Thyroidectomy Versus Medical Management for Euthyroid Patients With Hashimoto Disease and Persisting Symptoms

A Randomized Trial

Ivar Guldvog, MD, PhD; Laurens Cornelus Reitsma, MD*; Lene Johnsen, MD*; Andromeda Lauzike, MD; Charlotte Gibbs, MD; Eivind Carlsten, MD; Tone Hoel Lende, MD; Jon Kristian Harvendal, MD; Roald Omdal, MD, PhD; Jan Terje Kvaley, PhD; Geir Hoff, MD, PhD; Tomm Bernkev, PhD†; and Håvard Søiland, MD, PhD†

Background: Hashimoto disease is a chronic autoimmune thyroiditis. Despite adequate hormone substitution, some patients have persistent symptoms that may be the result of immunologic pathophysiology.

Objective: To determine whether thyroidectomy improves symptoms in patients with Hashimoto thyroiditis who still have symptoms despite having normal thyroid gland function while receiving medical therapy.

Design: Randomized trial. (ClinicalTrials.gov: NCT02319538)

Setting: Secondary care hospital in Norway.

Patients: 150 patients aged 18 to 79 years with persistent Hashimoto-related symptoms despite euthyroid status while receiving hormone replacement therapy and with serum antithyroid peroxidase (anti-TPO) antibody titers greater than 1000 IU/mL.

Intervention: Total thyroidectomy or medical management with hormone substitution to secure euthyroid status in both groups.

Measurements: The primary outcome was general health score on the Short Form-36 Health Survey (SF-36) at 18 months. Secondary outcomes were adverse effects of surgery, the other 7 SF-36 subscores, fatigue questionnaire scores, and serum anti-TPO antibody titers at 6, 12, and 18 months.

Results: During follow-up, only the surgical group demonstrated improvement: Mean general health score increased from 38 to 64 points, for a between-group difference of 29 points (95% CI, 22 to 35 points) at 18 months. Fatigue score decreased from 23 to 14 points, for a between-group difference of 9.3 points (CI, 7.4 to 11.2 points). Chronic fatigue frequency decreased from 82% to 35%, for a between-group difference of 39 percentage points (CI, 23 to 53 percentage points). Median serum anti-TPO antibody titers decreased from 2232 to 152 IU/mL, for a between-group difference of 1148 IU/mL (CI, 1080 to 1304 IU/mL). In multivariable regression analyses, the adjusted treatment effects remained similar to the unadjusted effects.

Limitation: Results are applicable only to a subgroup of patients with Hashimoto disease, and follow-up was limited to 18 months.

Conclusion: Total thyroidectomy improved health-related quality of life and fatigue, whereas medical therapy did not. This improvement, along with concomitant elimination of serum anti-TPO antibodies, may elucidate disease mechanisms.

Primary Funding Source: Telemark Hospital.

Ann Intern Med. doi:10.7326/M18-0284

See also:

Web-Only Supplement

Hashimoto disease, first described in 1912 (1) and recognized as an autoimmune disease decades later (2), is the most prevalent autoimmune disease worldwide (3, 4). It is a T-cell–mediated disease of unknown cause (5, 6), with elevated levels of serum antithyroid peroxidase (anti-TPO) antibody and proinflammatory cytokines (7). The disease process ultimately leads to hypothyroidism (6, 8, 9). Patients often report profound fatigue, poor sleep quality, muscle and joint tenderness, and dry mouth and eyes. These symptoms may respond only partly to adequate thyroid hormone substitution (10, 11). Hashimoto disease has 2 types: one in which symptoms benefit from levothyroxine supplementation (12) and another in which symptoms persist despite patients’ euthyroid status while receiving hormone substitution. Persistent symptoms are thought to be related to autoimmune disease rather than to hypothyroidism (13). To date, no specific treatment exists for patients whose symptoms persist despite adequate thyroid hormone replacement.

Complete removal of the antigenic tissue via total thyroidectomy has been hypothesized to attenuate the autoimmune response (14) and relieve symptoms (15). In an earlier study of patients who had surgery for bilateral goiter, we found that serum anti-TPO antibody levels decreased and general well-being and fatigue improved (16). The present randomized trial, The Norwegian Trial on Surgery for Hashimoto’s Disease, aimed to compare the outcomes of total thyroidectomy with those of medical therapy alone in patients with Hashimoto-related symptoms despite adequate thyroid hormone replacement (17).
METHODS

Design Overview

We conducted a randomized, open-label, controlled trial. Enrollment began on 13 February 2012 and closed on 10 April 2015, with follow-up through 15 December 2016.

Participants were randomly assigned to undergo total thyroidectomy with standard medical therapy or to receive standard medical therapy only. Patients in both groups maintained normal thyroid function throughout the trial and were free to receive symptomatic medical treatment (such as nonsteroidal anti-inflammatory drugs or paracetamol).

