Clinical Thyroidology / Original Paper

European
Thyroid Journal

Eur Thyroid J 2019;8:144–151
DOI: 10.1159/000496827

Received: August 29, 2018
Accepted after revision: December 27, 2018
Published online: March 25, 2019

Detecting True Change in the Hospital Anxiety and Depression Scale, SF-36, and Hypothyroid Score when Monitoring Patients with Subclinical Hypothyroidism

Jesper Karmisholt[a, c] Stig Andersen[b, c]

[a]Department of Medical Endocrinology, Aalborg University Hospital, Aalborg, Denmark; [b]Department of Geriatrics, Aalborg University Hospital, Aalborg, Denmark; [c]Clinical Institute, Aalborg University, Aalborg, Denmark

Keywords
Subclinical hypothyroidism · Symptom score · Monitoring · Reference change value · Biological variation

Abstract
Guidelines suggest that subclinical hypothyroid (SCH) patients with thyrotropin (TSH) between 4 and 10 mU/L and symptoms associated with hypothyroidism should receive L-T4 substitution treatment, be evaluated, and continue treatment if symptoms subside. The latter requires detecting a true change in symptoms, which can be calculated from within-person variation in symptom evaluation tools. This led us to assess within-person variation in hypothyroid symptoms, in mood-related symptoms, and quality of life in patients with untreated SCH in order to support the recommended evaluations. ***Method:*** The within-person coefficient of variation (CV) was estimated from 13 consecutive monthly evaluations in 15 patients with initial TSH between 5 and 12 mU/L and no trend in TSH. ***Results:*** The within-person CV was rather large for the Hospital Anxiety and Depression Scale (HADS) and Zulewski hypothyroid score at 41.6 and 60.9%, respectively. For quality of life the within-person CV was lower at 8.0% for the physical component summary and 8.7% for the mental component summary from the SF-36 questionnaire. The difference required between two measurements to detect a true change was 97% for mood-related symptoms (HADS) and 140% for hypothyroid symptoms. For quality of life (SF-36) the required difference was 20%. ***Conclusion:*** Score differences of almost 100% and higher were required to support a true change in mood (HADS) and hypothyroid symptom scores in untreated SCH patients. For quality of life a true change was detected at a 20% difference in SF-36 scores. The hypothyroid score and HADS questionnaire do not seem useful for the evaluation of individuals.

Introduction

Subclinical hypothyroidism (SCH) is a frequent condition defined by serum values of thyrotropin (TSH) above and thyroxine (T4) within the population-based reference ranges [1–3]. It may reflect mild thyroid failure with only TSH outside the reference range [4]. Thus, a number of organ-specific changes have been reported in patients with SCH [5].

KARGER

E-Mail karger@karger.com
www.karger.com/etj

Jesper Karmisholt, MD, PhD
Department of Medical Endocrinology
Aalborg University Hospital
DK–9000 Aalborg (Denmark)
E-Mail jsk@rn.dk

Quality of life is important to the individual patient, but placebo-controlled randomized treatment trials have shown small or no effect of L-T4 treatment on quality of life and other parameters [6, 7]. Still, guidelines from both Europe and North America suggest that SCH patients with TSH between 4 and 10 mU/L and symptoms associated with hypothyroidism should receive an L-T4 substitution treatment trial, be evaluated, and continue treatment if symptoms subside [8, 9].

Patients in the spectrum from subclinical to overt and severe hypothyroidism can experience a plethora of different symptoms also reported by otherwise healthy subjects [10, 11]. The focus has been on the occurrence and predictive value of symptoms and symptom scores while reports on variation and the reliability of such symptoms are lacking. This led us to assess the level of difference in symptoms needed between two tests for the difference to be clinically relevant and the level of difference that may be due to random variation of such symptoms. This was done by generation and application of data on variation in questionnaires on mood and hypothyroid symptoms and quality of life in untreated patients with SCH.

The aim of the present study was thus to assess the individual variation in questionnaire response and to provide a systematic evaluation of various hypothyroid symptoms, mood-related symptoms and quality of life in a cohort of patients with untreated SCH in order to assess the magnitude of change needed for a true change in hypothyroid symptoms.

