Screening and Treatment of Thyroid Dysfunction: An Evidence Review for the U.S. Preventive Services Task Force

J. Bruin Rugge, MD, MPH; Christina Bougatsos, MPH; and Roger Chou, MD

Background: In 2004, the U.S. Preventive Services Task Force found insufficient evidence to recommend thyroid screening.

Purpose: To update the 2004 U.S. Preventive Services Task Force review on the benefits and harms of screening and treatment of subclinical and undiagnosed overt hypothyroidism and hyperthyroidism in adults without goiter or thyroid nodules.

Data Sources: MEDLINE and Cochrane databases through July 2014.

Study Selection: Randomized, controlled trials and observational studies of screening and treatment.

Data Extraction: One investigator abstracted data, and a second investigator confirmed; 2 investigators independently assessed study quality.

Data Synthesis: No study directly assessed benefits and harms of screening versus no screening. For subclinical hypothyroidism (based on thyroid-stimulating hormone levels of 4.1 to 11.0 mIU/L), 1 fair-quality cohort study found that treatment of subclinical hypothyroidism was associated with decreased risk for coronary heart disease events versus no treatment. No study found that treatment was associated with improved quality of life, cognitive function, blood pressure, or body mass index versus no treatment. Effects of treatment versus no treatment showed potential beneficial effects on lipid levels, but effects were inconsistent, not statistically significant in most studies, and of uncertain clinical significance (difference, −0.7 to 0 mmol/L [−28 to 0 mg/dL] for total cholesterol levels and −0.6 to 0.1 mmol/L [−22 to 2 mg/dL] for low-density lipoprotein cholesterol levels). Treatment harms were poorly studied and sparsely reported. Two poor-quality studies evaluated treatment of subclinical hyperthyroidism but examined intermediate outcomes. No study evaluated treatment versus no treatment of screen-detected, undiagnosed overt thyroid dysfunction.

Limitation: English-language articles only, no treatment study performed in the United States, and small trials with short duration that used different dosage protocols.

Conclusion: More research is needed to determine the clinical benefits associated with thyroid screening.

Primary Funding Source: Agency for Healthcare Research and Quality.

An estimated 5% of women and 3% of men in the United States have subclinical thyroid dysfunction (1), and approximately 0.5% of the population may have undiagnosed overt thyroid disease (2, 3). Subclinical thyroid dysfunction is defined as elevated or low results on a thyroid-stimulating hormone (TSH) test (reference range, 0.45 to 4.50 mIU/L) in the setting of normal thyroid hormone levels. Overt thyroid disease is defined by the presence of abnormal thyroid hormone (free thyroxine, with or without triiodothyronine) levels (4, 5) (Table 1). In some studies, subclinical hypothyroidism is associated with increased risk for coronary artery disease (6, 7); congestive heart failure (8); and subclinical hyperthyroidism with increased risk for all-cause and coronary heart disease mortality, atrial fibrillation (9), and decreased bone density (5). Overt thyroid disease is associated with negative cardiovascular, musculoskeletal, dermatologic, gastrointestinal, and other effects, but clinical manifestations are highly variable and depend on the severity of thyroid abnormalities. Thyroid screening could identify persons with subclinical as well as undiagnosed overt thyroid dysfunction who could potentially benefit from treatment to reduce the risk for adverse health outcomes.

In 2004, the U.S. Preventive Services Task Force (USPSTF) found insufficient evidence to recommend for or against thyroid screening in asymptomatic, nonpregnant adults. Although the USPSTF concluded that subclinical hypothyroidism is a risk factor for overt thyroid disease, it found insufficient data to estimate effects of early treatment on clinical outcomes. A contemporaneous systematic review conducted for the American Thyroid Association, the American Association of Clinical Endocrinologists, and the Endocrine Society reached similar conclusions (5). Nonetheless, prescribing rates for thyroid medication in the United States have increased dramatically, from an estimated 49.8 million in 2006 to 70.5 million in 2010 (10). Among community-dwelling persons who are older than 65 years with subclinical hypothyroidism, the proportion receiving thyroid hormone has more than doubled, from 8.1% to 20.0%, between 1989 and 2005 (11).

This report was commissioned by the USPSTF to update its 2004 recommendation on thyroid screening. It builds on a 2011 comparative-effectiveness review funded by the Agency for Healthcare Research and Quality (12) and previous USPSTF reviews on identification and treatment of subclinical thyroid dysfunction (1, 13). Before updating its 2004 recommendation, the

See also:

Web-Only Supplements
USPSTF determined that in addition to subclinical thyroid dysfunction, screening could also identify undiagnosed overt thyroid disease (2, 3); therefore, the decision to screen should also consider the potential benefits and harms of identifying and treating undiagnosed overt disease. Therefore, this update differs from previous USPSTF reviews and the 2011 review in that it also addresses identification and treatment of undiagnosed overt thyroid disease.