All participants were given oral and written information about the study and provided written consent. An internist (C.G.) in the Department of Endocrinology informed all patients about the experimental nature of surgery for Hashimoto disease and the risks involved. The research protocol was approved by the Institutional Research Board and the Regional Ethics Committee of South East Norway (2011/2222-B). Communication failure caused a delay in ClinicalTrials.gov registration (NCT02319538) after 138 patients had already enrolled. However, no protocol adjustments were made or analyses performed between the start of enrollment and the registration date (18 December 2014).

Setting and Participants

Patients with Hashimoto disease and related symptoms despite adequate hormone substitution were referred to Telemark Hospital by their general practitioners to be considered for inclusion in the study. Telemark Hospital is a secondary care hospital in the South-Eastern Norway Health Region. Eligible patients had Hashimoto disease symptoms, were medically fit for surgery, could provide informed consent, were at least 18 years of age, had serum anti-TPO antibody levels greater than 1000 IU/mL (normal, <100 IU/mL), and had thyroid-stimulating hormone (TSH) levels above 3.5 mIU/L (normal range, 0.2 to 3.5 mIU/L) before hormone substitution. Patients with clinical euthyroid status and serum TSH, free triiodothyronine (fT₃), and free thyroxine (fT₄) levels within normal range, or with only slight aberrations, were eligible. In addition, patients’ symptoms had to be severe enough despite adequate hormone substitution that they were willing to consider surgery. Typical symptoms reported included fatigue, increased need for sleep with reduced sleep quality, joint and muscle tenderness, and dry mouth and eyes.

Before randomization, patients had ultrasonography changes compatible with Hashimoto disease, and suspicious, solitary lesions were clarified with cytology.

Randomization and Interventions

Patients were randomly assigned 1:1 in blocks of 8 patients. The randomization schedule was computer generated by an independent party in the Department of Clinical Science, University of Bergen, Norway. On the basis of this schedule, sequential numbers were assigned and kept in sealed envelopes. Whenever one of the authors (I.G.) enrolled a patient at Telemark Hospital, he telephoned the Department of Clinical Science at Bergen to obtain the patient’s allocation group. Baseline data were collected before randomization, including patient-reported outcome measures (PROMs).

Medical Therapy

All patients were surveilled closely for thyroid function and dosing of thyroid medication until they achieved euthyroid status before data were collected and surgery was performed (that is, PROM data and surgery). At baseline, patients in the surgery group received a median weekly levothyroxine dose of 618 μg (range, 157 to 1450 μg), whereas those in the control group received 550 μg (range, 88 to 1500 μg) (Table 1). Two patients received a combination of levothyroxine and lithium. One patient received the desiccated porcine product Armour Thyroid (Allergan) only, whereas 3 patients used it in combination with levothyroxine (for detailed dosages, see Table 1). During follow-up, the participants were assessed every third month and their thyroid medication was adjusted accordingly (Supplement Figure 1, available at Annals.org). Euthyroid was defined as the absence of hyper- or hypothyroid symptoms and signs combined with biochemical euthyroid status (that is, a serum TSH and serum fT₄ or fT₃ level within the normal range). A patient was defined as an outlier if he or she had a TSH level outside the normal range along with an abnormal fT₄ or fT₃ value.

Operative Procedure

Total thyroidectomy involved the complete removal of all visible thyroid tissue, with special focus on 3 sites: the angle at which the recurrent laryngeal nerve enters the cricothyroid membrane, the pyramidal lobe, and the hilus where the superior vessels enter the field. We used the NIM 3.0 nerve monitoring system (Medtronic) for intraoperative observation of the recurrent laryngeal nerves.

Surgical patients were hospitalized for 1 to 2 days after the procedure. All complications (including hemorrhage; infection; damage to the recurrent laryngeal nerve; and damage to the parathyroid glands, resulting in hypocalcemia) were recorded on case report forms and assessed at the scheduled follow-up appointments.

Outcomes and Follow-up

For both groups, follow-up visits occurred every 3 months for 18 months. At each visit, blood samples were drawn for analysis of serum anti-TPO antibody titers and TSH, fT₄, fT₃, calcium, and parathyroid hormone levels. Serum anti-TPO antibody titers were measured by the chemiluminescent immune assay ADVIA Centaur (Siemens). For patients with titers above the upper limit for quantification (>1300 IU/mL), the samples were diluted to measure the exact concentrations. No changes in any of these measurement methods occurred during the trial.
**PROMs**

Translated and validated questionnaires were used for PROMs, all of which were obtained at baseline and 6, 12, and 18 months. The PROMs were completed at the hospital, and the study nurse immediately checked the responses for completeness.