Subjects and Methods

Participants were initially diagnosed with SCH by their family physician and subsequently included in the study if TSH was between 5 and 12 mU/L and total T4 within the laboratory reference range (60–140 nmol/L) both at first measurement and on repeated testing 3 months later, as described previously [12]. The patients were recruited from May 2004 and followed until July 2006. The exclusion criteria were: previous thyroid disease, change in any type of medication during the past 3 months, medication or disease with influence on the thyroid, and pregnancy within the past 12 months. Twenty-one patients, all Caucasians, were included. The symptoms causing the initial contact with the family physician was a general feeling of tiredness, of unhappiness, of cold intolerance, of weight gain or paresthesia in 12 patients, and either atypical symptoms or just a routine general health check in the remainder.

Procedures

The participants were investigated monthly for a full year and all investigational sessions (n = 273) but one were performed by the same investigator (J.K.). Each session consisted of blood sampling, a clinical investigation, and the quality of life questionnaire. The latter was filled out by the patient and checked for completion by the investigator. The remaining questionnaires were distributed for completion at the participants' home and subsequently returned by postal delivery.

The questionnaires used were: Danish version 1 of the SF-36 [13], a Danish version of the Hospital Anxiety and Depression Scale (HADS) [14], and the hypothyroid score as published by Zulewski et al. [15].

SF-36 is a widely used general quality of life questionnaire where the output is organized into eight subscales (role limitation due to: physical function, pf; limitations due to physical problems, rp; limitations due to bodily pain, bp; limitations due to general health problems, gh; limitations due to reduced vitality, vt; limitations due to reduced social function, sf; limitations due to emotional problems, re; and limitations due to mental health problems, mh) or into two component summaries. The physical component summary (PCS) and mental component summary (MCS) are based on standard US reference values [16]. The mean (SD) values of the reference population are 50 (10) for the component summaries. On all SF-36 scales a higher score is better.

HADS measures the level of anxiety and depression without using anxiety and depression terms attributable to somatic symptoms. The output is organized as a total HADS score. More or severe symptoms give higher scores. A score above 15 may indicate a possible case with significant anxiety or depression.

The hypothyroid score is calculated based on the presence or absence of seven symptoms related to hypothyroidism and five hypothyroid signs. These are summarized into a total score with a maximum score of 12. The patients evaluated the seven symptoms (diminished sweating, hoarseness paraesthesia, dry skin, constipation, impairment of hearing, and weight increase) whereas the five hypothyroid signs (slow movements, delayed ankle reflex, coarse skin, and periorbital puffiness) were evaluated by a physician (J.K.). With this clinical score hypothyroid patients generally score five or more while euthyroid subjects score two or less [15]. In the original score women aged below 55 years received one extra point. This extra point has been omitted in the calculations. The evaluations were performed without knowledge of the results of the thyroid function tests at the current visit. Thyroid function at previous visits were known as thyroid function tests were performed at each visit according to protocol to detect and be able to respond to any critical deterioration in thyroid function during the study.

The participants attended the session after an overnight fast and the blood samples were drawn between 09:00 and 12:00 h. Serum was separated and stored at –20 °C until analysis.

Assays

The thyroid hormones were measured using an Electro-Chemi Luminescence Immuno Assay method on a Modular Analytics E170 (Roche, Mannheim, Germany). The assay characteristics provided by the manufacturer were (detection limit, reference range, single determination intra-assay CV%): fT4, 0.3 pmol/L, 12–22 pmol/L, 3.6%, and total T4 5.4 nmol/L, 60–140 nmol/L (laboratory reference range), 1.6%. For TSH: 0.005 mU/L, 0.27–4.2 mU/L, and 3.0% in samples with a mean TSH of 0.04 mU/L, 1.2% in samples with a mean TSH of 0.96 mU/L, and 1.1% in samples with a mean TSH of 9.37 mU/L. Thyroid peroxidase antibodies (TPO-Ab) were measured using anti-TPO KRYPTOR (BRAHMS, Henningsdorf, Germany) with analytical sensitivity of 10 U/mL, intra-assay CV% 4.2, and inter-assay CV% 9.7, as provided by the manufacturer.

