

1 **A systematic review of quality of thyroid-specific health-related quality of**
2 **life instruments recommends ThyPRO for patients with benign thyroid**
3 **diseases**

4
5
6 Order of Author: Carlos K.H. Wong, PhD*¹ , Brian H.H. Lang, MS, FRACS² , Cindy L.K.
7 Lam, MD¹

8 * First and correspondence Author

9 ¹ Department of Family Medicine and Primary Care, The University of Hong Kong

10 ² Division of Endocrine Surgery, Department of Surgery, The University of Hong Kong

11
12 **Abbreviated Title:** Recommendation for HRQOL Instruments in Thyroid

13 **Keywords:** systematic review; thyroid disease; quality of life, measurement property;
14 psychometrics; COSMIN

15 **Word Count:** 3224

16 **Number of figures and tables:** 5

17
18 *Correspondence author and person to whom reprint request should be addressed:*

19 Name: Carlos King Ho Wong, PhD, MPhil, BSc

20 Institution: Department of Family Medicine and Primary Care, The University of Hong Kong

21 Address: 3/F, Ap Lei Chau Clinic, 161 Ap Lei Chau Main Street, Ap Lei Chau, Hong Kong

22 Contact: +852-25185688 (tel); +852-28147475 (fax) carloshe@hku.hk (email)

23
24 Disclose Statement: The authors have nothing to disclose.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Abstract

Objective: To appraise the measurement properties of thyroid-specific health-related quality of life (HRQOL) instruments, and to provide recommendations on the choice of HRQOL instruments.

Study Design and Setting: Systematic review of English-language literature published between 1993 and 2015 identified psychometric studies involving patients with thyroid disease through a search of Pubmed, Web of Science, Embase, and OVID Medline. HRQOL instruments were graded on methodological quality and overall levels of evidence using a COSMIN checklist.

Results: After a review of 743 original studies, 23 studies reporting 14 standardized HRQOL instruments targeted for Graves’ ophthalmopathy (n=4), hypothyroidism (n=3), thyroid cancer (n=2), other thyroid disease (n=3) and non-thyroid tumor sites (n=2) were identified. Hypothesis testing was evaluated most frequently. The 84-item thyroid-specific patient-reported outcome measure (ThyPRO) and 16-item Graves' ophthalmopathy specific Quality of Life (GO-QOL) instruments were the most extensively evaluated instrument. The highest number of positive ratings in overall level of psychometric evidence was found in ThyPRO, GO-QOL and 11-item Thyroid Treatment Satisfaction Questionnaire (ThyTSQ) instruments.

Conclusions: The ThyPRO is recommended for the assessment of HRQOL in patients with benign thyroid diseases whilst measurement properties of GO-QOL and ThyTSQ are satisfactory in measuring HRQOL of patients with Graves’ ophthalmopathy and hypothyroidism, respectively.

Abstract Word Count: 200

50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70

What is new?

Key finding:

- ThyPRO is recommended to assess health-related quality of life (HRQOL) in patients with benign thyroid diseases according to quality assessment criteria
- Measurement properties of GO-QOL and ThyTSQ are satisfactory in measuring HRQOL in patients with graves' ophthalmopathy and hypothyroidism, respectively.

What this adds to what was known:

- No systematic review synthesized evidence on the critical appraisal of the measurement properties of generic and condition-specific HRQOL instruments that have been validated for use in patients with thyroid disease.

What is the implication, what should change now:

- There is a need for an improvement in the reporting quality of measurement properties in newly developed or translated HRQOL instruments for thyroid disease patients, in particular for thyroid cancer patients.
- Evaluation of measurement properties of HRQOL instruments, especially measurement error, criterion validity and cross-cultural validity, should be commenced.

71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94

Manuscript Text

Introduction

Thyroid diseases such as thyroid dysfunctions and cancers occur not uncommonly in the general population and have imposed an increased burden on our already over-stretched healthcare system[1-3]. This increased burden is attributed to the increased incidence as well as increased ascertainment and earlier diagnosis of thyroid diseases[4]. However, despite this increased burden, patients with thyroid diseases are expected to have a comparable life-expectancy as the normal population[5]. Therefore, health-related quality of life (HRQOL) has become a more clinically-relevant and an important outcome measurement for patients with thyroid diseases.

In accordance to classical literature, HRQOL is defined as a subjective and multi-dimensional construct of health and welling. Concepts of general health, physical functioning, psychological functioning and social functioning are fundamental determinants of HRQOL in general[6] or thyroid disease populations[7]. Studies so far[8-10] have revealed that persisting symptoms in thyroid dysfunctions and treatment following diagnosis of thyroid cancer had considerable impact on HRQOL. Impaired HRQOL in patients with thyroid dysfunctions were observed when being treated with antithyroid medication[11]. Relative to the normal population, thyroid cancer survivors suffered from significant HRQOL impairment[10]. Despite the availability of various thyroid-specific instruments[8-10], there has been no recommended, commonly-accepted or the best instrument in the HRQOL assessment of patients with thyroid diseases. One reason for this lack of the best instrument is

95 the fact that measurement properties of instruments used for thyroid diseases have not been
96 systematically assessed with agreed quality criteria to enable direct comparison[12].
97 Comprehensive assessment of measurement properties, including validity and reliability of
98 instruments, should have been undertaken by systematic review[13]. However, no systematic
99 review synthesized evidence on the critical appraisal of the measurement properties of
100 generic and condition-specific HRQOL instruments that have been validated for use in
101 patients with thyroid disease, especially for hypothyroidism, graves' ophthalmopathy and
102 thyroid cancer.

103

104 The aim of this paper was to conduct a systematic literature review on the
105 measurement properties of standardized HRQOL instruments for thyroid disease and to make
106 recommendations on the most appropriate HRQOL instrument(s) for a specific type of
107 thyroid disease through collective evidence from previous studies.