**METHODS**

**Key Questions and Analytic Framework**

We developed a review protocol and analytic framework (Supplement 1, available at www.annals.org) that included the following key questions:

1. Does screening for thyroid dysfunction reduce morbidity and mortality?
2. What are the harms of screening?
3. Does treating screen-detected overt or subclinical thyroid dysfunction improve: a) mortality and morbidity? or b) intermediate outcomes?
4. What are the harms of treating thyroid dysfunction detected by screening?

Detailed methods and data for this review, including search strategies, inclusion criteria, abstraction and quality rating tables, and evidence on benefits and harms of treatment of subclinical hyperthyroidism, are in the full report (14). The protocol was developed using a standardized process with input from the USPSTF, experts, and the public. The analytic framework addresses direct evidence on benefits and harms of thyroid screening, as well as benefits and harms of treatment of subclinical or overt thyroid dysfunction.

**Data Sources and Searches**

A research librarian searched MEDLINE, the Cochrane Central Register of Controlled Trials, and the Cochrane Database of Systematic Reviews from 2002 to mid-July 2014 for subclinical hypothyroidism and hyperthyroidism and without a previous date limitation for overt hypothyroidism and hyperthyroidism (Supplement 2, available at www.annals.org). Additional studies were identified from a review of reference lists of relevant articles and peer-review suggestions.

---

**Table 1. Classification of Thyroid Dysfunction: Biochemical Definition**

<table>
<thead>
<tr>
<th>TSH Level, by Condition</th>
<th>Thyroid Hormones</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overt hyperthyroidism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.1 mIU/L or undetectable</td>
<td>Elevated thyroxine or triiodothyronine</td>
<td></td>
</tr>
<tr>
<td>&gt;4.5 mIU/L</td>
<td>Low thyroxine</td>
<td></td>
</tr>
<tr>
<td>Subclinical hyperthyroidism</td>
<td>Normal thyroxine and triiodothyronine</td>
<td>Clearly low serum TSH</td>
</tr>
<tr>
<td>&lt;0.1 mIU/L</td>
<td>Normal thyroxine and triiodothyronine</td>
<td>Low but detectable</td>
</tr>
<tr>
<td>0.1–0.4 mIU/L</td>
<td>Normal thyroxine and triiodothyronine</td>
<td></td>
</tr>
<tr>
<td>Subclinical hypothyroidism</td>
<td>Normal thyroxine</td>
<td>Mildly elevated TSH</td>
</tr>
<tr>
<td>4.5–10.0 mIU/L</td>
<td>Normal thyroxine</td>
<td>Markedly elevated TSH</td>
</tr>
<tr>
<td>≥10 mIU/L</td>
<td>Normal thyroxine</td>
<td></td>
</tr>
</tbody>
</table>

TSH = thyroid-stimulating hormone.

* From references 4 and 5.

---

**Study Selection**

Two investigators independently evaluated each study at the title or abstract and full-text article stages to determine eligibility for inclusion. We included randomized trials and observational studies of thyroid screening versus no screening in adults (excluding pregnant women) without a history of thyroid dysfunction or obvious goiter, nodules, or symptoms, following the protocol. We also included studies of treatment versus no treatment in adults with subclinical or overt thyroid dysfunction. Screening was based on TSH testing, with follow-up testing of thyroid hormone levels (free thyroxine, with or without triiodothyronine). Studies of patients with subclinical hypothyroidism due to Hashimoto thyroiditis (based on antibody testing) were included if they did not describe enrollment of symptomatic patients. Clinical outcomes were cardiovascular end points (cardiovascular disease, coronary artery disease or congestive heart failure, and atrial fibrillation); fractures; measures of quality of life or cognitive function; and harms, including those related to overplacement (such as negative effects on bone mineral density or atrial fibrillation). Intermediate outcomes were effects on lipid levels, blood pressure, weight change, and bone mineral density.

We restricted inclusion to English-language articles and excluded studies published only as abstracts. The literature flow diagram is shown in Supplement 3 (available at www.annals.org).

