

REVIEW ARTICLE

Thyroid disease-specific quality of life questionnaires - A systematic review

Verena Uslar  | Caroline Becker | Dirk Weyhe | Navid Tabriz

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

Correspondence

Verena Uslar, Pius-Hospital Oldenburg,
Georgstr. 12, D-26121 Oldenburg,
Germany.

Email: verena.uslar@uni-oldenburg.de

Abstract

Introduction: Thyroid diseases are very common and rarely life-threatening. One of the main therapeutic goals is an improvement in quality of life, making it important to measure in clinical and research settings. The aim of this systematic review is to provide an overview of the currently available thyroid-specific quality of life questionnaires with regard to their validation quality in order to make recommendations for clinical use with a special focus on German questionnaires.

Methods: A systematic literature search was performed in Pubmed, Google Scholar and the Cochrane Library. A total of 904 studies were identified. After excluding duplicates, non-English- or German-language texts, full texts that were not freely available and studies with irrelevant content, 64 studies reporting on 16 different questionnaires were included in the analysis.

Results: Four questionnaires concerned benign thyroid diseases (ThyPRO, ThyPRO-39, Thy-R-HRQoL and Thy-D-QoL), six malignant thyroid diseases (THYCA-QoL, ThyCA-HRLQOL, EORTC-Thy34, MADSI-Thy, QOL-Thyroid and ThyCAT), and six endocrine orbitopathy (GO-QOL, GO-QLS, TED-QOL, STED-QOL, TAO-QoL and Ox-TED). Only five questionnaires were at least developed, if not validated, in German, and five were developed in more than two languages.

Conclusions: ThyPRO and the ThyPRO-39 are the best-evaluated questionnaires for benign thyroid diseases. Alternatively, in hypothyroid patients, the adequately validated Thy-D-QoL can be used. For malignant thyroid diseases, the choice should be made individually, as all six questionnaires (THYCA-QoL, ThyCA-HRQOL, EORTC-Thy34, MDASI-Thy, QOL-Thyroid and ThyCAT) have different strengths and weaknesses. The GO-QOL is the best-validated questionnaire in endocrine orbitopathy. However, the TED-QOL is also suitable as a short-screening questionnaire for these patients.

KEYWORDS

benign thyroid disease, COSMIN, endocrine orbitopathy, malignant thyroid disease

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Endocrinology, Diabetes & Metabolism* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Thyroid diseases are very common in the general population and their prevalence increase with age.¹ Common thyroid diseases include hypo- and hyperthyroidism, nodular goitre, thyroid cancer and autoimmune disorders such as Hashimoto thyroiditis or graves' disease with and without endocrine orbitopathy.² Due to the simple and widely available diagnostic tools, thyroid diseases are often detected early, although not every thyroid disease requires therapy.³

However, there is a consensus that quality of life (QoL) is negatively influenced by thyroid dysfunction, both in hyperthyroidism and hypothyroidism, and one of the main aims in the therapy of thyroid dysfunction should be at least preserving or, ideally, improving QoL.⁴⁻⁶ For this reason, the measurement of health-related QoL has become an important issue of interest, and many instruments have been developed to measure this outcome parameter. QoL is often defined as a multidimensional subjective construct containing the dimensions of general health, physical, psychological and social functioning. It can be best measured by patients themselves in form of questionnaires using patient-reported outcomes (PROs).⁷ Typical domains in QoL questionnaires include anxiety, impaired social life or overall quality of life. Those more general domains are usually complemented in thyroid disease-specific questionnaires. Based on the underlying thyroid diseases these thyroid-specific QoL questionnaires often contain domains like goitre symptoms, eye symptoms or tiredness.

Establishing relevant domains in a standardized manner should be part of the development process of each questionnaire. However, a crucial weak point of some questionnaires is this missing development step, and the lack of comprehensive assessment of measurement properties, such as validity and reliability, which prevents generalizability or comparability.⁸ Based on these shortcomings the COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments) initiative has developed a guideline for systematic reviews of PROs and a checklist for the evaluation of studies reporting on the development of PROs.⁹⁻¹² In 2016 a systematic review was conducted with regard to the quality of thyroid-specific PROs.¹³ This review judges the quality of the 14 thyroid-specific QoL questionnaires available at the time and emphasizes the need for high quality and standard reporting of the development of thyroid-specific QoL questionnaires. However, since then, new questionnaires have been developed.

Therefore, the aim of this systematic review is on the one hand the presentation of current thyroid disease-specific QoL questionnaires with regard to validity and reliability. On the other hand, this review focusses on the clinical usability of the respective questionnaires in order to make recommendations for clinical practice, especially with regard to validated questionnaires in the German language and in order to identify gaps regarding specific questions.

2 | METHODS

At all times during the preparation of this Systematic Review, the PRISMA guidelines were followed.

2.1 | Literature search

A systematic literature search has been conducted by two of the authors independently (C.B. and V.U.) in the time between 18.02.2021 and 16.03.2021. The PICO ((P) participants, (I) interventions and (C) comparators but not the outcomes) scheme was used and the databases Pubmed, the Cochrane Library and Google Scholar were searched to identify studies that assess the QoL of patients with thyroid disease through thyroid-specific questionnaires. Initially, The Medical Subject Heading 'quality of life' was combined with 'thyroid' AND 'questionnaire' for a first search, and additionally combined with 'questionnaire' AND ('thyroid dysfunction' OR 'thyroid disease' OR 'hyperthyroidism' OR 'hypothyroidism' OR 'graves' disease' OR 'hashimoto's disease' OR 'thyroid cancer' OR 'thyroid neoplasia' OR 'thyroid carcinoma' OR 'goitre' OR 'thyroid autonomy') for the second search in PubMed. Subsequently, the search was performed via the Cochrane Library and Google Scholar using the terms: 'thyroid', 'quality of life' and 'questionnaire'. For Google Scholar, the results of the first 5 pages (50 results) were selected and supplemented by another 101 results similar to the first source of Watt et al.¹⁴ The type of snowballing search possible with Google Scholar may not yield the exactly reproducible results as with Pubmed or the Cochrane Library. However, as a type of add-on like it was used here, it provides nonetheless valuable information. No time or other restrictions were applied to any search. Only English or German publications for which the full text was available were included in the search.

2.2 | Study selection, inclusion and exclusion criteria

The process of study selection and inclusion was performed by two reviewers independently (C.B. and V.U.). In the first step, the total number of studies was reduced by excluding duplicates. Then, all titles and subsequently all abstracts were screened and checked for relevance. Relevant studies were read in their entirety. All studies for which no abstract or no complete text was available were excluded. Only studies in English or German language were included. Furthermore, only studies with human study populations were considered. In addition, those that used only general QoL questionnaires, such as the SF-36 or the EORTC-QLQ-C30 questionnaire, were also excluded. Studies that used questionnaires that did not measure QoL were also not considered further. Finally, those studies that used a thyroid-specific questionnaire but contained little information were also not analysed if there were studies about the same questionnaire presenting more information. All guidelines, reviews and meta-analyses were excluded, but references were screened for further relevant studies. Consequently, only studies that reported on the development of a thyroid-specific QoL questionnaire were included. After this selection process the authors C.B. and V.U. conferred about their results and any differences were resolved by consensus. The remaining studies were read and analysed by both authors independently, and relevant information was extracted, implementing an excel spreadsheet. Again, any discrepancies were resolved by consensus.

Finally, the retrieved studies were categorized into the following three groups: (1) benign thyroid disease questionnaires, (2) malignant thyroid disease questionnaires and (3) endocrine orbitopathy questionnaires.

