Evaluation of new measures of the impact of hypothyroidism on quality of life and symptoms: the ThyDQoL and ThySRQ

Carolyn McMillan PhD[1], Clare Bradley PhD[1], Salman Razvi MD MRCP[2,3], Jolanta Weaver PhD FRCP[2,3]

[1]Royal Holloway, University of London, Egham, Surrey, UK

[2]Department of Endocrinology, Queen Elizabeth Hospital, Gateshead, UK
[3]School of Clinical Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK.

Corresponding Author: Prof Clare Bradley, Professor of Health Psychology, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK

Tel: +44-1784-443708

Fax: +44-1784-471168

E-mail: c.bradley@rhul.ac.uk

Running head: Hypothyroid-specific QoL & symptom measures

ABSTRACT

Objectives: This paper reports the psychometric properties of two new condition-specific questionnaires: 1) the 18-item ThyDQoL individualised measure of perceived impact of hypothyroidism on quality of life (QoL), and 2) the 15-item ThySRQ, in which patients rate symptom bother.

Methods: A cross-sectional survey was conducted of 110 adults with overt and subclinical hypothyroidism, 103 treated with thyroxine. Patients, the majority of whom (81%) were women, were recruited from primary care (57%) and from hospital clinics (43%). The mean age of patients was 55.1 (SD 14.3) years. Respondents rated personally applicable ThyDQoL life domains for importance and impact of hypothyroidism, and ThySRQ symptom bother.

Results: Completion rates were high (>98%). All 18 ThyDQoL domains were rated as negatively impacted by hypothyroidism and important for QoL. The ThyDQoL had high internal consistency reliability [Cronbach's alpha = 0.949, (N = 97)], factor analysis indicated that applicable domains could be combined into an overall Average Weighted Impact score (AWI-18), for which the sample mean, -3.11 (2.2), indicated considerable negative impact of hypothyroidism on QoL (maximum possible range -9 to +3). There is good preliminary evidence to justify shortening the ThyDQoLto 14 domain-specific items. For the ThySRQ Cronbach's alpha was 0.808, (N = 95). Highest symptom bother ratings were for hair problems, weight gain, depression, cold and tiredness.

Conclusions: Both the ThyDQoL and ThySRQ are highly acceptable to patients with hypothyroidism and have good internal consistency reliability. Their sensitivity to change now needs to be evaluated in clinical trials.

INTRODUCTION

Hypothyroidism (whereby the thyroid gland produces insufficient thyroid hormone) is a life-long chronic condition particularly prevalent in women and the elderly [1]. A recent large US survey found that 9.5% of the population had elevated serum levels (>5.1 mIU/L) of thyroid stimulating hormone (TSH), indicative of hypothyroidism (both clinical and subclinical) [2]. The disorder slows the metabolism, with classic symptoms of slowness of movement, tiredness, low energy, change in appearance and voice, weight gain, cold intolerance, poor libido and constipation. Neuromuscular symptoms such as muscle stiffness, cramps and easy fatigability are common. There may be cognitive problems including poor memory, attention deficits, difficulties with calculations, and slow reaction times. Hypothyroidism is also associated with high psychiatric morbidity, in particular depression and paranoid disorder [3].

The multiple aversive symptoms of hypothyroidism, both physiological and psychological, can have a considerable effect on quality of life (QoL). There is currently controversy about appropriate levels of thyroid hormone replacement, and greater understanding is needed for the reasons why some patients feel far from well on thyroid hormone replacement dosage that meets current guidelines [4]. However, psychosocial aspects have hitherto received less attention than clinical aspects [5], and multidimensional assessments of patient-reported outcomes are now needed for clinical trials and routine clinical use with individual hypothyroid patients. A new condition-specific, individualised questionnaire, the Underactive Thyroid-Dependent Quality of Life Questionnaire (ThyDQoL), has recently been designed [6] to measure the impact of hypothyroidism on different aspects of life. The present paper concerns the evaluation of this questionnaire's psychometric properties.

We also introduce a new self-completion measure of symptoms in hypothyroidism, the Underactive Thyroid Symptom Rating Questionnaire (ThySRQ), which assesses the degree to which patients are bothered by applicable symptoms. The ThySRQ was designed simultaneously with the ThyDQoL in interviews with patients [6], but hitherto has not been reported. There are instruments in place for the diagnosis of hypothyroidism on the basis of symptoms [7, 8] but they are for completion by health professionals. Other symptom measures exist [9-14], usually designed specifically for a clinical trial [12-15], but which do not appear to have been validated psychometrically, or designed on the basis of

interviews with patients. Although patients were involved in the identification of important symptoms in the development of the Chronic Thyroid Questionnaire [11], the measure itself does not appear to have undergone validation for completion by patients. Measures of symptom frequency or relative change [12, 13, 15] are not tapping how much 'bothered' a patient is by a symptom: a symptom might occur frequently, but not trouble the patient unduly and vice versa. There is a need for a short, validated self-completion measure of the most common hypothyroid symptoms that has been developed through interviews with patients. The present paper reports the validation of the new ThySRQ measure of perceived symptom severity and the evaluation of its psychometric properties.

METHODS

In this cross-sectional study, hypothyroid patients completed the ThyDQoL and ThySRQ instruments. Forty-seven patients were recruited from the diabetes and endocrinology clinic at Queen Elizabeth Hospital, Gateshead, UK (5% of whom were attending the hospital for their primary hypothyroidism, the remainder for other endocrine conditions including diabetes or secondary hypothyroidism, and some had developed hypothyroidism following treatment of hyperthyroidism and thus they did not have primary hypothyroidism). Sixty-three of the sample were recruited from three local primary care practices in which all patients were approached who fitted the inclusion criteria. The reasons patients were recruited from primary care as well as from a hospital setting were that a) people referred to hospital may have a more severe form of disease, and b) hypothyroidism is primarily managed in primary care and the vast majority of patients are seen there. Thus, combining the two would give a sample that included a substantial number of patients representative of those treated in primary care, as well as ensuring that more severely affected patients were included. The age range was 18+ years (no upper limit), and all patients had a diagnosis of hypothyroidism based on two blood tests taken at least three months apart. People known to have mental health problems likely to render them incapable of understanding and completing the questionnaires were not included. The Gateshead Local Research Ethics Committee gave approval for the research, which was carried out in compliance with the Helsinki declaration.