---

**Data Abstraction and Quality Assessment**

One investigator abstracted details about the study design, patient population, setting, screening method, interventions, data analysis, and results, and another investigator verified data abstraction for accuracy. Two investigators independently applied criteria developed by the USPSTF to rate the quality of each study as good, fair, or poor (15, 16). Discrepancies were resolved through a consensus process. For all studies, we evaluated applicability to populations likely to be encountered in primary care screening settings.
RESULTS

No study compared clinical benefits or harms in persons screened versus not screened for thyroid dysfunction (the first 2 key questions). We identified 11 trials (in 14 publications) and 1 retrospective cohort study on treatment of subclinical hypothyroidism (17–31). Two studies examined treatment of subclinical hyperthyroidism but were poor-quality and evaluated intermediate outcomes (32, 33); they are discussed in the full report (14). No study addressed treatment versus no treatment of screen-detected, undiagnosed overt hypothyroidism.

Three trials were rated good-quality (22, 27, 28), 6 trials (reported in 7 publications) were rated fair-quality (18–21, 23, 24, 26), and 4 trials (in 5 publications) were rated poor-quality (17, 30–33). The retrospective cohort study was rated fair-quality (29). Most of the poor-quality studies were characterized by poor reporting of methods (such as methods of randomization, allocation concealment, blinding, and reporting of attrition) rather than clearly inadequate methods.

None of the trials were conducted in the United States. The TSH values used to diagnose subclinical thyroid dysfunction at baseline ranged from 4.1 to 11.0 mIU/L. Mean patient ages ranged from 32 years to older than 70 years. Treatment of subclinical hypothyroidism was levothyroxine using different dosing regimens. Study duration ranged from 4 to 12 months, except for the cohort study, which analyzed data with up to a 7.6-year follow-up (29). Sample sizes in the trials ranged from 14 to 120 patients; the cohort study analyzed 4735 patients (29). Whereas most studies evaluated a placebo comparator, 3 used a no-treatment comparison (18, 19, 29).

Three trials (in 4 publications) specifically evaluated screen-detected populations (17, 20, 21, 27). Most other trials did not clearly report how patients were identified, other than that they were recruited from outpatient clinics. Trials generally reported that patients were newly diagnosed and excluded those with previous thyroid dysfunction or previous receipt of antithyroid medications.

Effectiveness of Treatment of Subclinical Hypothyroidism on Clinical Outcomes

Cardiovascular Events and Mortality

The 2004 USPSTF review (1, 13) included no study on the effects of treatment of subclinical hypothyroidism on risk for cardiac events or death. We identified 1 fair-quality, retrospective cohort study published since the 2004 USPSTF review on the effects of treatment of subclinical hypothyroidism (based on a single TSH level of >5.01 to 10.00 mIU/L) in 4735 persons aged 40 years or older in the United Kingdom on risk for cardiac events (mean follow-up, 7.6 years) (29) (Table 2). On the basis of an a priori categorization, analyses were stratified by age (40 to 70 years vs. >70 years), and analyses of the entire cohort were not reported. Approximately one half of the persons were treated with levothyroxine (mean dose, 75 µg/day).

After adjustment for age, sex, body mass index, socioeconomic status, total cholesterol level, smoking status, history of diabetes mellitus, index serum TSH level, and blood pressure, levothyroxine use versus no treatment was associated with lower risk for fatal or nonfatal ischemic heart disease events (4.2% vs. 6.6%; hazard ratio [HR], 0.61 [95% CI, 0.39 to 0.95]), all-cause mortality (3.4% vs. 6.4%; HR, 0.36 [CI, 0.19 to 0.66]), death due to circulatory diseases (1.4% vs. 2.4%; HR, 0.54 [CI, 0.37 to 0.92]), and cancer mortality (1.2% vs. 2.2%; HR, 0.59 [CI, 0.21 to 0.88]) in the younger age group (40 to 70 years) (29). In patients older than 70 years, there was no association between use of levothyroxine versus nonuse and risk for ischemic heart disease events (HR, 0.99 [CI, 0.59 to 1.33]), all-cause mortality (35.2% vs. 40.5%; HR, 0.71 [CI, 0.56 to 1.08]), or cancer mortality (4.6% vs. 6.5%; HR, 0.51 [CI, 0.24 to 1.09]). Potential limitations include the lack of adjustment for medica- tions to reduce risk for cardiovascular disease, although baseline data suggested no differences between treatment groups.

