Variation in Thyroid Function Tests in Patients with Stable Untreated Subclinical Hypothyroidism

Jesper Karmisholt, Stig Andersen, and Peter Laurberg

Objective: Knowledge of variation in thyroid function is important for interpretation of thyroid function tests. We aimed to describe intra-individual variation in thyroid function in patients with stable, untreated subclinical hypothyroidism (SCH) compared to euthyroid individuals to assess the importance of monitoring SCH patients.

Design: We measured thyrotropin (TSH), free thyroxine (fT4), and free triiodothyronine (fT3) monthly for 1 year in a longitudinal study of 15 untreated SCH patients with initial TSH 5–12 mU/L, without trends in TSH, and compared findings with results from 15 euthyroid individuals.

Main outcome: CV% was 17.0, 6.1, and 6.2 for TSH, fT4, and fT3, respectively. Overall CV% for TSH was lower in SCH patients than controls. Contrary to euthyroid individuals, CV% in SCH patients increased with rising mean TSH ($r^2 = 0.29$, $p = 0.04$). Individual disease set points were established with 45, 6, and 6 tests for TSH, fT4, and fT3, with 95% confidence. Differences required between two test results were 40%, 15%, and 15%, respectively, with 90% confidence.

Conclusion: Percent variation in TSH was lower in SCH than in euthyroid controls, but increased with higher mean TSH. The number of tests needed to establish disease set points was high. The difference required between two tests to be truly different was 40% for TSH and 15% for fT4 and fT3.

Introduction

Variation in thyroid function tests is important because it influences the reliability of reference ranges and the interpretation of test results (1). Variation in thyroid function tests was described in euthyroid individuals, which allowed for an estimate of the reliability of population-based reference ranges (2–7). Also, this provided new insight into the entity subclinical thyroid disease (6).

Subclinical hypothyroidism (SCH) is a state with elevated serum thyrotropin (TSH) and an estimate of serum thyroxine (T4) within reference limits of the assay. SCH is frequent in most population studied (8–11). The risk of progression to overt thyroid failure varied between studies, but thyroid function was stable in the majority of SCH patients in all studies (12–16).

Alterations in thyroid function tests in patients with SCH may be part of normal changes in thyroid function or be caused by a further decline in thyroid function. In patients with stable SCH, alterations in thyroid function tests originate from individual biological variation around a diseasespecific set point. This variation may consist of individual biological variation as seen in healthy individuals (1) and possibly added variation attributable to fluctuations in thyroid disease activity. The variation in thyroid function tests in SCH may differ from that observed in the euthyroid state, and thus alter the conditions for interpretation of thyroid function tests in SCH (1).

Knowledge of variation in thyroid function tests in SCH may improve the validity of interpretation of these tests, which is important for monitoring development of further thyroid failure in patients with SCH. However, no information is available on variation in thyroid function tests in patients with stable SCH.

We collected longitudinal data in a group of individuals with different degrees of SCH and identified a subgroup of patients with no trend in TSH over time. Data from these patients were used to describe the intra-individual variation in thyroid function tests in the SCH state and to calculate the number of tests needed to identify the individual level of thyroid dysfunction and the difference required for two test results to be significantly different in SCH.

Materials and Methods

Patients were referred to our department after finding SCH based on TSH at their family physician and included if TSH was between 5 and 12 mU/L and total T4 was within the...
Table 1. Characteristics of Individual Patients with Subclinical Hypothyroidism (SCH, n = 15) and a Group of Euthyroid (ET, n = 15) Individuals

