

RESEARCH

Validity and reliability of the German version of the shortened thyroid-specific quality of life questionnaire (ThyPRO-39de)

Navid Tabriz, Kilian Gloy, Astrid Schantzen, Dennis Fried, Dirk Weyhe and Verena Uslar 

University Hospital for Visceral Surgery, Pius-Hospital Oldenburg, Carl von Ossietzky University, Oldenburg, Germany

Correspondence should be addressed to V Uslar: verena.uslar@uol.de

Abstract

Objectives: Validation of a German version of the ThyPRO-39 questionnaire for quality of life (QoL) in patients with benign thyroid diseases.

Design: Internal consistency, retest reliability, and validity were to be assessed in a test-retest study.

Methods: The ThyPRO-39 was translated based on standard methodology. A sample of 98 patients with benign thyroid diseases was tested with the ThyPRO-39de and the generic EuroQol 5D-5L. Forty-four patients with stable symptoms after 2 weeks formed the repeated measures sample. Cronbach's alpha was calculated for the ThyPRO-39de composite score and for each disease-specific scale. Intraclass correlations between the original and the repeated measures sample were calculated for each scale as well as Pearson correlations between various ThyPRO scales and the EuroQol. *T*-tests were used to test for differences in the goiter and hyperthyroid symptom scales between relevant patient groups and other patients.

Results: Internal consistency was between satisfactory and good, except for two scales (tiredness and cosmetic complaints/appearance). The test-retest correlation was between 0.62 and 0.8 for most scales, but below 0.5 for two scales (tiredness and impaired social life). There were significant correlations between the EuroQol index score and most aspects of the ThyPRO-39de. Only the hyperthyroid symptoms scale was specific for the relevant patient group (Graves' disease).

Conclusion: The ThyPRO-39de may be recommended for use in clinical and research settings, especially with regards to the composite score. However, the underlying thyroid disease should always be kept in mind when interpreting the test results. A larger sample would be needed to implement further improvements.

Key Words

- ▶ quality of life
- ▶ benign thyroid disease
- ▶ EuroQol 5D-5L
- ▶ ThyPRO-39de
- ▶ questionnaire validation

Endocrine Connections
(2021) **10**, 1065–1072

Introduction

One of the main aims in medical therapy is the improvement or at least the maintenance of the quality of life (QoL). The quality of life – also referred to as health-related quality of life (HRQL) – is understood as a multidimensional construct that includes many relevant aspects of a patient's life such as physical, mental, and social factors. HRQL presents patients' subjective views of their current

health-related status. Therefore, the term patient-reported outcome measures (PRO) are often being used. For an accurate assessment of HRQL, the combination of generic and disease-specific questionnaires is recommended (1, 2).

Thyroid diseases are very common worldwide, and as depicted by the Papillon study, every third German adult shows pathological changes of the thyroid gland (3).

While not all of those affected need medical treatment, patients with benign thyroid diseases, such as non-toxic or toxic goiters, autoimmune disorders, or postoperative hypothyroidism, may be impaired in their QoL (2, 4). To determine patient's QoL in benign thyroid diseases, a thyroid disease-specific PRO questionnaire (ThyPRO) has been developed by Watt *et al.* in 2010 (5). In the following years, this questionnaire has been validated in different languages (1, 6, 7, 8) and a shorter version called ThyPRO-39 was developed (9). So far, the ThyPRO-39 has not been validated in the German language.

The aim of this study was to investigate the reliability of the translated German ThyPRO-39de version and to compare this disease-specific questionnaire with the generic EQ-5D-5L questionnaire, as a measure of external validity.

Methods

Patients

The research sample consists of 103 patients diagnosed with non-toxic and toxic nodular goiter, and Graves' disease. Patients with thyroid cancer were excluded. Five participants were excluded due to incomplete data. Therefore, $n=98$ patients were included in the analysis. A detailed descriptive analysis of the patient sample was performed, for the results see Table 1. Of these 98 patients, 91 participated in and completed the retest questionnaire.