2.3 | Extracted information

The following information was extracted from the relevant studies if available:

- Author(s)
- Publication year
- Type of thyroid disease
- Availability in different languages
- Number of items and scales (including general content), and how they are scored
- Internal consistency and test-retest reliability, including respective values
- If, and how external validity, construct validity and criterion validity was tested
- Availability of the questionnaire
- Questionnaire processing time

3 | RESULTS

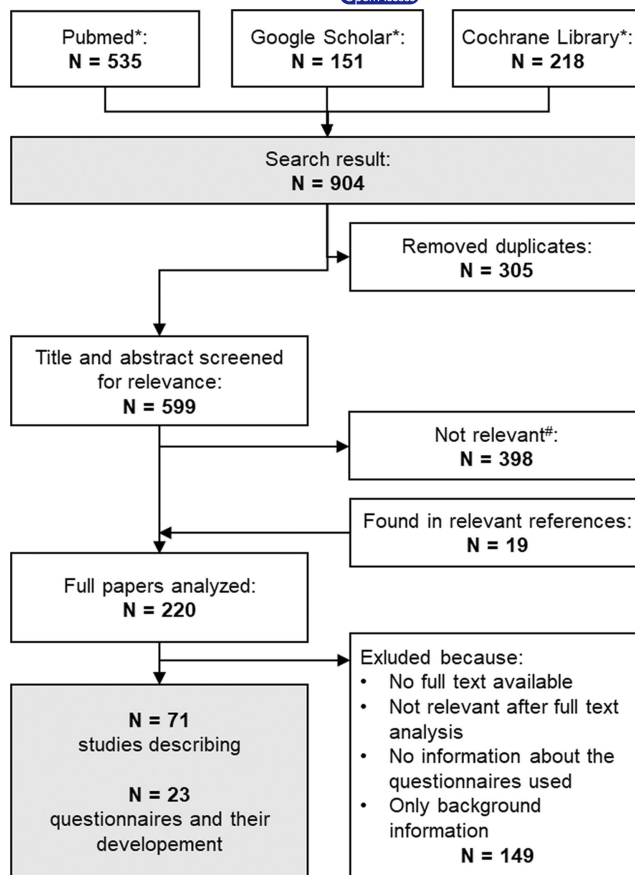
3.1 | Literature search

In [Figure 1](#) the process of literature search, literature identification and study selection is depicted. The PubMed search yielded a total of 535 studies. The search via Google Scholar and Cochrane Library revealed a total of 369 papers. After the removal of duplicates, a total of 599 studies were screened for relevance based on their title and abstract.

This produced 220 studies, which were then read by both aforementioned authors and were independently analysed with regard to the inclusion and exclusion criteria, i.e. studies that assessed the QoL of patients with thyroid disease through thyroid-specific questionnaires.

After further consensus, the selection was further narrowed to 71 studies, discussing the development of 16 different, relevant questionnaires. The questionnaires that did not fit the inclusion criteria were excluded because they either were more concerned with symptoms, therapy satisfaction or anxiety (ThySRQ, ThyTSQ, HCQ and WSCI-T) or they were not specific enough (the EORTC questionnaires). Also, the NEI-VFQ-25 was excluded, because even from the abstract it was obvious, that the questionnaire was, albeit thoroughly validated, not useful in a research or clinical setting.

Four out of the remaining 16 questionnaires were assigned to benign thyroid diseases including hypothyroidism and Graves' disease. Six questionnaires were assigned to malignant thyroid diseases, and six questionnaires were assigned to the third category 'endocrine



*see search strategies described in the text for specific information
#see inclusion and exclusion criteria for specific information

FIGURE 1 PRISMA flow chart

orbitopathy'. In the following sections, the identified questionnaires are presented in more detail.

3.2 | Benign thyroid disease QoL questionnaires


The following section describes the four questionnaires for benign thyroid diseases. An overview is presented in [Table 1](#).

3.2.1 | ThyPRO

The ThyPRO (Thyroid Patient Reported Outcome) questionnaire was developed and validated by Watt et al.¹⁵ It consists of 85 items, which are divided into a total of 13 scales and one single-item scale. Each item has five response options on a Likert scale ranging from '0 = no symptoms or problems' to '4 = severe symptoms or problems' based on the period of the last 4 weeks. The higher the score, the more strongly QoL is affected.

The 13 scales include four scales on specific physical symptoms, two on mental symptoms, three on well-being and functioning, and four on participation and social functioning. The single item asks about the impact of thyroid disease on overall QoL. Thus, different

TABLE 1 Overview of the questionnaires developed for benign thyroid diseases

	ThyPRO	ThyPRO-39	Thy-R-HRQoL	Thy-D-QOL
Authors	Watt et al.	Watt et al.	Kaniuka-Jakubowska et al.	McMillan
Type of disease	Benign thyroid disease	Benign thyroid disease	Euthyroid goitre	Hypothyroidism
No of items	85	39	49	20
Available languages	Danish, English, Dutch, Italian, Serbian, Indian, Swedish, Spanish, Filipino, Greek, Thai, Polish, Romanian	Danish, English, Romanian, Greek, German	English, Polish	English, German, Greek, Portuguese, Turkish
Internal consistency	>0.83	-	>0.9	0.949
Test-Retest-Reliability	>0.77	>0.75	-	-
External validity	SF-36, HADS	ThyPRO	-	SF-36, Beck
construct validity	Factor analysis, multitrait	Factor analysis	-	Factor analysis
Criterion validity	Known-groups, DIF	Known-groups	-	-
Availability	∅	∅		∅
Answering time	14 min	4 min	-	-

Note: -, No information available or not evaluated.

∅: Available from author;  available with the following link: <https://www.frontiersin.org/articles/10.3389/fendo.2018.00114/full#supplementary-material>.

aspects of QoL are represented by the individual scales. The four symptom scales capture specific symptoms and problems in hyperthyroidism, hypothyroidism, goitre and thyroid disease with eye involvement. Thus, the ThyPRO is suitable for the assessment of QoL in all benign thyroid diseases such as Graves' disease or Hashimoto's thyroiditis.

The ThyPRO was developed based on a systemic review,⁵ interviews with patients and professionals¹⁵ and cognitive interviews after operationalizing the problems into items.¹⁶ The preliminary questionnaire was adjusted after analysing construct validity, by 'multitrait scaling' analysis (convergence + discriminant validity) and reliability by Cronbach's alpha.¹⁷ Subsequently, clinical validity was evaluated by 'known-groups' and reliability by test-retest analysis of the final questionnaire.¹⁵ Dimensionality of the scales was confirmed by confirmatory factor analysis and the extent of differential item functioning was tested by ordinal regression.^{18,19} Sensitivity to changes after clinically relevant therapies was tested and compared with the generic quality of life questionnaire SF-36.²⁰ The ThyPRO was originally developed in Danish and English but has been translated and cross-culturally validated in many other languages such as Dutch, Indian, Italian, Serbian and Swedish.²¹ It has also been translated and validated in Greek, Filipino, Polish, Romanian, Spanish, Thai and German by other authors.²²⁻²⁶

3.2.2 | ThyPRO-39

The ThyPRO-39 questionnaire is the short version of the ThyPRO and has 39 items.^{14,27} These are divided into 12 scales and one single item. The median response time is 4 min, which is shorter than the long version at 14 min. The development of the short version

was divided into the steps of item selection, scale scoring and validation. Items were excluded for which previously missing scores were frequent, such as impaired sex life, which did not conform to item response theory, and for which differential item functioning or cross-cultural weaknesses occurred in the previous validation studies of the long version. An additional score comprising seven scales covering mental and social well-being and functioning was created, and the degree of agreement between the short and long scales was assessed using agreement plots and intraclass correlations.

Effect sizes and validity indices for response to the change in therapy were calculated for validation. Clinical validity was tested by the ability to discriminate between clinical patient groups. Test-retest reliability was also collected and checked with the long version. Each scale can be evaluated individually, but there is also an additional score that combines seven scales. The short version of the ThyPRO can assess QoL in many different benign thyroid diseases, and in many different languages, for instance, German, Romanian, Greek or Spanish.^{22,23,26,28}

3.2.3 | Thy-R-HRQoL

The Thy-R-HRQoL (Thyroid-Related Health-Related Quality of Life) questionnaire was created by Kaniuka-Jakubowska et al. and is suitable for assessing QoL in euthyroid goitre.²⁹ It consists of a total of 49 questions divided into 7 domains. The answer options range from '1 = definitely no' to '6 = definitely yes'. Consequently, the QoL decreases as the score increases. In terms of content, three of the domains reflect the influence of the disease on the shape of the throat, dyspnea and performing the social role, and three others reflect the severity of subjective symptoms such as difficulty breathing, foreign

body sensation in the throat, and difficulty with swallowing and hoarseness. The last domain asks about the influence of the disease on the course of comorbidities. Reliability was analysed by measuring internal consistency using Cronbach's alpha (>0.9). No further validation analyses were performed.