Quality of Life

The 2004 USPSTF review (1, 13) included 5 trials on the association between treatment of subclinical thyroid dysfunction and quality of life (22, 34–37). One trial found treatment of subclinical hypothyroidism associated with better quality of life in patients with recent Graves disease (34). The other 4 trials found no effects of treatment (22, 35–37). However, 3 of these trials would have been excluded from this update because patients were previously treated for thyroid dysfunction (34, 36) or because it enrolled mostly patients who were euthyroid (37).
Table 2. Subclinical Hypothyroidism Cardiovascular Events and Mortality*

<table>
<thead>
<tr>
<th>Study</th>
<th>Year (Reference); Study Design; and Study Duration</th>
<th>Country; Age, and TSH Level</th>
<th>Patients, n</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Razvi et al, 2012 (29); retrospective cohort study (database analysis); median, 7.6 y for 40–70 y age group and 5.2 y for &gt;70 y age group</td>
<td>United Kingdom</td>
<td>Aged 40–70 y: treated§ (median), 1634; not treated, 1459</td>
<td>Fatal and nonfatal ischemic heart disease events; all-cause mortality</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aged &gt;70 y: treated§ (median), 819; not treated, 832</td>
<td>Death due to circulatory diseases; death due to ischemic heart disease events</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Death due to malignant neoplasms; fatal and nonfatal cerebrovascular accident</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Atrial fibrillation</td>
<td></td>
</tr>
</tbody>
</table>

TSH = thyroid-stimulating hormone.
* The quality of all studies was fair.
† Multivariate-adjusted for age, sex, body mass index, socioeconomic deprivation score, total cholesterol, TSH level, smoking status, history of diabetes mellitus, systolic and diastolic blood pressure, and levothyroxine use as a time-dependent covariate.
‡ Adjusted for age and sex.
§ Received levothyroxine, 75 μg/d.
\[\text{Significant difference.}\]
¶ Ischemic heart disease, cerebrovascular accident, and peripheral vascular disease.

We identified 5 trials (3 good-quality [22, 27, 28], 1 fair-quality [21], and 1 poor-quality [17]) published since the 2004 USPSTF review on effects of treatment of subclinical hypothyroidism (TSH thresholds varied from >3.5 to >5.5 mIU/L using various doses of levothyroxine (mean, 50.0 to 109.7 μg/day) on measures of quality of life (Short Form-36 Health Survey, the General Health Questionnaire-30, the Beck Depression Inventory, the Hospital Anxiety and Depression Scale, and the Underactive Thyroid-Dependent Quality of Life Questionnaire) (Table 3). Sample sizes were less than 100 in all trials, mean age ranged from 45 to 74 years, and follow-up ranged from 4 to 12 months. No differences were found between treatment and placebo in any study. Three trials evaluated screen-detected populations (17, 21, 27).

Cognitive Function

Two trials included in the previous USPSTF review evaluated effects of treatment of subclinical hypothyroidism on cognitive function (35, 37). One trial that also included patients who were euthyroid (37) found no effect, and the second trial found a statistically significant but clinically small improvement in memory using a composite outcome in persons older than 55 years (35).

We identified 1 good-quality (27) and 1 fair-quality (21) trial published since the last USPSTF review that found no association between treatment with levothyroxine for subclinical hypothyroidism (defined as TSH levels >3.5 mIU/L [21] or TSH levels >5.5 mIU/L [27]) versus placebo and various measures of cognitive function after 12 months (Table 3). Both studies seemed to evaluate screen-detected populations. Mean ages were 62 to 63 years in 1 study (69 patients) (21) and 74 years in the other (94 patients) (27). The good-quality study (27) found no effects on cognitive skills and performance (Middlesex Elderly Assessment of Mental State score, 11.67 vs. 11.60; \( P = 0.57 \)), cognitive status (Mini-Mental State Examination score, 28.24 vs. 28.22; \( P = 0.18 \)), speed of cognitive processing and accounting (Speed and Capacity of Language Processing Test score, 1.29 vs. 0.84; \( P = 0.59 \)), or psychomotor tests of executive function (Trail Making Test, Part A or B).

Effectiveness of Treatment of Subclinical Hypothyroidism on Intermediate Outcomes

Blood Pressure

The previous USPSTF review (1, 13) included no study on the effect of treatment of subclinical hypothyroidism on blood pressure.

We identified 1 good-quality (28) and 2 fair-quality (24, 26) trials that found no effects on blood pressure between treatment versus no treatment of subclinical hypothyroidism (defined as TSH levels >3.6 mIU/L [24] or >4.0 mIU/L [28], or TSH levels greater than the normal limit [26]). Differences between treatment and placebo groups in mean systolic blood pressure ranged from −3 to −2 mm Hg and in mean diastolic blood pressure ranged from −3 to 0 mm Hg (Supplement 4, available at www.annals.org).

Lipids

The previous USPSTF review included 7 trials on the effect of treatment of subclinical hypothyroidism and effects on lipid profiles (22, 25, 34–36, 38, 39). Six trials found no improvement in lipid variables (22, 25, 34–36, 39), with 1 poor-quality trial in euthyroid patients reporting approximately a 5% improvement in low-density lipoprotein (LDL) cholesterol levels with 50 μg/day versus 25 μg/day of levothyroxine (38).