Eur Thyroid J 2019;8:144–151
DOI: 10.1159/000496827

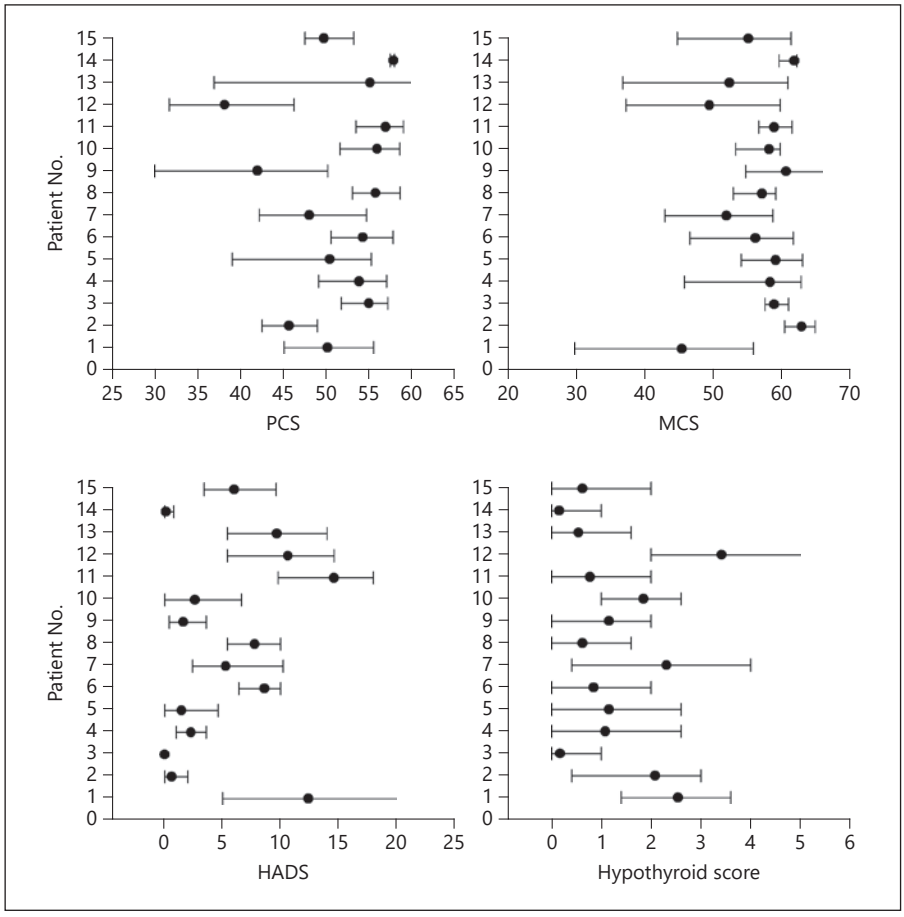

Fig. 1. Mean and 10–90 percentiles of measures of symptom scores in 15 individual patients. Abbreviations are as in Table 2.

Statistics and Calculations

Data in individuals and groups were tested for conformity to the normal distribution by the inspection of Q-Q plots. Data in individuals were normally distributed except in patient 15, as seen in Figure 1, where the questionnaire scores were skewed. Group data also followed a normal distribution (TSH after natural logarithmic transformation). A *t* test was used for between-group comparisons. Within-person (expressed as within-person coefficient of variation; CV) and between-person variation (expressed as between-person CV) were calculated using analysis of variation (ANOVA) according to recommendations [17]. Within-person correlations and between-person correlations were calculated using analysis of covariation (ANCOVA) and Pearson correlation on mean values [18, 19].

Data on within-person variation were used for calculations on reference change values (RCV). The formula developed for laboratory tests, $RCV = 2^{1/2} \times Z \times (CV_A^2 + CV_I^2)^{1/2}$, include analytical CV_A and within-subject CV_I. These are both contained within in the questionnaire. Hence, RCV was calculated as $CV \times Z \times \sqrt{2}$, where Z is the number of standard deviations fitting the probability, i.e., 1.96 for 95% probability [20]. We used a 90% probability of significant change ($Z = 1.65$) as a clinically relevant change. Missing questionnaire data were treated as the last information carried forward.