3.2.4 | Thy-D-QoL

The Underactive Thyroid-Dependent Quality of Life Questionnaire (Thy-D-QoL) was developed by McMillan et al.³⁰ It is based on the Audit of Diabetes-Dependent Quality of Life Questionnaire (ADDQoL) and was further developed with patient interviews into a hypothyroidism-specific questionnaire. At the beginning of the questionnaire, a question is asked about the current general QoL, which can be rated on a seven-point scale from '+3 = excellent' to '-3 = extremely poor'. This is followed by a question about how the QoL would be without hypothyroidism, which is answered on a scale from '-3 = much better' to '1 = worse'. The other 18 domains refer to the influence of hypothyroidism on individual aspects of QoL, such as leisure time, work, family, social environment, energy, appearance and depression. Again, these domains ask what the problems would be like if they did not have hypothyroidism. The response options for these 18 domains are again from '-3 = much better' to '1 = worse'. Each of the domains is then rated on how important the aspect was to the patients, from '3 = very important' to '0 = not at all important'. The two questions in a domain can then be multiplied to give a weighted domain score, which can range from '-9 = maximum negative' to '+3 = maximum positive'. However, the 18 domains can also be combined into an overall average weighted score. The lower the score, the worse the QoL. For internal validation, Cronbach's alpha was determined with a value of 0.949. A forced one-factor analysis was also performed, which confirmed that all domains can be combined into an average score. The focus of this questionnaire is clearly on QoL rather than symptoms, as the authors also developed a questionnaire for symptoms of hypothyroidism only, the Thyroid Symptom Rating Questionnaire (ThySRQ), and one for the effects of L-thyroxine therapy, the Thyroid Therapy Satisfaction Questionnaire (ThyTSQ).³¹ Quinque et al. performed a validation of the German version of McMillan's questionnaires. The German version of the Thy-D-QoL also has an underlying factor with factor loadings >0.4 , a Cronbach's alpha of 0.92, item-total correlation of >0.2 and low to moderate correlations with generic quality of life questionnaires such as the SF-36.³²

3.2.5 | Further questionnaires for benign thyroid diseases

In addition to the questionnaires mentioned above, further questionnaires for the assessment of QoL in benign thyroid diseases were found in the literature search. Some of these were not validated and only developed for one study specifically or only capture a partial

aspect of QoL, such as physical and mental symptoms. Four studies measured QoL in subclinical and manifest hypothyroidism with non-validated questionnaires designed for the author's own studies.³³⁻³⁶ Scerrino et al. compiled a questionnaire to measure QoL after total thyroidectomy in patients suffering from Graves' disease from various validated questionnaires such as the MOS-24, the SF-36 and the GO-QoL. This contained 36 items covering clinical, neurovegetative, psychosocial and daily aspects, and scar quality. Except for a pilot test, which was intended to check the understanding of the items but also the test-retest reliability, no analyses of validity were performed.⁴ McMillan et al. created, in addition to the Thy-D-QoL, the ThyTSQ, the Thyroid Treatment Satisfaction Questionnaire, and the ThySRQ, the Thyroid Symptom Rating Questionnaire. In terms of content, the ThyTSQ can be divided into two sections. One measures current treatment satisfaction with seven items, and one measures past treatment satisfaction with four items. The ThySRQ is a list of 15 symptoms for which occurrence and severity can be rated. Both questionnaires have a high reliability of 0.9 for the ThyTSQ and 0.8 for the ThySRQ.^{31,37} The Hyperthyroid Complaint Questionnaire (HCQ) was developed by Fahrenfort et al. in Dutch to quantify somatic and mental problems that occur in the context of hyperthyroidism.³⁸ It contains 31 items determined by patient interviews. Individual symptoms can be scored as to whether they are 'currently present', 'previously present' or 'not present at all'. The 'base scale' can be calculated from the number of all currently present symptoms and an 'additional scale' from the number of previously present symptoms. Item-total correlations were >0.4 for 29 items, and reliability, as measured by Cronbach's alpha, was 0.93. The Cooper questionnaire measures the extent of six symptoms associated with hypothyroidism and whether they have worsened or improved over time.³⁹ No validation is performed yet.






The literature search revealed two further questionnaires that have been used to assess QoL in thyroid diseases. They address QoL after thyroid surgery and have been used in both benign and malignant thyroid disease.^{6,40} In addition, Wilde et al. developed the German-language HPQ-40/28 questionnaire to assess QoL in hypothyroidism.⁴¹ The original version consists of 40 items, and the shortened one of 28 items. Factor analysis was performed to confirm the five scales depression and anxiety, loss of vitality, pain and cramps, neurovegetative symptoms and gastrointestinal symptoms. The reliability of these scales measured by Cronbach's alpha ranged from 0.74 to 0.88.

Lastly, Jaeschke et al. investigated the frequency of importance of patients' symptoms attributed to hypothyroidism.⁴² They established a wide variety of patient-specific complaints associated with hypothyroidism, which might be helpful to investigators interested in the development of thyroid-specific QoL questionnaires.


3.3 | Malignant thyroid disease QoL questionnaires

The following section describes the six questionnaires for malignant thyroid diseases. An overview is presented in [Table 2](#).

TABLE 2 Questionnaires for malignant thyroid diseases

	THYCA-QoL	ThyCa-HRLQOL	EORTC-Thy34	MADSI-Thy	QOL-Thyroid	ThyCAT
Authors	Husson et al.	Li et al.	Singers et al.	Gning et al.	Ferrell, Dow et al.	Aschebrook-Kilfoy et al.
No of items	24	22	34	25	30	10
Available languages	Dutch, Korean	English, Filipino	See EORTC website	English	English, Korean	English
Internal Consistency	0.45–0.82	0.73–0.87	-	0.76–0.92	0.74–0.94	-
Test-retest reliability	-	-	-	-	-	-
External validity	EORTC	EORTC	-	-	FACT-G (0.78)	NATCSS
Construct validity	Factor analysis, multitrait	Factor analysis, Divergence, convergence	-	-	-	Consistency with bifactor model
Criterion validity	-	-	-	ECOG	-	-
Availability						-
Answering time	-	-	15 min	-	-	<2 min

Note: -: No information available or not evaluated.

: available in the publication (ThyCa-HRQOL (45), MDASI-Thy (48)) or with the following links: THYCA-QoL (<https://informahealthcare.com/doi/full/10.3109/0284186X.2012.718445>), EORTC-Thy34 (<https://www.eortc.org/app/uploads/sites/2/2018/08/Specimen-THY34-English.pdf>), QOL-TV (<https://www.midss.org/sites/default/files/qol-thy.pdf>).

3.3.1 | THYCA-QoL

The THYCA-QoL (Thyroid Cancer-Specific Quality of Life) questionnaire was developed by Husson et al. as an add-on to a general QoL questionnaire, the EORTC-QLQ C30, to assess specific aspects of thyroid cancer patients.⁴³ The THYCA-QoL consists of 24 items covering patient's complaints for the past 4 weeks, which are assigned to seven scales. These scales cover neuromuscular symptoms, voice, concentration, local symptomatic symptoms, throat and mouth, psychological and sensory complaints. The response options are similar to those of the general EORTC questionnaire. Thus, there are four response categories ranging from '1 = not at all' to '4 = very much'. The evaluation also follows the scheme of the EORTC. An increasing score reflects a decreasing QoL.