Total Cholesterol

We identified 2 good-quality trials (22, 28), 6 fair-quality trials (18, 19, 23, 24, 26), and 1 poor-quality trial (30, 31) trial published since the 2004 USPSTF review on the effects of treatment of subclinical hypothyroidism on total cholesterol levels. Thyroid-stimulating hormone thresholds varied from greater than 3.6 to greater than 5 mIU/L, or “greater than the upper limit of normal” (26) (Supplement 4). In the 8 good- and fair-quality trials,
differences between treatment and no treatment in mean total cholesterol levels ranged from −0.7 to 0 mmol/L (−28 to 0 mg/dL). Three of the trials (45, 100, and 120 patients) reported statistically significant differences in mean total cholesterol levels of −0.3 mmol/L (−12 mg/dL) \( (P < 0.03) \) (23), −0.7 mmol/L (−28 mg/dL) \( (P = 0.03) \) (24), and −0.3 mmol/L (−12 mg/dL) \( (P < 0.001) \) (28). The poor-quality trial found treatment associated with slightly lower total cholesterol levels (difference in means, −0.2 mmol/L [−6 mg/dL], \( P = 0.03 \)) (30, 31).

### Low-Density Lipoprotein Cholesterol

Two good-quality trials (22, 28), 6 fair-quality trials (18–20, 23, 24, 26), and 1 poor-quality (30, 31) trial published since the previous USPSTF review evaluated the effect of treatment of subclinical hypothyroidism on LDL cholesterol levels (Supplement 4). In the 8 good- and fair-quality trials, differences between treatment and no treatment in mean LDL cholesterol levels ranged from −0.6 to 0.1 mmol/L (−22 to 2 mg/dL). Three of the trials (45, 100, and 120 patients) reported statistically significant differences in mean LDL cholesterol levels of −0.2 mmol/L (−8 mg/dL) \( (P < 0.001) \) (23), −0.6 mmol/L (−22 mg/dL) \( (P = 0.03) \) (24), and −0.3 mmol/L (−12 mg/dL) \( (P < 0.001) \) (28). The poor-quality trial found treatment associated with slightly lower LDL cholesterol levels (difference in means, −0.3 mmol/L [−12 mg/dL], \( P = 0.02 \)) (30, 31).

### High-Density Lipoprotein Cholesterol

We identified 2 good-quality trials (22, 28), 6 fair-quality trials (18–20, 23, 24, 26), and 1 poor-quality (30, 31) trial on the effect of treatment of subclinical hypothyroidism on triglyceride levels (Supplement 4). In the 8 good- and fair-quality trials, differences in means ranged from −0.4 to 0.1 mmol/L (−32 to 11 mg/dL). None of the trials found a significant difference between treatment and control in triglyceride values.

### Body Mass Index or Weight

No study in the 2004 USPSTF review assessed effects of treatment of subclinical thyroid dysfunction on body mass index or weight. We identified 2 good-quality (22, 28) and 4 fair-quality (19, 20, 24, 26) trials published since the 2004 USPSTF review on the effect of treatment of subclinical hypothyroidism (TSH thresholds varied from >3.5 to >5 mIU/L, or “greater than the normal range” [26]) on body mass index or weight (Supplement 4). None of the trials found a significant difference between treatment and control groups in body mass index or weight. Of 5 trials reporting body mass index, differences between treatment and placebo groups ranged from −1 to 1 kg/m². Of the 2 trials reporting weight, 1 found a difference in means of −1 kg (28) and 1 found a 0.1% difference in lean body weight (22).

### Harms of Treatment of Subclinical Hypothyroidism

The 2004 USPSTF report found very limited evidence on harms related to treatment of subclinical hypothyroidism. One good-quality trial of patients who developed subclinical hypothyroidism after treatment of Graves disease found that 4 of 17 persons randomly assigned to levothyroxine felt worse than 6 of 15 persons given placebo \( (P = 0.33) \) (34). Other studies reported 1 case of angina (35), 1 case of atrial fibrillation (35), decreased anxiety scores (22), decreased Short Form-36 Health Survey vitality scores (37), and 2 withdrawals due to adverse events (39).

Five trials (in 6 publications) published since the 2004 USPSTF review reported harms, but they were poorly assessed and reported, precluding reliable conclusions (17, 26–28, 30, 31). In addition, the studies were not designed or powered to assess long-term or serious harms, or harms related to overtreatment. One study reported “no indication of harms” (17), and another study stated that none of the patients reported adverse events requiring withdrawal or dose reduction.