The ThyDQoL

Design of the ThyDQoL was based on the Audit of Diabetes-Dependent Quality of Life (ADDQoL) [16, 17] and subsequent adaptations of the ADDQoL for people with renal failure [18], macular degeneration [19], and with adult hypopituitarism [20]. The ADDQoL was influenced by work on the development of an interview method of QoL assessment: the Schedule for the Evaluation of Individual Quality of Life [21]. Such individualised measures assess QoL from the viewpoint of individual patients themselves rather than by using externally imposed definitions of good or bad QoL. Thus, the ThyDQoL defines QoL as 'how good or bad you feel your life to be'. The new ThyDQoL measure was developed following interviews with 38 adults with hypothyroidism at three UK centres and, as the measure and its design are fully described elsewhere [6], only a brief description is provided here.

The first two items in the questionnaire provide an overview of the respondent's QoL. Question I (QI,), Present QoL, asks respondents to rate their present QoL on a 7-point scale from excellent to extremely bad. Question II (QII), Thyroid-dependent QoL, asks them to rate what their QoL would be if they did not have hypothyroidism, on a 5-point scale, very much better, much better, a little better, the same, worse, providing an overview measure of impact of hypothyroidism on QoL. There are 18 additional items covering different domains of life (Table 1), in which respondents assess the impact of hypothyroidism on the domain and then rate how important that domain is to their QoL. The ThyDQoL is also individualised by giving respondents the opportunity to indicate whether a particular domain is not applicable (N/A), e.g. work or family life.

Scoring

The domain impact ratings are scored from -3 to +1 (Fig. 1) (from greatest negative impact to positive impact of hypothyroidism on QoL), and importance ratings from 0 to 3 (from not at all important to very important). A weighted domain impact score is obtained by multiplying the domain's impact rating by the corresponding importance rating. An overall score for the questionnaire, the ThyDQoL Average Weighted Impact score (AWI) is obtained by summing all applicable weighted domain scores, before dividing by the number of domains applicable to the individual. Weighted domain scores and the ThyDQoL AWI score range from -9 to +3 (maximum negative to maximum positive weighted impact of hypothyroidism on the individual domain or on the average of all applicable domains). The overview

items are not weighted by importance ratings, but QII, Thyroid-dependent QoL, is scored in the same way as the domains, from -3 to +1. QI, Present QoL, however, is scored +3 to -3 from excellent to extremely bad, so that a higher score indicates better QoL.

Respondents may also indicate any other ways in which they perceive hypothyroidism to affect their QoL in a 'free comments' section at the end of the questionnaire. This allows for the addition of further domains to the questionnaire in the future, if required, as part of its continuing development.

Fig. 1: Example of a ThyDQoL domain item, which has a not applicable response option

4	*Do you have family / relatives?*
	Yes ☐ If *yes*, complete **(a)** and **(b)**.
	No ☐ If *no*, go straight to **Question 5**.

(a)	**If I did <u>not</u> have underactive thyroid, my family life would be:**
	-3* -2 -1 0 1
	☐ ☐ ☐ ☐ ☐
	very much much better a little better the same worse
	better

(b)	**My family life is:**
	3 2 1 0
	☐ ☐ ☐ ☐
	very important important somewhat not at all
			important important

*scores are not shown on the ThyDQoL questionnaire

Fig. 2: Example of a ThySRQ item and an abbreviated form of the 15 symptom questions.

Figure 2

1(a)	Have you felt tired in recent weeks?			Yes ☐	No ☐
(b)	If **yes**, how much does feeling tired bother you?	☐ not at all	☐ a little	☐ quite a bit	☐ very much

*Have you…(*symptom*)… in recent weeks?*

1. …. felt tired …
2. …. gained weight …
3. …. felt colder than other people …
4. …. had constipation …
5. …. had hair problems (e.g. hair loss, coarseness) …
6. …. had skin problems (e.g. dryness, coarseness) …
7. …. had nail problems (e.g. brittleness, flaking) …
8. …. had loss of appetite …
9. …. had hearing problems …
10. …. had voice problems (e.g. hoarseness, huskiness) …
11. …. had speech problems (e.g. slowness, inaccuracy) …
12. …. had memory problems …
13. …. had problems with your concentration …
14. …. felt giddy or dizzy …
15. …. felt depressed or low …

The Underactive Thyroid Symptom Rating Questionnaire (ThySRQ)

When designing the ThyDQoL a number of symptoms were initially included in the measure, on the basis of a review of the literature on hypothyroidism and discussions with clinicians. However, during the early stages of the process of conducting interviews with patients, the majority of these symptoms were found to be unsuitable for inclusion as ThyDQoL domains, being too specific in nature to be important for many aspects of life, e.g. voice problems, or because some patients were unsure whether they were attributable to hypothyroidism. Four symptoms were retained in the ThyDQoL, however, because patients mentioned them frequently and spontaneously as being particularly detrimental to their QoL: energy levels, weight, bodily discomfort and depression. As interviews

progressed it was decided to develop a separate questionnaire, the Underactive Thyroid Symptom Rating Questionnaire (ThySRQ), for completion by patients. ThySRQ design was influenced by design of the Diabetes Symptom Checklist [22] and the Asthma Symptoms Questionnaire [23]. Symptoms selected for inclusion were based on the literature, consultations with endocrinologists, and were the most frequently mentioned symptoms in the interviews. The method used to design the ThySRQ was the same as for the ThyDQoL [6], and the ThySRQ was finalised and piloted in semi-structured interviews with 10 patients conducted at two centres (London and Surrey): 2 men and 8 women, [mean age 49 (SD 14.3), range 32 – 67 years] while the ThyDQoL involved 18 interviews in Gateshead as well as 20 in London and Surrey.