In the first phase of the development, a literature review, focus group discussions and an evaluation of the relevance of the identified problems by experts and patients were conducted. In phase two, the final problem list was operationalized into questions. Finally, in phase three, first a pretest with a debriefing and then a second pretest to check the scales structure was performed. This was confirmed by factor analysis. The Cronbach's alpha values range from 0.45 to 0.82, which means that only four of the seven scales are above the limit (>0.7) for reliability. The THYCA-QoL was originally developed in Dutch and English and translated and validated in Korean by Jeong et al.⁴⁴

3.3.2 | ThyCa-HRLQOL

The Thyroid Cancer-Specific Health-Related Quality of Life Questionnaire (ThyCa-HRLQOL) was developed by Li et al. in

Filipino and English.⁴⁵ It contains 22 items, which are assigned to five scales. Thematically, the five scales include perceived anxiety, psychological distress, functionality, voice and throat complaints. The focus here is more on psychological and functional complaints, compared with the THYCA-QoL, which focuses on symptoms. This questionnaire was also developed as a specific extension for thyroid cancer patients to the general EORTC-QLQ C30 questionnaire.⁴⁶ Thus, the response period, the response options and also the scoring are the same as for the EORTC and THYCA-QoL Questionnaire.

The questionnaire is based on a literature review, focus group discussions and an assessment of the problems for relevance by patients and experts. Initially, a preliminary version was constructed as a pilot test. Subsequently, the resulting questionnaire has been analysed with respect to scaling, validation and reliability. Construct validity was tested by factor analysis, convergence and divergence validity by the Spearman correlation and the five-factor scale structure by a scree test. The results of the analysis indicate good validity and the Cronbach's alpha values of 0.73–0.87 depending on the scale indicate good reliability.

3.3.3 | EORTC-Thy34

The EORTC-Thy34 (Thyroid Cancer Module of the EORTC) is developed by Singer et al.⁴⁷ The questionnaire contains 34 items divided into 10 scale items: discomfort in the head and neck, fatigue, anxiety, hair problems, restlessness, influence on job or education, swallowing, worry about other things, tingling or numbness, voice and into six individual items: body image, cramps, dry mouth, temperature intolerance, joint pain, shoulder problems. The response options range on a four-point scale from '1 = not at all' to '4 = very much', which

refer to the past week and can be transformed into a score from 0 to 100. A high score correlates with a high symptom burden and is associated with low QoL. Since the beginning, patients from different regions of the world have been involved in the process of questionnaire development. As a result, it is already available in 15 languages: Arabic, Chinese, Dutch, English, French, German, Greek, Hebrew, Hindi, Japanese, Italian, Polish, Portuguese, Spanish and Tamil.

The questionnaire was based on two literature reviews and structured interviews with patients and professionals. The problems identified were rated by them for relevance and importance and shortened to a list of 47 items. This provisional item list was tested and evaluated with debriefing interviews. Criteria previously established according to EORTC guidelines determined which items were retained, modified or excluded. In addition, items were categorized into hypothetical scales, and Cronbach's alpha and item-scale correlations were evaluated. Psychometric testing to validate the final questionnaire is still to be conducted in the fourth phase of development.

3.3.4 | MDASI-Thy

The MDASI-Thy (Thyroid Cancer Module of the M. D. Anderson Symptom Inventory) was developed as a supplement to the MDASI, a questionnaire that measures symptoms and its impact on activities of daily living in carcinoma patients, by Gning et al. for thyroid cancer patients.⁴⁸ The MDASI-Thy is composed of six items with specific symptoms such as hoarseness, feeling hot, feeling cold, palpitations, difficulty in swallowing and diarrhoea, 13 items on general cancer symptoms from the original MDASI and six interference items that capture the impact of the disease on activities of daily living in the past 24 h. These include general activity, emotional state, influence on work and relationships with other people, walking and enjoyment of life.

Each item can be scored on an 11-point scale ranging from '0 = nonexistent/no influence' to '10 = as bad as you can get/totally bothers you'. Various scores can be calculated by averaging the scores of the included items. A 'symptom severity score' can be formed from the 19 symptom items, a thyroid sub-score from the six thyroid-specific items, with a larger score indicating a higher symptom burden and impact on daily life.

The thyroid module was created based on a literature review, focus group discussions with patients and experts, additional interviews and a 'cognitive debriefing' after completing the questionnaire. Reliability, measured by Cronbach's alpha, ranged from 0.76 for the Thyroid Scale to 0.92 for the Interference Scale. Furthermore, 'known-groups' validity was demonstrated, as the MDASI-Thy can be used to distinguish different patient groups.

3.3.5 | QOL-Thyroid

The Quality of Life - Thyroid (QOL-Thyroid) questionnaire was developed by Ferrell et al. and contains 30 items.^{49,50} These are

assigned to four scales, each reflecting an aspect of QoL. Physical well-being is measured with two items and 13 questions, psychological well-being with 13 items and 22 questions, social concerns with eight items and 14 questions, and spiritual well-being with seven items and questions. Response options range from '0 = worst outcome' to '10 = best outcome' on an ordinal 11-point scale, with items 1, 3, 10–23 and 27 scored inversely. This should also be considered in the evaluation when calculating an average score or a scale score.

The thyroid version of the questionnaire is based on a general questionnaire for assessing the QoL of patients with cancer, which was also developed by Ferrell et al. and the City of Hope Institute. This was based on a literature review, patient interviews and a pilot test. Psychometric evaluation of validity and reliability, with factor analysis, correlations to the validated FACT-G questionnaire,⁵¹ Cronbach's alpha and test-retest analysis showed good results.

Results on the validation of a previous thyroid version, which consisted of 41 items, were described by Dow et al.⁵² Cronbach's alpha values ranged from 0.74 to 0.94 on the scales and 0.97 overall. A moderate to high correlation of 0.78 was found with the FACT-G questionnaire. A translation of the English version into Korean was performed by Ryu et al. with subsequent validation.⁵³ For this, item convergence and discriminant validity were measured by Pearson correlation, internal consistency by Cronbach's alpha, test-retest reliability by intraclass correlation and external validity by Spearman correlation of an item with the VHI-30 questionnaire.

3.3.6 | ThyCAT

The special feature of ThyCAT is that it is not a classical questionnaire but a computer adaptive test to measure the QoL of patients suffering from thyroid cancer. It was developed in English by Aschebrook-Kilfoy et al.⁵⁴ The questions are taken from the questionnaire of the NATCSS, North American Thyroid Cancer Survivorship Study. The NATCSS questionnaire includes the City of Hope Quality of Life Thyroid Version questionnaire by Ferrell et al., and other open-ended QoL questions, and includes a total of 75 items.⁵⁵ The ThyCAT consists of 58 of these questions. Each participant answers different items. Depending on how the first question is answered, a provisional score is formed. This decides which question will be asked next to get the most information. Using a linear regression model, it was confirmed that the ThyCAT scores predicted the QoL scores of the entire NATCSS questionnaire with a precision of 0.96. Validity was demonstrated by the degree of agreement of ThyCAT with the bifactor model on which it is based.

3.3.7 | Further questionnaires for thyroid cancer

The questionnaire of Emmanouilidis et al. has a focus on QoL after thyroidectomy and radio ablation therapy.⁵⁶ Pak et al. focus on symptoms and complaints associated with radio ablation therapy and

hypothyroidism.⁵⁷ The Hebrew questionnaire created by Dagan et al. (TQOL, Thyroid Quality of Life Instrument) consists of a total of 15 items that were composed of items from the University of Washington Quality of Life Instrument for Head and Neck Tumours and the City of Hope Quality of Life Thyroid Version by Ferrell et al. described earlier.⁵⁸

3.4 | QoL questionnaires for endocrine orbitopathy

The following section describes the six questionnaires for endocrine orbitopathy. An overview is presented in Table 3.