A p value <0.05 was considered statistically significant. The Statistical Package for the Social Sciences version 15.0 (SPSS Inc., Chicago, IL, USA) and Excel 2007 (Microsoft Corp., Redmond, WA, USA) were used for the calculations.

Results

Twenty-one patients (2 male) were enrolled. Four patients were tobacco users. One of the patients became overtly hypothyroid with a steep increase in TSH and commenced L-T4 treatment. One patient had normal TSH on all samples. Four patients had a significant individual trend in TSH during the study. This trend was defined as a statistically significant slope in TSH with time (in months) using linear regression. These 6 patients (no males) were excluded from the RCV calculations. The 15 patients with stable TSH had a mean TSH during the study period ranging from 4.4 to 12.3 mU/L. Table 1 lists the baseline characteristics of patients with and without a

Table 1. Baseline values of the included patients

	Baseline		*p* value of difference
	without TSH trend (*n* = 15)	with TSH trend (*n* = 4)	
Age, years	57.5 (12.5)	56.2 (12.5)	0.84
BMI	27.6 (4.53)	30.2 (6.25)	0.32
TSH, mU/L	6.4 (1.9)	6.2 (2.1)	0.80
fT4, pmol/L	14.2 (2.0)	13.1 (1.4)	0.28
TPO-Ab, U/L	7,202 (8,877)	4,851 (8,305)	0.61

Values are presented as the mean (SD). Data from the patients without a TSH trend were used for the calculations on variations in symptoms. BMI, body mass index; TSH, thyrotropin; fT4, free thyroxine; TPO-Ab, Thyroid peroxidase antibodies.

Table 2. Values at baseline and during the 13 monthly investigations in patients without a significant trend in TSH (*n* = 15) during the 13 measurements

	Baseline in patients without TSH trend	During the study (13 investigations) in patients without TSH trend
PCS	50.9 (6.27)	51.2 (6.85)
MCS	53.0 (9.50)	56.3(6.63)
HADS	6.3 (5.1	5.6 (5.2)
Hypothyroid score	1.3 (0.49)	1.2 (0.36)
TSH, mU/L	6.4 (1.9)	7.1 (2.8)
fT4, pmol/L	14.2 (2.0)	13.6 (1.9)
TPO-Ab, U/L	7,202 (8,877)	7,038 (9,095)

Values are presented as the mean (SD). PCS, physical component summary from the SF-36 questionnaire; MCS, mental component summary from the SF-36 questionnaire; HADS, Hospital Anxiety and Depression Scale; Hypothyroid score, a score of 7 symptoms related to hypothyroidism and 5 hypothyroid signs according to Zulewski et al. [15]; TSH, thyrotropin; fT4, free thyroxine; TPO-Ab, thyroid peroxidase antibodies.

trend in TSH during the study. Two patients were TPO-Ab negative. At baseline there were no statistical differences between the patients with and without a trend in TSH.

Few data were missing. The SF-36 data were complete. For HADS, 10 (3.7%) values were missing in 7 different patients and for the hypothyroid score 3 (1.1%) values were missing in 3 different patients.

Table 2 gives the mean data at baseline and during the study for each of the questionnaires. Figure 1 shows the individual values (individual mean and 10–90 percentiles) of the two SF-36 component scales, the HADS questionnaire, and the hypothyroid score of the 13 measurements in the 15 patients without TSH trend. The hypothyroid score displayed a marked floor-effect with 73% of the score values below 2.

