Thyroid hormone state and quality of life at long-term follow-up after randomized treatment of Graves' disease

Mirna Abraham-Nordling, Göran Wallin, Göran Lundell and Ove Töring

Department of Surgery, Institution of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Solna, Stockholm, Sweden, 1Department of Oncology, Institution of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Solna, Stockholm, Sweden and 2Division of Endocrinology, Department of Medicine, Institution of Clinical Research and Education, Sodersjukhuset, Stockholm, Sweden

(Correspondence should be addressed to M Abraham-Nordling who is now at Department of Surgery, Department of Molecular Institution of Surgery, Karolinska University Hospital, 17176 Stockholm, Sweden; Email: mirna.nordling@ds.se)

Abstract

Objective: In a 14–21 year follow-up of health-related quality of life (HRQL) outcome of 179 patients after randomized treatment of Graves' disease (GD) with surgical, medical or radioiodine, we found no differences. The HRQL for Graves' patients, however, was lower compared with a large age- and sex-matched Swedish reference population. We have now studied whether the reported HRQL-scores by Medical Outcome Study 36-item Short-Form Health Status Survey (SF36) and quality of life 2004 (QoL2004) answers were related to the thyroid hormone state of the patient.

Methods: This report comprises 91 of the original patients in which both the results of SF36 and QoL2004 questionnaire as well as serum thyroid hormones and current use of l-thyroxine treatment were available.

Results: A large number of the patients had low or undetectable serum TSH concentrations. SF36 scores and answers to QoL2004 questionnaires were not correlated to TSH levels or associated with suppressed TSH. A low free triiodothyronine was weakly associated with a low GH score (P < 0.02) and elevated thyrotropin receptor antibody with a low physical component summary (P < 0.02).

Conclusion: HRQL do not seem to be influenced by the thyroid hormone state of the patient including subclinical thyrotoxicosis. It is possible that the personality of GD patients as such may have resulted both in the development of GD and lower HQRL scores later on in life. Alternatively, the generic SF36 may not be a proper instrument to detect relevant differences in HRQL related to the thyroid state.

European Journal of Endocrinology 156 173–179

Introduction

Several studies have shown that patients with Graves’ disease (GD) and Graves’ ophthalmopathy have diminished health-related quality of life (HRQL) during and after treatment (1–6).

We have previously compared the outcome of treatment with surgery, antithyroid drugs (ATD) or 131I in a prospective, randomized study (7). Increased risk for the development or worsening of Graves’ ophthalmopathy after radioiodine was observed (7). No treatment-related differences in sick-leave or satisfaction with the treatment were observed (8). No significant differences in HRQL were found between the three treatments as estimated with a specific questionnaire developed for the purpose (‘QoL1996’). 3 years after initial treatment (9). Even after 14–21 years, there were no significant differences in HRQL by the first validated comparison published, which was related to the treatment modality for Graves’ hyperthyroidism (surgical, medical or 131I) (10). As instruments for HRQL in the long-term follow-up, the validated Medical Outcome Study 36-item Short-Form Health Status Survey (SF36) questionnaire combined with an updated version of QoL1996: ‘QoL2004’ was used (10).

The history of GD, however, did have a negative influence on HRQL on long-term compared with a healthy, age-matched reference population, especially with regard to mental performance and ‘vitality’ (10). Slight changes in the thyroid hormone state such as in subclinical thyrotoxicosis or hypothyroidism may lead to diminished HRQL (11).