---

**Table 2**—Continued

<table>
<thead>
<tr>
<th>Events in Treated vs. Nontreated Groups, %</th>
<th>Multivariate-Adjusted Hazard Ratio (95% CI)†</th>
<th>Adjusted Hazard Ratio (95% CI)‡</th>
<th>Events in Treated vs. Nontreated Groups, %</th>
<th>Multivariate-Adjusted Hazard Ratio (95% CI)†</th>
<th>Adjusted Hazard Ratio (95% CI)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 vs. 6.6</td>
<td>0.61 ( (0.39-0.95) )</td>
<td>0.64 ( (0.41-0.98) )</td>
<td>12.7 vs. 10.7</td>
<td>0.99 ( (0.59-1.33) )</td>
<td>1.03 ( (0.98-1.83) )</td>
</tr>
<tr>
<td>3.4 vs. 6.4</td>
<td>0.36 ( (0.19-0.66) )</td>
<td>0.43 ( (0.30-0.78) )</td>
<td>35.2 vs. 40.5</td>
<td>0.71 ( (0.56-1.08) )</td>
<td>0.91 ( (0.65-1.46) )</td>
</tr>
<tr>
<td>1.4 vs. 2.4</td>
<td>0.54 ( (0.37-0.92) )</td>
<td>0.61 ( (0.37-1.91) )</td>
<td>17.1 vs. 18.3</td>
<td>0.91 ( (0.56-1.46) )</td>
<td>0.87 ( (0.43-1.37) )</td>
</tr>
<tr>
<td>1.0 vs. 1.7</td>
<td>0.43 ( (0.19-2.05) )</td>
<td>0.55 ( (0.38-1.19) )</td>
<td>5.5 vs. 6.3</td>
<td>1.04 ( (0.56-1.93) )</td>
<td>1.12 ( (0.66-2.05) )</td>
</tr>
<tr>
<td>1.2 vs. 2.2</td>
<td>0.59 ( (0.21-0.88) )</td>
<td>0.61 ( (0.36-0.95) )</td>
<td>4.6 vs. 6.5</td>
<td>0.51 ( (0.24-1.09) )</td>
<td>0.73 ( (0.34-1.61) )</td>
</tr>
<tr>
<td>3.4 vs. 3.0</td>
<td>1.03 ( (0.51-2.13) )</td>
<td>1.09 ( (0.75-1.89) )</td>
<td>17.7 vs. 17.9</td>
<td>0.81 ( (0.31-2.12) )</td>
<td>1.11 ( (0.45-2.01) )</td>
</tr>
<tr>
<td>2.0 vs. 2.3</td>
<td>0.76 ( (0.26-1.73) )</td>
<td>0.87 ( (0.59-1.44) )</td>
<td>8.1 vs. 7.7</td>
<td>0.98 ( (0.54-1.76) )</td>
<td>1.23 ( (0.69-1.58) )</td>
</tr>
</tbody>
</table>
40
discussion

The evidence we reviewed is summarized in Table 4. As in the 2004 USPSTF review, we found no direct evidence on effects of thyroid screening versus no screening on clinical outcomes. The scope of this update was expanded to include detection and treatment