Out of these 91 patients, only those with self-reported stable disease symptoms and no change in medication were included in the test-retest analysis ($n=44$).

The study was approved by the medical ethics committee of the University of Oldenburg (2017/105) and informed written consent was obtained from all participants, and it is registered with the German Clinical Trials Registry (DRKS; ID: DRKS00017195).

Questionnaires

ThyPRO-39de

The ThyPRO questionnaire is a well-validated questionnaire for measuring thyroid-related QoL and was developed by Watt and colleagues (5). The original version with its 85 items is rather long for clinical use, so a shorter version with 39 items was developed in 2015, which is now recommended in a clinical setting (9). The ThyPRO-39 summarizes 13 scales, namely: goiter symptoms, hyper- and hypothyroid symptoms, eye symptoms, tiredness, cognitive impairment, anxiety, depressivity, emotional susceptibility, impairment in social and daily life, cosmetic complaints, and one scale covering the influence of the disease on the overall HRQL. The ThyPRO scales are each summarized and transformed to a range of 0–100 according to a specific algorithm in order to maintain comparability with the 85-item version of ThyPRO using a spreadsheet provided by Torquil Watt. Three items (3b, 6g, and 7h) are inversed questions, which have to be changed accordingly if the original spreadsheet for the calculation of the scores

Table 1 Demographic data of the patient sample.

	Total ($n=98$)	Graves' disease ($n=29$)	Goiter or other nodular changes ($n=69$)
Sex			
Male (n ; %)	21 (21.4)	4 (13.8)	17 (24.6)
Female (n ; %)	77 (78.6)	25 (86.2)	52 (75.4)
Age, mean (s.d.; range)	51.2 (14.0; 18–80)	43.9 (15.3; 18–74)	54.3 (12.2; 30–80)
BMI, mean (s.d.; range) (kg/m ²)	27.9 (5.2; 17.9–44.4)	27.0 (5.5; 17.9–41)	28.3 (5.1; 18.4–44.4)
Current treatment			
L-thyroxine (n ; %)	9 (9.2)	0	9 (13.0)
Antithyroid drugs (n ; %)	34 (34.7)	28 (96.6)	6 (8.7)
Iodine (n ; %)	3 (3.1)	0	3 (4.3)
None (n ; %)	52 (53.1)	1 (3.4)	51 (73.9)
Preoperative thyroid volume ^a			
Volume (mL) mean (s.d.; range)	50.5 (7–380)	31.1 (7–87)	59.4 (10–380)
Above upper reference value (n ; %)	78 (79.6)	22 (75.9)	56 (81.2)
Hormonal values			
TSH (μIU/mL) mean (s.d.; range)	1.3 (1.89; 0–13)	1.06 (1.99; 0–7.16)	1.39 (1.86; 0–13)
T3 (pg/mL) mean (s.d.; range)	3.93 (2.50; 0.99–24.42)	4.33 (1.81; 0.99–8.32)	3.77 (2.72; 1.79–24.42)
T4 (ng/dL) mean (s.d.; range)	3.68 (5.13; 0.19–25.5)	3.68 (5.52; 0.51–25.50)	3.69 (5.02; 0.19–18.72)

TSH Ref, 0.27–4.20; T3 Ref, 2.57–4.43; T4 Ref, 0.93–1.70.

^aData available only for 94 out of 98 patients; normal range for thyroid volume: men < 25 mL; women < 18 mL.

or the script for SPSS is not used. Increasing scores indicate worsening symptoms and thus, a decreasing QoL.

In addition to the 13 individual scales, there is a composite scale that reflects the psychological and social aspects of QoL (9). The composite score includes all scales except the symptom-specific scales (goiter symptoms, hyper- and hypothyroid symptoms, eye symptoms).