The ThySRQ has 15 symptoms with a 4-point symptom bother scale measuring perceived symptom severity, and is introduced: "This questionnaire asks you about symptoms that can be associated with under-active thyroid and that you may have experienced in recent weeks". Respondents provide a bother rating for applicable symptoms, indicating how much the symptom bothers them from not at all, a little, quite a bit, very much (scoring 0, 1, 2 and 3 respectively). Any patients who report that they do not have a symptom are given a symptom bother rating of zero (not at all bothered). See Fig. 2 for an example of an item, and a list of the 15 symptoms. The term 'in recent weeks', whilst not specific, has been found, during interviews or in questionnaire development work, to be interpreted as meaning the past three to six weeks. Although it would be possible to specify a time period (e.g. the past four weeks), in practice patients are not able to remember so precisely when symptoms occur and such specificity may give an illusion of precision that does not exist.

Statistical Analyses

Normality of distributions was determined through investigation of histograms and z (skew) scores (comparing the skewness value with zero using the z distribution), where z (skew) scores between ±2.58 are indicative of normality [24].

No data from respondents selecting a N/A response option on the ThyDQoL would normally be included in factor and reliability analyses. N/A responses are treated as missing by the statistical

package used (SPSS for Windows, Release 10.0). Furthermore, if the SPSS default of list-wise deletion of missings is used, all cases that have any missing values across all 18 items are lost to analysis, so considerable data could be lost. From 3 to 50% of volunteers selected N/A on items where the option was available, N/A responses were therefore coded as zero for factor and reliability analyses, with pair-wise deletion of data missing for other reasons, as in the original development of the ADDQoL [16]. Coding a N/A response for a particular domain as zero implies that the domain has no impact on QoL. Note that any genuinely missing values (as opposed to N/A responses) are not substituted for in this way. The minimum acceptable Cronbach's alpha coefficient of internal consistency reliability was taken as 0.8 [25] and acceptable corrected item-total correlations were >0.2 [26]. To assess the effects of missing domain data (impact or importance ratings) on the reliability of the ThyDQoL, reliability analyses were run sequentially, deleting the strongest item each time, (i.e. deleting the item having the lowest 'alpha if item deleted' and therefore contributing most to the scale's internal reliability) [27]. Factor structure was explored using Principal Components as an extraction method with Varimax rotation. Loadings greater than 0.3 are, by convention, considered salient and one of 0.6 is considered high [28] but salient loadings were here taken as ≥0.4 to ensure findings were robust. Correlations between questionnaire variables were explored using Pearson's r and Spearman's rho for parametric and non-parametric data respectively.

Construct validity was not specifically assessed, as there were no a priori hypotheses about strength or direction of correlations or subgroup differences in relation to biomedical variables. Sub-group analyses were nevertheless undertaken and explored differences in treatment and degree of hypothyroidism (subclinical vs. overt) using t-tests or Mann-Whitney tests for parametric and non-parametric data respectively, and differences in TSH level (within the reference range, above or below this range) using Univariate Analysis of Variance. Two-tailed significance was applied throughout. To reduce Type 1 error, the Holm's sequential Bonferroni procedure for multiple tests [29] was applied whereby the number of multiple tests divides alpha, but the denominator decreases by one for each significant result.

Initial Analyses To Determine Whether The Four Recruitment Samples Could Be Combined.

There is a possibility that subgroup differences associated with data from the four recruitment sources might create artifactual correlations in the combined sample, if the subgroup mean scores differed consistently across items. This was investigated by converting the scores to standardised z-scores within each subgroup before combining the sub-sets of scores. Such standardisation renders group means identical and thereby removes the possibility of correlations caused by subgroup differences (a method adopted when developing the ADDQoL [16]). Forced one-factor solutions on the raw scores, and then on the z-scores, produced two sets of factor loadings that were then compared using correlation and regression analyses. There was no significant difference between loadings of standardised and raw scores: the correlation of 0.998 was close to 1, the constant was 0.0, and the slope of the regression line (1.001) did not differ significantly from 1, [$t(15) = 60.28$, $p > 0.05$]. Thus, initial analyses demonstrated that the four recruitment groups could be treated as one sample for reliability and factor analyses (for which a larger N is desirable).

RESULTS

The Study Sample And Questionnaire Completion Rates

Of the 128 patients invited to complete the ThyDQoL and ThySRQ, 110 accepted and returned completed questionnaires (participation rates of 90% and 83% at the hospital clinic and primary care respectively). The characteristics (ethnicity and age) of responders were similar to non-responders, but there was a significant sex difference (p = 0.042), as the ratio of non-responders to responders was 0:21 (men) but 18:89 (women). There were four sources of recruitment: hospital clinic (47 patients), and three primary care practices, providing 9, 6, and 48 patients respectively. The mean age of the sample was 55.1 years, and mean duration of hypothyroidism was 8.3 years. The ratio of women to men was 9:2, similar to the sex ratio of thyroid disease found in the general US population in a recent study [30]. The great majority of patients had autoimmune hypothyroidism (N = 98, 89%) of whom 11 patients had subclinical hypothyroidism. The great majority (N = 103, 94%) were receiving thyroxine replacement. Measurements of TSH levels were available for 95 patients (86%) (median 2.4 mu/L), but in the case of free thyroxine (FT4) only 49 patients (45%) (mean 17.4 pmol/L). See Table 2.