In this report, we have therefore studied whether the reported QoL scores in SF36 and QoL2004 answers may have been influenced by the current thyroid hormonal status or the level of thyroxine (T4) substitution. Our hypotheses were that surrogate markers for the thyroid state of the patient such as serum thyroid-stimulating hormone (TSH), free T4 (fT4), free triiodothyronine (fT3), total T3 or of thyroid immunity (TRab or TPOab) were associated with or could predict the HRQL scores.
Material and methods

Subjects

In the present follow-up, 172 of the initial 179 patients between the ages of 20 and 55 years were included (10). The patients had been referred to our units between 1983 and 1990 due to Graves’ hyperthyroidism (7). Through May 2004, 172 of the 179 randomized patients were alive and were asked to participate in the study. Two patients were too ill (psychotic n = 1, cerebrovascular stroke n = 1) to be able to answer the questionnaires. Out of 179 patients, 147 answered the questionnaires, of which 145 were possible to evaluate. The blood samples were available in 113 of the 145 patients. Out of 113 patients, 22 had the blood samples analysed at another hospital with different reference ranges, had received more than one treatment for hyperthyroidism or the SF36 or QoL2004 response was lacking. These 22 patients are therefore excluded from the investigation. Thus, in total, the present report comprises 91 subjects in whom both blood analyses as well as SF36 or QoL2004 were available.

The study groups

Since there were no significant sex- and age-related differences in the results of HRQL questionnaires in our previous study (10), we have grouped the patients based on the mode of initial treatment only. Thus, the present report consists of the surgical group (n = 38), medical group (n = 17) and radioiodine group 131I (n = 27). The ‘additional 131I group’ consists of nine subjects, who in addition to the initial medical or surgical treatment had been given 131I (n = 9; Table 1).

Follow-up

The period from the last follow-up at 3 years after initiation of treatment (9) and until 2003. The follow-up time therefore consists of 14–21 years after initial treatment for Graves’ hyperthyroidism.

Study design

A self-assessment questionnaire was sent to each patient, which included (a) the Swedish version of the validated generic Medical Outcome Study (MOS) 36-item SF36; (b) some disease-specific treatment-related questions from ‘Qol.1996’ that we have used in the previous 3 years follow-up (9); (c) additional disease-specific items which addressed occurrence of other autoimmune diseases, osteoporosis, cardiac diseases or psychosocial events which are not covered by SF36; and (d) the patients were also asked to give a blood sample. The questions in (c) and (d) are comprised in the disease-specific ‘Qol.2004’. The generic HRQL instrument MOS SF36 is validated and well established (10, 12–15). From four physical (PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health) and four mental items (VT, vitality; SF, social functioning; RE, role-emotional; MH, mental health), a standardized physical component summary (PCS) and a mental component summary (MCS) were calculated which represent the deviation from the reference population in Sweden (14, 15) (for further details of the study design, SF36 and QoL2004, please see Ref. (10) with appendix.).

The study was approved by the local ethics committee of the Karolinska Institutet (KI no. 03-232).

Laboratory assessment

TSH was determined by chemiluminescent immunoassay (Beckman Coulter, Fullerton, CA, USA, reference range 0.2–4.0 mU/l; intraassay coefficient of variation (CV) value, 3.1%; interassay CV value, 3.86%). The lower detection limit is 0.01 mU/l. FT3 was determined by time-resolved fluoroimmunoassay (AutoDELFIA Triiodothyronine, Wallac Oy, Turku, Finland), reference range 4.0–7.0 pmol/l; intraassay CV value, 3.9%, interassay CV value, 4.4%. T3 was determined by chemiluminescent immunoassay (Beckman Coulter, reference range 1.1–2.5 nmol/l). FT4 was determined by time-resolved fluoroimmunoassay (AutoDELFIA Triiodothyronine, Wallac Oy, Finland), reference range 1.1–2.5 nmol/l. Antibodies to thyroid peroxidase (anti-TPO) was determined by chemiluminescent immunoassay (Nichols Advantage, San Clemente, CA, USA, reference range <2 kU/l; intraassay CV value, 3%, interassay CV value,

