Hypothyroid symptoms and the likelihood of overt thyroid failure: a population-based case–control study

Allan Carlé1,2, Inge Bülow Pedersen1, Nils Knudsen3, Hans Perrild3, Lars Ovesen4 and Peter Laurberg1,5

1Department of Endocrinology, Aalborg University Hospital, DK-9000 Aalborg, Denmark, 2Diagnostic Centre, Region Hospital Silkeborg, Silkeborg, Denmark, 3Department of Endocrinology, Bispebjerg University Hospital, Copenhagen, Denmark, 4Department of Internal Medicine, Slagelse Hospital, Slagelse, Denmark and 5Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Abstract

Background: It is generally accepted that patients suffering from hypothyroidism may express few symptoms, but this has not been studied in a population-based study design.

Objectives: To study the array of symptoms as they are reported in newly diagnosed overt autoimmune hypothyroidism using a population-based case–control design.

Methods: Patients with new overt autoimmune hypothyroidism (n = 140) and their individually matched thyroid disease-free controls (n = 560) recruited from the same population underwent a comprehensive program and self-reported a number of symptoms. We identified the symptoms associated with overt hypothyroidism and calculated positive (LR+) and negative (LR–) likelihood ratios as well as diagnostic odds ratios (DORs) as measures for the association between disease state and symptoms.

Results: Among 34 symptoms investigated, 13 symptoms were statistically overrepresented in hypothyroidism. Hypothyroid patients suffered mostly from tiredness (81%), dry skin (63%), and shortness of breath (51%). Highest DORs (95% CI) were reported for tiredness (5.94 (3.70–9.60)), hair loss (4.58 (2.80–7.51)), and dry skin (4.09 (2.73–6.16)). A hypothyroidism-component-score was defined as the number of hypothyroidism-associated symptoms (range: 0–13). LR+ for participants with a hypothyroidism-component-score of 0 was 0.21 (0.09–0.39), meaning that the post-test probability was lowered to 21% of what it was before asking for symptoms. LR+ for scores of 1–2/3/4–6/7–9/10–13 were: 0.47 (0.30–0.72)/1.16 (0.70–1.87)/1.90 (1.29–2.45)/3.52 (2.30–5.36)/6.29 (2.30–17.7).

Conclusions: None of the individual symptoms of hypothyroidism had high LRs or DORs. Thus, neither the presence nor absence of any individual hypothyroidism symptom was reliable in the decision making of who should have their thyroid function tested. Therefore, even minor suspicion should lead to a blood test.

Introduction

We previously calculated the lifetime risk for developing overt hypothyroidism in Denmark to be 2.3%, with a threefold excess in women (3.5 vs 1.0% in men) (1). Thus, hypothyroidism is a common disorder, and healthcare providers often have to consider which patients should have their thyroid function tested in order to discover undetected hypothyroidism.

Hypothyroidism may present with a variety of symptoms reflecting the hormonal insufficiency in different organs (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15). Often symptoms develop insidiously and may go unnoticed for prolonged periods, as already described more than 100 years ago by Murray who claimed that ‘patients in the earlier stages of thyroidal disease may go about for
a considerable period without the true nature of the disorder being recognized' (16).

Several studies have reported low prevalence of symptoms and signs in hypothyroidism, and similar prevalence of symptom has often been recorded in healthy control subjects. Various symptom scores have been introduced, some of which weigh each symptom equally (3, 4, 5, 17), whereas others apply different weights to the complaints (6, 7, 8, 9, 10, 11, 12). However, these scores tend to have quite low sensitivity and specificity values in statistical workout, and the accuracy for identifying subjects with undiagnosed hypothyroidism is low.

We carried out a study recruiting patients newly diagnosed with overt autoimmune hypothyroidism in the Danish population and aimed to study the relative frequency and importance of various symptoms. Furthermore, we studied the same symptoms in individually matched volunteers recruited from the same population in order to analyze symptoms that may be statistically associated with overt hypothyroidism. Finally, we examined whether the total number of symptoms may be helpful in the clinical setting to decide who should have their thyroid function tested.