The German version of the ThyPRO questionnaire, ThyPRO-39de, was obtained from Torquil Watt via email. It was translated to German according to the standard methodology for translation and linguistic validation for PROs (10, 11).

To ensure that missing data did not distort the results, only participants with complete ThyPRO-39de data were included (this reduced the sample from originally 103 to 98 participants).

EuroQol 5D-5L (EQ 5D-5L)

The EQ 5D-5L is well validated generic questionnaire to determine the general QoL. Today, it is available in over 130 languages, and in Germany, it has even been used as an evaluation instrument in thyroid diseases (12, 13, 14, 15). However, we hypothesize that it might be too generic to really capture thyroid-related QoL.

The EQ 5D-5L consists of two parts: a descriptive EQ-5D system with five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each of which has patients rate aspects of their quality of life on a five-point Likert scale. These scales are then summarized in an index value, where a high score represents a better QoL and vice versa. The second part of EQ 5D-5L consists of a visual analog scale (EQ VAS). The scale ranges from 'worst health you can imagine' (0) to 'best health you can imagine' (100) and allows a quantitative measurement of QoL.

Statistics

All statistical analyses were performed using R (16). Figure 1 was created with the same program.

Reliability

Internal consistency of ThyPRO-39de was assessed using Cronbach's α coefficient. An alpha coefficient of 0 indicates the total absence of consistency, while a coefficient of 1 corresponds to perfect internal consistency. However, since 100% internal consistency implies redundant information, values above 0.9 but below 1.0 are considered

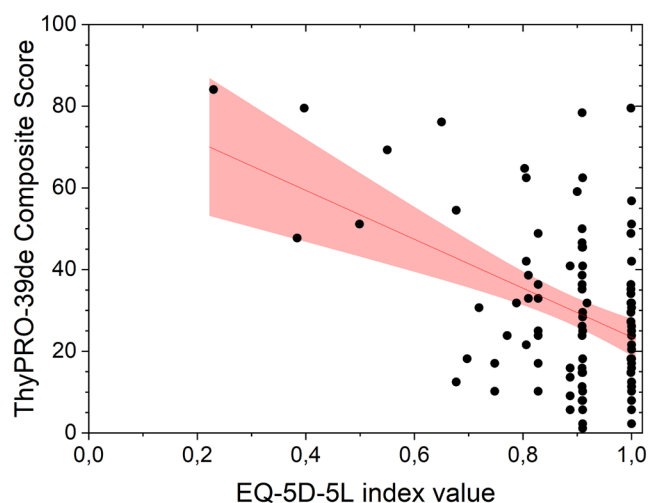


Figure 1

Scatterplot of the composite score from the ThyPRO-39-de and the index score from the EQ 5D-5L. The line represents the regression of y on x, and the gray area denotes the 95% CI.

optimal. Values below 0.7 are not acceptable (17, 18). Cronbach's alpha was calculated once for the entire questionnaire and then for every disease-specific scale separately. To test the quality of each item, a descriptive analysis of response distributions was performed, and the scale-specific alpha without the item in question was determined.

The test-retest reliability was assessed by repeating the questionnaire with the same participants 2 weeks after the initial test. A question regarding the development of the patient's symptoms was added, and only those who reported stable symptoms were included in the analysis. A Pearson correlation was then performed between each scale's transformed scores from the first and the second questionnaire.

Validity

To test the external validity of the ThyPRO-39de, a Pearson correlation was performed between the index value from the generic EQ 5D-5L and some scales of the ThyPro-39de. Specifically, the composite score and the scales goiter symptoms, hyper- and hypothyroid symptoms, and eye symptoms were correlated with the EQ's index value. This was done to reflect both the non-specific and the thyroid-specific aspects of QoL. The generic EQ 5D-5L questionnaire was used to demonstrate the added value of using a thyroid-specific patient-reported QoL instrument. We hypothesized a low correlation between both questionnaires. This would demonstrate the additional value of the more time-consuming ThyPRO-39de.