| Table 1 Overview of the groups and investigated parameters. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                  | Surgery n=38    | Medical n=17    | Radioiodine n=27 | Additional radioiodine n=9 |
| Sex (M/F)        | 4/34            | 3/14            | 2/25            | 2/7             |
| Age              | 59 ± 8          | 61 ± 9          | 64 ± 5          | 65 ± 6          |
| Follow-up years  | 17.6 ± 1.6      | 18 ± 1.8        | 17.6 ± 1.7      | 18.3 ± 1.5      |
| L-Thyroxine (μg/day) | 141.6 ± 36     | 116.7 ± 58      | 142.3 ± 29      | 169.4 ± 59      |
| TRab < 8 U/l     | 6.6 ± 4         | 6.5 ± 7         | 10.5 ± 25       | 7.8 ± 10        |
| TSH (mU/l)       | 0.9 ± 2         | 1.7 ± 2         | 0.5 ± 1         | 0.3 ± 0         |
| free T4 (pmol/l) | 19.3 ± 7        | 13.8 ± 3        | 19.6 ± 3        | 19.7 ± 3        |
| T4 (nmol/l)      | 140.2 ± 37      | 101.8 ± 24      | 142.3 ± 26      | 141.8 ± 26      |
| free T3 (pmol/l) | 5.6 ± 2         | 5.2 ± 1         | 5.6 ± 1         | 5.5 ± 1         |
| T3 (nmol/l)      | 1.7 ± 0         | 1.8 ± 0         | 1.7 ± 0         | 1.6 ± 0         |

Values given are means and s.d.
5%) and thyrotropin receptor antibody (TRab) was determined by radioreceptor antibody assay (BRAHMS, TRAB human, Henningsdorf, Germany, reference range < 8 U/l).

**Statistics**

The data were analysed using the software Statistica TM (Statsoft, Tulsa, OK, USA). Non-parametrical statistics were used (Kruskal-Wallis ANOVA by ranks, Spearman’s rank correlations, Mann–Whitney U-test and χ²-test) for further details cf. Ref. (10)). The association between the subscales SF36 and the variables; the four treatment groups, dose of ³¹I-T4, TRab, antibodies to thyroid peroxidase (TPOab), TSH, T4, fT4, T3, fT3 were performed using stepwise logistic regression analyses for ordinal response variables, a proportional odds model. The subscales in SF36 were categorized into four categories, according to percentiles 0–25, 25–50, 50–75 and 75–100%. The association between the QoL2004 ‘Do you feel well’ and the independent variables above was analysed by a stepwise logistic regression for binary response. Software used was SAS System 9.1 (SAS Institute Inc., Cary, NC, USA).

**Results**

**PCS and MCS in relation to serum TSH**

In the Surgical ATD- and ¹³¹I-treated groups, a substantial number of patients had a PCS or MCS score below the reference mean 50, independent of their serum TSH value (Fig. 1). However, it is apparent that most of the lower scores both for PCS and MCS are seen with suppressed or low-normal TSH values in the surgical and ¹³¹I-treated groups, which both had a large proportion of patients with suppressed (< 0.2 mU/l) or undetectable (<0.01 mU/l) serum TSH values. However, many subject with suppressed TSH also had PCS and MCS scores above the average, which were unrelated to the serum TSH concentrations (Fig. 1). The dissociation between TSH and PCS or MCS was seen in all four groups. There was no significant difference between the four groups in PCS (P = 0.97, n = 91, n.s.) or in MCS (P = 0.76, n = 91, n.s.) (ANOVA). The following analyses are therefore performed on the whole study group.

Forty-four patients (48%) had TSH within the reference range (0.2–4.0 mU/l) and PCS of 45.7 ± 11.5 (mean ± s.d.); range 18.7–57.6 and MCS of 47.1 ± 12.2; range 15.9–60.9. Forty-three patients (47%) had suppressed TSH (<0.2 mU/l) and PCS of 46.3 ± 10.7; range 23.5–62.4 and MCS of 47.2 ± 10.6; range 14.4–60.9. There was no statistical difference between the normal versus the suppressed TSH group in median PCS or median MCS (Mann–Whitney U-test). The remaining four patients had elevated TSH with no consistent pattern in PCS or MCS.