Table 3 lists the within-subject correlation between thyroid function (TSH) and the different questionnaires. Apart from HADS, which showed a negative correlation to TSH, none of the within-person correlations were significant. The significant and negative within-person correlation between TSH and HADS suggests a decrease in TSH correlates to an increase in the HADS score in individuals. Change in none of the other parameters correlated to change in TSH in individual patients. The between-subject correlation evaluates whether high values in one parameter correlate to high (or low) values in another parameter. These data are not shown, but no significant between-subject correlations were found. Correlations between fT4 and the questionnaires were also tested, and showed no significant within-person or between-person correlations. The within-subject variation given in Table 3 was used for the calculation of RCV. The large individual variation in the HADS and the hypothyroid score imply that rather large changes between two measurements were required to be 90% confident that the difference was a true change and not due to chance. The changes required were 96.8% for the HADS and 140% for the hypothyroid score. The RCV of the two SF-36 component scores were calculated as 18.7% (PCS) and 20.3% (MCS), respectively (Table 3). The RCV for the SF-36 subscales ranged from 18% (sf) to 20% (pf), 20% (gh), 24% (mh), 37% (gh), 41% (bp), 48% (re), and 51% (rp).

Finally, we compared the individual variation between patients older (*n* = 6) and younger (*n* = 9) than 60 years of age. Individual variations in PCS and HADS were identical. The MCS was slightly lower (CV 6.3% vs. CV 10.1%) and hypothyroid score slightly higher (CV 67.3% vs. CV 54.8%) in patients above 60 years of age, with limited impact on RCV.

Discussion

This is the first report of systematically evaluated within-subject variation of symptom scores in patients with mild, untreated SCH. Many symptoms of hypothyroidism are common to healthy subjects too, but may improve on treatment in the SCH patient. Hence, symptoms need

Table 3. Within- and between-person variation and the calculated change needed for true change (RCV) of various symptom and quality of life scores in patients with subclinical hypothyroidism

Variable	Within-person correlation (parameter vs. TSH)	*p* value of within person correlation	Within-person CV, %	Between-person CV, %	RCV, %
PCS	0.04	0.64	8.0	11.1	18.7
MCS	0.05	0.50	8.7	8.2	20.3
HADS	–0.16	**0.03**	41.6	85.2	97.0
Hypothyroid score	–0.11	0.16	60.9	73.3	142
TSH	NA	NA	23.6	32.8	55.1
fT4	–0.19	<0.001	6.41	12.4	15.0

See Statistics and Calculations for details of the calculations. The bold *p* value is significant. CV, coefficient of variation; RCV, reference change value (relevant difference required between two measurements); PCS, physical component summary from the SF-36 questionnaire; MCS, mental component summary from the SF-36 questionnaire; HADS, Hospital Anxiety and Depression Scale; Hypothyroid score, a score of 7 symptoms related to hypothyroidism and 5 hypothyroid signs according to Zulewski et al. [15]; TSH, thyrotropin (analytical variation is included in the CV calculation); fT4, free thyroxine (analytical variation is included in the CV calculation); NA, not applicable.

monitoring, i.e., repeated measurement, after starting treatment. The use of symptoms scores could be helpful when evaluating the patient and knowledge of variation in symptom scores is valuable as the probability of an observed change being due to chance depends on the spontaneous individual variation of the measure. Data are now available and the change needed for true change is provided.

Patients with SCH and unremitting TSH above 10 mU/L have increased risk of a cardiovascular event and mortality in observational studies [21]. Thus, patients with this grade of SCH should receive L-thyroxine substitution therapy [8, 9]. Patients with TSH less than 10 mU/L have lower morbidity risk, and lower, if any, benefit of L-thyroxine therapy [7, 22, 23]. Consequently, the need for careful and individually tailored work-up, discussion of treatment goal, and evaluation of effect to weigh up the potential benefit of the treatment is paramount in patients with TSH below 10 mU/L.

In clinical practice, as in this study, patients are often diagnosed with SCH based on a blood sample taken when presenting at a family physician with various common symptoms, such as a general feeling of tiredness, depression, discontent, weight gain, constipation, etc. In SCH patients with symptoms such as tiredness it is recommended to start a treatment trial of L-thyroxine substitution. The risk of having such symptoms accompanied by a slightly elevated TSH by chance is considerable as two to three symptoms of hypothyroidism are seen in 30% of biochemically euthyroid women [24]. If a treatment trial is suggested in such a patient, it is necessary to decide on a goal for the treatment and time period to achieve this goal prior to commencing treatment.