Subjects and methods

The Danish Investigation on Iodine Intake and Thyroid Diseases (DanThyr) was established in 1997. A population of 538 734 citizens was under surveillance during the study period March 1997 to December 2000 (2 027 208 person-years of observation).

Patients

Two geographical areas were chosen for this study, an area in and around the city of Aalborg (population, n=311 102) and in Copenhagen (n=227 632). A register was linked to the diagnostic laboratories responsible for all thyroid function testing performed. The subjects identified by the register as possible new cases with a first time elevated serum TSH and a low thyroxine (T4) estimate were individually evaluated. The physician requesting the blood sample was contacted, and hospital registers and records as well as later thyroid function tests performed were scrutinized in order to verify or disprove incident overt autoimmune hypothyroidism (1). The diagnosis was verified, if patients had sustained hypothyroidism 3 weeks or later and/or T4 substitution therapy was initiated.

In the study period, we identified, verified, and classified 578 patients with incident primary autoimmune overt hypothyroidism (Fig. 1). Staff were only available in some periods, where we invited patients newly diagnosed with new autoimmune hypothyroidism to participate in an investigational program as described in detail previously (18). Of the 247 patients invited, 147 (59.5%) gave full participation. Patients joined our comprehensive investigational program on average 50 (interquartile range (IQR): 36–78) days after the first thyroid function test suggesting overt hypothyroidism.

The surveillance program and the use of a register linked to laboratory databases (19), the diagnostic criteria used to identify patients, the algorithm for final verification, and the classification of thyroid failure into overt autoimmune hypothyroidism have been presented previously (1).

Controls

We contacted the institution responsible for the Civil Registration System, in which all Danish civilians are registered with a unique identification number. The subjects were randomly selected from the register and invited to participate in two surveys if they lived within the two study areas. For each patient, we included from two simultaneously ongoing surveys four randomly selected control subjects matched on sex, age, and region (Fig. 1). Survey I, a cross-sectional population study, was performed in Aalborg and Copenhagen in 1997–1998 (20). Women aged 18–22, 25–30, 40–45, and 60–65 years and men aged 60–65 years participated (total, n=4649). In order to obtain four controls for each hypothyroid patient, we additionally invited from the same population in 1998 (survey II; n=684) four control subjects falling outside these age categories. For the present study, we finally selected controls with serum TSH between 0.2 and 5.0 mU/l and with no previous thyroid disease. We were able to find 560 control subjects (survey I+II; n=218+342) matching 140 patients diagnosed with autoimmune hypothyroidism. We excluded from the present study, seven cases, to whom we were not able to identify four age-matching controls.

Questionnaires

Patients filled out questionnaires comprising information on symptoms experienced (n=38, see Supplementary Data, see section on supplementary data given at the end of this article) before hypothyroidism was diagnosed. The controls answered the same questionnaires that involved questions about neck discomfort, respiratory complaints,
cardiac and gastrointestinal symptoms, symptoms elucidating hair loss or dry skin, a number of psychiatric symptoms, and some other symptoms. Two questionnaires were used. In the ‘Formal questionnaire’ (symptoms, n Z 16; see Supplementary Data) participants were asked for the presence of a number of symptoms previously reported to be associated with hypothyroidism in literature (F1–F8). We added a number of symptoms comprising psychiatric complaints (F9–F11), and a number of symptoms which may be present in patients presenting thyroid enlargement (F12–F16). Participants also answered a ‘Mental vulnerability questionnaire’ (21, 22), in which additional symptoms (n Z 22) were addressed. Some of these questions (M2, M6, and M14) were already encountered by the ‘Formal questionnaire’, and the question about feeling unwell (M15) was regarded unspecific and also addressed in various other questions. Thus, 34 questions were included in our final analysis (F1–F16, M1, M3–M5, M7–M13, and M16–M22).