Sixteen patients had no T4 substitution. Two of those had TSH <0.2 mU/l and one had TSH >4 mU/l. Figure 2 shows the PCS and MCS results for the 16

![Figure 1](https://example.com/figure1.png)

**Figure 1** PCS and MCS in relation to log serum TSH in the treatment groups. The shaded area indicates the reference range for TSH (0.2–4.0 mU/l, log reference range –0.69 to 0.6). Please note the different scales on the X-axis.

www.eje-online.org
subjects without T4 substitution. There were no significant differences in the distribution of PCS ($P = 0.91$, n.s. $\chi^2$) or MCS ($P = 0.15$; n.s.) above or below the reference average of 50 for the population in relation to suppressed or normal TSH.

**PCS and MCS in thyroxine-substituted patients**

Seventy-five patients had T4 as thyroid hormone replacement therapy. The relation between serum TSH and PCS or MCS therefore has been analysed. In 31 cases (41%), the result of substitution was euthyroidism with normal serum TSH and in other cases ($n = 41$, 55%) suppressed TSH with subclinical exogenous thyrotoxicosis (Fig. 3a and b).

**PCS** Of patients with suppressed TSH, 24 (59%) had a PCS below average (50) for the reference population and 17 (41%) had a PCS above average (Fig. 3a). Of patients with normal serum TSH, 14 (45%) had a PCS below the reference average and 17 (55%) above 50. Three patients (4%) had elevated serum TSH. There was no significant difference between the median PCS in the normal versus the suppressed TSH group 45.1 $\pm$ 12.7; $n = 31$ versus 45.9 $\pm$ 10.7; $n = 41$ (Mann–Whitney U-test).

The number of PCS above or below 50 for patients with TSH $< 0.2$ mU/l was not statistically significantly different from the distribution of PCS above or below 50 for patients with TSH between 0.2 and 4.0 mU/l ($P = 0.26$, $\chi^2$).

**MCS** Suppressed TSH ($< 0.2$ mU/l) was seen in 41 of the 75 patients (55%) (Fig. 3b). Of those, 20 patients (49%) had MCS below average for the reference population ($< 50$) and 21 (51%) had a score $> 50$. Of the 31 patients with a normal TSH, 12 (39%) had MCS $< 50$ and 19 (61%) had MCS $> 50$. There was no significant difference between the median MCS in the normal versus the suppressed TSH group (46.9 $\pm$ 13.1; $n = 31$ vs 47.4 $\pm$ 10.9; $n = 41$, Mann–Whitney U-test).

The number of MCS above or below 50 for patients with TSH $< 0.2$ mU/l was not statistically significantly different from the distribution of MCS above or below 50 for patients with TSH between 0.2 and 4.0 mU/l ($P = 0.39$, $\chi^2$).

**Free T4, total T3, free T3, TRab and TPOab, and daily l-thyroxine dose**

MCS was positively correlated with serum-fT3 in the whole group ($R = 0.26$, $n = 89$, $P < 0.02$, Spearman rank). No correlation was found between fT3 and PCS. In addition, no other statistically significant correlations were seen between PCS or MCS and TSH, fT4, daily dose of T4, TRab or TPOab (Fig. 4).

**SF36: multivariate analyses by stepwise logistic regression**

Multivariate analyses of possible associations between PCS as well as MCS and TSH, dose of l-T4, fT4, total T3, fT3 and TPOab, TRab in the four treatment groups were performed. The analyses confirmed the absence of correlation between PCS or MCS scores and serum TSH, fT4, total T3, fT3 and TPOab.

However, there seems to be a relation between TRab and PCS since the odds ratio (OR) for lower PCS scores were 4.4 higher than for the patients who had TRab $\geq 8$ U/l (95% confidence interval 1.2–15.4) compared with the ones who had normal TRab ($< 8$ U/l) ($P < 0.02$).

Multivariate analyses were also performed for the eight subdomains in SF36. Only a weak relation between fT3 and GH was observed but not for PCS.

![Figure 2](https://www.eje-online.org)
A low fT3 was associated with a low score in the GH subdomain ($P < 0.02$).