Table 3. Subclinical Hyperthyroidism Quality of Life and Cognitive Function

<table>
<thead>
<tr>
<th>Study, Year (Reference); Study Design; Study Duration; and Country</th>
<th>Mean Age and Mean TSH Level (Levothyroxine vs. Placebo)</th>
<th>Patients Receiving Intervention, n</th>
<th>Results in Levothyroxine Versus Placebo</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abu-Helal et al, 2010 (17); RCT crossover (at 2 mo); 4 mo; United Kingdom</td>
<td>58 y overall (NR by group) 4.1–9.0 mIU/L (mean NR)</td>
<td>Placebo: 53</td>
<td>QOL: Odds of patients feeling better while receiving levothyroxine vs. placebo: TSH level &gt;4.0 mIU/L: 21 vs. 16 patients; odds, 1.3 TSH level &gt;4.5 mIU/L: 17 vs. 7 patients; odds, 2.4 TSH level &gt;5.0 mIU/L: 12 vs. 5 patients; odds, 2.4 TSH level &gt;5.5 mIU/L: 11 vs. 4 patients; odds, 2.8 TSH level &gt;6.0 mIU/L: 8 vs. 2 patients; odds, 4.0</td>
<td>Poor</td>
</tr>
<tr>
<td>Jorde et al, 2006 (21); RCT; 12 mo; Norway</td>
<td>62 vs. 63 y 5.8 vs. 5.3 mIU/L</td>
<td>Levothyroxine, 109.7 μg for 12 mo (mean): 36 Placebo: 33</td>
<td>Mean GHQ-30 score (SD): 1.9 (3.3) vs. 1.2 (2.0); P = NS Mean BDI score (SD): 4.3 (3.6) vs. 3.3 (4.0); P = NS Mean change in GHQ score (mean): −2 (CI, −3 to −1); P = NS HADS, anxiety score: −1 (95% CI, −1 to −1); P = NS HADS, depression score: −1 (CI, −3 to 1); P = NS GHQ-30 score: 2 (CI, −5 to 7); P = NS</td>
<td>Fair</td>
</tr>
<tr>
<td>Kong et al, 2002 (22); RCT; 6 mo; United Kingdom</td>
<td>53 vs. 45 y 8.0 vs. 7.3 mIU/L</td>
<td>Levothyroxine for 6 mo (mean NR): 23 Placebo: 17</td>
<td>Mean change in levothyroxine group minus mean change in placebo group: HADS, anxiety score: 1 (95% CI, −1 to −1); P = NS HADS, depression score: 1 (95% CI, −1 to 1); P = NS</td>
<td>Good</td>
</tr>
<tr>
<td>Parle et al, 2010 (27); RCT; 12 mo; United Kingdom</td>
<td>73.5 vs. 74.2 y 6.6 vs. 6.6 mIU/L</td>
<td>Levothyroxine, 50 μg for 12 mo (median): 52 Placebo: 42</td>
<td>Mean HADS, depression score (SD): 3.55 (0.27) vs. 3.37 (0.31); P = 0.82</td>
<td>Good</td>
</tr>
<tr>
<td>Razvi et al, 2007 (28); RCT crossover (at 2.8 mo); 5.5 mo; United Kingdom</td>
<td>53.5 vs. 54.2 y 5.4 vs. 5.3 mIU/L</td>
<td>Levothyroxine, 100 μg for 12 wk: 50 Placebo: 50</td>
<td>Mean ThyDQoL score (SD): −1.1 (1.0) vs. −1.2 (0.9); P = 0.24 Mean SF-36, sex score (SD): −2.3 (2.7) vs. −2.7 (2.8); P = 0.18 Mean SF-36, motivation score (SD): −3.6 (2.7) vs. −3.7 (2.7); P = 0.16 Mean SF-36, worries score (SD): −2.5 (3.0) vs. −2.8 (2.9); P = 0.23 Mean weighted effect of all 18 QOL domains (SD): −2.7 (2.4) vs. −2.8 (2.3); P = 0.45</td>
<td>Good</td>
</tr>
</tbody>
</table>

Cognitive function

<table>
<thead>
<tr>
<th>Study, Year (Reference); Study Design; Study Duration; and Country</th>
<th>Mean Age and Mean TSH Level (Levothyroxine vs. Placebo)</th>
<th>Patients Receiving Intervention, n</th>
<th>Results in Levothyroxine Versus Placebo</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jorde et al, 2006 (21); RCT; 12 mo; Norway</td>
<td>62 vs. 63 y 5.8 vs. 5.3 mIU/L</td>
<td>Levothyroxine, 109.7 μg for 12 mo (mean): 36 Placebo: 33</td>
<td>Mean composite cognitive function score (SD): 1.5 (3.7) vs. −0.9 (4.8); P = NS Mean Trail Making Test, Part A, psychomotor test of executive function score (SD): 39.0 (14.8) vs. 44.1 (17.7); P = NS Mean Trail Making Test, Part B, psychomotor test of executive function score (SD): 94.62 (49); P = NS</td>
<td>Fair</td>
</tr>
<tr>
<td>Parle et al, 2010 (27); RCT; 12 mo; United Kingdom</td>
<td>73.5 vs. 74.2 y 6.6 vs. 6.6 mIU/L</td>
<td>Levothyroxine, 50 μg for 12 mo (median): 52 Placebo: 42</td>
<td>Mean MEAMS, cognitive skills and performance score (SD): 11.67 (0.09) vs. 11.60 (0.11); P = 0.57 Mean MMSE, cognitive status score (SD): 28.24 (0.38) vs. 28.22 (0.43); P = 0.18 Mean SCOLP Test, speed of cognitive processing and accounting score (SD): 1.29 (0.30) vs. 0.84 (0.35); P = 0.59 Mean Trail Making Test, Part A, psychomotor test of executive function score (SD): 45.33 (2.63) vs. 46.78 (3.05); P = 0.52 Mean Trail Making Test, Part B, psychomotor test of executive function score (SD): 100.65 (7.75) vs. 114.11 (9.07); P = 0.95 Mean Trail Making Test, Part B-Part A, psychomotor test of executive function score (SD): 54.55 (6.80) vs. 67.27 (7.97); P = 0.86</td>
<td>Good</td>
</tr>
</tbody>
</table>

BDI = Beck Depression Inventory; GHQ = General Health Questionnaire; HADS = Hospital Anxiety and Depression Scale; MEAMS = Middlesex Elderly Assessment on Mental State; MMSE = Mini-Mental State Examination; NR = not reported; NS = not significant; QOL = quality of life; RCT = randomized, controlled trial; SCOLP = Speed and Capacity of Language Processing; SF-36 = Short Form-36 Health Survey; ThyDQoL = Underactive Thyroid-Dependent Quality of Life Questionnaire; TSH = thyroid-stimulating hormone.