The ‘Mental vulnerability questionnaire’ has its origin from a 22-item screening score of primarily psychophysiological symptoms originally presented by Langner (22), and later validated in a group of 6546 Danish subjects attending the cross-sectional Inter99 Study (21). The validation process revealed that the 22-item ‘Mental vulnerability questionnaire’ was superior to four other scales in both sexes (21).

Participants also had their height and weight measured, and BMI was calculated according to the Quetelet index (kg/m²).

**Blood specimen analyses**

Blood was drawn at investigation and serum kept frozen at −20 °C. After study end, thyroid peroxidase antibodies (TPOAb) and thyroglobulin antibodies (TgAb) from participants’ serum specimens were measured in random order (18, 23). The subjects with antibody concentrations above the functional sensitivity given by the manufacturer were regarded as antibody-positive (TPOAb+, > 30 kU/l; TgAb+, > 20 kU/l). TSH was measured in serum drawn from controls. For all patients diagnosed with overt autoimmune hypothyroidism.
hypothyroidism, our laboratory monitoring database had registered serum TSH and T4 concentrations at disease identification before i-T4 treatment was started. Details on TSH and T4 analyses have been given previously (1, 19).

**Statistical analyses**

For calculations and statistical analyses, we used IBM Statistical Package for Social Sciences version 15.0 (SPSS). Medians and IQR (25–75% ranges) were calculated. The percentages of patients and controls suffering from each of the 34 symptoms were calculated and frequencies were compared using Pearson’s \( \chi^2 \)-test with Bonferroni’s correction for multicomparison test \( (P<1-34/0.95, \text{i.e. } <0.0015) \) to identify hypothyroidism-associated symptoms. Similar identification of hypothyroidism-associated symptoms has been reported in several other studies (5, 8, 10, 13). We calculated a ‘simple hypothyroidism-component-score’ for each of the 700 participants with the number of hypothyroidism-associated symptoms reported. In addition, we calculated a ‘weighed hypothyroidism-component-score’ by adding up weighed points for each of the 13 symptoms according to the principle by Charlson et al. (24). For symptoms overrepresented in hypothyroidism by a diagnostic odds ratio (DOR) in the range of 2–2.99/3–3.99/4–4.99/5+, we applied 1/2/3/4 points to the weighed additive score and 1/2/4/8 points to the weighed exponential score (for DOR, shown later). For both the simple and the two-weighted scores, we tested the power of discrimination by means of the area under a receiver-operated characteristic curve (AUROC; values of \(<0.6/0.6/>0.7/>0.8/)>, 0.9 indicate that the model is worthless/poor/fair/good/excellent) and the calibration by means of Hosmer–Lemeshow (HL) \( \chi^2 \) goodness-of-fit test (the higher the \( P \) value, the better the test) and Nagelkerke’s \( r^2 \).

We tested whether the hypothyroidism-component-score was correlated with serum TSH, tri-iodothyronine \( (T_3) \), and T4 in controls and in hypothyroid patients at time of diagnosis, and also whether it was correlated with the duration of symptoms among patients (Spearman’s correlation coefficient, \( r_s \)).

When people seek medical care due to one or more symptoms, they have a certain a priori probability \( (P) \) or odds \( (\text{odds} = \log(P/1-P)) \) for suffering from a well-defined disease. The important task for the health care provider is to bring this probability in the direction of zero (disease not present) or one (disease proven). Therefore, the patient is asked about a number of possible symptoms. The risk or the odds for the patient having hypothyroidism is altered depending on the presence of these symptoms. If a hypothyroidism-associated symptom is present, the probability for hypothyroidism to be present becomes higher, and the factor by which the probability is raised is called the positive likelihood ratio \( (LR+) \). LR+ can be calculated from the percentage of ill people with a given symptom divided by the percentage of healthy controls reporting the same symptom. Conversely, if a certain symptom is absent, the odds for having hypothyroidism become lower with a factor that is called the negative likelihood ratio \( (LR-) \). The pre-test odds of hypothyroidism multiplied by the LR give the post-test odds of hypothyroidism (25):

\[
\text{Odds}_{\text{post-test}} = \text{Odds}_{\text{pre-test}} \times LR(+/–)
\]