Health-related quality of life was measured with the generic Short Form-36 Health Survey (SF-36) (18), which is made up of 36 questions, each of which is scored from 0 to 100. The SF-36 has 8 domains: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health. Fatigue was measured with the fatigue questionnaire, from which 2 dimensions are calculated. First, the total fatigue dimension comprises 11 questions assessed on a scale of 0 to 3, resulting in a score

<p>| Table 1. Baseline Variables for the 2 Treatment Groups Included in the Analyses* |
|-------------------------------|-------------------------------|-------------------------------|-------------------------------|</p>
<table>
<thead>
<tr>
<th>Variable</th>
<th>Surgery Group (n = 73)</th>
<th>Patients in Surgery Group With Missing Data, n</th>
<th>Control Group (n = 74)</th>
<th>Patients in Control Group With Missing Data, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (SD), y</td>
<td>48.0 (13)</td>
<td>49.2 (12)</td>
<td>48.0 (13)</td>
<td>49.3 (12)</td>
</tr>
<tr>
<td>Range, y</td>
<td>18-74</td>
<td>18-79</td>
<td>18-74</td>
<td>18-79</td>
</tr>
<tr>
<td>Sex (male/female), n/n</td>
<td>5/68</td>
<td>7/67</td>
<td>5/68</td>
<td>7/67</td>
</tr>
<tr>
<td>Mean BMI (SD), kg/m²</td>
<td>25 (7)</td>
<td>26 (8)</td>
<td>25 (7)</td>
<td>26 (8)</td>
</tr>
<tr>
<td>Cardiovascular disease, n (%)</td>
<td>11 (15)</td>
<td>23 (31)</td>
<td>11 (15)</td>
<td>23 (31)</td>
</tr>
<tr>
<td>Nonthyroid autoimmune disease, n (%)†</td>
<td>19 (26)</td>
<td>18 (24)</td>
<td>19 (26)</td>
<td>18 (24)</td>
</tr>
<tr>
<td>Psychiatric disorder, n %‡</td>
<td>10 (14)</td>
<td>14 (19)</td>
<td>10 (14)</td>
<td>14 (19)</td>
</tr>
<tr>
<td>Other disease, n %§</td>
<td>16 (22)</td>
<td>16 (22)</td>
<td>16 (22)</td>
<td>16 (22)</td>
</tr>
<tr>
<td>Median serum anti-TPO antibody (IQR), IU/mL</td>
<td>2232 (1278–4263)</td>
<td>2052 (1204–3791)</td>
<td>2232 (1278–4263)</td>
<td>2052 (1204–3791)</td>
</tr>
<tr>
<td>Median serum TSH (IQR), mIU/L</td>
<td>1.7 (0.5–2.6)</td>
<td>1.3 (0.5–1.9)</td>
<td>1.7 (0.5–2.6)</td>
<td>1.3 (0.5–1.9)</td>
</tr>
<tr>
<td>Median serum fT3 (IQR), pmol/L</td>
<td>4.6 (4.3–4.8)</td>
<td>4.6 (4.3–5.0)</td>
<td>4.6 (4.3–4.8)</td>
<td>4.6 (4.3–5.0)</td>
</tr>
<tr>
<td>Median serum fT4 (IQR), pmol/L</td>
<td>16 (14-19)</td>
<td>16 (15-19)</td>
<td>16 (14-19)</td>
<td>16 (15-19)</td>
</tr>
<tr>
<td>Median serum Ca (IQR)</td>
<td>4.1 (3.0–5.4)</td>
<td>4.1 (3.0-5.8)</td>
<td>4.1 (3.0-5.4)</td>
<td>4.1 (3.0-5.8)</td>
</tr>
<tr>
<td>Median serum PTH (IQR), pmol/L</td>
<td>2.34 (2.27-2.41)</td>
<td>2.35 (2.31-2.42)</td>
<td>2.34 (2.27-2.41)</td>
<td>2.35 (2.31-2.42)</td>
</tr>
<tr>
<td>Median serum Armour (IQR)</td>
<td>618 (350-700)</td>
<td>550 (400-800)</td>
<td>618 (350-700)</td>
<td>550 (400-800)</td>
</tr>
<tr>
<td>Range, µg</td>
<td>175-1450</td>
<td>88-1500</td>
<td>175-1450</td>
<td>88-1500</td>
</tr>
<tr>
<td>Mean quality-of-life scores (SD)</td>
<td>37 (21)</td>
<td>39 (20)</td>
<td>37 (21)</td>
<td>39 (20)</td>
</tr>
<tr>
<td>General health</td>
<td>69 (21)</td>
<td>65 (19)</td>
<td>69 (21)</td>
<td>65 (19)</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>19 (30)</td>
<td>25 (32)</td>
<td>19 (30)</td>
<td>25 (32)</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>42 (21)</td>
<td>40 (22)</td>
<td>42 (21)</td>
<td>40 (22)</td>
</tr>
<tr>
<td>Vitality</td>
<td>22 (15)</td>
<td>25 (20)</td>
<td>22 (15)</td>
<td>25 (20)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>46 (27)</td>
<td>49 (26)</td>
<td>46 (27)</td>
<td>49 (26)</td>
</tr>
<tr>
<td>Role emotional</td>
<td>35 (39)</td>
<td>42 (39)</td>
<td>35 (39)</td>
<td>42 (39)</td>
</tr>
<tr>
<td>Mental health</td>
<td>57 (20)</td>
<td>58 (18)</td>
<td>57 (20)</td>
<td>58 (18)</td>
</tr>
<tr>
<td>Mean total fatigue score (SD)</td>
<td>23 (5)</td>
<td>23 (6)</td>
<td>23 (5)</td>
<td>23 (6)</td>
</tr>
<tr>
<td>Chronic fatigue, n %</td>
<td>60 (82)</td>
<td>62 (84)</td>
<td>60 (82)</td>
<td>62 (84)</td>
</tr>
</tbody>
</table>