Healthcare professionals prompted patients recruited at the hospital to complete any missed items, however, completion rates for the 63 questionnaire sets collected at the primary care clinics (where patients were not prompted) were: 1) ThyDQoL: impact ratings (98.5%), importance ratings (98.1%). The data from two patients (one from the hospital clinic and the other from primary care) were excluded from ThyDQoL psychometric analyses owing to large numbers of missing responses (half or more of the 18 items missing and large numbers of N/A responses), providing a sample of 108; and 2) ThySRQ: 99.9% (parts a and b). The data from all 110 patients were available for ThySRQ analyses.

ThyDQoL

Domains reported as most severely (and negatively) impacted by hypothyroidism were: motivation (mean weighted impact = -4.84), weight (-4.51), and depression (-4.36), (maximum possible range -9 to +3) (Table 1). The majority of respondents (71%) perceived that hypothyroidism had had a negative impact on overall QoL (mean QII, Thyroid-dependent QoL, was -1.25), and QI, Present QoL, was rated as between good and neither good nor bad (mean 0.89) (Table 1). Transformations of any non-normal distributions left three items with slightly skewed distributions [z (skew) ranging from 2.8 to 4.0].

The 18-item ThyDQoL had very high internal consistency reliability [Cronbach's alpha = 0.949, standardised item alpha = 0.95, N = 97 (N/A responses coded as zero)]. All corrected item-total correlations were satisfactory (>0.2, range 0.49 to 0.88), the lowest being item 16, depression, which also detracted slightly from overall scale alpha (0.950). An unforced factor analysis produced two components (after Varimax rotation) with eigenvalues >1, accounting for 55.4% and 7.3% of the variance. Several items double loaded and there was no readily interpretable pattern of factor loadings (Table 3). Forcing the items onto greater numbers of factors did not clarify the picture either. The scree plot indicated only one strong factor. A forced 1-factor analysis showed all 18 items had satisfactory loadings (>0.5) on the single factor (accounting for 55.4% of the variance) (Table 3), indicating that all applicable domains could be combined into an overall Average Weighted Impact score (AWI-18), reflecting a maximum of 18 weighted domain impact scores. The AWI-18 mean score of -3.11 (2.2) indicated overall negative perceived impact of hypothyroidism on QoL. Calculation of ThyDQoL AWI-18 was reliable at alpha = 0.9 with a maximum of five items missing and reliable at alpha = 0.8 with a maximum 10 missing items. It is therefore reasonable to calculate AWI-18 providing at least half of the items (i.e. domain-specific items) have been completed.

The correlation between AWI-18 and QII, Thyroid-dependent QoL, [r = 0.69, p <0.001, N = 108], was sufficiently high for QII to replace the full ThyDQoL for some purposes. There were no a priori hypotheses regarding any subgroup differences, and the study was not powered to detect any, and indeed no significant treatment sub-group differences were found. However, it is interesting to note that the difference in QII, Thyroid-dependent QoL, between those with overt and subclinical hypothyroidism approached significance (Bonferroni correction required p value 0.05/3 = 0.017): those with overt (treated) hypothyroidism tended towards greater perceived negative impact of hypothyroidism on QoL, [mean = −1.31 (1.0), N = 98] than those with subclinical hypothyroidism, [mean = −0.7 (0.67), N = 10]; t(13.4) = -2.57, p = 0.023.

Twenty-two patients responded in the free comments section but when their comments were analysed, it was found that they were expanding on their responses to existing questionnaire domains,

or were referring to aspects of symptoms or treatment covered in the ThySRQ or the Underactive Thyroid Treatment Satisfaction Questionnaire [6] respectively. No new domains were indicated.

ThySRQ

The most common reported symptoms were tiredness (90% of respondents), feeling depressed or low, and skin problems (Table 4). The mean number of symptoms reported per patient was 7.4 (3.2), N = 107, with six patients indicating they had 13 or more symptoms. Almost one third reported being very much bothered by symptoms of tiredness, weight gain or depression when applicable. The highest mean bother ratings were for hair, weight gain, depressed, and cold. When data were examined from the 57 respondents who had either no comorbidities (N = 42), or whose only comorbidities were likely to be perceived by patients as symptomless (hypertension and hypercholesterolaemia, N = 15), the picture was only a little different: the most common reported symptoms were tiredness, depression, problems with concentration and skin, weight gain, memory and cold (full results not reported).

Exploratory psychometric analyses were performed on the ThySRQ scale formed by the 15 symptom bother ratings although the questionnaire, being a number of disparate symptoms, was not expected to have high internal consistency reliability or any particular factor structure. ThySRQ internal consistency reliability was unexpectedly high: Cronbach's alpha = 0.808, standardised alpha = 0.81, N = 95, with all corrected item-total correlations >0.2. Only two items would increase alpha if deleted from the scale (items 5, hair and 9, hearing), and by no more than 0.002 (full results not shown). Results of unforced and forced factor analyses were unclear, with insufficient evidence of satisfactory subgroups of symptoms that could produce reliable subscale scores. Forced 1-factor analyses showed 12/15 items loading >0.44 but constipation, hair and hearing fell short of this (0.35 to 0.28) indicating that a total score is not fully justified. This result was not unexpected.

Chi-Square tests found no significant differences in symptom frequency of thyroxine-treated and untreated patients nor were significant correlations found between TSH (a tissue marker for severity of hypothyroidism) and ThySRQ bother ratings, once the Bonferroni correction had been applied.