It is crucial to render and keep the patient euthyroid. This may seem obvious, but has proved to be difficult in clinical practice with 20–40% of the patients being overtreated [25, 26]. Such induced hyperthyroidism, even in the subclinical range, increases the risk of atrial fibrillation [27], bone loss [28], and all-cause mortality [29]. The potential harms associated with the risk of overtreatment should be outweighed by potential benefits. This could improve quality of life. Thus, a second goal of a treatment trial is to relieve symptoms that may be related to hypothyroidism in the subclinical range. The change in symptoms needed to detect a true change in symptoms depends on the variation in symptoms. The present study shows that when measured monthly during a year, there is quite a large variation in symptom scores of hypothyroidism and mood, both within and between individuals. A change of around 140% in symptom score for hypothyroidism using the Zulewski score, and of 95% using the HADS score, is needed in order to be 90% certain that the symptom change is a true change and exceeds what is expected due to random variation in these symptoms. This is further detailed in Figure 2, which illustrates that a difference of 50% between two measurements corresponds to a 40 and 60% probability, for HADS and hypothyroid score, respectively, that the change seen is due to chance alone. The within-person variations are much lower (around 8%) for quality of life measured using SF-36 and

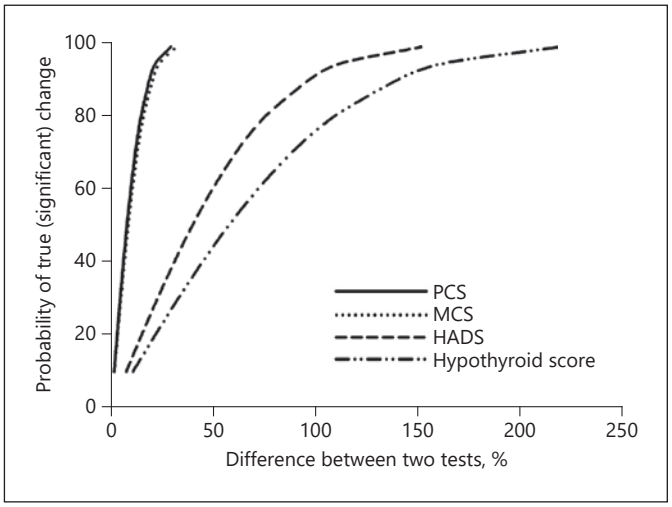

Fig. 2. Probability of change plot illustrating how a difference between two tests (*x*-axis) correspond to the probability of the change being a true change (*y*-axis) and not due to chance. Abbreviations are as in Table 2.

a difference of 8% between two tests of either PCS or MCS corresponds to a 50% probability that the change is due to chance. Conversely, a 90% chance of true change is seen with 18.7 and 20.3% differences between two measurements using SF-36 PCS and MCS.

The relevant time span before re-measuring was not evaluated in the present study. Data may suggest that it should at the earliest be performed when the patient's thyroid function tests have been within the reference range for 3 months. In most cases an evaluation 6–12 months after starting treatment would thus be relevant.

Apart from the HADS score, there was no correlation between change in thyroid function and changes in symptoms in untreated individuals with SCH. Neither did we find a correlation between high TSH and more symptoms in the group of SCH patients. Our results thus support the findings reported in a meta-analysis on SCH [6].

Variation of symptoms did not differ with age. This suggests that the results are applicable over a broad age span despite the fact that symptoms seem to become less distinct with age [30]. Recent randomized treatment trials focusing on elderly patients with SCH [7, 31] suggest that small changes for the better in thyroid function do not cause clinically relevant improvement in thyroid symptoms or in quality of life in this age group.

Our study had strengths and limitations. It was designed to assess biological variation in SCH as reported previously [12]. The meticulous approach to biochemical evaluation was extended into another physiological domain allowing the generation and application of data on variation in symptoms scores in SCH taking an established mathematical approach [17] into a different domain. A further strength of the present study was the many and comparable investigational sessions where the patients were studied at the same time of the day, in the same room, meeting the same physician at each session, thereby limiting "investigational variation." An important additional strength is the high degree of completeness of data. Finally, the number participants in the study comply with the recommended number of participants for a repeated measurement study [17].