- anti-TPO = antithyroid peroxidase antibody; BMI = body mass index; Ca = calcium; fT3 = free triiodothyronine; fT4 = free thyroxine; IQR = interquartile range; PTH = parathyroid hormone; TSH = thyroid-stimulating hormone.
- * Data are reported as numbers and percentages for dichotomous variables. For continuous variables, means and SDs are reported for normally distributed variables and medians and IQRs (i.e., 25% to 75% percentile) for non-normally distributed variables.
- † Other autoimmune diseases, such as rheumatoid arthritis, diabetes mellitus, and ulcerative colitis.
- ‡ Psychiatric disorders are defined as mental illnesses treated by a specialist.
- § All other diseases that may influence general well-being (e.g., allergy, skin diseases, neurologic diseases).
- ‖ Armour and Armour Thyroid (Allergan) are tablets containing desiccated porcine thyroid glands, which provide the patient with the natural T3:T4 ratio.
- ¶ Monotherapy with either levothyroxine or Armour. SD was not calculated in the Armour group because of the small number of patients.
- ** SDs were not calculated because of the small number of patients in these subgroups.
from 0 to 33 (19). Second, the chronic fatigue dimension is calculated as follows: For each question on the total fatigue scale, a score of 0 or 1 is assigned a value of 0 on a new dichotomized scale, and a score of 2 or 3 on the total fatigue scale is assigned a value of 1. The dichotomized scale provides a score from 0 to 11, with higher scores corresponding to more severe fatigue-related symptoms or problems. On the basis of results from the original validation study (20), the Norwegian validation study (21), and general consensus (19), chronic fatigue was defined as a dichotomized score of 4 or greater and a duration of 6 months or longer.

All patients achieved biochemical and clinical euthyroid status before they began working on the PROM instruments. Patients in the surgery group had their vocal cords examined preoperatively and again on the first postoperative day. Although the examinations were performed at Telemark Hospital’s ear, nose, and throat department, this facility was not otherwise involved in the study. Laryngoscopy was performed every 4 months in patients with signs of laryngeal nerve injury. The preinclination and preoperative information was provided by a physician and surgeon working together.

**Main Outcome Measures**

The primary outcome was SF-36 general health subscore at 18 months. This end point was chosen as the most relevant variable to capture subjective health perception assessed directly by the patients. Secondary outcomes were serum anti-TPO antibody titters (measured at baseline and then every 3 months), the other 7 SF-36 subscores, total fatigue score, and frequency of chronic fatigue, all of which were assessed at 0, 6, 12, and 18 months. No changes were made to the trial’s outcome measures after the study began. Safety outcomes, such as recurrent laryngeal nerve palsy and hypocalcemia requiring calcium and vitamin D supplements for more than 1 year, were recorded.
Statistical Analysis

The main consideration in determining the sample size for this study was having sufficient power to detect relevant differences in the SF-36 general health score. With 75 patients in each group and an estimated dropout rate of 10%, we had 80% power to detect a clinically relevant difference in SF-36 general health score of 0.5 SD, which equals 11 points assuming an SD of 22, as reported in the literature (22, 23).

Data were analyzed with SPSS Statistics, version 24.0 for Macintosh (IBM), and R, version 3.4.3 (The R Foundation for Statistical Computing) (24).

A scheduled interim analysis of recurrent laryngeal nerve effects and hypocalcemia was performed after 75 patients had been randomly assigned (14). The difference in symptom intensity between the groups during follow-up also was considered in deciding whether to pursue or discontinue the study. In the surgery group, 1 of 35 patients (2.9%) had recurrent nerve palsy (1.4% of the nerves at risk) and none had hypocalcemia.

To examine the efficacy of surgery, we excluded from our analyses 2 patients in the surgery group who wanted to postpone the operation. The analyses were based on the remaining randomly assigned patients who had any data for the efficacy outcomes or any follow-up for safety outcomes. One patient in the control group did not complete the baseline PROM questionnaires before being lost to follow-up and therefore was excluded.

For the primary outcome variable, general health, we fitted a linear mixed-effects model with the general health score at the 4 visits as the response variable, time as a categorical variable, and time-by-group interaction. The model was fitted by maximum likelihood. From this model, we estimated within-group differences from baseline to 18 months and between-group differences in scores at 6, 12, and 18 months. We also calculated the mean score with 95% CI in each treatment group at each time point and calculated t tests for the difference between the groups at 18 months.

Most of the aforementioned analyses were also performed for the secondary outcome measures. For the dichotomous outcome variable chronic fatigue, 95% CIs for the proportions and the differences in proportions were calculated at each time point by using the Wilson score method (25). The χ² test was used to identify differences between groups at each time point, and the McNemar test was used to determine differences from baseline to 18 months. For the highly skewed data for serum anti-TPO antibody titers, we used bootstrapping to calculate CIs for medians and the Mann–Whitney test to identify differences between groups, and we used a log-transform when this variable was used in regression analyses.