Relationships Between The ThySRQ And ThyDQoL Questionnaires

When both ThyDQoL and ThySRQ questionnaires are administered to patients it would be prudent, if analysing relationships between ThyDQoL AWI and ThySRQ symptom bother ratings, to exclude patients' responses to the four ThyDQoL symptom domains (energy, weight, bodily discomfort and depression) from the calculation of the AWI score, i.e. a maximum of 14 items should be included in the AWI-14 score. Analyses showed a very high internal consistency reliability of the 14-item ThyDQoL (Cronbach's alpha = 0.944, N = 99), with lowest corrected item-total correlation of 0.57. All loadings on a forced 1-factor analysis were >0.6. This provided evidence for the reliability and validity of calculating overall AWI-14. Mean ThyDQoL AWI-14 for the sample was -2.86 (2.22), slightly less negative than -3.11 (2.2) for the full 18-item AWI-18, as expected.

When correlating AWI-14 with ThySRQ bother ratings, the strongest significant negative correlations were found (in decreasing order) with tiredness, concentration, depressed, weight gain, cold, speech and skin (Table 5), indicating that increased symptom bother was associated with greater negative impact of hypothyroidism on QoL. The correlations with AWI-14 were very similar to those with the full 18-item ThyDQoL AWI-18, (Table 5) indicating that concerns about spuriously high correlations between ThySRQ and AWI-18 were not supported. Correlations between symptoms and AWI scores were low to moderate (range -0.1 to -0.6), an indication that symptoms and QoL are distinct and separate patient-reported outcomes. A stepwise multiple regression analysis was conducted to evaluate how well the bother ratings for the five symptoms with the highest correlations with QI, Present QoL, i.e. tiredness, depressed, cold, concentration and giddy (Table 5), predicted Present QoL. Only tiredness was a significant predictor of Present QoL: [$F_{(1,98)}$ = 31.1, $p < 0.001$]. The sample multiple correlation coefficient was 0.49, indicating that 24% of variance in QI could be accounted for by tiredness: the greater the bother rating for tiredness, the worse the present QoL.

Moderate significant negative correlations were found between the number of applicable symptoms and ThyDQoL overview items QI, Present QoL (r = -0.42), and QII, Thyroid-dependent QoL (r = -0.43)

(p <0.001, N = 105) indicating decreased present QoL, and greater negative impact of hypothyroidism on QoL, with increasing numbers of symptoms.

DISCUSSION

The ThyDQoL is a new individualised measure of the perceived impact of hypothyroidism on QoL, with high completion rates indicating good acceptability to respondents. Internal consistency reliability was excellent and factor analyses highly supportive of the combination of all 18 weighted domains into a single average score, the ThyDQoL AWI-18 score, which indicated moderately severe perceived negative impact of hypothyroidism on QoL in this sample. Useful information can be elicited from the single overview items. On average, respondents' present QoL (QI) was perceived to be between good and neither good nor bad and QII, concerning impact of hypothyroidism on QoL, could provide an approximate substitute for the full ThyDQoL for some purposes, (e.g. when respondent burden is of particular concern), as the correlation between QII and overall ThyDQoL AWI-18 was high (r = 0.69). Analysing individual domains separately can provide rich information. For example, we found that, despite most patients receiving thyroxine replacement to normalise TSH levels, hypothyroidism still had a marked negative impact, particularly on motivation, weight, feelings of depression, bodily discomfort and energy, which may well impact on several other domains, such as spare-time activities or work. Over 70% of respondents reported feeling depressed or having bodily discomfort in recent weeks. Content validity was supported, as no new ThyDQoL domains emerged when analysing the free comments section, and all existing domains detected negative impact of hypothyroidism, with mean weighted domain impact scores of -2 or less, as shown in Table 1.

The new Underactive Thyroid Symptom Rating Questionnaire measures both symptom frequency and the degree to which patients are bothered by symptoms associated with hypothyroidism. All 15 symptoms were experienced recently by at least some patients and there was a wide spread of frequencies of responses from not at all bothered to very much bothered. Internal consistency reliability was unexpectedly high, but factor analysis did not support the calculation of a total symptom bother score. Analysis of the most frequently reported symptoms (tiredness, depression, weight gain, and problems with skin, concentration and memory) and bother ratings, confirmed the typical picture

of hypothyroidism: physiological symptoms connected with slowing of the metabolism and also reduced psychological well-being and impaired cognitive functioning. Although measures of health status are often and erroneously described as QoL measures [31, 32], in this study symptoms (an aspect of health status) and QoL were distinct and separate patient-reported outcomes, as indicated by low to moderate correlations between symptom bother ratings and ThyDQoL AWI.

The ThyDQoL includes four symptom items that are also covered on the ThySRQ symptom measure, namely energy, weight, bodily discomfort and depression. We recommend that these four items be excluded from the calculation of ThyDQoL AWI when the questionnaire is administered with the ThySRQ, to avoid duplication and spurious correlations between the two questionnaires. The equivalent symptoms on the ThySRQ (tiredness, depressed, weight gain and cold) still had some of the strongest correlations with AWI-14, suggesting that the impact of those symptoms on QoL is reflected in ratings of many domains, and symptom-specific domains are not required in the ThyDQoL, although they are likely to increase face and content validity where no symptom measure is used. If troublesome symptoms can be reduced, QoL of patients may be improved [33]. The data suggested that patients were most bothered by hair problems, weight gain, depression, feeling cold, and tiredness and, of these symptoms, depression, feeling cold and tiredness were significantly correlated with ratings of present QoL. An optimal strategy for treating individual patients may be to identify individual patients' most bothersome symptoms using the ThySRQ and the impact of hypothyroidism on their QoL using the ThyDQoL and target efforts according to an individual's needs and priorities. However, in order to improve the QoL of their patients in general, the present results suggest that healthcare professionals could most usefully direct their efforts to alleviating the symptoms of depression, feeling cold and tiredness. Symptoms can be selected for the ThySRQ to suit the purposes of a particular clinical trial. Further symptoms may need to be added if required (e.g. side effects of a new drug under evaluation), although the rationale of any deletions or additions would need to be justified.