**QoL2004**

From the QoL2004 HRQL questionnaire, there also seemed to be no relation between the level of serum TSH and the response to the well-being questions ‘Do you feel well now’ (11). Of 89 patients, 69 (78%) among the combined treatment group answered ‘yes’ to the question (Table 2). Of those, 32/69 (46%) had a suppressed TSH ($<0.2$ mU/l), of which 27 (39%) had TSH below 0.1 mU/l and 35 (51%) had normal TSH. Among those who answered ‘no’ to the question

![Figure 3](image1.png)  
**Figure 3** PCS (a) and MCS (b) in the patients with L-thyroxine substitution ($n=75$) in relation to log serum TSH. The shaded area indicates the reference range for TSH (0.2–4.0 mU/l, log reference range $-0.69$ to 0.6).

![Figure 4](image2.png)  
**Figure 4** PCS and MCS in the patients with TSH levels below 0.1 mU/l ($n=34$) and thyroxine substitution versus the L-thyroxine substitution dose/day. The scores at dose 150 μg/day are shifted to ease the interpretation of the figure.
Table 2  Self-reported answers to the QoL2004 questionnaire among the subjects where blood sample was also available (TSH: n=89; free T4 (fT4): n=84).

<table>
<thead>
<tr>
<th>Do you feel well now?</th>
<th>N</th>
<th>%</th>
<th>&lt;0.01</th>
<th>0.01–0.1</th>
<th>0.1–0.2</th>
<th>0.2–4.0</th>
<th>&gt;4.0</th>
<th>L</th>
<th>N</th>
<th>H</th>
<th>L</th>
<th>N</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>69</td>
<td>78</td>
<td>0</td>
<td>27</td>
<td>5</td>
<td>35</td>
<td>2</td>
<td>2</td>
<td>57</td>
<td>6</td>
<td>1</td>
<td>43</td>
<td>21</td>
</tr>
<tr>
<td>No</td>
<td>20</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>4</td>
</tr>
</tbody>
</table>

The reference range for serum TSH is 0.2–4.0 mU/l and for fT4 10–20 pmol/l. L, low; N, normal and H, high; L < 10 pmol/l; N = 10–20; H > 20 pmol/l.

‘Do you feel well now’, 9/20 (45%) had suppressed TSH, 7 (35%) had TSH below 0.1 mU/l and 10 had a normal TSH value. There is no significant difference in the number of answers ‘yes’ and ‘no’ between suppressed and normal serum TSH ($\chi^2$, n.s.) (Table 2). Neither did the level of fT3 or fT4 significantly influence the ‘yes’ versus ‘no’ responses ($\chi^2$, n.s.).

QoL2004: multivariate analyses  A similar analysis, as for SF36, was performed for the answer to the question ‘Do you feel well now’ classified as ‘yes’ or ‘no’ with the same variables as above. No significant correlations were observed between any of the variables and the answer to the question.

Discussion

We have recently reported that GD has negative consequences for the patients’ HRQL even after 14–21 years, especially with regard to mental performance and ‘vitality’ (10). However, the mode of treatment for Graves’ hyperthyroidism whether surgical, medical or radioiodine, had little impact on HRQL in the long-term.

In the present study, we have analysed whether the thyroid state of the subjects at the time of response to the questionnaires could explain why the GD patients had lower HRQL compared with a large, normal Swedish reference population (10). Provided the subjects have normal pituitary and hypothalamic function, and we have no reason to believe otherwise, a normal serum TSH is the most valid and recognized surrogate marker for a normal thyroid state and has therefore been used.

However, we found no association between serum TSH and PCS or MCS. Among subjects with suppressed TSH, there was the same proportion of MCS and PCS scores above as below the average of 50 for the reference population. And that also holds for patients with TSH within the reference range. The stepwise logistic regression analyses gave the same result. Taken together, PCS or MCS or subdomain score do not seem to be substantially influenced by thyroid hormone state when TSH was used as an indicator (Figs 1–3).