To determine whether the effect of surgery on general health score was influenced by a difference in clinically relevant variables, the same linear mixed-effects regression model described earlier also was fitted with the following set of covariates considered to be clinically relevant: the baseline covariates age, sex, body mass index, cardiovascular diseases, nonthyroid autoimmune diseases, and psychiatric diseases and the time-dependent covariates serum TSH, serum calcium, serum parathyroid hormone, and logarithm of serum anti-TPO antibody titers. We made the same adjusted analysis for the total fatigue score.

In the plots in which we report the mean health-related quality-of-life variables for each treatment group over time, we added a line showing age- and sex-adjusted scores considered normal for the Norwegian population (21, 22).

<table>
<thead>
<tr>
<th>Time</th>
<th>Surgery*</th>
<th>Control*</th>
<th>Difference*</th>
<th>P Value</th>
<th>n Surg n Cont</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36 general health score†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>38 (34-41)</td>
<td>37 (32-41)</td>
<td>20 (15-26)</td>
<td>&lt;0.001</td>
<td>72, 74</td>
</tr>
<tr>
<td>6 mo</td>
<td>57 (53-61)</td>
<td>37 (32-41)</td>
<td>20 (15-26)</td>
<td>&lt;0.001</td>
<td>71, 72</td>
</tr>
<tr>
<td>12 mo</td>
<td>64 (59-68)</td>
<td>37 (32-42)</td>
<td>27 (20-33)</td>
<td>&lt;0.001</td>
<td>70, 70</td>
</tr>
<tr>
<td>18 mo</td>
<td>64 (59-68)</td>
<td>35 (30-39)</td>
<td>29 (22-35)</td>
<td>&lt;0.001</td>
<td>73, 70</td>
</tr>
<tr>
<td>Total fatigue score†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>23 (22-24)</td>
<td>23 (22-24)</td>
<td>7.2 (5.6-8.8)</td>
<td>&lt;0.001</td>
<td>73, 74</td>
</tr>
<tr>
<td>6 mo</td>
<td>16 (14-17)</td>
<td>23 (22-24)</td>
<td>7.2 (5.6-8.8)</td>
<td>&lt;0.001</td>
<td>73, 72</td>
</tr>
<tr>
<td>12 mo</td>
<td>15 (14-17)</td>
<td>23 (21-24)</td>
<td>7.5 (5.6-9.3)</td>
<td>&lt;0.001</td>
<td>72, 70</td>
</tr>
<tr>
<td>18 mo</td>
<td>14 (13-16)</td>
<td>24 (22-25)</td>
<td>9.3 (7.4-11.2)</td>
<td>&lt;0.001</td>
<td>73, 70</td>
</tr>
<tr>
<td>Frequency of chronic fatigue‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>83 (76-88)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>73, 74</td>
</tr>
<tr>
<td>6 mo</td>
<td>36 (26-48)</td>
<td>77 (66-86)</td>
<td>41 (26-55)</td>
<td>&lt;0.001</td>
<td>72, 71</td>
</tr>
<tr>
<td>12 mo</td>
<td>33 (23-46)</td>
<td>73 (61-82)</td>
<td>40 (24-54)</td>
<td>&lt;0.001</td>
<td>70, 70</td>
</tr>
<tr>
<td>18 mo</td>
<td>35 (25-46)</td>
<td>74 (62-83)</td>
<td>39 (23-53)</td>
<td>&lt;0.001</td>
<td>72, 70</td>
</tr>
</tbody>
</table>

n Surg and n Cont = number of patients with complete data in the surgery and control groups, respectively, at each time point; SF-36 = Short Form-36 Health Survey.

† Scored as percentages of patients, and the CIs were calculated by using the Wilson score procedure.

<table>
<thead>
<tr>
<th>Table 2. Estimated Health-Related Quality-of-Life Scores in Each Group Over Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>SF-36 general health score†</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>6 mo</td>
</tr>
<tr>
<td>12 mo</td>
</tr>
<tr>
<td>18 mo</td>
</tr>
<tr>
<td>Total fatigue score†</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>6 mo</td>
</tr>
<tr>
<td>12 mo</td>
</tr>
<tr>
<td>18 mo</td>
</tr>
</tbody>
</table>

n Surg and n Cont = number of patients with complete data in the surgery and control groups, respectively, at each time point; SF-36 = Short Form-36 Health Survey.

† Scored as percentages of patients, and the CIs were calculated by using the Wilson score procedure.

‡ Reported data are percentages of patients, and the CIs were calculated by using the Wilson score procedure.
The study protocol was developed after discussions with patients with Hashimoto disease, who provided feedback in meetings with the local branch of the Thyroid Patient Association in Norway. A newsletter will disseminate the study’s results to the participants.