3.4.1 | GO-QOL

The original Graves Ophthalmopathy Quality of Life Questionnaire (GO-QOL) was developed in Dutch by Terwee et al.⁵⁹ With a total of 16 items, it measures 'visual functioning' on the one hand (eight items) and psychosocial effects of the altered appearance on the other hand (also eight items). The visual functioning scale can be answered on a three-point scale with 'not limited, a little limited, strongly limited' and the appearance scale with 'strongly, a little, not at all', each referring to the past week. For scoring, the eight questions in each case are summed and transformed into a scale from 0 to 100, where 0 represents the worst health status and 100 the best. To create the GO-QOL, relevant items from existing vision-specific quality of life questionnaires such as the VF-14, ADVS and VR-SIP were used, based on discussions from patients and experts. In addition, a questionnaire with open-ended questions related to

symptoms and limitations due to the disease was completed by patients and subsequently analysed. A pretest was conducted before using the GO-QOL. The two-scale structure was confirmed by factor analysis. Validity has been demonstrated by correlations with individual scales of other questionnaires and with age, sex and clinical measures. Both scales show good reliability as measured by Cronbach's alpha values of 0.89 and 0.87. This was also confirmed by test-retest analysis, with intraclass correlations of 0.83 and 0.87 (59). Further analyses measured good long-term validity and found that clinical change was well-captured by the GO-QOL.^{60,61}

Over time, modified versions of the questionnaire have been published in Croatian, Danish, Dutch, English, French, German, Greek, Italian, Russian, Spanish and Turkish. It was also adapted for the Australian population.⁶²

3.4.2 | GO-QLS

The Graves' Ophthalmopathy Quality of Life Scale (GO-QLS) was developed by Yeatts et al.⁶³ A total of nine items include restrictions on social functioning, 'distance activities' such as recognizing faces in a room or watching a movie, 'near activities' such as finding something on a crowded shelf and restrictions on general vision such as the appearance of the eyes and the influence of symptoms on well-being. The two general items can be answered on a five-point scale, the remaining items on a six-point scale. A high score represents a more impaired QoL.

In a first step, a 105-item questionnaire was assembled from SF-12, DSQL, NEI-VFQ-51 and disease-specific questions on visual

TABLE 3 Questionnaires developed for endocrine orbitopathy

	GO-QOL	GO-QLS	TED-QOL	STED-QOL	TAO-QoL	Ox-TED
Authors	Terwee et al.	Yeatts et al.	Fayers et al.	Wong et al.	Tehrani et al.	Insull et al.
No of items	16	9	3	10	90	7
Available languages	See EUGOGO website	English	English, Korean	Chinese	German	English
Internal consistency	0.89V 0.97 A	0.89	-	-	0.63	-
Test-retest reliability	0.83V, 0.87 A	-	0.74V, 0.87 A, 0.81 LQ	-	-	-
External validity	MOS-24, SIP	NEI-VFQ 25	GO-QOL, GO-QLS	-	-	-
Construct validity	Factor analysis	Factor analysis	-	Factor analysis, Rasch analysis	-	-
Criterion validity	✓	✓	✓	✓	(✓)	-
Availability					∅	
Answering time	3,1 min	2,4 min	1,6 min	-	-	-

Note: -: No information available or not evaluated.

✓: Correlation with clinical parameters (NOSPECS classification, CAS or Herthel index).

: available in the document (GO-QOL (59), GO-QLS (65), TED-QOL (66), STED-QOL <https://tvst.arvojournals.org/article.aspx?articleid=2705521>, Ox-TED (69)).

∅: Inquiries to the author.

acuity.⁶⁴ For the GO-QLS, only those items were selected that had matching response scales, few missing values, correlated best with disease severity and discriminated best between mild and moderate disease severity. Unidimensionality was demonstrated by factor analysis. The validity of the final GO-QLS was tested by correlation with the NEI-VFQ-25 and clinical disease severity scores. The Cronbach's alpha value was 0.89.

3.4.3 | TED-QOL

The TED-QOL (Thyroid eye disease Quality of Life Questionnaire) developed by Fayers et al. is the shortest with a total of three items.⁶⁵ The items enquire about the influence of endocrine orbitopathy on general QoL, on daily activities and on satisfaction with the own appearance. The extent of the influence can be evaluated on an 11 point scale, where 0 stands for 'no influence' and 10 for 'totally disturbing'. Each item is evaluated with a separate score. There is no composite score for all three items. The values of each item can be linearly transformed to a score of '0 = worst possible outcome' and '100 = best possible outcome'. Interviews with experts and patients and a pilot test were performed. A high positive correlation of the scales of the TED-QOL with the corresponding scales of the GO-QOL and the GO-QLS questionnaires indicates good validity. Correlation with clinical scores such as the VISA score was also tested. Test-retest reliability was demonstrated by intraclass correlation coefficients ranging from 0.74 to 0.87.

3.4.4 | STED-QoL

The STED-QoL (Singapore Thyroid Eye Disease Quality of Life Questionnaire) includes 10 items and was developed by Wong et al. specifically for Chinese patients with endocrine orbitopathy.⁶⁶ The items represent four aspects of QoL with the domains activity limitations, comfort, psychological and social. The questions are answered on a four-point Likert scale from '0 = all the time' to '4 = not at all'. Consequently, a high score represents a good QoL. A total of six psychosocial items can be evaluated separately in a subscale. Qualitative analyses of focus group discussions were conducted to create the STED-QoL. The resulting questionnaire was pilot-tested and modified according to psychometric analyses. For validation, a factor analysis was performed to confirm unidimensionality, and a Rasch analysis was performed to test the 'item response theory'. Furthermore, the ability of the questionnaire to discriminate between different disease activities was tested using ANOVA analysis. Evaluations of reliability have not yet been performed.

3.4.5 | TAO-QoL

The TAO-QoL (Thyroid Associated Ophthalmology Quality of Life) questionnaire was developed by Tehrani et al.⁶⁷ The aim was to

develop a questionnaire for German patients with endocrine orbitopathy, which measures the QoL also after surgical therapy. It contains 90 items and was designed in collaboration with ophthalmologists and endocrinologists. Four categories ranging from '1 = maximum satisfaction' to '4 = minimum satisfaction' are available for response. From all answers, the average can be calculated for a total score, where low values indicate a high QoL. The validation was indirectly demonstrated by a correlation between the QoL scores and clinical parameters (Hertel value). However, a correlation value was not available. Reliability, measured by Cronbach's alpha, was low with a value of 0.63.

3.4.6 | Ox-TED quality of life score

The Ox-TED (Oxford Thyroid Eye Disease) Quality of Life Score was created by Insull et al.⁶⁸ The questionnaire consists of seven questions about the influence of the disease, the therapy, the changed appearance on the general QoL and on daily activities. The questions can each be answered with a scale from '1 = does not bother' to '10 = very bothersome'. All scores were added together for evaluation. The total score ranges from 7 to 100. The greater this score, the greater the influence on QoL. Information on validity and reliability is not available.

3.4.7 | Further questionnaires for endocrine orbitopathy

Two more questionnaires could be identified that did not have validity and reliability testing. Finamor et al. created a 10-item questionnaire called GO-HRQL with three response categories ('0 = not impaired, 0.5 = somewhat impaired and 1 = very impaired').⁶⁹ The score ranges from '0 = minimal' to '10 = maximum' impact on QoL. It includes psychosocial aspects such as change in appearance and influence on self-esteem and social contacts, and visual function aspects such as walking or reading. The second questionnaire was developed by Sisson et al.⁷⁰ It consists of four items with the four response categories '0 = none to 3 = strong', which ask for the pain in the eyes, changed appearance and visual acuity.

4 | DISCUSSION

The aim of this systematic review was to provide an overview of currently available thyroid-specific QoL questionnaires with a focus on the quality of the studies by considering the validation and reliability process. In contrast to the work of Wong et al.,¹³ the focus is not so much on the assessment of the quality of the questionnaires, as this is described clearly and in detail there. Instead, the focus is more on the clinical applicability of the questionnaires and the identification of gaps regarding specific questions.

In this systematic review, a total of 16 questionnaires could be identified, and others that capture only partial aspects of QoL. Since the assessment of QoL depends on the underlying thyroid disease, the questionnaires are divided into three main groups.