LRs combine the conventional test properties of sensitivity and specificity into a combined index which is often more useful in the clinical setting than sensitivity and specificity (26):

\[
\begin{align*}
LR+ &= \text{sensitivity}/(1 – \text{specificity}) \\
LR– &= (1 – \text{sensitivity})/\text{specificity}
\end{align*}
\]

In general, an LR around one provides no further information, an LR between 2 and 5 or between 0.2 and 0.5 gives little further information, whereas LRs in the interval 5–10 or 0.1–0.2 change the pre- to post-test probability for disease moderately (26). Optimally, LRs should be >10 or <0.1 (26).

DORs (26) may be calculated from a 2×2 table as ‘a×d/b×c’ or by:

\[
\text{DOR} = (\text{sensitivity} \times \text{specificity})/((1 – \text{sensitivity}) \times (1 – \text{specificity})) = \text{LR + }/\text{LR–}
\]

**Ethical approval**

This study was approved by Regional Ethics Committees in North Jutland and Copenhagen. Registry permission was obtained from the Danish Data Protection Agency. All participants gave their written informed consent. No conflicts of interest have occurred during implementation or completion of the study.

**Results**

**Patients and controls**

The baseline characteristics of patients newly diagnosed with autoimmune overt hypothyroidism \( (n=140) \) and their gender-, age-, and region-matched controls free of
thyroid disease \((n = 560)\) are given in Table 1. At disease onset, hypothyroid patients had a wide range of biochemical disruption with median serum TSH of 54.5 mU/l and low but very dispersed serum T4 levels.

### Symptoms in patients and controls

The percentage of patients and controls reporting the 34 symptoms are given in Table 2. Mor et al (1993) for the patients reported tiredness, dry skin, and shortness of breath. The LR+ and LR− showed large differences, with high LR+ for symptoms such as anterior neck pain and hair loss (Table 2). Low LR− was observed for tiredness and dry skin. The highest DORs, reflecting the over-representation of symptoms in hypothyroid patients compared with controls, were calculated for tiredness, hair loss, and dry skin.

We specified which among the symptoms under investigation were specific for hypothyroidism. After Bonferroni correction, we identified 13 ‘hypothyroidism-associated symptoms’: globulus sensation, difficulty swallowing, anterior neck pain, wheezing, shortness of breath, palpitations, constipation, hair loss, dry skin, restlessness, mood lability, tiredness, and vertigo.

Figure 2 shows how many of the 13 symptoms were reported by cases and controls. The number of symptoms are equivalent to the ‘simple hypothyroidism-component-score’ (possible range: 0–13 points). At diagnosis, cases had a median of five symptoms (IQR: 3–7), whereas controls had a median of two (0–4) symptoms. Only 5.7% of cases were free of ‘hypothyroidism-associated symptoms’, whereas 70.0% of controls reported one or more symptoms associated with hypothyroidism. The simple hypothyroidism score and the two-weighed scores (additional and exponential) were compared by means of the discriminative power (AUROC) and calibration (HL). We found statistically indifferent AUROC values by comparing the simple (AUROC (95% CI); 0.76 (0.73–0.80)) and the two-weighed scores (AUROC for both models was 0.77 (0.73–0.82)). The calibration test of the simple score model revealed a LH goodness-of-fit P value of 0.73 indicating a fairly good test, not different from the P values obtained in the weighed score models. Nagelkerke’s \(r^2\) was 0.25, indicating that one-fourth of the disease state variation (being a hypothyroid patient or a healthy person) was explained by the variation in symptom expression leaving 75% of the variation unexplained.