A limitation is that the patients did not undergo neuropsychological investigation. A comprehensive neuropsychological work-up could potentially provide further information regarding why the patients answered the questionnaires in the way they did. What was due to the patient's mental constitution and what was related to changes in thyroid function? Clinicians evaluate individual patients and consider symptoms unique for a particular individual rather than using patient reported outcome measures or quality of life questionnaires designed for use on groups. Systematic and consecutive evaluation of the symptom triggering the initial blood sample was not performed in the present study. We do not have information on macro-TSH or polymorphism of the TSH-receptor in the participants, which could have been present. These conditions are exceedingly rare, and all but 2 participants had TPO-Ab and confirmed SCH by the inclusion protocol. Still, even if present in a participant, such a condition is unlikely to alter the conclusions. Finally, a now commonly used comprehensive questionnaire with focus on thyroid symptoms was not fully developed when our study was planned and conducted [32]. A questionnaire with a potentially lower floor-effect than the thyroid symptoms score used in the present study could provide information regarding the co-variation between thyroid hormone function and symptoms, which was absent in the present study. An extension of the present study with focus on the symptom or symptoms which triggered the individual patient to contact a doctor and using the more recent thyroid questionnaire is thus recommended.

Conclusion

Evaluation of symptoms in SCH using tools with symptom scores may help clinical decision making. We evaluated the validity and usefulness of such tools by ap-

plication of classical clinical test methodology. We found that the clinical relevant difference required between two measurements was rather large for mood-related symptoms (HADS) and hypothyroid symptoms (Zulewski), at 95 and 140%, respectively. For quality of life (SF-36) the required difference was 20%. We found no significant correlation between changes in these measures and changes in thyroid function in individual patients. Thus, the validity and usefulness of symptom scores is questionable for the HADS and Zulewski tools, while the better performance by SF-36 supports this as a candidate for helping clinical decision making, possibly even in individual patients.

Acknowledgments

The late Peter Laurberg was the principal organizer and co-author for the protocol of the project. Unrestricted public grants were kindly provided by "Speciallæge Heinrich Kopp's legat" and "Det Obelske Familiefond."

Statement of Ethics

Ethical approval by the Regional Ethics Committee in North-Jutland and Viborg County, Denmark, was obtained prior to study onset. All participants provided written informed consent.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

1 Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, 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 Feb;87(2):489–99.

2 Teng W, Shan Z, Teng X, Guan H, Li Y, Teng D, et al. Effect of iodine intake on thyroid diseases in China. N Engl J Med. 2006 Jun;354(26):2783–93.

3 Garmendia Madariaga A, Santos Palacios S, Guillén-Grima F, Galofré JC. The incidence and prevalence of thyroid dysfunction in Europe: a meta-analysis. J Clin Endocrinol Metab. 2014 Mar;99(3):923–31.

4 Andersen S, Pedersen KM, Bruun NH, Laurberg P. Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. J Clin Endocrinol Metab. 2002 Mar;87(3):1068–72.

5 Biondi B, Cooper DS. The clinical significance of subclinical thyroid dysfunction. Endocr Rev. 2008 Feb;29(1):76–131.

6 Villar HC, Saconato H, Valente O, Atallah AN. Thyroid hormone replacement for subclinical hypothyroidism. Cochrane Database Syst Rev. 2007 Jul;3:CD003419.

7 Stott DJ, Rodondi N, Kearney PM, Ford I, Westendorp RG, Mooijaart SP, et al.; TRUST Study Group. Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. N Engl J Med. 2017 Jun;376(26):2534–44.

8 Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI, et al.; American Association Of Clinical Endocrinologists And American Thyroid Association Taskforce On Hypothyroidism In Adults. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Thyroid. 2012 Dec;22(12):1200–35.

9 Pearce SH, Brabant G, Duntas LH, Monzani F, Peeters RP, Razvi S, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. Eur Thyroid J. 2013 Dec;2(4):215–28.

10 Carlé A, Pedersen IB, Knudsen N, Perrild H, Ovesen L, Laurberg P. Hypothyroid symptoms and the likelihood of overt thyroid failure: a population-based case-control study. Eur J Endocrinol. 2014 Nov;171(5):593–602.

