life Res 2007;16(1):115-129.
22. Martin M, Blaisdell B, Kwong JW, Bjorner JB. The Short-Form Headache Impact Test (HIT-6) was psychometrically equivalent in nine languages. J Clin Epidemiol 2004;57(12):1271-1278.
23. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, Gundy C, Koller M, Petersen MA, Sprangers MA, Group EQoL, the Quality of Life Cross-Cultural Meta-Analysis G: Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. Health Qual Life Outcomes 2010;8:81.
24. Thissen D SL, Gerrard M. Beyond group-mean differences: The concept of item bias. Psychological bulletin 1986:99(1):118-128.
25. Swaminathan AP RJ. Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement 1990:27(4):361-370.
26. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D, Group PC. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care 2007;45(4):22-31.
27. Teresi JA, Fleishman JA. Differential item functioning and health assessment. Qual Life Res 2007;16(1):33-42.
28. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gershon R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013;22(8):1889-1905.
29. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, Hareendran A, Wiklund I, Chong LY, von Maltzahn R, Cohen L, Molsen E. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health 2009;12(4):430-440.
30. Cronbach L.J. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
31. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education 2011;2:53-55
32. BD Z. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense 1999.
33. Crane PK GL, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. Med Care 2006;44(3):115-123.
34. Cook KF, Teal CR, Bjorner JB, Cella D, Chang CH, Crane PK, Gibbons LE, Hays RD, McHorney CA, Ocepek-Welikson K, Raczek AE, Teresi JA, Reeve BB. IRT health outcomes data analysis project: an overview and summary. Qual Life Res 2007;16(1):121-132.
35. Jodoin MG, Gierl, M. J. Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Applied Measurement in Education 2001;14(14):329-349.
36. Holland PW, Thayer DT. Differential item performance and the Mantel-Haenzel procedure. In H Wainer & H I Braun (Eds), Test validity Hillsdale, NJ 1988:129-145.
37. Rwhen Z. When Do Item Response Function and Mantel-Haenszel Definitions of Differential Item Functioning Coincide? Journal of Educational Statistics 1990:15(13):185-197.
38. French BF, Maller SJ. Iterative purification and effect size use with logistic regression for differential item functioning detection. Educational and Psychological Measurement 2007;67(63): 373-393.
39. Bukvic B, Zivaljevic V, Sipetic S, Diklic A, Tausanovic K, Paunovic I. Validation and cross-cultural adaptation of the questionnaire ThyPRO in thyroid patients in Serbia. Vojnosanitetski pregled 2015;72(7):583-588.
40. Watt T, Groenvold M, Hegedus L, Bonnema SJ, Rasmussen AK, Feldt-Rasmussen U, Bjorner JB. Few items in the thyroid-related quality of life instrument ThyPRO exhibited differential item functioning. Qual Life Res 2014;23(1):327-338.
41. Watt T, Groenvold M, Deng N, Gandek B, Feldt-Rasmussen U, Rasmussen AK, Hegedus L, Bonnema SJ, Bjorner JB. Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. Health Qual Life Outcomes 2014;12:126.