### Symptoms and likelihood of hypothyroidism

Figure 3 depicts how the simple hypothyroidism-component-score was associated with LR+ , LR−, and DOR. Subjects with 0–2 symptoms had positive LR−, negative LR+, and negative DOR, suggesting that the
The probability of hypothyroidism was lowered by the presence of only up to two symptoms. Subjects reporting three symptoms had almost the same probability of being a case or a control, and the likelihood of hypothyroidism was not altered from average by such reporting. On the other hand, participants who reported more than three symptoms had a probability higher than average of being a patient suffering from hypothyroidism. In participants

Table 2 Symptoms experienced by patients with autoimmune hypothyroidism (n = 140) and euthyroid controls matched on age, sex and region of inhabitancy (n = 560).

<table>
<thead>
<tr>
<th>Symptoms experienced by patients with autoimmune hypothyroidism</th>
<th>Percentage of subjects with symptoms (95% CI)</th>
<th>Likelihood ratios (LR)</th>
<th>Diagnostic odds ratio (DOR) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasms</td>
<td>Cases</td>
<td>Controls</td>
<td>Positive (LR+) (95% CI)</td>
</tr>
<tr>
<td>Anterior neck pain</td>
<td>16 (11–21)</td>
<td>4 (3–6)</td>
<td>3.50 (1.95–6.22)</td>
</tr>
<tr>
<td>Difficulty swallowing</td>
<td>29 (23–36)</td>
<td>11 (10–13)</td>
<td>2.59 (1.78–3.70)</td>
</tr>
<tr>
<td>Globulon sensation</td>
<td>36 (29–43)</td>
<td>17 (15–19)</td>
<td>2.09 (1.54–2.80)</td>
</tr>
<tr>
<td>Respiratory complaints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wheezing</td>
<td>27 (21–34)</td>
<td>12 (11–14)</td>
<td>2.20 (1.51–3.15)</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>51 (44–59)</td>
<td>29 (28–31)</td>
<td>1.74 (1.40–2.13)</td>
</tr>
<tr>
<td>Cough at daytime</td>
<td>24 (18–31)</td>
<td>16 (14–18)</td>
<td>1.46 (0.99–2.10)</td>
</tr>
<tr>
<td>Cough in morning</td>
<td>14 (9–19)</td>
<td>11 (10–12)</td>
<td>1.22 (0.72–2.00)</td>
</tr>
<tr>
<td>Cardiac symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palpitations</td>
<td>35 (28–42)</td>
<td>20 (18–22)</td>
<td>1.70 (1.25–2.26)</td>
</tr>
<tr>
<td>Gastrointestinal symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constipation</td>
<td>39 (32–46)</td>
<td>17 (15–19)</td>
<td>2.29 (1.70–3.04)</td>
</tr>
<tr>
<td>Decreased appetite</td>
<td>20 (14–26)</td>
<td>10 (8–11)</td>
<td>2.06 (1.30–3.20)</td>
</tr>
<tr>
<td>Hair–skin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hair loss</td>
<td>30 (24–36)</td>
<td>9 (7–10)</td>
<td>3.50 (2.36–5.14)</td>
</tr>
<tr>
<td>Dry, sensitive skin</td>
<td>63 (55–70)</td>
<td>29 (27–31)</td>
<td>2.16 (1.78–2.56)</td>
</tr>
<tr>
<td>Psychic symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restlessness</td>