One limitation of the study was that only two patients were recruited from ethnic minorities, although the incidence of hypothyroidism is similar in Whites and Asians but lower in Afro-Caribbeans. While the clinical manifestation of hypothyroidism would be expected to be similar for all ethnic groups, it is

possible that different ethnic groups have varying perceptions of the effects of the disorder on their QoL, or they might be bothered to varying degrees by the symptoms, but this study was not designed to obtain data of this nature.

CONCLUSIONS

The ThyDQoL and ThySRQ are two new condition-specific questionnaires measuring the perceived impact of hypothyroidism on QoL, and bother ratings for 15 symptoms respectively, and allowing multidimensional assessments of the effects of treatment for hypothyroidism. Both questionnaires are performing well with good acceptability to the great majority of respondents. The ThyDQoL has excellent internal consistency reliability, domains can be analysed separately if required, and an overall score calculated. Additional symptoms may be added to the ThySRQ to suit the requirements of different studies. Sensitivity to change of both measures now needs to be evaluated in clinical trials.

ACKNOWLEDGEMENTS

ACCESS TO QUESTIONNAIRES

For access to and licence to use the ThyDQoL and ThySRQ questionnaires, contact the copyright holder, Clare Bradley PhD, Professor of Health Psychology, Health Psychology Research, Royal Holloway, University of London, Egham, Surrey, TW20 0EX. Email: c.bradley@rhul.ac.uk. There is no charge for their use in publicly or charitably funded academic and medical research, other than a small administration charge, which is waived for student projects. Charges are applied for commercial use.

REFERENCES

1. Hueston WJ. Treatment of hypothyroidism. Am Fam Physician 2001; 64:1717-24.

2. Canaris GJ, Manowitz NR, Mayor G, et al. The Colorado thyroid disease prevalence study. Arch Intern Med 2000; 160:526-34.

3. Wiersinga WM. Adult Hypothyroidism. In: Thyroid Disease Manager, 2002. Available from: http://www.thyroidmanager.org. Last accessed January 25, 2007.

4. Kaplan MM, Sarne DH, Schneider AB. In search of the impossible dream? Thyroid hormone replacement therapy that treats all symptoms in all hypothyroid patients. J Clin Endocrinol Metab 2003;88:4540-2.

5. Sonino N, Fava GA. Psychological aspects of endocrine disease. Clin Endocrinol (Oxf) 1998; 49:1-7.

6. McMillan CV, Bradley C, Woodcock A, et al. Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. Thyroid 2004;14:916-25.

7. Billewicz WZ, Chapman RS, Crooks J, et al. Statistical methods applied to the diagnosis of hypothyroidism. Q J Med 1969;38:255-66.

8. Zulewski H, Muller B, Exer P, et al. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. J Clin Endocrinol Metab 1997;82:771-6.

9. Cooper DS. Combined T4 and T3 therapy - back to the drawing board. JAMA 2003;290:3002-4.

10. Jaeschke R, Guyatt G, Cook D, et al. Spectrum of quality of life impairment in hypothyroidism. Qual Life Res 1994;3:323-27.

11. Jaeschke R, Guyatt G, Gerstein H, et al. Does treatment with L-thyroxine influence health status in middle-aged and older adults with subclinical hypothyroidism? J Gen Intern Med 1996;11:744-79.

12. Canaris GJ, Steiner JF, Ridgway EC. Do traditional symptoms of hypothyroidism correlate with biochemical disease? J Gen Intern Med 1997;12:544-50.

13. Saravanan P, Chau WF, Roberts N, et al. Psychological well-being in patients on 'adequate' doses of l-thyroxine: results of a large, controlled community-based questionnaire study. Clin Endocrinol (Oxf) 2002;57:577-85.

14. Clyde PW, Harari AE, Getka EJ, et al. Combined levothyroxine plus liothyronine compared with levothyroxine alone in primary hypothyroidism: a randomized controlled trial. JAMA 2003; 290:2952-8.

15. Cooper DS, Halpern R, Wood LC, et al. L-Thyroxine therapy in subclinical hypothyroidism. A double-blind, placebo-controlled trial. Ann Intern Med 1984;101:18-24.

16. Bradley C, Todd C, Gorton T, et al. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Qual Life Res 1999;8:79-91.

17. Bradley C, Speight J. Patient perceptions of diabetes and diabetes therapy: assessing quality of life. Diabetes Metab Res Rev 2002;18:S64-9.

18. Bradley C. Design of a renal-dependent individualised quality of life questionnaire: RDQoL. Adv Perit Dial 1997;13:116-20.

19. Mitchell J, Bradley C. Design of an individualised measure of the impact of macular disease on quality of life (the MacDQoL). Qual Life Res 2004;13:1163-75.

20. McMillan CV, Bradley C, Gibney J, et al. Preliminary development of the new individualized HDQoL questionnaire measuring quality of life in adult hypopituitarism. J Eval Clin Pract 2006; 12:501-14.

21. McGee HM, O'Boyle CA, Hickey A, et al. Assessing the quality of life of the individual: The SEIQoL with a healthy and a gastroenterology unit population. Psychol Med 1991;21:749-59.

22. Grootenhuis PA, Snoek FJ, Heine RJ, et al. Development of a Type 2 Diabetes Symptom Checklist: a measure of symptom severity. Diabet Med 1994;11:253-61.

23. Steen N, McColl E. Developing and testing symptom-based outcome measures. In: Hutchinson A, McColl E, Christie MJ, Riccalton CL, (eds.). Health Outcomes for Primary and Outpatient Care. Chur, Switzerland: Harwood Academic, 1995.