108

109 **Methods**

110

111 Literature Search Methods

112

113 *Search Engines and Strategies*

114

115 Systematic literature searches were conducted in databases of PubMed, Web of
116 Science using Web of Knowledge platform, Embase and MEDLINE using OVID searching
117 platform, to identify studies that investigated the HRQOL of patients with thyroid diseases.
118 The Medical Subject Heading (MESH) 'quality of life' term was combined with 'thyroid

119 disease', 'thyroid cancer', 'thyroid neoplasm', 'thyroid carcinoma' or 'hypothyroidism' for
120 our target population. Studies were limited to English language, and the years between 1993
121 and 2015. The earliest year that assessed the HRQOL of thyroid dysfunction was published in
122 1994[14] so as to limit the starting year of searching as 1993. Systematic searches were
123 conducted in May 1st, 2015 with electronic search strategies shown in Appendix 1. No
124 additional hand search was done. After the initial check for duplicated articles, the abstracts
125 of remaining articles were screened to rule out the literature reviews, meta-analyses, clinical
126 guidelines, study protocols, editorials, letters, commentaries, case reports, and conference
127 proceedings that were not recognized as original articles. Articles were also excluded if no
128 abstract was available.

129

130 *Inclusion and Exclusion Criteria*

131

132 The eligibility criteria of studies were 1) to involve original articles 2) to use
133 standardized self-reported HRQOL instruments with items rating on point Likert scales or on
134 visual analogue scales, 3) to carry out in human subjects, 4) to evaluate the measurement
135 properties of HRQOL instruments referring to a study population consisting of thyroid
136 disease patients, and 5) to develop HRQOL instruments referring to a study population
137 consisting of thyroid disease patients. Articles without available full-text were excluded.
138 Instruments that focus on at least one general concept (i.e. general health, physical
139 functioning, psychological functioning and social functioning) of HRQOL were included;
140 while instruments measured symptom scales only were excluded. Two reviewers (CW and
141 BL) independently screened the eligibility criteria of study titles, abstracts, selected full-texts,
142 and reference lists of the studies retrieved by the literature search. To standardize the
143 appraisal criteria amongst the two reviewers, the methodological quality of the included

144 studies was assessed using the Consensus-based Standards for the selection of health
145 Measurement Instruments (COSMIN) checklist[13, 15-17]. This checklist was previously
146 adopted in the evaluation of cancer-specific HRQOL instruments[18-21]. Assessment of the
147 methodological quality per property was performed by two reviewers independently.
148 Disagreements regarding the procedures of database search, study selection and eligibility
149 were resolved by discussion.

150

151 *Quality Criteria of Measurement Properties*

152

153 According to the COSMIN taxonomy and definitions[15], the review evaluated the
154 nine internationally agreed measurement properties including: 1) internal consistency, 2)
155 reliability, 3) measurement error, 4) content validity, 5) structural validity, 6) hypothesis
156 testing, 7) cross-cultural validity, 8) criterion validity, and 9) responsiveness. The first three
157 measurement properties are in the subset of reliability category, whereas the subsequent six
158 measure properties are contained in the validity category. Internal consistency, referring to
159 the degree of the interrelatedness among the items within subscales or domains, was
160 supported if the Cronbach's alpha was equal to or greater than 0.70 and the factor analysis
161 was conducted with adequate sample size for the support of uni-dimensionality of the scales.
162 Reliability was supported if the test-retest reliability coefficient represented by intraclass
163 correlation coefficient (ICC) and weighted kappa was equal to or greater than 0.70 between
164 two administrations over short period of time among subjects with stable health condition.
165 Measurement error was considered adequate if the smallest detectable change was smaller
166 than the minimal important change, or if the minimal important change reached the limits of
167 agreement. Content validity refers to the extent to which the instrument includes the most

168 relevant and important aspects of the HRQOL concept. Construct validity was assessed by
169 testing a priori hypotheses specific to the expected correlations between scores representing
170 similar concepts or expected differences in scores between known groups. Criterion validity
171 was considered to be present if the gold standard for HRQOL measure existed as the full-
172 length version and tested for the correlations with the shortened version of the instrument.
173 Responsiveness was examined using different statistics to detect changes in the construct
174 being measured over time.

175

176 *Data Synthesis on Methodological Quality Evaluation*

177

178 For each study, each measurement property was rated as ‘adequate’ (+, positive sign)
179 or ‘not adequate’ (-, negative sign) if the quality criterion was met or was not met for each
180 measurement property respectively. If the information given to the measurement property
181 was unclear or ambiguous, it was rated as ‘doubtful’ (?). Given no information was found on
182 that measurement property, zero (0) rating was assigned to that quality assessment. The
183 measurement properties of HRQOL instruments were evaluated based on the explicit quality
184 criteria proposed by Terwee et al.[12]. A summary of the quality criteria for measurement
185 properties of HRQOL instruments is presented in the Appendix 2. To consolidate the grading
186 of measurement properties of multiple instruments, the overall rating for a measurement
187 property was synthesized by taking the quality ratings of each measurement study,
188 consistency of results between studies, and its evidence level for measurement properties.
189 One of the five possible rating options representing ‘strong’ (+++ or ---), ‘moderate’ (++ or --
190), ‘limited’ (+ or -), ‘conflicting’ (+/-) or ‘unknown’ (?) were assigned if the measurement

191 property of instrument was graded at least one. Rating summary of the overall levels of
192 evidence for the quality of each measurement property is displayed in Appendix 3.

193

194 **Results**

195

196 Figure 1 shows the process of literature identification, screening for eligibility and
197 selection of studies during the literature search presented in a Preferred Reporting Items for
198 Systematic Reviews and Meta-Analyses (PRISMA) flow diagram[22]. The literature search
199 was completed in May 2015 and identified a total of 2108 potentially relevant studies
200 (PubMed: 266; Web of Science: 545; MEDLINE: 486; Embase: 811) that met the searching
201 criteria in four bibliographic databases. After the removal of duplicated (n=835) and non-
202 original articles (n=530) by abstract screening, the abstract content of 743 studies were
203 reviewed for eligibility. Twenty-four articles were reviewed in full text for eligibility. We
204 included 23 studies which investigated the measurement properties of HRQOL instruments in
205 patients with thyroid disease inclusive of hypothyroidism, hyperthyroidism, Graves'
206 ophthalmopathy (GO) or thyroid cancer.