<td>33 (26–40)</td>
<td>17 (15–18)</td>
<td>1.97 (1.42–2.67)</td>
</tr>
<tr>
<td>Mood lability</td>
<td>46 (38–53)</td>
<td>20 (19–22)</td>
<td>2.23 (1.72–2.84)</td>
</tr>
<tr>
<td>Shyness, sensitive person</td>
<td>34 (27–41)</td>
<td>22 (20–24)</td>
<td>1.55 (1.14–2.05)</td>
</tr>
<tr>
<td>Easily irritated by noise</td>
<td>26 (20–33)</td>
<td>17 (15–19)</td>
<td>1.54 (1.08–2.16)</td>
</tr>
<tr>
<td>Problems working under supervision</td>
<td>25 (19–32)</td>
<td>17 (15–18)</td>
<td>1.52 (1.05–2.16)</td>
</tr>
<tr>
<td>Bad mood every day</td>
<td>6 (3–9)</td>
<td>3 (2–3)</td>
<td>2.16 (0.85–5.29)</td>
</tr>
<tr>
<td>Anxiety attacks</td>
<td>9 (6–14)</td>
<td>5 (4–6)</td>
<td>1.87 (0.94–3.65)</td>
</tr>
<tr>
<td>Problems getting friends</td>
<td>7 (4–11)</td>
<td>4 (3–5)</td>
<td>1.82 (0.82–3.93)</td>
</tr>
<tr>
<td>Prefer being alone</td>
<td>20 (14–26)</td>
<td>13 (12–15)</td>
<td>1.47 (0.95–2.21)</td>
</tr>
<tr>
<td>Sensation of stress</td>
<td>15 (10–21)</td>
<td>10 (9–12)</td>
<td>1.46 (0.88–2.37)</td>
</tr>
<tr>
<td>Emotional lability on small stuff</td>
<td>17 (12–23)</td>
<td>12 (10–13)</td>
<td>1.45 (0.91–2.26)</td>
</tr>
<tr>
<td>Sensation of being misunderstood</td>
<td>7 (3–11)</td>
<td>4 (3–5)</td>
<td>1.50 (0.66–3.29)</td>
</tr>
<tr>
<td>Bad nerves</td>
<td>8 (4–13)</td>
<td>6 (5–7)</td>
<td>1.34 (0.65–2.67)</td>
</tr>
<tr>
<td>Anxious thoughts</td>
<td>20 (14–26)</td>
<td>17 (15–18)</td>
<td>1.16 (0.77–1.73)</td>
</tr>
<tr>
<td>Prefer to decide</td>
<td>47 (40–55)</td>
<td>44 (40–46)</td>
<td>1.08 (0.87–1.32)</td>
</tr>
<tr>
<td>hands easily shake</td>
<td>10 (6–16)</td>
<td>11 (10–12)</td>
<td>0.93 (0.53–1.54)</td>
</tr>
<tr>
<td>Often taking medicine</td>
<td>21 (15–28)</td>
<td>23 (22–25)</td>
<td>0.90 (0.61–1.30)</td>
</tr>
<tr>
<td>Other symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiredness</td>
<td>81 (74–86)</td>
<td>41 (40–43)</td>
<td>1.96 (1.72–2.18)</td>
</tr>
<tr>
<td>Vertigo</td>
<td>24 (18–30)</td>
<td>11 (9–12)</td>
<td>2.23 (1.48–3.32)</td>
</tr>
<tr>
<td>Sleeplessness</td>
<td>27 (20–34)</td>
<td>19 (18–21)</td>
<td>1.37 (0.97–1.91)</td>
</tr>
<tr>
<td>Headache</td>
<td>22 (16–29)</td>
<td>16 (15–18)</td>
<td>1.32 (0.88–1.92)</td>
</tr>
<tr>
<td>Generalized pain</td>
<td>47 (40–55)</td>
<td>39 (37–41)</td>
<td>1.21 (0.97–1.48)</td>
</tr>
</tbody>
</table>

*P<0.05, †P<0.01, ‡P<0.001 (χ²-test).

*Positive likelihood ratio (LR+), referred in Subjects and methods for definition.

*Negative likelihood ratio (LR–), referred in Subjects and methods for definition.

*Diagnosis odds ratio (DOR), referred in Subjects and methods for definition.

*Information was obtained from the ‘Formal Questionnaire’ (referred in Supplementary Data).