**Role of the Funding Source**
Telemark Hospital funded the study but did not play a role in the reporting of the study.

**RESULTS**

**Study Population**
Of 191 patients referred for eligibility assessment, 41 did not meet the inclusion criteria and 150 (137 women and 13 men, all white) were randomly assigned to 1 of the 2 groups. Two patients in the surgery group who postponed surgery and 1 patient in the control group who dropped out of the study did not have any
follow-up data (Figure 1). The analyses included 147 patients (135 women and 12 men), 144 of whom provided follow-up data at all time points. Three patients in the control group had partial follow-up data. In a few cases, data were missing for some of the variables at some of the time points. Numbers of patients with complete data are reported in all tables and figures. Table 1 provides baseline characteristics for the surgery and control groups. Distribution of baseline values for TSH, fT₄, and fT₃ are shown in Supplement Figure 1. No pa-
tients crossed over from one group to the other (Table 1).

Health-Related Quality of Life
For the surgery group, the linear mixed-effects model estimated an improvement in mean SF-36 general health score from 38 points at baseline to 64 points at 18 months—that is, an increase of 26 points (95% CI, 21 to 31 points). The control group’s score decreased from 38 to 35 points, a change of −3 points (CI, −8 to 2 points). At 18 months, an estimated mean difference of 29 points (CI, 22 to 35 points) in general health score was observed between the groups (Table 2). By 6 months, a statistically significant difference in general health score between groups was apparent (Figure 2, A, and Table 2). The same pattern—that is, substantial improvement in the surgery group versus no change in the control group—also was seen for the other SF-36 dimensions (Figure 2, B to H, and Supplement Tables 1 and 2, available at Annals.org). Individual general health trajectories in each group are displayed in Supplement Figure 2 (available at Annals.org).

Fatigue
From baseline to 18 months, estimated mean total fatigue score decreased from 23 to 14 points, a reduction of 9 points (CI, 7 to 10 points), in the surgery group but remained essentially unchanged at 23 to 24 points, a change of 1 point (CI, −1 to 2 points), in the control group. The estimated mean difference at 18 months was 9.3 points (CI, 7.4 to 11.2 points) (Figure 3, top, and Table 2). The proportion of patients reporting chronic fatigue declined from 82% to 35% in the surgery group, whereas a nonsignificant change from 84% to 74% was seen in the control group. At 18 months, an estimated mean difference of 39 percentage points (CI, 23 to 53 percentage points) was seen between groups (Figure 3, bottom; Table 2; and Supplement Table 1).

Serum Anti-TPO Antibody Titers
At 18 months, serum anti-TPO antibody titers were normal in almost all patients in the surgery group (Figure 4, top and bottom) and modestly reduced in the control group (Figure 4, bottom). The median baseline values of serum anti-TPO antibodies were 2232 IU/mL (interquartile range [IQR], 1278 to 4263 IU/mL) and 2052 IU/mL (IQR, 1204 to 3791 IU/mL) for the surgery and control groups, respectively. The median serum anti-TPO antibody titer level after surgery was reduced to 152 IU/mL (IQR, 100 to 286 IU/mL) after 18 months but remained high, at a median of 1300 IU/mL, in the control group, for an estimated difference in medians of 1148 IU/mL (CI, 1080 to 1304 IU/mL) (Figure 4, bottom). The median percentage reduction in serum anti-TPO antibody titers in the surgery group during the 18-month follow-up was 92% (IQR, 87% to 96%).

Adjustment for Covariates
Adjusted linear mixed-effects models including clinically relevant covariates were developed for the general health and total fatigue scores and showed minimal changes in estimated effects compared with unadjusted results (Supplement Table 3 and Supplement Table 4).

Figure 3. Fatigue in the surgery and control groups.

The surgery and control groups are depicted by solid and dashed lines, respectively, and 95% CIs are represented by vertical lines at each time point. The mean age- and sex-adjusted value for the Norwegian background population is indicated by a dotted line, with 95% CIs indicated by the shaded areas on both sides of the line. The x-axis shows time in months, and the numbers below represent patients in the surgery and control groups, respectively at each time point. Top. Mean observed total fatigue scores and 95% CIs in patients with Hashimoto thyroiditis, derived by using the fatigue questionnaire. Total fatigue scores are shown on the y-axis (scale, 0 to 100). The P value reported at 18 mo is for an independent sample t test for the difference in mean score between the surgery and control groups. Bottom. Frequency of chronic fatigue and 95% CIs among patients with Hashimoto thyroiditis, derived by using the fatigue questionnaire. The y-axis shows the percentage of patients with chronic fatigue. The P value reported at 18 mo is for a χ² test for the difference in proportions between the surgery and control groups.
Adverse Events, Surgical Complications, and Surgical Specimens

In the control group, 1 patient had a stroke and 1 patient died in a car accident (Figure 1). In the surgery group, 3 patients had postsurgical infections (4.1%), 1 of which was initiated by a 4-mm injury to the trachea and subcutaneous emphysema; the other 2 were wound infections. Three patients (4.1%) had long-standing hypocalcemia. No clinically relevant bleeding was observed. Four patients (5.5%) had unilateral recurrent laryngeal nerve palsy at examination 4 and 12 months after surgery. All 4 were followed by an ear, nose, and throat specialist for 12 months, and their voice quality improved either spontaneously or after training with a speech therapist.

An experienced pathologist examined all surgical specimens; histologically, they showed characteristic Hashimoto changes. Papillary cancer was diagnosed in 2 patients, 1 with an 8-mm lesion and 1 with several 1-mm lesions in both lobes.