207

208 This review identified 14 condition-specific instruments: two instruments in the
209 European Organization for Research and Treatment of Cancer (EORTC) module (QLQ-C30
210 and QLQ-H&N35)[23], three instruments in Health Psychology Research Group and nine
211 instruments in other modules. The 84-item thyroid-specific patient-reported outcome measure
212 (ThyPRO)[24-29] was the most frequently evaluated HRQOL condition-specific instrument,
213 followed by the 16-item Graves' ophthalmopathy specific Quality of Life (GO-QOL)[30-33],
214 20-item Thyroid-Dependent Quality of Life Questionnaire (ThyDQoL)[34-36] and 11-item

215 Thyroid Treatment Satisfaction Questionnaire (ThyTSQ)[34, 36, 37]. Table 1 illustrates the
216 general characteristics of the evaluated HRQOL instruments. A vast majority (8/14, 57.1%)
217 of the HRQOL instruments evaluated were for thyroid dysfunctions (3 for
218 hypothyroidism[34-39], 1 for hyperthyroidism[40] and 4 for GO[30-33, 38, 39, 41, 42]). One
219 was for benign thyroid disease[24-29] and another one was for awaiting thyroidectomy[43])
220 whereas about one-seventh (2/14, 14.3%) was designed for thyroid cancer patients[44, 45]
221 with one for cancer survivors exclusively. The remainders were designed for a wider scope of
222 target population such as head and neck cancer (n=1)[23] and general cancer (n=1)[23].
223 About two-third (9/14, 64.3%) of evaluated HRQOL instruments were developed in
224 European countries, and one-third (5/14, 35.7%) in North America. The total number of items
225 varied from 3[38, 39] to 105[41] whilst the number of subscales or domains varied from zero
226 (single-item questions only)[38, 39] to 23[41].

227

228 *Characteristics of included instruments*

229 Characteristics of 23 eligible articles included in this review are summarized in Table
230 2. Articles were predominately conducted in countries located in Europe[24-32, 34-37, 40, 45]
231 and North America[38, 41-44], and the only one was conducted in Asia[39]. Most of the
232 instruments were evaluated in the language versions of English[34-38, 43, 44], Danish[24-29]
233 and Dutch[30-32, 40, 45].

234

235 *Methodological Quality of Each Study*

236

237 Eligible studies were weighted on nine measurement properties based on the aforesaid
238 criteria on a 4-point Likert scale: ‘poor’, ‘fair’, ‘good’ and ‘excellent’. Table 3 summarizes
239 the details of methodological quality of measurement properties evaluated in each study.
240 Most studies evaluated the properties of hypothesis testing and internal consistency, whereas
241 the criterion validity, responsiveness and measurement error were not often informed. Of
242 those properties evaluated, ‘excellent’ methodological quality was found in internal
243 consistency, content validity, structural validity and hypothesis testing. Content validity was
244 rated the best as seven out of eight studies were weighted as ‘excellent’. Most of the studies
245 were rated as ‘good’ or ‘excellent’ methodological quality in structural validity (9/13, 69.2%),
246 hypothesis testing (14/24, 58.3%) and responsiveness (2/2, 100%). Less than half of the
247 studies were rated as ‘good’ or ‘excellent’ methodological quality in internal consistency
248 (8/17, 47.1%), reliability (2/5, 40%), and cultural validity (2/5, 40%).

249

250 *Overall quality of measurement properties*

251

252 The overall levels of evidence per measurement property and HRQOL instrument are
253 reported in Table 4. None of the HRQOL instruments has been assessed and rated on all nine
254 measurement properties recommended by the quality assessment criteria. Most of instruments
255 assessed overall levels of evidence in internal consistency, content validity and hypothesis
256 testing. Evidence of criterion validity was unknown. Measurement error was only assessed in
257 GO-QOL instrument[31]; whereas the criterion validity was only assessed in Thyroid eye
258 disease quality of life (TED-QOL)[38, 39] in which no convincing argument that gold
259 standard is “gold” and therefore rated as unknown. Among 14 evaluated instruments,
260 ThyPRO, GO-QOL and ThyTSQ achieved positive evidence on six, six and five

261 measurement properties, respectively. Two instruments, Underactive Thyroid Symptom
262 Rating Questionnaire (ThySRQ)[35, 36] and THYCA-QoL[45], achieved moderate evidence
263 of inadequate structural validity.

264

265 **Discussions**

266

267 Despite the need for high-quality thyroid-specific HRQOL instruments, to our
268 knowledge, no previous studies have ever reviewed and compared the measurement property
269 of generic HRQOL instruments. This systematic review evaluated the measurement
270 properties of 14 standardized HRQOL instruments used in thyroid diseases patients among 23
271 eligible studies identified in the full-text assessment stage. The most frequently evaluated
272 HRQOL instrument was the ThyPRO[24-29] instrument, with six studies described the
273 measurement properties. ThyPRO demonstrated strong adequate evidence in internal
274 consistency, content validity, and structural validity, and moderate adequate evidence in
275 hypothesis testing and cross-cultural validity. Five out of nine measurement properties were
276 evaluated in ThyPRO instrument, reflecting the most comprehensive measurement property
277 assessment among HRQOL instruments for use in thyroid disease patients. When assessing
278 HRQOL of GO patients, the GO-QOL[30-33] instrument was rated positive evidence in six
279 of nine measurement properties. In light of the HRQOL assessment of patients with
280 hypothyroidism, the ThyTSQ instrument was rated moderate to strong positive evidence in
281 five out of nine measurement properties. Overall measurement property of ThyTSQ was
282 superior to that of ThyDQoL and ThySRQ instruments. In overall, the highest number of
283 positive sign as an indication of satisfactorily adequate measurement property was obtained

284 in the following three HRQOL instruments: ThyPRO (+: 15), GO-QOL (+:14) and ThyTSQ
285 (+: 13) instruments.

286

287 Notwithstanding, measurement properties of included instruments were evaluated in
288 at most two countries, limiting the transferability and wide applicability of instruments across
289 continents. For instance, ThyTSQ that was recommended for use in hypothyroidism was
290 validated in the UK and Germany. However, researchers should interpret these results with
291 caution as it was not validated in other languages and therefore, the non-English and non-
292 German versions would require further evaluation. While ThyPRO yielded the best overall
293 performance in measurement properties among existing instruments, it was only validated in
294 Denmark. ThyPRO in language versions other than Danish speaking patients has not been
295 evaluated, and such results may over-represent evidence on the excellent quality and
296 performance of ThyPRO instrument in Danish patients with thyroid disease. However, the
297 latest published evidence[46] filled the knowledge gap on cross-cultural validity from an
298 international study in which ThyPRO was assessed in seven countries. Despite its strengths in
299 measurement properties, its length may be an administrative burden to respondents calling for
300 development of an abbreviated 39-item version of the ThyPRO[47].