*Information was obtained from the ‘Mental Vulnerability Questionnaire’ (referred in Supplementary Data).
reporting ≥10 symptoms, a DOR of 6.74 (2.36–19.7) was calculated. We compared the two extremes of the figure, namely those participants with no symptoms and those with ≥10 symptoms. A subject with ≥10 hypothyroidism-associated symptoms was 32.0 (8.6–126) more likely to be a hypothyroid patient than if free of any symptoms.

Symptoms and biochemical status

In patients, we found no association between the hypothyroidism-component-score and serum TSH ($r_s = -0.018$, $P=0.83$), serum T3 ($r_s = -0.096$, $P=0.26$), or serum T4 ($r_s = -0.088$, $P=0.31$). Thus, the hypothyroidism-component-score was not correlated with the biochemical markers at disease onset. We only found borderline association ($r_s = 0.193$, $P=0.047$) between the simple hypothyroidism-component-score and duration of symptoms preceding the diagnosis among those patients reporting any symptoms.

Discussion

Symptoms in hypothyroidism

We identified 13 symptoms to be significantly associated with hypothyroidism and found that 81% of overtly hypothyroid patients expressed tiredness as the most prominent symptom, whereas this was present in 41% of all euthyroid controls matched on age, sex, and region. Other studies also report that tiredness/fatigue/lethargy was the most common of all symptoms in hypothyroidism (4, 14, 27, 28, 29, 30). Interestingly, tiredness only ranked with the 11th highest LR+ but the lowest LR−. This indicates that the presence of tiredness does not increase the likelihood of hypothyroidism much. However, it is unlikely that a patient suffers from hypothyroidism if tiredness is not present. Similarly, hypothyroidism is not very likely in the absence of dry skin. The 81% prevalence of tiredness in hypothyroidism was high compared with its prevalence in a number of other diseases: 36% in Parkinson’s disease (31), 71% in chronic obstructive pulmonary disease (32), 78% in adult coeliac disease (33), and 84% in sarcoidosis (34). The exact mechanism leading to tiredness in hypothyroidism is unknown.
Possibly this and also some of the other symptoms may be caused by a combination of various suboptimally functioning organs.

Dry skin was reported by 63% of the patients. This ranged from 39 to 95% in other studies reporting on symptoms in hypothyroidism (2, 5, 6, 10, 14, 17). Hair loss was observed in 30% of all cases, but this symptom had a high diagnostic odds value because only 9% of euthyroid controls reported that complaint. The LR+ of the hypothyroidism-associated symptoms identified ranged from 1.70 with regards to the sensation of palpitations up to 3.50 for hair loss and anterior neck pain.

Symptom score models in hypothyroidism

We tested the model for predicting hypothyroidism, and the AUROC of 0.76 (0.72–0.80) is only indicative for a ‘fair’ discrimination. However, this is somewhat higher than the AUROC values between 0.64 and 0.72 as reported by Canaris et al. (5, 13). The introduction of the two-weighed scores did not improve our model.

Clinical use of symptoms expressed

Several symptoms were overrepresented among hypothyroid patients, and by combining the symptoms we found that the best symptom discrimination in this study was observed in subjects with ≤1 or ≥10 symptoms. Only 7.9% of the hypothyroid patients had ≥10 symptoms (13.6% had ≤1 symptoms), and only 47.1% of the healthy controls had ≤1 symptoms (1.3% had ≥10 symptoms). Thus, the overall applicability of the symptom-score in the clinical setting is low. Only around one-fourth (Nagelkerke’s $r^2 = 0.25$ for the simple score model) of the variation in disease state can be explained by variations in symptom expression. This indicates, together with an only fairly high HL $P$ value of 0.73, that at maximum only a fair statistical association is present between symptom expression and the probability for suffering from overt autoimmune hypothyroidism. Nonetheless, a blood test is the gold standard for diagnosing overt hypothyroidism, and such thyroid function testing should always be carried out on wide indications as symptom expression often is too vague.