DISCUSSION

We believe that this is the first randomized controlled trial to demonstrate improvement in health-related quality of life and fatigue and normalization of serum anti-TPO antibody titer levels after complete removal of the diseased thyroid gland in patients with histologically verified Hashimoto disease. Some patients with this disorder report symptoms despite adequate treatment with levothyroxine (10), and at baseline, patients in this trial exhibited clinically relevant impairment in health-related quality-of-life and fatigue scores compared with the general Norwegian population (22). After we defined clinical relevance as a change in SF-36 domain scores greater than 0.5 SD (23), we found clinically relevant improvements in all SF-36 dimensions in the surgery group. After 18 months, the general health score in the surgery group improved to a level that compared favorably with the mean score for the age- and sex-adjusted Norwegian population without Hashimoto disease (22). In addition, fatigue scores were almost normalized when compared with the Norwegian population (21). In contrast, no changes were seen in health-related quality of life or fatigue scores in the control group.

We hypothesize that the improvement in symptoms may be related to normalization of serum anti-TPO antibody titers through 2 pathways. First, serum anti-TPO antibodies may lead to cross-reaction with other tissues. In patients with oligoarthritis, anti-TPO antibodies have been detected in synovial fluid 1 year before appearing in serum, revealing Hashimoto disease (26). A possibility exists that activated anti-TPO antibody–producing lymphocytes may leave the thyroid gland and invade other, distant tissue, contributing to inflammation and nonhypothyroid symptoms (11, 27, 28). Second, clearance of se-
rum anti-TPO antibody may occur in parallel with a reduction in other immunologic mediators (7, 29–32). Of interest, higher levels of the T-lymphocyte–derived proinflammatory cytokines interferon-γ and tumor necrosis factor-α have been observed in persons who have Hashimoto disease and serum anti-TPO antibody titers greater than 1000 IU/mL than in those without Hashimoto disease (7). Moreover, Krysiak and Okopień (33) demonstrated an approximately 50% reduction in serum levels of interferon-γ, tumor necrosis factor-α, interleukin-1β, and interleukin-2, as well as C-reactive protein, when both levothyroxine and the immunosuppressive agent selomethionine were given for 6 months to euthyroid patients with Hashimoto disease. However, that study did not examine symptoms, although it did find correlations between reduced serum anti-TPO antibody titers and decreased serum levels of all the aforementioned cytokines as well as C-reactive protein (33). Thus, the postoperative reduction in serum anti-TPO antibody titers is probably a surrogate marker for the postoperative modifications of the systemic inflammatory process that takes place after a meticulously performed total thyroidectomy in these patients (7, 33). However, serum anti-TPO antibody titers will still serve as a useful proxy for a complete macroscopic extirpation of the thyroid gland with removal of the target antigens that drive the immunologic process in Hashimoto thyroiditis. The present study shows that clin­ically relevant improvements in both health-related quality of life and fatigue may be seen after total thyroidectomy in patients with Hashimoto disease, regardless of how much preoperative serum anti-TPO antibody titers are elevated beyond the 1000 IU/mL threshold.

Strengths of the present study include its randomized design, the histologic verification of Hashimoto disease, good follow-up with few dropouts, and the use of validated generic PROM instruments. Diagnostic misclassification due to lack of histologic verification may have been the reason Promberger and colleagues (15) did not demonstrate any postoperative improvement in health-related quality of life when the Hashimoto disease diagnosis was based solely on serum anti-TPO antibody titers (15).

Nevertheless, our study has some limitations. First, like symptom relief, the PROM parameters were “soft” end points completely dependent on the patient’s subjective judgment in this nonblinded trial. Including a more disease-specific PROM instrument, such as ThyPRO (thyroid-related patient-reported outcome) (34, 35), would have been preferable (36); however, a validated Norwegian translation of ThyPRO was not available when this study began. Second, the surgical procedure itself may produce a strong placebo effect (37, 38). The participants in our study were advised to consider surgery because of severe, unspecific symptoms, including extreme tiredness, sleep disturbance, and arthralgia or myalgia. These patients may be characterized as “end of the road” in terms of available treatment options and may be considered a highly select group with clear-cut symptoms and high motivation for surgery. Blinding of the treatment groups, however, was not possible for ethical reasons (39). Third, follow-up ended at 18 months, a duration that might be considered short. Published literature, however, suggests that improved health-related quality-of-life scores in patients who receive placebo will probably normalize within 6 months of follow-up (40); therefore, a persistent placebo effect after 18 months is an unlikely explanation for our findings. Nonetheless, further studies, ideally with longer follow-up, should be encouraged. Finally, one might hypothesize that an increased focus on symptomatic relief through medication in the non-surgery group may have inadvertently increased the risk for drug side effects resembling Hashimoto-like symptoms (such as tiredness) and strengthened the comparative effect of thyroidectomy during follow-up. However, after correcting for use of medications with a greater than 1 in 100 risk for relevant side effects, we could not document such an interaction.