301

302 Development of instruments in EORTC group has involved multiple countries,
303 following their standard module development guidelines. However, other instruments have
304 not been examined in multiple countries altogether in the stage of content analysis, item
305 selection and piloting testing. Involvement of patients from a range of countries set out to
306 facilitate the back-and-forth evolution through qualitative assessments in early stage of
307 development of HRQOL instrument, which may be certainly cultural-specific and subject to

308 linguistic modification. Furthermore, disparities in sampling of thyroid disease patients are of
309 concerns. Two studies[23, 44] (EORTC and MDASI-THY instruments) had recruited
310 considerably more male patients, of those samples may not be representative of thyroid
311 cancer with predominant female population. Moreover, hypothyroidism patients in two
312 studies[35, 37] (ThyDQoL, ThyTSQ and ThySRQ instruments assessed) were not free of
313 thyroid cancer due to the fact that thyroid cancer patients undergoing treatment may cause
314 hypothyroidism complications. HRQOL measures in those studies may be partly explained
315 by their experiences from thyroid cancer, not merely from hypothyroidism.

316

317 For thyroid cancers, only two instruments (MDASI-THY[44] and THYCA-QoL[45])
318 were identified in this review. Although THYCA-QoL had more positive signs than MDASI-
319 THY, overall measurement property standard of the former instrument was not superior to
320 that of latter instrument due to negative evidence of structural validity as a result of <50%
321 variance explained by 24 items of THYCA-QoL instrument. Collectively, no existing
322 instruments targeted to thyroid cancer had adequate level of positive evidence on their
323 measurement properties. The EORTC and Functional Assessment of Chronic Illness Therapy
324 (FACIT) groups are two core measurement tools of cancer-specific HRQOL used in clinical
325 trials. Although a few studies[7, 48, 49] measured HRQOL using FACIT instruments, the
326 measurement properties of instruments in FACIT measurement system was not assessed for
327 thyroid cancer patient. In FACIT measurement system based on official manual[50] and
328 website[51], to date, it was lacking in specific modules for measurement of thyroid-related
329 functioning, symptom problems, and quality of life. Likewise in EORTC module[52], the
330 development of 'QLQ-THY' is currently underway to meet a pressing need for available use
331 in thyroid cancer patients exclusively, although one study[23] has demonstrated evidence of
332 using the two instruments in EORTC group in patients inclusive of thyroid cancer. Thus, the

333 newly-developed QLQ-H&N35 instrument was recently superseded by the updated and
334 revised QLQ-H&N43[53] instrument which explicitly ruled out thyroid cancer as target
335 population of instrument. This review urged to benchmark on COSMIN checklist for high-
336 quality reporting standard in evaluative studies of these upcoming HRQOL instruments.

337

338 *Limitations*

339

340 Several limitations of this review should be noted. Firstly, the inclusions of full
341 papers within the scope of MESH terms during database searching have results in relevant
342 instruments being missed. Discrepancies in searched records between databases reflected the
343 missing of relevant studies in some databases. The plausible explanation was that only MESH
344 term ‘quality of life’ was used for searching the HRQOL studies in PubMed database,
345 yielding a lower number of records found in PubMed compared to other databases. Several
346 long-lasting cancer-specific HRQOL instruments targeted to general cancer (e.g. FACT-G),
347 hypothyroidism (e.g. Chronic Thyroid Questionnaire) and thyroid cancer (e.g. QOL-Thyroid)
348 were not covered at the stage of database search. QOL-Thyroid instrument was developed by
349 City of Hope National Medical Center[7], and dropped from this review because
350 measurement properties were not reported in a full paper. Thus, Chronic Thyroid
351 Questionnaire[14] was dropped due to absence of measurement properties described in the
352 abstract. Secondly, the review was not decisive to recommend one single thyroid-specific
353 HRQOL instrument for the use in patients with thyroid disease, regardless of diagnosis with
354 thyroid cancer or not. Choices of instruments were recommended with respect to
355 hypothyroidism, GO and benign thyroid disease. There were no available instruments

356 specific to HRQOL measurement in patients with different types of thyroid disease, i.e.
357 euthyroid and thyroid nodules.

358

359 **Conclusions**

360

361 After reviewing the current literature and critically examining published HRQOL
362 instruments, the ThyPRO, GO-QOL and ThyTSQ were the three with the greatest number of
363 positive ratings according to nine quality assessment criteria of measurement properties. The
364 ThyPRO was recommended to assess HRQOL in patients with benign thyroid disease whilst
365 the overall measurement properties of GO-QOL and ThyTSQ were satisfactory in measuring
366 HRQOL in hypothyroidism and GO patients, respectively. Therefore, this review emphasized
367 the need for high-quality and standard reporting of measurement properties of the existing
368 and newly-developed HRQOL instruments for thyroid dysfunctions or cancers.

369

370 **Acknowledgements: None**

371

372 **Funding: None**

373

374 **Conflict of interest statement: None declared**

375

376 **Abbreviations:** HRQOL=health-related quality of life; MESH=Medical Subject Heading;
377 COSMIN=Consensus-based Standards for the selection of health Measurement Instruments;
378 PRISMA=Preferred Reporting Items for Systematic Reviews and Meta-Analyses;
379 ICC=intraclass correlation coefficient; ThyDQoL=Thyroid-Dependent Quality of Life

Abbreviated Title: Recommendation for HRQOL Instruments in Thyroid

380 Questionnaire; ThySRQ=Underactive Thyroid Symptom Rating Questionnaire;
381 ThyTSQ=Thyroid Treatment Satisfaction Questionnaire; HCQ=Hyperthyroidism Complaint
382 Questionnaire; TED-QOL=Thyroid eye disease quality of life; GO-QOL=Graves'
383 ophthalmopathy specific Quality of Life; GO-QLS=Graves Ophthalmopathy Quality-of-Life
384 Scale; NEI VFQ-25=25-item National Eye Institute Visual Function Questionnaire; WSCI-
385 T=Western Surgical Concern Inventory–Thyroid; MDASI-THY=M. D. Anderson Symptom
386 Inventory Thyroid Cancer module; EORTC=European Organization for Research and
387 Treatment of Cancer; QLQ=Quality-of-Life Questionnaire; ThyPRO=thyroid-specific
388 patient-reported outcome measure;