As serum TSH showed a skew distribution, with 38% of the serum TSH concentrations below 0.02 mU/l, we also analysed whether PCS, MCS and the subdomain scores (PF, RP, BP, GH and VT, SF, RE, MH) could be predicted by serum-fT4, total T3 or the fT3 concentrations. No association was seen for fT4, total T3 or fT3 in MCS or PCS but we did see a weak association between a low serum-fT3 and lower GH scores.

A weak association between elevated TRab and lower PCS was also observed. The association between TRab and PCS is a new observation but difficult to interpretate until further studies substantiate this finding.

Consistent evidence indicates that both exogenous and endogenous subclinical hyperthyroidism reduces the HRQL (11). However, in the present study, the presence of exogenous subclinical hyperthyroidism did not significantly reduce the HRQL by SF36 scores as outcome.

The QoL2004 questionnaire contained the question ‘Do you feel well now’ stated at the end of GD-related questions (10). Although the QoL2004 questionnaire has not been validated, we assume that the answers reflect the subject’s apprehension of their present thyroid state and therefore included the response in the present study. With this reservation in mind, we found no relation between serum TSH, fT4, total T3 or fT3 and the response pattern. Again, the thyroid hormonal state was not reflected by the QoL2004 instrument.

The present study represents a subgroup of the previously reported GD group used in the first validly performed comparison of HRQL and in which no differences between the three treatment modalities were found (10). The inclusion of subjects in the present study was based on the availability of serum thyroid hormone analyses and response to SF36 and QoL2004. However, the results from the present subgroup corroborate our previous results of similar HRQL pattern despite the mode of treatment (10).

Several questions can be raised. First of all, the absent relation between the thyroid hormonal state and the SF36 and QoL2004 scores indicates that the diminished HRQL on long-term in GD are not related to the thyroid hormone levels and therefore may have other explanations. It is conceivable that GD patients, for example, are more susceptible to various forms of stress (which is a well-known eliciting factor for GD), which expectedly may impact QoL and HQRL despite treatment for GD. Alternatively, it can be imagined that the personality of the subject as such may imply difficulties in adapting to
many aspects of life even before GD and perhaps this is partly responsible for the GD, so that the subject many years after treatment has lower HRQL scores, as in the present study. It is therefore a question whether the lower HRQL scores later on are disease specific or may be caused by confounding factors, which are not related to the patient’s history of GD as such. Recently, lower HRQL using SF36 has also been observed, for example in asymptomatic, untreated primary hyperparathyroidism with borderline hypercalcaemia, which supports this view (16). Secondly, measuring HRQL in a particular disease such as GD pose several problems. It is possible that the validated but generic MOS SF36 and unvalidated QoL2004 are too insensitive instruments for a proper evaluation of HRQL in patients with a history of GD. The generic MOS SF36 poses general questions on health status and can be used for comparison of HRQL aspects in different diseases (12). SF36 does not cover the adequate physical, mental, cognitive, social well-being and other GD-specific items in particular as well as a thyroid disease-specific questionnaire would have done. One such thyroid disease-specific instrument is the ‘Hyperthyroid Complaint Questionnaire’ (HCQ) but the HCQ and other instruments, however, are not available in Swedish and has not been validated in a Swedish reference population (5, 6).

Conclusions

Our previous observations of reduced HRQL in patients with a history of GD compared with age- and sex-matched healthy Swedes seem not to be explained by the patient’s thyroid hormone status at the time of follow-up. HRQL scores by SF36 below or above the average value for the Swedish reference population was seen independently of the serum concentrations of TSH, fT4 or total T3. fT3 was weakly correlated with GH. The findings are somewhat unexpected and could indicate that GD patients may represent a special group of subjects, who would have produced a low SF36 score independently of their previous GD. Alternatively, the SF36 and QoL2004 instruments, which are not thyroid disease specific, may not be sensitive enough.

Acknowledgements

We wish to thank Professor B Hamberger for his interest, important and valuable help throughout the study. We wish also to thank Elisabeth Berg, LIME Karolinska Institutet for valuable statistical assistance. The study was supported by the Swedish Research Council and the Karolinska Institutet.

References