24. Tabachnik BG, Fidell LS. Using Multivariate Statistics. New York: Harper and Row, 1983.

25. Todd C, Bradley C. Evaluating the design and development of psychological scales. In: Bradley C, ed. Handbook of Psychology and Diabetes: A Guide to Psychological Measurement in Diabetes Research and Practice. Chur, Switzerland: Harwood Academic Publishers, 1994.

26. Kline P. A Handbook of Psychological Testing. London: Routledge, 1993.

27. Mitchell J, Bradley C. Psychometric evaluation of the 12-item Well-being Questionnaire for use with people with macular disease. Qual Life Res 2001;10:465-73.

28. Kline P. An Easy Guide to Factor Analysis. London: Routledge, 1994.

29. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6:65-70.

30. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab 2002;87:489-99.

31. Bradley C. Importance of differentiating health status from quality of life. Lancet 2001; 57:7-8.

32. Razvi S, McMillan CV, Weaver JU. Instruments used in measuring symptoms, health status and quality of life in hypothyroidism: a systematic qualitative review. Clin Endocrinol (Oxf) 2005;63:617-24.

33. Bradley C. Measuring quality of life in diabetes. In: Marshall SM, Home PD, Rizza RA, (eds.). The Diabetes Annual 10 Amsterdam: Elsevier Science, 1996.

34. Helfand M, Redfern CC. Screening for thyroid disease: An update. Ann Intern Med 1998; 129:144-58.

Table 1: ThyDQoL. Item wording and descriptive statistics (maximum N = 108)

No.	Item wording (response option)	Abbreviation	N/A	Missing	Mean* (SD)	Range	Median
QI	In general my present quality of life is…(excellent – extremely bad)	Present QoL	-	0	0.89 (1.02)	-2 to 3	1
QII	If I did not have underactive thyroid, my quality of life would be…(very much better – worse)	Thyroid-dependent QoL	-	0	-1.25 (0.99)	-3 to 0	-1
	If I did not have underactive thyroid, …						
1	…I would enjoy the things I do in my spare time…(very much more – less)	spare time	-	1	-2.49 (2.71)	-9 to 0	-2
2[†]	…my working life would be…(very much better – worse)	working life	54	0	-3.46 (3.30)	-9 to 0	-3
3[†]	…my holidays would be…(very much better – worse)	holidays	18	0	-2.37 (2.63)	-9 to 0	-2
4[†]	…my family life would be…(very much better – worse).	family life	3	2	-2.74 (2.93)	-9 to 3	-3
5	…my social life would be…(very much better – worse)	social life	-	2	-2.00 (2.57)	-9 to 0	-1
6[†]	…my closest personal relationship would be…(very much better – worse)	relationship	20	0	-2.91 (3.18)	-9 to 0	-2
7[†]	…my sex life would be…(very much better – worse)	sex life	32	1	-2.69 (3.12)	-9 to 0	-2
8	…physically I could do…(very much more – less)	physically do	-	3	-3.23 (2.82)	-9 to 0	-2
9[‡]	…my energy levels would be…(very much higher – lower)	energy	-	1	-4.21 (3.11)	-9 to 0	-4
10	…the speed I could do things would be…(very much faster – slower)	speed do	-	0	-2.78 (2.52)	-9 to 0	-2
11	…getting out and about (e.g. shopping, short trips) would be…(very much easier – more difficult)	get out and about	-	0	-2.60 (2.90)	-9 to 0	-2
12	…I could handle my household tasks…(very much better – worse)	household tasks	-	2	-3.37 (2.89)	-9 to 0	-3
13	…my physical appearance would be…(very much better – worse)	physical appearance	-	0	-3.24 (3.24)	-9 to 0	-2

No.	Item wording (response option)	Abbreviation	N/A	Missing	Mean* (SD)	Range	Median
14[†,‡]	…my weight would be…(very much better – worse)	weight	22	0	-4.51 (2.97)	-9 to 0	-4
15[†,‡]	…my experience of bodily discomfort would be…(very much less – greater)	bodily discomfort	19	0	-4.29 (2.81)	-9 to 2	-4
16[†,‡]	…I would feel depressed or low…(very much less – more)	depression	31	1	-4.36 (2.73)	-9 to 0	-4
17[†]	…my motivation to do things would be…(very much greater – less)	motivation	13	0	-4.84 (2.79)	-9 to 0	-4
18	…my feelings about the future (e.g. worries, hopes) would be…(very much better – worse)	future	-	1	-2.87 (2.86)	-9 to 0	-2.5

QoL: quality of life; N/A: not applicable; SD: Standard deviation. *Mean ThyDQoL weighted domain impact score. [†]Indicates items for which an opening question allows the respondent to indicate that the item is not applicable to them and, if so, move on to the next item. [‡]Item excluded from calculation of AWI-14. Maximum possible score ranges: QI, Present QoL (+3 to -3), QII, Thyroid-dependent QoL (-3 to +1), 18 domain items (-9 to +3).

ThyDQoL © Prof Clare Bradley: 22.7.03. Standard UK English (rev. 22.7.03). Health Psychology Research, Dept of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK.

Table 2: Characteristics of study participants (N = 110)

	Mean (SD)	Range	N
Mean age (years)	55.1 (14.3)	23 – 84	110
Mean duration of hypothyroidism (years)	8.3 (8.9)	0.2 – 57	103
Age at leaving full-time education (years)	16.4 (2.8)	14 – 27	94
BMI (kg/m^2)	28.6 (6.4)	18 – 48.8	99
TSH (mu/L)	3.9 (5.0)	0.02 – 28.7	95
FT4 (pmol/L)	17.4 (4.4)	8.9 – 28.4	49

	N
Ratio women to men	89:21
Ratio White to non-white	108:2
Hypothyroidism:	
Autoimmune	87
Secondary to treatment for thyroid cancer or hyperthyroidism	9
Secondary to lithium treatment	1
Secondary to amiodarone treatment	2
Subclinical (autoimmune)[*]	11
Treatment:	
Thyroxine-treated	103
Not receiving thyroxine	7[†]

BMI: body mass index; TSH: thyroid stimulating hormone; FT4: free thyroxine. *Subclinical hypothyroidism defined as TSH level> 4mu/L and FT4 level in normal range (9 to 25 pmol/L) [34]. †All with subclinical hypothyroidism. The following comorbid conditions occurred: hypertension (26 patients), osteoarthritis (10), ischaemic heart disease (9), asthma (8), Type II diabetes (8), rheumatic disorder (7), hypercholesterolaemia (6), Type I diabetes (6), cancer (4), pregnancy (4), depression (4), Addison's disease (3), pernicious anaemia (3), coeliac disease (2). Any other disorders were present in a maximum of one patient and are not reported. Several patients had multiple comorbid conditions.