11 Canaris GJ, Steiner JF, Ridgway EC. Do traditional symptoms of hypothyroidism correlate with biochemical disease? J Gen Intern Med. 1997 Sep;12(9):544–50.

12 Karmisholt J, Laurberg P. Serum TSH and serum thyroid peroxidase antibody fluctuate in parallel and high urinary iodine excretion predicts subsequent thyroid failure in a 1-year study of patients with untreated subclinical hypothyroidism. Eur J Endocrinol. 2008 Feb;158(2):209–15.

13 Bjorner JB, Damsgaard MT, Watt T, Groenvold M. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. J Clin Epidemiol. 1998 Nov;51(11):1001–11.

14 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983 Jun;67(6):361–70.

15 Zulewski H, Müller B, Exer P, Miserez AR, Staub JJ. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. J Clin Endocrinol Metab. 1997 Mar;82(3):771–6.

16 McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993 Mar;31(3):247–63.

17 Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. Crit Rev Clin Lab Sci. 1989;27(5):409–37.

18 Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 2—Correlation between subjects. BMJ. 1995 Mar;310(6980):633–633.

19 Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1—Correlation within subjects. BMJ. 1995 Feb;310(6977):446.

20 Fraser CG. Biological Variation. From Principles to Practice. Washington: AACC Press; 2001.

21 Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical Hypothyroidism and the Risk of Coronary Heart Disease and Mortality. JAMA. 2010;304(12):1365–74.

22 Jorde R, Waterloo K, Storhaug H, Nyrnes A, Sundsfjord J, Jenssen TG. Neuropsychological function and symptoms in subjects with subclinical hypothyroidism and the effect of thyroxine treatment. J Clin Endocrinol Metab. 2006 Jan;91(1):145–53.

23 Razvi S, Ingoe L, Keeka G, Oates C, McMillan C, Weaver JU. The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial. J Clin Endocrinol Metab. 2007 May;92(5):1715–23.
24 Carlé A, Pedersen IB, Knudsen N, Perrild H, Ovesen L, Laurberg P. Gender differences in symptoms of hypothyroidism: a population-based DanThyr study. Clin Endocrinol (Oxf). 2015 Nov;83(5):717–25.
25 Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado thyroid disease prevalence study. Arch Intern Med. 2000 Feb;160(4):526–34.
26 Somwaru LL, Arnold AM, Joshi N, Fried LP, Cappola AR. High frequency of and factors associated with thyroid hormone over-replacement and under-replacement in men and women aged 65 and over. J Clin Endocrinol Metab. 2009 Apr;94(4):1342–5.
27 Sawin CT, Geller A, Wolf PA, Belanger AJ, Baker E, Bacharach P, et al. Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. N Engl J Med. 1994 Nov;331(19):1249–52.
28 Faber J, Galløe AM. Changes in bone mass during prolonged subclinical hyperthyroidism due to L-thyroxine treatment: a meta-analysis. Eur J Endocrinol. 1994 Apr;130(4):350–6.
29 Selmer C, Olesen JB, Hansen ML, von Kappelgaard LM, Madsen JC, Hansen PR, et al. Subclinical and overt thyroid dysfunction and risk of all-cause mortality and cardiovascular events: a large population study. J Clin Endocrinol Metab. 2014 Jul;99(7):2372–82.
30 Carlé A, Pedersen IB, Knudsen N, Perrild H, Ovesen L, Andersen S, et al. Hypothyroid Symptoms Fail to Predict Thyroid Insufficiency in Old People: A Population-Based Case-Control Study. Am J Med. 2016 Oct;129(10):1082–92.
31 Parle J, Roberts L, Wilson S, Pattison H, Roalfe A, Haque MS, et al. A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: the Birmingham Elderly Thyroid study. J Clin Endocrinol Metab. 2010 Aug;95(8):3623–32.
32 Watt T, Hegedüs L, Groenvold M, Bjorner JB, Rasmussen AK, Bonnema SJ, et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. Eur J Endocrinol. 2010 Jan;162(1):161–7.