Additionally, to evaluate known-group validity, the hyperthyroid symptom scale from the ThyPRO-39de was compared between patients with and without Graves' disease, using an independent *t*-test ($P=0.05$). Based on the literature, a valid questionnaire should show a worse QoL for patients with Graves' disease (7). The same procedure was undertaken for patients with a goiter (i.e. men with sonographically confirmed thyroid size > 25 mL and women with a thyroid size > 18 mL) in comparison to patients with a thyroid size in the normal range and the goiter symptom scale.

Results

The descriptive analysis of the demographic data from the patient sample is summarized in Table 1.

Cronbach's alpha was between 0.558 and 0.910 for the individual scales and 0.901 for the composite score. Table 2 shows a complete list of each scale and the test-retest correlations, and Table 3 for the analysis of all individual items.

There is a significant negative correlation between the composite scale from the ThyPRO-39de and the index value from the EQ 5D-5L ($P=0.003$, $r=-0.299$ see Fig. 1 for visualization). Almost all of the thyroid symptom-specific scales also correlated significantly with the index value (goiter symptoms: $P=0.021$, $r=-0.232$; hyperthyroid symptoms: $P=0.032$, $r=-0.216$; hypothyroid symptoms: $P=0.158$, $r=-0.144$; eye symptoms: $P=0.009$, $r=-0.261$).

The independent *t*-test between patients with and without Graves' disease revealed a significant difference between the groups ($P=0.033$, $t(96)=2.162$). Patients with Graves' disease had a significantly higher hyperthyroid score, indicating worse QoL (mean = 38.1, s.d. = 17.4), than patients without Graves' disease (mean = 21.2, s.d. = 16.5). For $n=94$ patients, we acquired ultrasound values prior to surgery. Patients with a thyroid size above the reference value (>25 mL for men; >18 mL for women) had comparable scores on the goiter symptoms scale ($n=79$; thyroid volume (mean and range): 57.5 mL (19–380); goiter symptoms scale: mean = 24.8, s.d. = 20.3) to patients with normally sized thyroids ($n=15$; thyroid volume (mean and range): 14.4 mL (7–24.3); goiter symptoms scale: mean = 26.0, s.d. = 17.6; unpaired *t*-test: $t(91)=0.210$; $P=0.834$).

Discussion

The overall internal consistency of all ThyPRO-39de items was satisfactory with $\alpha=0.94$. This seems to indicate that all items measure the same concept to a high degree, which would conceivably be the general quality of life. Almost all disease-specific scales have sufficient values for Cronbach's alpha, meaning above 0.7, except for the scales tiredness and cosmetic complaints/appearance.

The alpha of tiredness could have been improved above the critical value of 0.7 by excluding the item 3b ('Hatten Sie innerhalb der letzten vier Wochen ein energiegeladenes Gefühl?'/ 'Did you feel energetic within the last four weeks?').

Table 2 Overview of the mean, s.d., median, range (min–max), Cronbach's alpha values, and the test-retest correlations for all ThyPRO-39de scales and the index value from the EQ 5D-5L.

Scale	Mean	s.d.	Median	Range	Alpha	Test-retest correlation ^a
ThyPRO-39de subscales						
Goiter symptoms	25.3	19.8	23.0	2–84	0.885	0.599
Hyperthyroid symptoms	26.2	18.4	23.0	2–71	0.644	0.617
Hypothyroid symptoms	20.9	20.6	18.8	0–75	0.724	0.660
Eye symptoms	19.3	17.8	14.0	1–78	0.755	0.448
Tiredness	53.3	22.3	50.0	0–100	0.558	0.257
Cognitive complaints	19.8	22.5	14.0	1–95	0.910	0.750
Anxiety	31.1	25.5	26.0	1–96	0.893	0.521
Depressivity	28.5	21.4	22.0	0–89	0.740	0.638
Emotional susceptibility	34.4	26.0	28.0	1–95	0.814	0.753
Impaired social life	17.2	25.5	0.0	0–92	0.878	0.402
Impaired daily life	21.5	28.5	15.0	0–98	0.900	0.768
Cosmetic complaints/appearance	33.4	24.2	28.0	1–87	0.649	0.625
Overall quality of life	33.4	34.0	25.0	0–100	–	0.773
Composite score	30.3	19.6	26.7	1.1–84.1	0.901	0.756
EQ index value	0.88	0.15	0.91	0.23–1	–	–0.010