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Sex (male/female)</th>
<th>Age (yr)</th>
<th>BMI (kg/m²)</th>
<th>TSH (mU/mL) Mean SD</th>
<th>fT₄ (pmol/L) Mean SD</th>
<th>fT₃ (pmol/L) Mean SD</th>
<th>TPOAb (pos/neg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>f</td>
<td>58</td>
<td>34.9</td>
<td>12.3 (4.4)</td>
<td>11.7 (1.2)</td>
<td>4.3 (0.3)</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>f</td>
<td>71</td>
<td>24.1</td>
<td>6.9 (1.2)</td>
<td>11.7 (0.8)</td>
<td>4.7 (0.4)</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>f</td>
<td>73</td>
<td>24.7</td>
<td>6.4 (1.1)</td>
<td>12.7 (0.6)</td>
<td>4.7 (0.4)</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>f</td>
<td>78</td>
<td>30.3</td>
<td>5.3 (0.3)</td>
<td>13.3 (0.4)</td>
<td>4.3 (0.1)</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>f</td>
<td>51</td>
<td>33.1</td>
<td>6.6 (1.7)</td>
<td>13.3 (1.0)</td>
<td>5.3 (0.4)</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>f</td>
<td>52</td>
<td>26.8</td>
<td>9.5 (2.4)</td>
<td>14.3 (1.1)</td>
<td>4.9 (0.3)</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>m</td>
<td>62</td>
<td>39.3</td>
<td>6.4 (0.7)</td>
<td>12.3 (1.1)</td>
<td>5.5 (0.3)</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>f</td>
<td>68</td>
<td>27.8</td>
<td>6.4 (0.7)</td>
<td>14.6 (0.4)</td>
<td>5.3 (0.3)</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>f</td>
<td>65</td>
<td>21.9</td>
<td>5.3 (1.0)</td>
<td>13.6 (1.0)</td>
<td>4.4 (0.2)</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>f</td>
<td>56</td>
<td>25.6</td>
<td>4.4 (0.6)</td>
<td>14.7 (0.8)</td>
<td>4.6 (0.3)</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>f</td>
<td>27</td>
<td>25.2</td>
<td>7.3 (2.1)</td>
<td>12.7 (0.8)</td>
<td>5.2 (0.2)</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>f</td>
<td>41</td>
<td>26.5</td>
<td>5.6 (1.2)</td>
<td>12.5 (0.7)</td>
<td>5.0 (0.4)</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>f</td>
<td>55</td>
<td>25.9</td>
<td>11.6 (1.6)</td>
<td>11.9 (0.6)</td>
<td>5.2 (0.2)</td>
<td>+</td>
</tr>
<tr>
<td>ET, all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean, (SD)</td>
<td></td>
<td></td>
<td>41.0 (25.1)</td>
<td>1.3 (1.3)</td>
<td>12.2 (22 pmol/L)</td>
<td>3.1–6.8 pmol/L</td>
<td></td>
</tr>
</tbody>
</table>

Laboratory reference ranges 0.3–4.2 mU/L 12–22 pmol/L 3.1–6.8 pmol/L

*BMI, body mass index; TSH, thyrotropin; fT₄, free thyroxine; fT₃, free triiodothyronine; TPOAb, thyroid peroxidase antibody. TSH, fT₄, and fT₃ values are averages of 13 consecutive monthly measurements.

Statistics and computations

Conformity to normal distribution of data in individuals was tested using Shapiro–Wilks test and inspection of Q-Q plots. The two groups were only limited skewed and nontransformed data were used for calculations as recommended (19). Coefficients of variation (CV) were calculated as the standard deviation (SD) divided by the mean and expressed...
as percentages. Mann–Whitney test was used for comparison between the groups. Linear regression analysis was done with measures of variation (SD and CV) as response variables and mean values of TSH, fT4, and fT3 as explanatory variables.

The patients varied around a set point for their individual disease state. The number of samples needed to establish this disease set point was calculated using the formula recommended for use when estimating the number of specimens required in biochemical measures: 

\[ n = \left( \frac{Z \times CV_{\text{total}}}{D} \right)^2 \]  

(19). \( Z \) was the number of standard deviations set to define the confidence interval (CI); i.e., \( Z = 1.96 \) for a 95% CI. \( CV_{\text{total}} \) included \( CV_{\text{preanalytical}} \), \( CV_{\text{analytical}} \), and \( CV_{\text{biological}} \), thus \( CV_{\text{total}} \) was equal to the measured CV in our data. Median \( CV_{\text{total}} \) was used for calculations. \( D \) was percent closeness to homeostatic set point; i.e., the precision level.

The difference required for two test results to be significantly different was calculated from: 

\[ Z \times CV_{\text{total}} \times 2^{1/2} \]  

according to recommendations (20). Again, \( Z \) was confidence interval; i.e., \( Z = 1.65 \) for 90% CI and 1.96 for 95% CI.