Table 3: ThyDQoL unforced, and forced 1-factor analyses[*]

	Unforced		Forced 1-factor
	1	2	
1. spare time	0.417	0.743	0.814
2. working life		0.743	0.608
3. holidays		0.783	0.740
4. family life	0.546	0.637	0.835
5. social life		0.794	0.809
6. relationship		0.690	0.729
7. sex life		0.690	0.642
8. physically do	0.640	0.644	0.908
9. energy	0.678	0.534	0.859
10. speed do	0.744	0.449	0.849
11. get out and about	0.636	0.485	0.795
12. household tasks	0.742		0.762
13. physical appearance	0.569		0.654
14 weight	0.753		0.603
15. bodily discomfort	0.669		0.643
16. depression	0.445		0.529
17. motivation	0.756		0.821
18. future	0.719		0.671

[*] Principal Component extraction with Varimax Rotation, Not Applicable responses coded as zero.

Loadings ≥0.4 are shown.

Table 4: ThySRQ symptom bother ratings

	Symptom* % (N)	not at all bothered (N)	a little bothered (N)	quite a bit bothered (N)	very much bothered (N)	Mean (SD) [range used]	Median
1. tiredness	90% (99)	2	18	46	33	2.11 (0.77) [0 – 3]	2
2. weight gain	59% (65)	3	13	16	33	2.22 (0.93) [0 – 3]	3
3. cold	53% (58)	1	10	28	19	2.12 (0.75) [0 – 3]	2
4.constipation	42% (46)	1	22	11	12	1.74 (0.88) [0 – 3]	1.5
5. hair	44% (48)	0	13	10	25	2.25 (0.86) [1 – 3]	3
6. skin	60% (66)	1	24	15	26	2.00 (0.91) [0 – 3]	2
7. nails	46% (51)	2	22	17	10	1.69 (0.84) [0 – 3]	2
8. appetite	22% (24)	3	10	6	5	1.54 (0.98) [0 – 3]	1
9. hearing	33% (36)	3	16	4	13	1.75 (1.05) [0 – 3]	1
10. voice	24% (26)	5	9	4	8	1.58 (1.14) [0 – 3]	1
11. speech	25% (27)	1	9	9	8	1.89 (0.89) [0 – 3]	2
12. memory	54% (59)	1	22	12	24	2.00 (0.93) [0 – 3]	2
13. concentration	56% (61)	1	20	14	26	2.07 (0.91) [0 – 3]	2
14. giddy	46% (51)	0	19	12	19	2.00 (0.88) [1 – 3]	2
15. depressed	73% (80)	1	20	27	32	2.13 (0.83) [0 – 3]	2

SD: Standard deviation. *Symptom experienced in recent weeks.

Table 5: Correlations between ThySRQ and ThyDQoL questionnaires

ThySRQ Symptom	Correlations with AWI-18	Correlations with AWI-14	Correlations with QI, Present QoL
1. tiredness	-0.61, p <0.001[*]	-0.61, p <0.001[*]	-0.44, p <0.001[*]
2. weight gain	-0.47, p <0.001[*]	-0.43, p <0.001[*]	-0.15, p = 0.13, n.s.
3. cold	-0.38, p <0.001[*]	-0.37, p <0.001[*]	-0.30, p <0.01
4. constipation	-0.27, p <0.01	-0.26, p <0.01	-0.19, p = 0.05
5. hair	-0.21, p = 0.029	-0.22, p = 0.022	-0.11, p = 0.27, n.s.
6. skin	-0.28, p <0.01	-0.28, p <0.01	-0.12, p = 0.23, n.s.
7. nails	-0.11, p = 0.263, n.s.	-0.11, p = 0.26, n.s.	-0.16, p = 0.11, n.s.
8. appetite	-0.17, p = 0.083, n.s.	-0.17, p = 0.085, n.s.	-0.34, p <0.001*
9. hearing	-0.11, p = 0.25, n.s.	-0.11, p = 0.25, n.s.	-0.11, p = 0.28, n.s.
10. voice	-0.26, p <0.01	-0.27, p <0.01	-0.04, p = 0.71, n.s.
11. speech	-0.29, p <0.01	-0.29, p <0.01	-0.03, p = 0.73, n.s.
12. memory	-0.25, p = 0.011	-0.26, p = 0.008	-0.18, p = 0.07, n.s.
13. concentration	-0.49, p <0.001[*]	-0.50, p <0.001[*]	-0.29, p <0.01
14. giddy	-0.24, p = 0.012	-0.23, p = 0.015	-0.27, p <0.01
15. depressed	-0.47, p <0.001[*]	-0.45, p <0.001[*]	-0.40, p <0.001[*]

QoL: quality of life; n.s.: not significant; AWI-18/ AWI-14: Average Weighted Impact score of the ThyDQoL questionnaire (18 items or 14 items respectively). All correlations are Spearman's rho and those which would remain significant after Bonferroni correction requiring minimum p value of 0.0033 are marked *. N ranged from 103-108 respondents.