^aIntraclass correlation coefficient.

Table 3 Descriptive analysis of each item per scale and for the items used in the composite score. Shown are the relative distributions of the answers, and the value which Cronbach's alpha would assume if the specific item was removed from the scale.

Short description	Response (%)					Cronbach's alpha, if item removed from subscale	Cronbach's alpha, if item removed from composite score (original alpha: 0.901)
	Not at all	A little	Some	Quite a bit	Very much		
Goiter symptoms (original alpha: 0.885)							
1a Sense of fullness in the neck	34	17	23	21	4	0.787	
1c Pressure in throat	40	18	24	14	3	0.815	
1h Discomfort swallowing	42	17	24	15	1	0.898	
Hyperthyroid symptoms (original alpha: 0.644)							
1l Trembling hands	61	14	12	12	0	0.470	
1m Increased sweating	37	13	19	19	11	0.555	
1n Palpitations	33	14	31	15	7	0.463	
1t Upset stomach	69	12	9	5	4	0.750	
Hypothyroid symptoms (original alpha: 0.724)							
1q Sensitive to cold	62	13	10	10	4	0.804	
1cc Swollen hands/feet	60	21	13	5	0	0.611	
1dd Dry skin	33	26	20	14	7	0.626	
1ee Itchy skin	56	15	17	7	4	0.566	
Eye symptoms (original alpha: 0.755)							
1w Grittiness in eyes	46	22	18	10	3	0.750	
1x Reduced sight	50	22	19	6	2	0.580	
1bb Sensitive to light	56	18	16	5	4	0.649	
Tiredness (original alpha: 0.558)							
2a Been tired	16	17	35	19	12	0.155	0.894
2c Difficulty getting motivated	32	21	28	10	9	0.124	0.892
3b Energetic	48	22	15	12	2	0.823	0.907
Cognitive complaints (original alpha: 0.910)							
4a Problems remembering	53	20	16	7	3	0.867	0.911
4b Slow or unclear thinking	56	28	10	3	3	0.869	0.894
4f Difficulty concentrating	39	30	21	5	5	0.875	0.895
Anxiety (original alpha: 0.893)							
5b Afraid or anxious	38	27	21	10	4	0.844	0.893
5c Felt tension	34	19	22	19	5	0.879	0.893
5e Uneasy	36	31	19	10	4	0.823	0.894
Depressivity (original alpha: 0.740)							
6a Sad	48	26	15	9	2	0.481	0.892
6e Unhappy	55	23	11	10	0	0.605	0.895
6g Self-confident	17	21	19	39	3	0.848	0.894
Emotional susceptibility (original alpha: 0.814)							
7c Easily stressed	21	21	16	26	15	0.720	0.916
7d Mood swings	24	24	24	15	11	0.646	0.903
7h Felt in control	6	9	15	24	45	0.845	0.893
Impaired social life (original alpha: 0.878)							
8a Difficulty being with others	62	16	9	9	3	0.842	0.892
8b A burden to others	72	14	7	4	2	0.841	0.922
8c Conflicts with others	71	11	10	6	1	0.801	0.895
Impaired daily life (original alpha: 0.900)							
9a Difficulty managing daily life	60	12	10	15	2	0.856	0.895
9c Difficulty participating in life	71	12	8	8	0	0.874	0.894
9e Everything takes longer	54	26	3	11	6	0.836	0.895
Cosmetic complaints/appearance (original alpha: 0.649)							
11a Disease affects appearance	23	21	22	21	11	0.417	0.892
11d Bothered by people looking	81	8	3	6	2	0.555	0.893
11e Influence on clothes worn	67	15	6	9	2	0.640	0.893
12 Overall impact on QoL	39	23	10	20	7	-	0.891