A \( p \) value of \(< 0.05 \) was considered significant. The Statistical Package for the Social Sciences version 11.0 (SPSS, Inc, Chicago, IL) and Excel 2003 (Microsoft Corp., Redmond, WA) were used.

**FIG. 1.** Thirteen consecutive monthly measurements of serum thyrotropin (TSH), free thyroxine (fT4), and free triiodothyronine (fT3) in 15 subclinical hypothyroidism (SCH) patients, sorted by increasing mean values. Each dot represents one monthly measurement. Patient numbers refers to patient 10 in Table 1.
Results

Table 1 shows the characteristics of the patients and mean values of the 13 consecutive monthly thyroid function tests in each individual. Table 1 also includes mean values of the group of 15 euthyroid individuals used for comparison (6).

Among the SCH patients TSH followed a normal distribution in all but two patients (number 1 and 6). FT4 and fT3 were normally distributed in all but two (patients 5 and 15).

In the group of healthy euthyroid subjects, all variables followed the normal distribution.

Variation in thyroid function test in stable SCH patients

Figure 1 depicts the 13 measurements in each of the 15 SCH patients sorted by increasing mean values for TSH (upper panel), fT4 (middle panel), and fT3 (lower panel).

SD of individual serum TSH ranged from 0.3 to 4.7 mU/L with mean SD 1.7 and median SD 1.1 mU/L. CVtotal ranged from 6% to 36% with a median and interquartile range (IQR) which is the 25th–75th percentile range of 17% (13–21%). SD increased with increasing mean TSH levels (linear regression, \( p < 0.001 \)) as did variation corrected for mean value (CV%) (linear regression, \( p = 0.04 \)). \( R^2 \) from regression of TSH mean and SD was 0.66 compared to 0.29 for TSH mean and CV%.

In the euthyroid individuals median SD of TSH was 0.29 mU/L (IQR 0.20–0.36 mU/L) whereas median (IQR) of CVtotal was 23% (18–28%). CVtotal for TSH was lower in the SCH patients compared to the euthyroid individuals (Mann–Whitney, \( p = 0.03 \)), but there was no change in CV with changing mean TSH in the euthyroid individuals, as can be read from Fig. 2.

SD ranged from 0.36 to 1.24 pmol/L for fT4 and from 0.12 to 0.43 pmol/L for fT3, and did not increase with increasing mean values (linear regression, fT4, \( p = 0.65 \); fT3, \( p = 0.45 \)). Median CVtotal for fT4 and fT3 was 6.1% and 6.2% with IQR of 5.2–7.4% and 4.5–7.4%, respectively.

Tools for monitoring SCH patients

Table 2 shows the number of samples needed to assess the disease-specific set point for TSH, fT4, and fT3 for different precision levels in patients with stable SCH. A median of 12 (IQR, 7–26) samples were needed in SCH patients to be 95% confident of disease set point for serum TSH with a 90% precision level, while a precision level of 80% required a median of three (IQR, 2–6) samples. The number of samples needed for fT3 or fT4 disease-specific set point was lower. For fT4 or fT3, two samples established the individual disease set point with a 90% precision and a 95% significance level.

Figure 3 shows the probability of a significant change in test results with increasing relative differences between two specimens of TSH and of fT4 or fT3. A lower CV% shifts the curve to the left whereas a higher CV% shifts the curve to the right; the IQR is shown for TSH. To be 90% confident of a significant difference requires a median 40% difference between two TSH tests and 15% difference between two tests of fT4 or fT3.

Discussion

Variation in TSH secretion from the pituitary gland is generated by the hypothalamic circadian and pulsatile

Table 2. Number of Monthly Samples Needed to Calculate Disease-Specific Set Point for Serum TSH and Thyroid Hormones with Different Precision Levels in Patients with Stable Untreated Subclinical Hypothyroidism

<table>
<thead>
<tr>
<th>Precision level (D)</th>
<th>No. of TSH tests</th>
<th>No. of fT4 or fT3 tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80% significance (Z)</td>
<td>95% significance (Z)</td>
</tr>
<tr>
<td>60%</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>70%</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>80%</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>90%</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>95%</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>99%</td>
<td>474</td>
<td>1110</td>
</tr>
</tbody>
</table>