389

390 **Reference**

- 391 [1] Madariaga AG, Palacios SS, Guillén-Grima F, Galofré JC. The Incidence and Prevalence of
392 Thyroid Dysfunction in Europe: A Meta-Analysis. *The Journal of Clinical Endocrinology &*
393 *Metabolism*. 2014;99:923-31.
- 394 [2] Morris LGT, Sikora AG, Tosteson TD, Davies L. The Increasing Incidence of Thyroid Cancer:
395 The Influence of Access to Care. *Thyroid*. 2013;23:885-91.
- 396 [3] Lang BHH, Wong CK, Chan CT. Initial attributable cost and economic burden of clinically-
397 relevant differentiated thyroid cancer: A health care service provider perspective. *European Journal of*
398 *Surgical Oncology (EJSO)*. 2015;41:758-65.
- 399 [4] Leese GP, Flynn RV, Jung RT, MacDonald TM, Murphy MJ, Morris AD. Increasing prevalence
400 and incidence of thyroid disease in Tayside, Scotland: the Thyroid Epidemiology Audit and Research
401 Study (TEARS). *Clinical Endocrinology*. 2008;68:311-6.
- 402 [5] Links TP, van Tol KM, Jager PL, Plukker JTM, Piers DA, Boezen HM, et al. Life expectancy in
403 differentiated thyroid cancer: a novel approach to survival analysis. *Endocrine-Related Cancer*.
404 2005;12:273-80.
- 405 [6] Wilson IB, Cleary PD. Linking Clinical Variables With Health-Related Quality of Life. *JAMA*.
406 1995;273:59-65.
- 407 [7] Dow KH, Ferrell BR, Anello C. Quality-of-Life Changes in Patients with Thyroid Cancer After
408 Withdrawal of Thyroid Hormone Therapy. *Thyroid*. 1997;7:613-9.
- 409 [8] Razvi S, McMillan CV, Weaver JU. Instruments used in measuring symptoms, health status and
410 quality of life in hypothyroidism: a systematic qualitative review. *Clinical Endocrinology*.
411 2005;63:617-24.
- 412 [9] Watt T, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, Bjorner JB, et al. Quality of life
413 in patients with benign thyroid disorders. A review. *European Journal of Endocrinology*.
414 2006;154:501-10.
- 415 [10] Husson O, Haak HR, Oranje WA, Mols F, Reemst PH, van de Poll-Franse LV. Health-related
416 quality of life among thyroid cancer survivors: a systematic review. *Clinical Endocrinology*.
417 2011;75:544-54.
- 418 [11] Elberling T, Rasmussen A, Feldt-Rasmussen U, Hording M, Perrild H, Waldemar G. Impaired
419 health-related quality of life in Graves' disease. A prospective study. *European Journal of*
420 *Endocrinology*. 2004;151:549-55.
- 421 [12] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria
422 were proposed for measurement properties of health status questionnaires. *Journal of Clinical*
423 *Epidemiology*. 2007;60:34-42.
- 424 [13] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the
425 methodological quality in systematic reviews of studies on measurement properties: a scoring system
426 for the COSMIN checklist. *Qual Life Res*. 2012;21:651-7.
- 427 [14] Jaeschke R, Guyatt G, Cook D, Harper S, Gerstein HC. Spectrum of quality of life impairment in
428 hypothyroidism. *Qual Life Res*. 1994;3:323-7.
- 429 [15] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN
430 study reached international consensus on taxonomy, terminology, and definitions of measurement
431 properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*.
432 2010;63:737-45.
- 433 [16] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN
434 checklist for assessing the methodological quality of studies on measurement properties of health
435 status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539-49.
- 436 [17] Mokkink L, Terwee C, Knol D, Stratford P, Alonso J, Patrick D, et al. The COSMIN checklist
437 for evaluating the methodological quality of studies on measurement properties: A clarification of its
438 content. *BMC Medical Research Methodology*. 2010;10:22.
- 439 [18] Lockett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, et al. Choosing between the
440 EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical
441 research: issues, evidence and recommendations. *Annals of Oncology*. 2011;22:2179-90.