For benign thyroid disease, four specific questionnaires were identified. Of these, three have been evaluated in sufficient quality to be used for studies. These include the ThyPRO, the ThyPRO-39 and the Thy-D-QoL.^{14,15,30} The ThyPRO questionnaire is well-validated and has high reliability for assessing the QoL of benign thyroid diseases and should be used for questions examining QoL as the main outcome. A comparison between the different benign thyroid diseases can also be made through its different subscales. A disadvantage of the ThyPRO is its length of 85 items and long completion time. This could result in incomplete patients' responses. The ThyPRO-39 is much shorter and faster to answer, which is certainly an advantage over the long version. Another advantage of the short version is that items that previously showed weaknesses in cross-cultural validation or differential item functions were excluded. The validity of the Thy-D-QoL by McMillan et al. was evaluated in fewer steps compared with both versions of the ThyPRO, but it still shows good quality. The Thy-D-QoL could be used in research with hypothyroidism as a focus or in the clinical setting, as it is even shorter than the ThyPRO-39. For the goitre-specific questionnaire by Kaniuka-Jakubowska, the Thy-R-HRQoL, validity has not been tested.²⁹ Nevertheless, it has high reliability and could be used as an additional questionnaire for questions investigating the effects of goitre. One advantage is that with a total of 49 items it addresses possible problems caused by goitre very specifically. The Thy-R-HRQoL and the ThyPRO-39 were developed after the publication by Wong et al.¹³ Thus, Wong's recommendation for the use of the ThyPRO to assess QoL in benign thyroid disease is also confirmed by this review. However, it should be noted that a validated German version of the ThyPRO-39 questionnaire is now available, which can be used for the research on benign thyroid diseases.²⁸

For malignant thyroid diseases, six QoL questionnaires were identified. Three of the questionnaires have been developed as specific supplements to the valid, general EORTC questionnaire. These include the THYCA-QoL, the ThyCa-HRQoL and the EORTC-Thy34 module. Overall, none of the questionnaires was evaluated with regard to all possible evaluation aspects. However, the EORTC-Thy34 should be a well-evaluated tool in the foreseeable future. Another important aspect is the availability of validated translations. This is the case only in Dutch and Korean for the THYCA-QoL and in English and Filipino for the ThyCa-HRQoL. The EORTC Thy34 offers an advantage here, as it has been developed in 15 languages, including German. The focus of the MDASI-Thy questionnaire is more on symptoms and their effects than on other domains of QoL. Thus, this is more suitable for questions with a corresponding focus and not for questions that want to investigate, for example, psychosocial aspects.

The only psychometric tests used in the MDASI-Thy evaluation were the analysis of internal consistency and the 'validity of known-groups'. A major weakness of the City of Hope Institute QoL-Thyroid

is that the validation by Ferrell et al. was only performed for a general questionnaire, on which the specific versions for malignant thyroid disease are based. Subsequently, only Cronbach's alpha values and the correlation of a scale with the corresponding one of the FACT-G questionnaire were checked by Dow et al.^{51,52} This questionnaire was the first to be developed and has been used by many studies. The use of this questionnaire is supported by the fact that its content covers not only physical but also psychological, social and spiritual aspects. The ThyCAT is the only computer adaptive test. A big advantage is that it is very short with less than 2 min of answer time for 10 questions. Disadvantages are that it requires a smartphone or computer to access the software to complete and reliability has not been tested. In terms of content, it is based on the QoL-Thyroid. No test-retest analysis was performed on any of the questionnaires. In addition, criterion validity was only examined for the MDASI-Thy. Similarly, response time or completion rates are an important aspect to evaluate the questionnaires. Response times were only reported for the ThyCAT and the EORTC-Thy34, thus a comparison with other questionnaires in this respect is not possible. In conclusion, each of the six available questionnaires has strengths but also weaknesses. A recommendation for a single questionnaire cannot be given, as individual consideration should be given to which drawbacks have the least impact on use in the planned study or the clinical setting. For a quick screening, the ThyCAT is certainly a good option. If the focus is more on specific symptoms, the MDASI-Thy could be the right choice. If different dimensions of QoL or a healthy control group are required, the EORTC questionnaire with a specific module will be a good choice.⁴⁷ If this module is sufficiently validated in the future, it will probably become the gold standard for assessing QoL in malignant thyroid disease research.

Out of the six specific questionnaires that were identified on the topic of QoL in endocrine orbitopathy, only three are useful for clinical and research use. The TAO-QoL and Ox-Ted do not have sufficient quality for use, and the STED-QoL was developed exclusively for Asian patients. The GO-QoL questionnaire by Terwee et al. is the most commonly used disease-specific QoL questionnaire. The GO-QoL, on the other hand, has a focus on the impact of impaired vision, and less on consequences due to altered appearance. In contrast to the two questionnaires, the three items of the TED-QoL are less specific and superficial but quick to answer. Therefore, the TED-QoL is well-suited for a quick screening on general limitations of QoL or in studies in which QoL is measured as a secondary outcome parameter. A comprehensive analysis of aspects affecting QoL is not possible with the TED-QoL, in contrast to the GO-QoL and the GO-QoL. Response times and completion rates for these three questionnaires are reported by Fayers et al.⁶⁵ The completion times are short at 3 min or less for all three questionnaires.

The GO-QoL is the best-validated questionnaire. Reliability and external, construct and criterion validity are high. An advantage over the TED-QoL and GO-QoL is that GO-QoL is the only questionnaire for which longitudinal validity has been confirmed. Thus, a comparison of QoL before and after therapies is possible with the GO-QoL. Nevertheless, both the GO-QoL and the TED-QoL are sufficiently

validated for clinical and research use. An additional advantage of the GO-QOL is the available validated translations for various languages, which can be downloaded from the site of the European Group on Graves Orbitopathy (EUGOGO). A point of criticism is the fact that the item 'impairment while cycling' is not relevant for some cultures. Thus, it was replaced by Park et al. in creating an Australian version.⁶² The GO-QLS and the TED-QOL have not yet been translated and validated in German. Only the TAO-QoL by Tehrani et al. is available in German but in insufficient quality.

In the course of time, several reviews have been made on the assessment of QoL in endocrine orbitopathy.⁷¹⁻⁷⁴ Since the most recent reviews by Bartalena et al. and Lee et al. in 2019 and 2020, no new recording instruments or further validations of the questionnaires listed here have been published. An approach for future research could be either the validation of German versions of the GO-QLS and the TED-QoL, and an improvement of the German TAO-QoL, in order to have validated survey instruments available to assess QoL in patients with endocrine orbitopathy in clinical settings or studies with a German-speaking population.

There are certain limitations to this review. Firstly, we did not provide a quality appraisal per se or an analysis of bias. However, as this is no typical systematic review, we did not deem it necessary. Also, the tables detailing the different questionnaires and our comments with regards to the different questionnaires in the discussion section provide enough data and information to allow the reader to make an informed decision on which questionnaire might be relevant for them. Secondly, we did not include questionnaires dealing only with the aspect of anxiety. The respective questionnaires are mentioned, but since anxiety is only one aspect influencing the multi-faceted QoL, we deemed it beyond the scope of this review to go into detail there.

Our review indicates that currently for malignant thyroid diseases and for endocrine orbitopathy there is still some work to be done to establish a clinically usable and rigorously validated questionnaire. In case of malignant diseases, this will probably be no issue anymore in the foreseeable future, when the thyroid-specific EORTC questionnaire is available. However, as mentioned above, the currently available questionnaires for endocrine orbitopathy are not sufficiently validated or do not cover all aspects of thyroid-specific QoL. Thus, future research should cover this currently existing gap.

5 | CONCLUSIONS

This review offers a guide for researchers in need of a thyroid-specific Quality of Life questionnaire as it gives an overview of currently available questionnaires and the way they were validated, with a focus on their clinical applicability. For different benign thyroid diseases, the ThyPRO and its short version ThyPRO-39 are the most suitable questionnaires for clinical use. In hypothyroidism, the ThyD-QoL can be used as an alternative. The choice of a QoL questionnaire for malignant thyroid diseases should be decided individually.