*Variation was similar for fT4 and fT3, as was the number of samples needed. Number of samples needed to estimate disease-specific set point was calculated as \( n = (Z \times CV_{total}/D)^2 \). Z refers to number of standard deviations required for a confidence interval (CI), i.e., \( Z = 1.96 \) for a 95% CI. CVtotal was the measured CV in our data. D is percent closeness to disease set point, i.e., the precision level (19). TSH, thyrotropin; fT4, free thyroxine; fT3, free triiodothyronine.
This diversity in variation between TSH and thyroid hormones was independent of mean values. We studied a cohort of patients with untreated SCH without a significant trend in TSH during a year. These patients displayed higher standard deviations (absolute variation) in TSH compared to euthyroid individuals. On the other hand, the relative variation (standard deviations relative to mean values, CV%) was lower in the SCH patients. Both the absolute and the relative variation in TSH were higher with higher mean TSH in the SCH patients. This was in contrast to the euthyroid individuals where no difference in the variation was seen with rising mean TSH.

It could be speculated that this progressive increase in the relative variation of TSH with more pronounced thyroid failure in stable SCH is caused by either a higher variation of hypothalamic signalling or an increased sensitivity of the thyrotropes in the pituitary gland and an increased production capacity for TSH related to lower levels of thyroid hormones. The fall in relative circadian variation in TSH in thyroid failure (21) favors the latter rather than the former mechanism. Another mechanism that could add to the increasing variation in the SCH state is that patients with more pronounced thyroid failure and higher TSH have higher variation in disease activity between months.

While variation in TSH was higher with higher mean TSH, variations in T₃ and T₄ were independent of mean values. This diversity in variation between TSH and thyroid hormones is consistent with different secretion mechanisms from the pituitary and the thyroid gland (22–24).

The majority of our screened patients had stable SCH, in accordance with the findings in a number of previous studies (12–16). In such patients, thyroid function tests vary around a disease-specific set point. This disease set point was established with 95% confidence with six monthly T₄ and T₃ tests, while six monthly measurements established TSH disease set point with 86% confidence. For TSH, two to three tests described the disease set point with around 80% confidence, while 95% confidence required 45 samplings. A large number of samples may be achieved through repeated sampling in subjects suspected of developing SCH, but is cumbersome and will hardly prove valuable in general use.

Progression of SCH to overt hypothyroidism is an important issue. The overall risk of progression to overt hypothyroidism has been settled in population studies (25) while the issue is more tricky in the individual patient. To guide evaluation we found that a 40% difference was required between two test results to indicate a true difference for TSH in the range between 4.4 and 12 mU/L. Due to more narrow variation in T₄ and T₃, the difference required for similar changes in T₄ and T₃ was approximately 15%.

We used the longitudinal design. This allows a description of the intra-individual variation in the individual participants in the SCH state, where the between-individual variation is determined by the selection of participants. Thus, the cross-sectional design should not be used in the study of variation in thyroid function in SCH patients.

The euthyroid individuals used for comparison were all men, while the majority of SCH patients were women. However, as no difference was found between genders in variation in thyroid function tests in previous studies (1,23) we were confident with the comparisons. Our calculations did not include between-batch analytical variations. This would increase variation slightly, but is limited compared to the large biological variation (1), and would not alter the conclusions.

In conclusion, the relative variation in TSH was lower in SCH than in the euthyroid individuals used for comparison. Variation in TSH increased with increasing thyroid failure, while variation in thyroid hormones was unaltered in these patients with stable SCH with TSH initially up to 12 mU/L, as judged by monthly samplings during 1 year. Furthermore, for TSH levels below 12 mU/L, a second TSH had to be at least 40% different from a previous test to suggest a difference with 90% confidence. Similarly, T₄ and T₃ had to differ by 15% to denote a significant difference.

Acknowledgments

We are indebted to laboratory technicians Ingeborg Leegaard, Anne-Mette Christensen, and Anette Godske for invaluable assistance with biochemical analyses, and to Roche Diagnostics for providing the test kits.

References


Address reprint requests to:
Jesper Karmisholt, M.D.
Dept. of Endocrinology, Medicine
Aalborg Hospital, Aarhus University Hospital
Hobrovej 18-22
9000 Aalborg
Denmark

E-mail: jsk@rn.dk