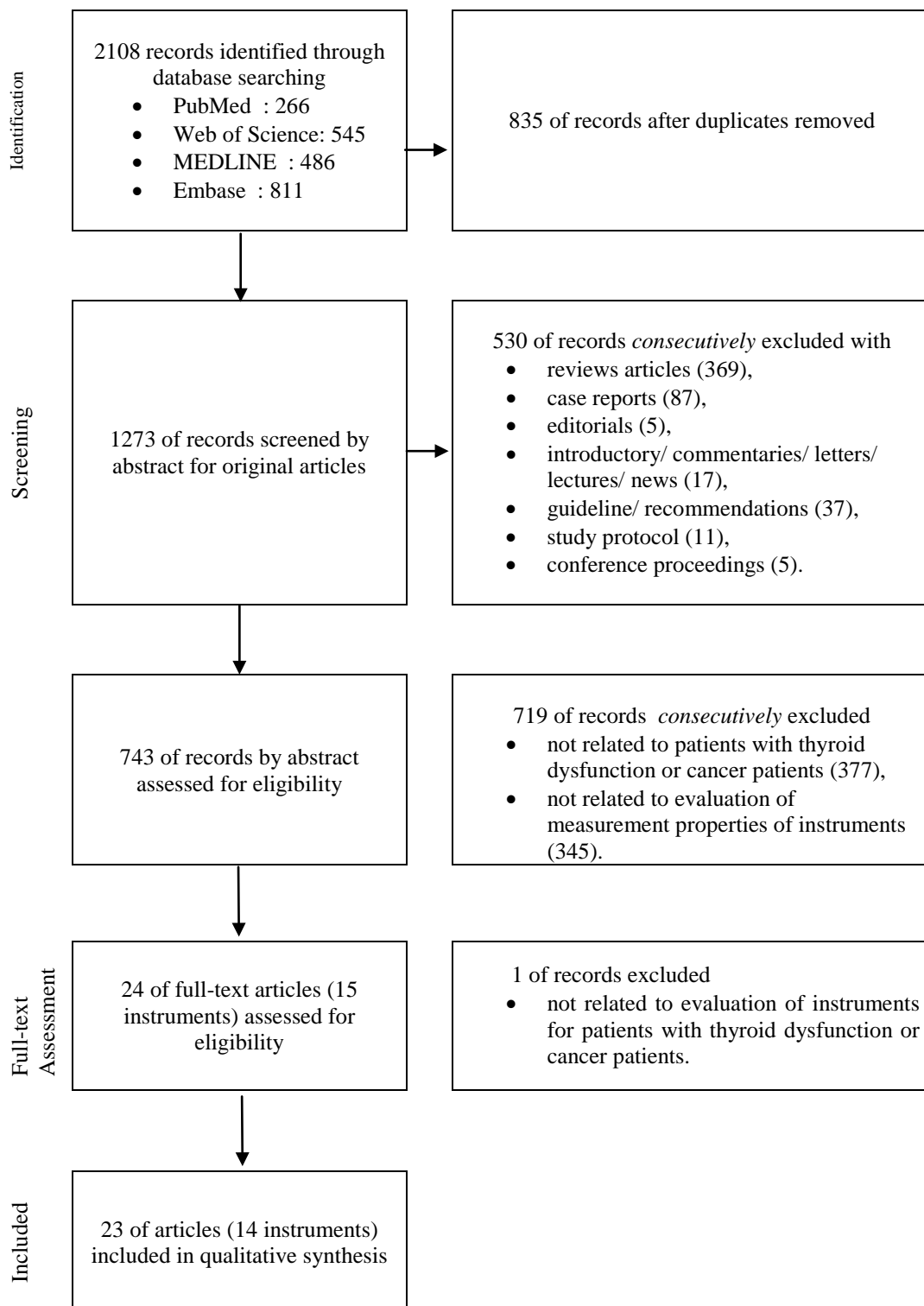
- 442 [19] Wong CKH, Chen J, Yu CLY, Sham M, Lam CLK. Systematic review recommends the
443 European Organization for Research and Treatment of Cancer colorectal cancer-specific module for
444 measuring quality of life in colorectal cancer patients. *Journal of Clinical Epidemiology*.
445 2015;68:266-78.
- 446 [20] Wheelwright S, Darlington A-S, Hopkinson J, Fitzsimmons D, White A, Johnson C. A
447 systematic review of health-related quality of life instruments in patients with cancer cachexia.
448 *Support Care Cancer*. 2013;21:2625-36.
- 449 [21] Pusic A, Cemal Y, Albornoz C, Klassen A, Cano S, Sulimanoff I, et al. Quality of life among
450 breast cancer patients with lymphedema: a systematic review of patient-reported outcome instruments
451 and outcomes. *J Cancer Surviv*. 2013;7:83-92.
- 452 [22] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews
453 and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology*. 2009;62:1006-12.
- 454 [23] Carrillo JF, Ortiz-Toledo MA, Salido-Noriega Z, Romero-Ventura NB, Ochoa-Carrillo FJ,
455 Onate-Ocana LF. Validation of the Mexican Spanish version of the EORTC QLQ-H&N35 instrument
456 to measure health-related quality of life in patients with head and neck cancers. *Annals of Surgical
457 Oncology*. 2013;20:1417-26.
- 458 [24] Watt T, Rasmussen AK, Groenvold M, Bjorner JB, Watt SH, Bonnema SJ, et al. Improving a
459 newly developed patient-reported outcome for thyroid patients, using cognitive interviewing. *Qual
460 Life Res*. 2008;17:1009-17.
- 461 [25] Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedus L, et al. Establishing
462 construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): An initial
463 examination. *Qual Life Res*. 2009;18:483-96.
- 464 [26] Watt T, Hegedus L, Groenvold M, Bjorner JB, Rasmussen AK, Bonnema SJ, et al. Validity and
465 reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *European Journal of
466 Endocrinology*. 2010;162:161-7.
- 467 [27] Watt T, Cramon P, Hegedus L, Bjorner JB, Bonnema SJ, Rasmussen AK, et al. The thyroid-
468 related quality of life measure ThyPRO has good responsiveness and ability to detect relevant
469 treatment effects. *Journal of Clinical Endocrinology & Metabolism*. 2014;99:3708-17.
- 470 [28] Watt T, Groenvold M, Deng N, Gandek B, Feldt-Rasmussen U, Rasmussen AK, et al.
471 Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. *Health and
472 Quality of Life Outcomes*. 2014;12.
- 473 [29] Watt T, Groenvold M, Hegedus L, Bonnema SJ, Rasmussen AK, Feldt-Rasmussen U, et al. Few
474 items in the thyroid-related quality of life instrument ThyPRO exhibited differential item functioning.
475 *Qual Life Res*. 2014;23:327-38.
- 476 [30] Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease
477 specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. *Br J
478 Ophthalmol*. 1998;82:773-9.
- 479 [31] Terwee CB, Gerding MN, Dekker FW, Prummel MF, van der Pol JP, Wiersinga WM. Test-retest
480 reliability of the GO-QOL: a disease-specific quality of life questionnaire for patients with Graves'
481 ophthalmopathy. *Journal of Clinical Epidemiology*. 1999;52:875-84.
- 482 [32] Terwee CB, Dekker FW, Mourits MP, Gerding MN, Baldeschi L, Kalmann R, et al.
483 Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life
484 questionnaire (GO-QOL) after different treatments. *Clinical Endocrinology*. 2001;54:391-8.
- 485 [33] Choi YJ, Lim HT, Lee SJ, Lee SY, Yoon JS. Assessing Graves' ophthalmopathy-specific quality
486 of life in Korean patients. *Eye (London, England)*. 2012;26:544-51.
- 487 [34] McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU. Design of new questionnaires to
488 measure quality of life and treatment satisfaction in hypothyroidism. *Thyroid*. 2004;14:916-25.
- 489 [35] McMillan C, Bradley C, Razvi S, Weaver J. Evaluation of new measures of the impact of
490 hypothyroidism on quality of life and symptoms: the ThyDQoL and ThySRQ. *Value in Health*.
491 2008;11:285-94.
- 492 [36] Quinque EM, Villringer A, Kratzsch J, Karger S. Patient-reported outcomes in adequately treated
493 hypothyroidism - insights from the German versions of ThyDQoL, ThySRQ and ThyTSQ. *Health &
494 Quality of Life Outcomes*. 2013;11:68.
- 495 [37] McMillan C, Bradley C, Razvi S, Weaver J. Psychometric evaluation of a new questionnaire
496 measuring treatment satisfaction in hypothyroidism: The ThyTSQ. *Value in Health*. 2006;9:132-9.