Patients with Hashimoto thyroiditis are thought to be particularly prone to recurrent nerve palsy and hypocalcemia after surgery because of extensive inflammatory changes in the surgical field (36, 41). When comparing complication rates in the literature, one should consider patient selection, type of surgical procedure, and surgical skills. Publication bias may further contribute to underestimation of complication rates (42). In generalized series of thyroid surgery, permanent injury to the recurrent laryngeal nerve varied from 0.3% in patients with simple, benign unilateral thyroid lesions up to 30% in those undergoing repeat surgery (42, 43). In a series of 3660 patients who had routine thyroid surgery, the Scandinavian Thyroid Quality Registry reported permanent recurrent nerve injuries in 4.1% and permanent hypocalcemia in 4.4% (44). In our study with only severely affected patients, a complication rate of 5.5% for recurrent laryngeal nerve injuries (2.7% of the nerves at risk) and 4.1% for permanent hypocalcemia may be considered a reasonable tradeoff between risk and benefit. Thus, total thyroidectomy seems to be a safe and effective treatment option for patients who have Hashimoto disease with severe Hashimoto-related symptoms, if these symptoms are not relieved by adequate hormone replacement therapy alone. Because symptom relief is the dominant goal for many patients, we encourage further studies using symptoms as the end point, with longer follow-up, and including differentiation among subgroups of patients with Hashimoto disease (45). In addition, studies of immunologic mechanisms in patients with surgically treated Hashimoto disease are warranted.

In conclusion, total thyroidectomy in patients with histologically verified Hashimoto disease, a heavy nonhypothyroid symptom burden, and serum anti-TPO antibody titers greater than 1000 IU/mL improved health-related quality of life and fatigue. The concomitant elimination of serum anti-TPO antibodies warrants further studies on disease mechanisms.
H.S.; University of Stavanger and Stavanger University Hospital, Stavanger, Norway (J.T.K.); and Telemark Hospital, Skien, Norway, Vestfold Hospital, Tønsberg, Norway, and University of Oslo, Oslo, Norway (T.B.).

Acknowledgment: The authors thank study nurse Rosemarie Ditlefsen for her diligence in following the study participants. Without her engagement with these patients, this study would have been impossible to carry out. The authors also thank the ear, nose, and throat department staff for performing pre- and postoperative laryngoscopy for all patients who received surgical treatment.

Financial Support: From Inge Steenslands Stiftelse, Stavanger, Norway, and Telemark Hospital.

Disclosures: Dr. Hoff reports personal fees from Amgen Norway during the conduct of the study. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M18-0284.

Reproducible Research Statement: Study protocol, statistical code, and data set: Available from Dr. Guldvog (e-mail, ivaguld@online.no). Patient consent was not obtained, but the presented data are anonymized and the risk for identification is low.

Corresponding Author: Ivar Guldvog, MD, PhD, Department of Breast and Endocrine Surgery, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway; e-mail, ivaguld@online.no.

Current author addresses and author contributions are available at Annals.org.

References


Current Author Addresses: Drs. Guldvog, Reitsma, and Lauzike: Department of Breast and Endocrine Surgery, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway. Dr. Johnsen: Department of Clinical Medicine, University of Bergen, Jonas Lies v 65, N-5021 Bergen, Norway, or Department of Breast and Endocrine Surgery, Stavanger University Hospital, Helse Stavanger HF, P.O. Box 8100, N-4068 Stavanger, Norway. Dr. Gibbs: Department of Endocrinology, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway. Dr. Carlsen: Department of Pathology, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway. Dr. Lende: Department of Breast and Endocrine Surgery, Stavanger University Hospital, Helse Stavanger HF, P.O. Box 8100, N-4068 Stavanger, Norway, or Department of Clinical Medicine, University of Bergen, Jonas Lies v 65, N-5021 Bergen, Norway. Dr. Narvestad: Department of Gastrointestinal Surgery, Stavanger University Hospital, Helse Stavanger HF, P.O. Box 8100, N-4068 Stavanger, Norway. Dr. Omade: Clinical Immunology Unit, Department of Internal Medicine, Helse Stavanger HF, Stavanger University Hospital, P.O. Box 8100, N-4068 Stavanger, Norway, or Department of Clinical Science, University of Bergen, Jonas Lies v 87, N-5021 Bergen, Norway. Dr. Kvaløy: Department of Mathematics and Physics, University of Stavanger, P.O. Box 8600 Forus, N-4036 Stavanger, Norway, or Department of Research, Stavanger University Hospital, PO Box 8100, N-4068 Stavanger, Norway. Dr. Hoff: Department of Research, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway. Dr. Bernkleb: Department of Research, Telemark Hospital, Ulefossvegen 55, N-3710 Skien, Norway; Department of Research, Vestfold Hospital, PO Box 2168, N-3103 Tønsberg, Norway, or Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, PO Box 1171, Blindern, N-0318 Oslo, Norway. Dr. Soiland: Department of Breast and Endocrine Surgery, Stavanger University Hospital, Helse Stavanger HF, P.O. Box 8100, N-4068 Stavanger, Norway, or University of Bergen, Department of Clinical Science, Jonas Lies v 87, N-5021 Bergen, Norway.