A similar result was reported for the Chinese adaptation of the ThyPRO-39, where the same item correlated with the tiredness scale only weakly (7). At the present time, it is difficult to be sure why this item is inconsistent with the rest of the scale. The inverse coding of the item is the only quality that obviously separates it from the others. Inversely coded items are often included to reduce floor effects, although this often comes at the cost of reliability (7, 19). Therefore, it comes as no surprise to see this effect here. However, it should be noted that the other scales with inversely coded items (depressivity and emotional susceptibility, items 6g and 7h, respectively) have acceptable alpha-values. Therefore, it seems to not just only be a problem of patients not noticing the need to switch their answering pattern. Although, it should also be noted that for the latter scales, removal of the inverse item would improve the value for Cronbach's alpha. Based on the present analysis, item 3b and perhaps all inverse items should be treated with caution.

The scale cosmetic complaints/appearance can not be improved by excluding a single item. This implies that all three items measure slightly different, yet related aspects of QoL. All relate to the appearance of the patient, but they address self-image, perceived external perception, and behavioral changes, respectively. These subtle differences might explain the scale's comparatively low alpha. Interestingly, the Spanish version of the ThyPRO-39 also suffered from the low alpha in the cosmetic complaints scale (8), but adaptations into other languages did not (7, 20). These circumstances make it difficult to determine whether there is a problem with the scale itself or its German adaptation.

In international comparison, the ThyPRO-39de was performing between average and well regarding the internal consistency. It was worse than the Romanian version, which had all scales above an alpha of 0.7 (19), performed just as well as the Spanish version, which also had two scales below 0.7 (8) and was better than the Chinese version, which had 6 scales below an alpha of 0.7 (7).

Most scales had a high test-retest correlation, indicating good reliability, though not as good as the original version of the ThyPRO-39 (9). Some imprecision in this measurement is impossible to avoid. All aspects measured by the ThyPRO are capable of rapid fluctuation, and even people who judge their overall symptoms to be similar may have experienced changes that reflect on specific scales. This is why the low retest-correlation in the scales tiredness and impaired social life are difficult to judge. The low correlation in tiredness could, of course, be related to the low internal consistency.

The significant correlations between the composite scale and (three out of four) thyroid symptom-specific scales on the one side and the EQ's index value on the other side show a strong connection between the test values of both questionnaires. This may indicate validity for the ThyPRO-39de since the measurement of QoL seems to have succeeded. However, the values for Pearson's *R* are only moderate for the composite scale and small for the thyroid symptom-specific scales. There are, therefore, aspects captured by the ThyPRO-39de which are not reflected in the EQ 5D-5L, and this lack of overlap is strongest in the thyroid-specific scales. This indicates that for thyroid-related research the ThyPRO-39de might be a more valid choice than the generic, albeit shorter EQ 5D-5L questionnaire. In addition, for future comparisons between studies using the ThyPRO-39de, it should be kept in mind to look at the distribution of specific diseases, especially Graves' disease vs others, which might explain differences in QoL between studies.

The hyperthyroid symptoms scale was capable of successfully differentiating between patients with Graves' disease and patients with other diagnoses, even though our patient collective was, in general, very well adjusted with regards to their medication. This seems to indicate that near-normal values of TSH, T3, and T4 do not automatically equal low symptomatic burden.