For the assessment of QoL in patients with endocrine orbitopathy, the GO-QOL is best validated. As a short screening questionnaire in the clinic or for questions in which QoL is a secondary outcome, the TED-QoL seems to be an appropriate alternative.

AUTHOR CONTRIBUTIONS

Verena Nicole Uslar: Data curation (equal); investigation (equal); methodology (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (supporting); writing – review and editing (lead). **Caroline Becker:** Data curation (equal); formal analysis (equal); investigation (equal); visualization (equal); writing – review and editing (equal). **Dirk Weyhe:** Conceptualization (equal); project administration (supporting); resources (lead); writing – review and editing (supporting). **Navid Tabriz:** Conceptualization (lead); methodology (equal); project administration (lead); supervision (equal); validation (equal); writing – original draft (lead); writing – review and editing (equal).

ACKNOWLEDGEMENTS

Open access funding enabled and organized by ProjektDEAL.

FUNDING INFORMATION

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

CONFLICT OF INTEREST

The authors have no relevant financial or nonfinancial interests to disclose.

ETHICAL APPROVAL

The conducted research is not related to either human or animal use.

ORCID

Verena Uslar  <https://orcid.org/0000-0003-3252-2076>

REFERENCES

- Faggiano A, Del Prete M, Marciello F, Marotta V, Ramundo V, Colao A. Thyroid diseases in elderly. *Minerva Endocrinol.* 2011;36(3):211-231.
- Vanderpump MP. The epidemiology of thyroid disease. *Br Med Bull.* 2011;99:39-51.
- Leese GP, Flynn RV, Jung RT, Macdonald TM, Murphy MJ, Morris AD. Increasing prevalence and incidence of thyroid disease in Tayside, Scotland: the thyroid epidemiology audit and research study (TEARS). *Clin Endocrinol.* 2008;68(2):311-316.
- Scerrino G, Morfino G, Paladino NC, et al. Does thyroid surgery for Graves' disease improve health-related quality of life? *Surg Today.* 2013;43(12):1398-1405.
- Watt T, Groenvold M, Rasmussen AK, et al. Quality of life in patients with benign thyroid disorders. A Review. *Eur J Endocrinol.* 2006;154(4):501-510.
- Wojewoda B, Juzwizyn J, Rosińczuk-Tonderys J. Quality of life after surgical treatment of thyroid gland diseases. *Rocz Akad Med Białymst.* 2005;50(Suppl 1):133-136.
- Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med.* 1993;118(8):622-629.

8. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
9. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.
10. Mokkink LB, De Vet HC, Prinsen CA, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-1179.
11. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
12. Terwee CB, Prinsen CA, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159-1170.
13. 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.
14. Watt T, Bjorner JB, Groenvold M, et al. Development of a short version of the thyroid-related patient-reported outcome ThyPRO. *Thyroid*. 2015;25(10):1069-1079.
15. Watt T, Hegedüs L, Groenvold M, et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *Eur J Endocrinol*. 2010;162(1):161-167.
16. Watt T, Rasmussen AK, Groenvold M, et al. Improving a newly developed patient-reported outcome for thyroid patients, using cognitive interviewing. *Qual Life Res*. 2008;17(7):1009-1017.
17. Watt T, Bjorner JB, Groenvold M, et al. Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. *Qual Life Res*. 2009;18(4):483-496.
18. Watt T, Groenvold M, Deng N, et al. Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. *Health Qual Life Outcomes*. 2014;12:126.
19. Watt T, Groenvold M, Hegedüs L, et al. Few items in the thyroid-related quality of life instrument ThyPRO exhibited differential item functioning. *Qual Life Res*. 2014;23(1):327-338.
20. Watt T, Cramon P, Hegedüs L, et al. 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.
21. Watt T, Barbesino G, Bjorner JB, et al. Cross-cultural validity of the thyroid-specific quality-of-life patient-reported outcome measure, ThyPRO. *Qual Life Res*. 2015;24(3):769-780.
22. Boronat M, González-Lleó A, Rodríguez-Pérez C, et al. Adaptation and cross-cultural validation of the Spanish version of the thyroid-related quality-of-life patient-reported outcome questionnaire. *Endocrinol Diabetes Nutr*. 2018;65(9):500-507.
23. Mintziori G, Watt T, Veneti S, et al. ThyPROgr: the Greek edition of the ThyPRO questionnaires for patients with benign thyroid diseases. *Hormones (Athens)*. 2018;17(1):107-112.
24. Sawicka-Gutaj, N., Watt, T., Sowiński, J., Gutaj, P., Waligórska-Stachura, J., & Ruchała, M. (2015). ThyPROpl--The Polish version of the thyroid-specific quality of life questionnaire ThyPRO. *Endokrynol Pol*, 66(4), 367-80. ThyPROpl – polska wersja kwestionariusza oceny jakości życia u pacjentów z chorobami tarczycy
25. Piomchai P, Chaiudomsom S, Wijakkanalan P, Watt T. Validity and reliability of the Thai version of the thyroid-related patient-reported outcome-A thyroid-specific quality of life questionnaire. *Int Arch Otorhinolaryngol*. 2021;25(1):e92-e97.
26. Zahan AE, Watt T, Pascanu I, et al. The Romanian version of the Thyroid-related patient-reported outcomes ThyPRO and ThyPRO-39. Translation and Assessment of reliability an cross-cultural validity. *Acta Endocrinol (Buchar)*. 2018;14(2):192-200.
27. Watt T, Cramon P, Frenzl DM, Ware JE Jr, Group, T. Assessing health-related quality of life in patients with benign non-toxic goitre. *Best Pract Res Clin Endocrinol Metab*. 2014;28(4):559-575.
28. Tabriz N, Gloy K, Schantzen A, Fried D, Weyhe D, UsLAR V. Validity and reliability of the German version of the shortened thyroid-specific quality of life questionnaire (ThyPRO-39de). *Endocr Connect*. 2021;10(9):1065-1072.
29. Kaniuka-Jakubowska S, Lewczuk A, Majkovicz M, et al. Nontoxic goiter (NTG) and radioiodine: what do patients think about it? Quality of life in patients with NTG before and after 131-I therapy. *Front Endocrinol (Lausanne)*. 2018;9:114.
30. McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU. Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. *Thyroid*. 2004;14(11):916-925.
31. McMillan C, Bradley C, Razvi S, Weaver J. Evaluation of new measures of the impact of hypothyroidism on quality of life and symptoms: the ThyDQoL and ThySRQ. *Value Health*. 2008;11(2):285-294.
32. Quinque EM, Villringer A, Kratzsch J, Karger S. Patient-reported outcomes in adequately treated hypothyroidism – insights from the German versions of ThyDQoL, ThySRQ and ThyTSQ. *Health Qual Life Outcomes*. 2013;11:68.
33. Clyde PW, Harari AE, Getka EJ, Shakir KM. Combined levothyroxine plus liothyronine compared with levothyroxine alone in primary hypothyroidism: a randomized controlled trial. *JAMA*. 2003;290(22):2952-2958.
34. Luster M, Felbinger R, Dietlein M, Reiners C. Thyroid hormone withdrawal in patients with differentiated thyroid carcinoma: a one hundred thirty-patient pilot survey on consequences of hypothyroidism and a pharmacoeconomic comparison to recombinant thyrotropin administration. *Thyroid*. 2005;15(10):1147-1155.
35. Mitchell AL, Hegedüs L, Žarković M, Hickey JL, Perros P. Patient satisfaction and quality of life in hypothyroidism: an online survey by the british thyroid foundation. *Clin Endocrinol*. 2021;94(3):513-520.
36. Saravanan P, Chau WF, Roberts N, Vedhara K, Greenwood R, Dayan CM. Psychological well-being in patients on 'adequate' doses of l-thyroxine: results of a large, controlled community-based questionnaire study. *Clin Endocrinol*. 2002;57(5):577-585.
37. McMillan C, Bradley C, Razvi S, Weaver J. Psychometric evaluation of a new questionnaire measuring treatment satisfaction in hypothyroidism: the ThyTSQ. *Value Health*. 2006;9(2):132-139.
38. Fahrenfort JJ, Wilterdink AM, van der Veen EA. Long-term residual complaints and psychosocial sequelae after remission of hyperthyroidism. *Psychoneuroendocrinology*. 2000;25(2):201-211.
39. Cooper DS, Halpern R, Wood LC, Levin AA, Ridgway EC. L-thyroxine therapy in subclinical hypothyroidism. A double-blind, placebo-controlled trial. *Ann Intern Med*. 1984;101(1):18-24.
40. Kasemsiri P, Trakulkajornsak S, Bamroong P, Mahawerawat K, Piromchai P, Ratanaaneekchai T. Comparison of quality of life between patients undergoing trans-oral endoscopic thyroid surgery and conventional open surgery. *BMC Surg*. 2020;20(1):18.
41. Wilde D, Wilken L, Stamm B, et al. The HPQ-development and first Administration of a Questionnaire for Hypoparathyroid patients. *JBMR Plus*. 2020;4(1):e10245.
42. Jaeschke R, Guyatt G, Cook D, Harper S, Gerstein H. Spectrum of quality of life impairment in hypothyroidism. *Qual Life Res*. 1994;3(5):323-327.
43. Husson O, Haak HR, Mols F, et al. Development of a disease-specific health-related quality of life questionnaire (THYCA-QoL) for thyroid cancer survivors. *Acta Oncol*. 2013;52(2):447-454.
44. Jeong Y, Choi J, Ahn AL, et al. Validation of the Korean version of the thyroid cancer-specific quality of life questionnaire. *Ann Surg Treat Res*. 2015;89(6):287-294.
45. Li RJ, Jimeno C, Sandoval MA, Cabungcal AC, Ogbac R, Uy GL. Development and validation of a thyroid cancer-specific health-related quality of life questionnaire for adult filipinos with