- 497 [38] Fayers T, Dolman PJ. Validity and reliability of the TED-QOL: a new three-item questionnaire to
498 assess quality of life in thyroid eye disease. *Br J Ophthalmol*. 2011;95:1670-4.
- 499 [39] Son BJ, Lee SY, Yoon JS. Evaluation of thyroid eye disease: quality-of-life questionnaire (TED-
500 QOL) in Korean patients. *Canadian journal of ophthalmology Journal canadien d'ophtalmologie*.
501 2014;49:167-73.
- 502 [40] Fahrenfort JJ, Wilterdink AM, van der Veen EA. Long-term residual complaints and
503 psychosocial sequelae after remission of hyperthyroidism. *Psychoneuroendocrinology*. 2000;25:201-
504 11.
- 505 [41] Patrick Yeatts R. Quality of life in patients with graves ophthalmopathy. *Transactions of the*
506 *American Ophthalmological Society*. 2005;103:368-411.
- 507 [42] Bradley EA, Sloan JA, Novotny PJ, Garrity JA, Woog JJ, West SK. Evaluation of the National
508 Eye Institute visual function questionnaire in Graves' ophthalmopathy. *Ophthalmology*.
509 2006;113:1450-4.
- 510 [43] Abdul-Sater L, Henry M, Majdan A, Mijovic T, Franklin JH, Brandt MG, et al. What Are
511 Thyroidectomy Patients Really Concerned About? *Otolaryngology-Head and Neck Surgery*.
512 2011;144:685-90.
- 513 [44] Gning I, Trask PC, Mendoza TR, Harle MT, Gutierrez KA, Kitaka SA, et al. Development and
514 Initial Validation of the Thyroid Cancer Module of the M. D. Anderson Symptom Inventory.
515 *Oncology*. 2009;76:59-68.
- 516 [45] Husson O, Haak HR, Mols F, Nieuwenhuijzen GA, Nieuwlaat WA, Reemst PH, et al.
517 Development of a disease-specific health-related quality of life questionnaire (THYCA-QoL) for
518 thyroid cancer survivors. *Acta Oncologica*. 2013;52:447-54.
- 519 [46] Watt T, Barbesino G, Bjorner J, Bonnema S, Bukvic B, Drummond R, et al. Cross-cultural
520 validity of the thyroid-specific quality-of-life patient-reported outcome measure, ThyPRO. *Qual Life*
521 *Res*. 2015;24:769-80.
- 522 [47] Watt T, Bjorner JB, Groenvold M, Cramon P, Winther KH, Hegedüs L, et al. Development of a
523 Short Version of the Thyroid-Related Patient-Reported Outcome ThyPRO. *Thyroid*. 2015;25:1069-79.
- 524 [48] Chow SM, Au KH, Choy TS, Lee SH, Yeung NY, Leung A, et al. Health-Related Quality-of-
525 Life Study in Patients With Carcinoma of the Thyroid After Thyroxine Withdrawal for Whole Body
526 Scanning. *The Laryngoscope*. 2006;116:2060-6.
- 527 [49] Taieb D, Baumstarck-Barrau K, Sebag F, Fortanier C, De Micco C, Loundou A, et al. Health-
528 related quality of life in thyroid cancer patients following radioiodine ablation. *Health and Quality of*
529 *Life Outcomes*. 2011;9:33.
- 530 [50] Cella D. *Manual of the Functional Assessment of Chronic Illness Therapy (FACIT)*
531 *Measurement System*. Version 4. Evanston IL: Center on Outcomes, Research & Education (CORE),
532 Evanston Northwestern Healthcare and Northwestern University.; 1997.
- 533 [51] FACIT.org. *FACIT Questionnaires*. 2015.
- 534 [52] EORTC QOL Module for Thyroid Cancer (QLQ-THY). EORTC Quality of Life Department.
- 535 [53] Singer S, Araújo C, Arraras JI, Baumann I, Boehm A, Brokstad Herlofson B, et al. Measuring
536 quality of life in patients with head and neck cancer: Update of the EORTC QLQ-H&N Module,
537 Phase III. *Head & Neck*. 2015;37:1358-67.

538

539

540 Figure 1. PRISMA Flow Diagram of the literature search and selection process
541



542

543

Abbreviated Title: Recommendation for HRQOL Instruments in Thyroid

Table 1. General Characteristics of the HRQOL Instruments and Their Evaluated Measurement Properties in Eligible Articles of Psychometric Evaluations

Instrument	Target Disease Population	Country of Development	Number of Items	Number of Subscales/Domains	Number of Single Item	Response Options	Reference
ThyDQoL	Hypothyroidism	UK	20	18	2	5-point (1 overall item) 7-point (1 overall item) 13-point (18 items)	[34-36]
ThySRQ	Hypothyroidism	UK	15	15	0	4-point	[35, 36]
ThyTSQ	Hypothyroidism	UK	11	2	0	7-point	[34, 36, 37]
HCQ	Hyperthyroidism	Netherlands	31	4	5	2-point (present/not)	[40]
WSCI-T	Pre-thyroidectomy	Canada	18	3	0	9-point	[43]
GO-QOL	Graves' ophthalmopathy	Netherlands	16	2	0	3-point	[30-33]
GO-QLS	Graves' ophthalmopathy	USA	105	23	0	2-/3-/5-/6-/7-/11-point	[41]
NEI VFQ-25	Graves' ophthalmopathy	USA	25	12	0	5-point (13 items) 6-point (12 items)	[42]
TED-QOL	Graves' ophthalmopathy	Canada	3	0	3	11-point	[38, 39]
EORTC QLQ-C30 (Version 3)	All cancer	European countries	30	9	6	4-point (28 items) 7-point (2 global items)	[23]
EORTC QLQ-H&N35	Head and neck cancer	European countries	35	7	11	4-point	[23]
MDASI-THY	Thyroid cancer	USA	25	6	0	11-point	[44]
ThyPRO	Benign thyroid disease	Denmark	84	13	0	5-point	[24-29]
THYCA-QoL	Thyroid cancer survivors	Netherlands	24	7	0	4-point	[45]

Note:

ThyDQoL=Thyroid-Dependent Quality of Life Questionnaire; ThySRQ=Underactive Thyroid Symptom Rating Questionnaire; ThyTSQ=Thyroid Treatment Satisfaction Questionnaire; HCQ=Hyperthyroidism Complaint Questionnaire; TED-QOL=Thyroid eye disease quality of life; GO-QOL=Graves' ophthalmopathy specific Quality of Life; GO-QLS=Graves Ophthalmopathy Quality-of-Life Scale; NEI VFQ-25=25-item National Eye Institute Visual Function Questionnaire; WSCI-T=Western Surgical Concern Inventory–Thyroid; MDASI-THY=M. D. Anderson Symptom Inventory Thyroid Cancer module; EORTC=European Organization for Research and Treatment of Cancer; QLQ=Quality-of-Life Questionnaire; ThyPRO=thyroid-specific patient-reported outcome measure;

Table 2. Description of Eligible Articles

Reference	Evaluated instrument(s)	Year of Publication	Country	Language	Study Population	Sample Size
[34]	ThyDQoL and ThyTSQ	2004	UK	English	Hypothyroidism	38
[35]	ThyDQoL and ThySRQ	2008	UK	English	Hypothyroidism	110
[36]	ThyDQoL, ThyTSQ and ThySRQ	2013	Germany	German	Hypothyroidism	101
[37]	ThyTSQ	2006	UK	English	Hypothyroidism	103
[40]	HCQ	2000	Netherlands	Dutch	Hyperthyroidism	303
[43]	WSCI-T	2011	Canada	English	Pre-thyroidectomy patients	148
[30]	GO-QOL	1998	Netherlands	Dutch	Graves' ophthalmopathy	70
[31]	GO-QOL	1999	Netherlands	Dutch	Graves' ophthalmopathy	93
[32]	GO-QOL	2001	Netherlands	Dutch	Graves' ophthalmopathy	164
[33]	GO-QOL	2012	Korea	Korean	Graves' ophthalmopathy	98
[41]	GO-QLS and NEI VFQ-25	2005	USA	English	Graves' ophthalmopathy	256
[42]	NEI VFQ-25	2006	USA	English	Graves' ophthalmopathy	30
[38]	TED-QOL, GO-QOL and GO-QLS	2011	Canada	English	Graves' ophthalmopathy	100
[39]	TED-QOL	2014	Korea	Korean	Graves' ophthalmopathy	90
[23]	EORTC QLQ-C30 (Version 3) and QLQ-H&N35	2013	Mexico	Spanish	Head and neck cancer including tumor location of thyroid carcinoma invasive to aerodigestive tract	193
[44]	MDASI-THY	2008	USA	English	Thyroid cancer patients undergoing hormonal therapy	60
[24]	ThyPRO	2008	Denmark	Danish	Thyroid disease patients including thyroid swelling and dysfunction of the gland	31
[25]	ThyPRO	2009	Denmark	Danish	Benign thyroid disease	907
[26]	ThyPRO	2010	Denmark	Danish	Benign thyroid disease	907
[28]	ThyPRO	2014	Denmark	Danish	Benign thyroid disease	907
[27]	ThyPRO	2014	Denmark	Danish	Benign thyroid disease	435
[29]	ThyPRO	2014	Denmark	Danish	Benign thyroid disease	838
[45]	THYCA-QoL	2013	Netherlands	Dutch	Thyroid cancer survivor	306

Abbreviated Title: Recommendation for HRQOL Instruments in Thyroid

Table 3. Methodological Quality of Each Study per Measurement Property and HRQOL Instrument

Instrument /Reference	Measurement Properties								
	Internal Consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	Responsiveness
ThyDQoL									
[34]				Excellent					
[35]	Good				Good	Fair			
[36]	Poor ‡				Poor	Good	Poor		
ThySRQ									
[35]	Good			Excellent	Excellent	Fair			
[36]	Good				Good	Good	Good		
ThyTSQ									
[34]				Excellent					
[37]	Excellent				Excellent	Fair			
[36]	Good				Good	Good	Good		
HCQ									
[40]	Poor †					Fair			
WSCI-T									
[43]	Excellent			Excellent	Excellent	Fair			
GO-QOL									
[30]	Poor ‡			Excellent	Poor	Good			
[31]	Fair	Good	Good		Fair	Good			
[32]						Excellent			Excellent
[33]	Poor †					Good			
GO-QLS									
[41]	Poor §				Poor	Good			
NEI VFQ-25									
[42]				Good		Fair			

Abbreviated Title: Recommendation for HRQOL Instruments in Thyroid

TED-QOL						
[38]		Fair			Fair	Fair
[39]			Excellent		Fair	Poor
EORTC QLQ-C30 (V3)						
[23]	Poor †	Fair			Fair	
EORTC QLQ-H&N35						
[23]	Poor †	Fair			Fair	Fair
MDASI-THY						
[44]	Poor †		Excellent		Good	
ThyPRO						
[24]			Excellent			
[25]	Good				Excellent	
[26]		Good			Good	
[28]				Good	Excellent	
[27]						Good
[29]				Excellent	Excellent	
THYCA-QoL						
[45]	Good		Good	Good	Good	

Note:

† No exploratory or confirmatory factor analysis was performed.

‡ A small sample size, defined as sample size was less than the number of items times five, was performed in factor analysis.

§ Internal consistency statistic was not calculated for each subscale separately.

Abbreviated Title: Recommendation for HRQOL Instruments in Thyroid

Table 4. Overall Levels of Evidence per measurement Property and HRQOL Instrument

Instrument	Measurement Properties								Reference
	Internal Consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	
ThyDQoL	++			+++	++	++	?		[34-36]
ThySRQ	++			+++	--	++	++		[35, 36]
ThyTSQ	+++			+++	+++	++	++		[34, 36, 37]
HCQ	?					+			[40]
WSCI-T	+++			+++	+++	+			[43]
GO-QOL	+	++	++	+++	?	+++		+++	[30-33]
GO-QLS	?				?	++			[41]
NEI VFQ-25				++		+			[42]
TED-QOL		+		+++		+	?	?	[38, 39]
EORTC QLQ-C30 (V3)	?	?				+			[23]
EORTC QLQ-H&N35	?	?				+	?		[23]
MDASI-THY	?			+++		++			[44]
ThyPRO	++	++		+++	+++	+++		++	[24-29]
THYCA-QoL	++			++	--	++			[45]

Note: Overall levels of evidence: +++/--- , strong evidence positive/negative result; ++/-- , moderate evidence positive/negative result; +/- , limited evidence positive/negative result; ?, unknown due to poor methodological quality.