Strengths and limitations

The main strength of this study is the inclusion of overtly autoimmune hypothyroid patients as they are successively diagnosed in the population. We have previously illustrated that many studies based on patients referred to hospital may be hampered by referral-bias (35). However, we cannot exclude that this study may have introduced selection bias, as some patients may go for long periods with undetected hypothyroidism. If lifelong, they will not be included in this study and some selection may have taken place. Such patients may only suffer from minor complaints, and a more vague symptom presentation might theoretically be noticed if we had included such patients.

An important limitation of our study is the rather low participation rate of 59.5% among patients. Participants were younger and were more likely to reside in the Aalborg area, which may hamper the external study validity.

Although patients were invited shortly after overt hypothyroidism was diagnosed, we cannot exclude some bias in recalling symptoms.

From literature we know that not all patients’ complaints are resolved even months after serum TSH and $T_4$ are normalized (36, 37, 38). However, as patients were only asked for symptom presentation at disease onset, we cannot answer how well symptoms were relieved after $T_4$ treatment.

This study investigated a large number of symptoms suggestive of thyroid disease including signs of mental vulnerability. We used the ‘Mental vulnerability questionnaire’ as it had been validated in a Danish setting (21). No specific time frame of the psychiatric symptoms was defined in contrast to the 12-month time frame in the ‘Formal questionnaire’. Therefore, any comparison of symptom prevalence obtained from the two questionnaires should be performed cautiously.

The number of symptoms explored in this study is considerable but does not cover all possible symptoms of hypothyroidism. For example, we had no information on cold intolerance or decreased sweating (2, 5, 14, 17, 27, 28). In a study by Watt et al. (30), symptoms were ranked according to their individual importance, and cold intolerance was the third most compelling complaint. On the other hand, cold intolerance was not experienced more often in overtly hypothyroid patients than in controls in the study reported by Zulewski et al. (10).

We chose to investigate only the presentation of symptoms at disease onset, thus leaving out the role of signs suggestive of hypothyroidism (3, 6, 7, 8, 9, 10), such as changes in weight, blood pressure, heart rate, eye signs, and the doctors, perceptions of skin and hair changes.

In conclusion, we performed the first population-based study and found that hypothyroidism presented with a broad spectrum of symptoms. However, many of those symptoms were also experienced by volunteer
control subjects. Therefore, even a minor suspicion from the clinician should lead to a thyroid function test, as many undiagnosed patients present only few symptoms.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/EJE-14-0481.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This study was part of DanThyr, and it was supported by the following grants: IMK General Foundation; The Danish Council for Independent Research; Ministry of Food, Agriculture and Fisheries; the Danish Agency for Science, Technology and Innovation, Institute for Clinical Medicine, University of Aarhus; and Aase og Ejnar Danielsen Foundation.

Author contribution statement

Acknowledgements
The authors express their gratitude to the general practitioners in Copenhagen and Northern Jutland, and to clinical chemical laboratories at Aalborg University Hospital, Bispebjerg University Hospital, Frederiksberg Hospital, as well as Laboratory of General Practitioners in Copenhagen for their helpful collaboration in identifying patients.

References
6 Wayne EJ. Clinical and metabolic studies in thyroid disease. BMJ 1960 1 78–90. (doi:10.1136/bmj.1.5166.78)
11 Crilly M. Thyroxine Adherence Study, School of Primary Care, Faculty of Medicine, University of Manchester, 2003.
16 Murray GR. An address on the signs of early disease of the thyroid gland. BMJ 1909 1 381–384. (doi:10.1136/bmj.1.2511.381)


30 Watt T, Hegedus L, Rasmussen AK, Groenvold M, Bonnema SJ, Bjorner JB & Feldt-Rasmussen U. Which domains of thyroid-related quality of life are most relevant? Patients and clinicians provide complementary perspectives *Thyroid* 2007 **17** 647–654. (doi:10.1089/thy.2007.0069)