- differentiated thyroid cancer. *J ASEAN Federation Endocrine Soc.* 2016;31(2):87-96.
46. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI J Natl Cancer Inst.* 1993;85(5):365-376.
 47. Singer S, Jordan S, Locati LD, et al. The EORTC module for quality of life in patients with thyroid cancer: phase III. *Endocr Relat Cancer.* 2017;24(4):197-207.
 48. Gning I, Trask PC, Mendoza TR, et al. Development and initial validation of the thyroid cancer module of the M.D. Anderson Symptom Inventory. *Oncology.* 2009;76(1):59-68.
 49. Ferrell, B.R. Quality of Life: Thyroid Version. 2000.
 50. Ferrell BR, Dow KH, Grant M. Measurement of the quality of life in cancer survivors. *Qual Life Res.* 1995;4(6):523-531.
 51. Cella DF, Tulskey DS, Gray G, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol.* 1993;11(3):570-579.
 52. Dow KH, Ferrell BR, Anello C. Quality-of-life changes in patients with thyroid cancer after withdrawal of thyroid hormone therapy. *Thyroid.* 1997;7(4):613-619.
 53. Ryu CH, Park B, Ryu J, et al. Development and evaluation of a Korean version of a thyroid-specific quality-of-life questionnaire scale in thyroid cancer patients. *Cancer Res Treat.* 2018;50(2):405-415.
 54. Aschebrook-Kilfoy B, Ferguson BA, Angelos P, Kaplan EL, Grogan RH, Gibbons RD. Development of the ThyCAT: A clinically useful computerized adaptive test to assess quality of life in thyroid cancer survivors. *Surgery.* 2018;163(1):137-142.
 55. Aschebrook-Kilfoy B, James B, Nagar S, et al. Risk factors for decreased quality of life in thyroid cancer survivors: initial findings from the north American thyroid cancer survivorship study. *Thyroid.* 2015;25(12):1313-1321.
 56. Emmanouilidis N, Schrem H, Winkler M, Klempnauer J, Scheumann GFW. Long-term results after treatment of very low-, low-, and high-risk thyroid cancers in a combined setting of thyroidectomy and radio ablation therapy in euthyroidism. *Int J Endocrinol.* 2013;2013:769473.
 57. Pak K, Kim SJ, Shin S, et al. Impact of age and sex on the quality of life following radioactive iodine ablation in patients with thyroid cancer. *Nuklearmedizin.* 2017;56(5):177-183.
 58. Dagan T, Bedrin L, Horowitz Z, et al. Quality of life of well-differentiated thyroid carcinoma patients. *J Laryngol Otol.* 2004;118(7):537-542.
 59. Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. *Br J Ophthalmol.* 1998;82(7):773-779.
 60. Terwee CB, Dekker FW, Mourits MP, et al. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. *Clin Endocrinol.* 2001;54(3):391-398.
 61. Terwee CB, Gerding MN, Dekker FW, Prummel MF, van der Pol JP, Wiersinga WM. Test-retest reliability of the GO-QOL: a disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy. *J Clin Epidemiol.* 1999;52(9):875-884.
 62. Park JJ, Sullivan TJ, Mortimer RH, Wagenaar M, Perry-Keene DA. Assessing quality of life in Australian patients with Graves' ophthalmopathy. *Br J Ophthalmol.* 2004;88(1):75-78.
 63. Yeatts RP. Quality of life in patients with graves ophthalmopathy. *Trans Am Ophthalmol Soc.* 2005;103:368-411.
 64. Bradley EA, Sloan JA, Novotny PJ, Garrity JA, Woog JJ, West SK. Evaluation of the National eye Institute visual function questionnaire in Graves' ophthalmopathy. *Ophthalmology.* 2006;113(8):1450-1454.
 65. Fayers T, Dolman PJ. Validity and reliability of the TED-QOL: a new three-item questionnaire to assess quality of life in thyroid eye disease. *Br J Ophthalmol.* 2011;95(12):1670-1674.
 66. Wong MHY, Fenwick E, Aw AT, Lamoureux EL, Seah LL. Development and validation of the Singapore thyroid eye disease quality of life questionnaire. *Transl Vis Sci Technol.* 2018;7(5):14.
 67. Tehrani M, Krummenauer F, Mann WJ, Pitz S, Dick HB, Kahaly GJ. Disease-specific assessment of quality of life after decompression surgery for Graves' ophthalmopathy. *Eur J Ophthalmol.* 2004;14(3):193-199.
 68. Insull EA, Sipkova Z, David J, Turner HE, Norris JH. Early low-dose rituximab for active thyroid eye disease: an effective and well-tolerated treatment. *Clin Endocrinol.* 2019;91(1):179-186.
 69. Finamor FE, Martins JR, Nakanami D, Paiva ER, Manso PG, Furlanetto RP. Pentoxifylline (PTX)—an alternative treatment in Graves' ophthalmopathy (inactive phase): assessment by a disease specific quality of life questionnaire and by exophthalmometry in a prospective randomized trial. *Eur J Ophthalmol.* 2004;14(4):277-283.
 70. Sisson JC, Schipper MJ, Nelson CC, Freitas JE, Frueh BR. Radioiodine therapy and Graves' ophthalmopathy. *J Nucl Med.* 2008;49(6):923-930.
 71. Bartalena L, Wiersinga WM. Proposal for standardization of primary and secondary outcomes in patients with active, Moderate-to-Severe Graves' Orbitopathy. *Eur Thyroid J.* 2020;9(Suppl 1):3-16.
 72. Estcourt S, Quinn AG, Vaidya B. Quality of life in thyroid eye disease: impact of quality of care. *Eur J Endocrinol.* 2011;164(5):649-655.
 73. Lee THB, Sundar G. Quality of life in thyroid eye disease: A systematic review. *Ophthalmic Plast Reconstr Surg.* 2020;36(2):118-126.
 74. Wiersinga WM, Prummel MF, Terwee CB. Effects of Graves' ophthalmopathy on quality of life. *J Endocrinol Investig.* 2004;27(3):259-264.

How to cite this article: UsLAR V, Becker C, Weyhe D, Tabriz N. Thyroid disease-specific quality of life questionnaires - A systematic review. *Endocrinol Diab Metab.* 2022;5:e357. doi: [10.1002/edm2.357](https://doi.org/10.1002/edm2.357)