In contrast, we found no difference between patients with normal and enlarged thyroids concerning the goiter symptoms scale. This could be due to the fact that we divided the groups by thyroid volume (>25 mL for men; >18 mL for women) and not by symptoms. These results indicate that thyroid volume might not correlate with goiter symptoms *per se*. On the one hand, goiter symptoms might occur in patients with normal-sized thyroids due to the localization of the nodules, and on the other hand, patients with large thyroids may be asymptomatic for a long time.

Regrettably, the sample did not allow for more such tests for validity and specificity, which is one of the limitations of this study. Generally, a larger sample would have been preferable for these tests. While 98 patients are enough for considerations of external validity and internal consistency, the necessity of reducing the sample to the specific patient groups impedes the specificity tests. These results should, therefore, be considered with caution. In addition, it was only possible to divide patients by preoperative thyroid size for the test of the goiter symptoms scale. A comparison of patients with goiter referred for treatment with patients without goiter would maybe have yielded other results. Also, the ThyPRO-39de is not validated for use with patients

with malignant diseases. Therefore, our conclusions are addressed only to patients with benign thyroid diseases. Taken together, the results of all scales but the composite score should be regarded with caution since more detailed known-groups comparisons are missing.

The strengths include our relatively homogenous patient collective with most patients being euthyroid. Also, our collective very well reflects the typical collective of patients with benign thyroid diseases in specialized German hospitals.

Conclusion

The German adaptation of the ThyPRO-39 is an internally consistent and externally valid questionnaire. The transition into a different cultural and linguistic background did not come without issues, and two or three scales within the test might need slight corrections in a future version, once larger samples are available. For now, it is a useful questionnaire, especially with regards to the composite score, that well reflects a patient's QoL and captures the aspects of thyroid diseases that impact QoL better than generic questionnaires could. Based on this study, the ThyPRO-39de may be recommended for clinical use, with the composite score as the main outcome and the other scale scores used as exploratory or informatory outcomes. The interpretation of the test results, however, should always be based on a separate diagnosis of the underlying thyroid disease.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

References

- 1 Watt T, Barbesino G, Bjorner JB, Bonnema SJ, Bukvic B, Drummond R, Groenvold M, Hegedüs L, Kantzer V, Lasch KE, *et al.* Cross-cultural validity of the thyroid-specific quality-of-life patient-reported outcome measure, ThyPRO. *Quality of Life Research* 2015 **24** 769–780. (<https://doi.org/10.1007/s11136-014-0798-1>)
- 2 Watt T, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedüs L, Bjorner JB & Feldt-Rasmussen U. Quality of life in patients with benign thyroid disorders. A review. *European Journal of Endocrinology* 2006 **154** 501–510. (<https://doi.org/10.1530/eje.1.02124>)
- 3 Verburg FA, Grelle I, Tatschner K, Reiners C & Luster M. Prevalence of thyroid disorders in elderly people in Germany. *Nuclear Medicine* 2017 **56** 9–13. (<https://doi.org/10.3413/Nukmed-0852-16-10>)
- 4 Elberling TV, Rasmussen AK, Feldt-Rasmussen U, Hørding M, Perrild H & Waldemar G. Impaired health-related quality of life in Graves' disease. A prospective study. *European Journal of Endocrinology* 2004 **151** 549–555. (<https://doi.org/10.1530/eje.0.1510549>)
- 5 Watt T, Hegedüs 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. *European Journal of Endocrinology* 2010 **162** 161–167. (<https://doi.org/10.1530/EJE-09-0521>)
- 6 Mintzioti G, Watt T, Veneti S, Panagiotou A, Pournaras DJ, Feldt-Rasmussen U, Rasmussen ÅK, Hegedüs L, Bonnema SJ, Bjorner JB, *et al.* ThyPROgr: the Greek edition of the ThyPRO questionnaires for patients with benign thyroid diseases. *Hormones* 2018 **17** 107–112. (<https://doi.org/10.1007/s42000-018-0015-7>)
- 7 Wong CKH, Choi EPH, Woo YC & Lang BHH. Measurement properties of ThyPRO short-form (ThyPRO-39) for use in Chinese patients with benign thyroid diseases. *Quality of Life Research* 2018 **27** 2177–2187. (<https://doi.org/10.1007/s11136-018-1857-9>)
- 8 Boronat M, González-Lleó A, Rodríguez-Pérez C, Feldt-Rasmussen U, López-Plasencia Y, Rasmussen ÅK, Alberiche-Ruano MP, Hegedüs L, Alvarado-Martel D, Bonnema SJ, *et al.* Adaptation and cross-cultural validation of the Spanish version of the thyroid-related quality-of-life patient-reported outcome questionnaire. *Endocrinología, Diabetes y Nutrición* 2018 **65** 500–507. (<https://doi.org/10.1016/j.endinu.2018.06.010>)
- 9 Watt T, Bjorner JB, Groenvold M, Cramon P, Winther KH, Hegedüs L, Bonnema SJ, Rasmussen ÅK, Ware Jr JE & Feldt-Rasmussen U. Development of a short version of the thyroid-related patient-reported outcome ThyPRO. *Thyroid* 2015 **25** 1069–1079. (<https://doi.org/10.1089/thy.2015.0209>)
- 10 Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moynihan CM, McLeod LD, *et al.* ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research* 2013 **22** 1889–1905. (<https://doi.org/10.1007/s11136-012-0344-y>)
- 11 Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, Hareendran A, Wiklund I, Chong LY, von Maltzahn R, *et al.* 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 in Health* 2009 **12** 430–440. (<https://doi.org/10.1111/j.1524-4733.2008.00471.x>)
- 12 Schmitz-Winnenthal FH, Schimmack S, Lawrence B, Maier U, Heidmann M, Buchler MW & von Frankenberg M. Quality of life is not influenced by the extent of surgery in patients with benign goiter. *Langenbeck's Archives of Surgery* 2011 **396** 1157–1163. (<https://doi.org/10.1007/s00423-011-0822-7>)
- 13 Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Loucks L, Khudairy AA & Noseworthy TW. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Quality of Life Research* 2015 **24** 1775–1784. (<https://doi.org/10.1007/s11136-014-0910-6>)
- 14 EuroQol Research Foundation. EQ-5D-5L | about 2020 [version 18APR2017]. (available at: <https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/>)
- 15 Tabriz N, Usklar VN, Tabriz I & Weyhe D. Quality of life is not affected by thyroid surgery in nontoxic benign goitre in long-term surveillance – a prospective observational study. *Endocrinology, Diabetes and Metabolism* 2020 **3** e00115. (<https://doi.org/10.1002/edm2.115>)
- 16 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010. (available at <https://www.R-project.org/>)

- 17 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951 **16** 297–334. (<https://doi.org/10.1007/BF02310555>)
- 18 Tavakol M & Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2011 **2** 53–55. (<https://doi.org/10.5116/ijme.4dfb.8dfd>)
- 19 Watt T, Groenvold M, Deng N, Gandek B, Feldt-Rasmussen U, Rasmussen ÅK, Hegedüs L, Bonnema SJ & Bjorner JB. Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. *Health and Quality of Life Outcomes* 2014 **12** 126. (<https://doi.org/10.1186/s12955-014-0126-z>)
- 20 Zahan AE, Watt T, Pascanu I, Rasmussen AK, Hegedüs L, Bonnema SJ, Feldt-Rasmussen U, Bjorner JB, Nadasan V, Boila A, *et al.* The Romanian version of the thyroid-related patient-reported outcomes ThyPRO and ThyPRO-39. Translation and assessment of reliability and cross-cultural validity. *Acta Endocrinologica* 2018 **14** 192–200. (<https://doi.org/10.4183/aeb.2018.192>)

Received in final form 19 July 2021

Accepted 6 August 2021

Accepted Manuscript published online 6 August 2021