(26). One study reported no difference between treatment versus placebo in risk for withdrawal due to adverse events after 12 months (9.6% vs. 14.3%; P = 0.49) (27). Two other trials (100 and 60 patients) reported 0 (28) or 2 (30, 31) withdrawals due to adverse events in patients with subclinical hypothyroidism.

Discussion

As in the 2004 USPSTF review, we found no direct evidence on effects of thyroid screening versus no screening on clinical outcomes. The scope of this update was expanded to include detection and treatment
Table 4. Summary of Evidence*

<table>
<thead>
<tr>
<th>Previous Report Findings</th>
<th>Studies Identified in Update</th>
<th>Limitations</th>
<th>Consistency</th>
<th>Applicability</th>
<th>Summary of Findings</th>
<th>Overall Quality†</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ 1. Does screening for thyroid dysfunction reduce morbidity and mortality?</td>
<td>No studies</td>
<td>No studies</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KQ 2. What are the harms of screening?</td>
<td>No studies</td>
<td>No studies</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KQ 3a. Does treating screen-detected overt or subclinical thyroid dysfunction improve mortality and morbidity?</td>
<td>Subclinical hypothyroidism</td>
<td>Cardiovascular events, coronary artery disease, and heart failure</td>
<td>1 retrospective cohort study</td>
<td>Did not adjust for use of aspirin, lipid-lowering therapy, or cardiovascular medications</td>
<td>Study population in United Kingdom</td>
<td>1 fair-quality retrospective cohort study found treatment for subclinical hypothyroidism associated with decreased risk for cardiac events, cancer, and all-cause mortality in adults aged 40-70 y but not in those aged &gt;70 y. However, this study had methodological limitations, including failure to adjust for some important confounders. The findings could represent a true effect or a spurious association as a result of residual confounding.</td>
</tr>
<tr>
<td>Overall quality of life</td>
<td>Only 1 of 5 trials found improvement in quality of life; most studies evaluated patients previously treated for Graves disease</td>
<td>5 RCTs</td>
<td>Trials were small and of short duration</td>
<td>Consistent</td>
<td>Study populations in Norway and United Kingdom</td>
<td>Levothyroxine associated with no effect on quality of life using various measures</td>
</tr>
<tr>
<td>Changes in cognition</td>
<td>1 of 2 trials found a statistically significant improvement in memory in persons aged &gt;55 y that the authors described as &quot;small and of questionable clinical importance&quot;</td>
<td>2 RCTs</td>
<td>Trials were small and of short duration</td>
<td>Consistent</td>
<td>Study populations in Norway and United Kingdom</td>
<td>Levothyroxine associated with no effect on cognitive function using various measures</td>
</tr>
<tr>
<td>Overt hypothyroidism‡</td>
<td>Not assessed</td>
<td>No studies</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KQ 3b. Does treating screen-detected overt or subclinical thyroid dysfunction improve intermediate outcomes?</td>
<td>Subclinical hypothyroidism</td>
<td>Blood pressure changes</td>
<td>3 RCTs</td>
<td>Studies were small, of limited duration, and used different cutoffs for TSH and different dosing protocols</td>
<td>Study populations in Italy, Japan, United Kingdom</td>
<td>Levothyroxine associated with no effect on systolic blood pressure (difference range, −3 to −2 mm Hg) or diastolic blood pressure (difference range, −3 to 0 mm Hg)</td>
</tr>
</tbody>
</table>

Continued on following page
of screen-detected, undiagnosed overt thyroid disease, but we found no studies of treatment versus no treatment, probably because treatment is considered the standard of care for this condition.

Evidence on benefits and harms of treatment was largely restricted to patients with subclinical hypothyroidism. Despite the potential association between subclinical hypothyroidism and cardiovascular disease and congestive heart failure (6–8), there is no clear evidence that treatment improves clinical outcomes. Although 1 fair-quality retrospective cohort study found treatment of subclinical hypothyroidism associated with decreased risk for cardiac events, cancer, and all-cause mortality in adults aged 40 to 70 years (29), it was an observational study with potential methodological limitations, including a lack of adjustment for some important confounders. As in the 2004 USPSTF review, evidence from newer trials found that treatment of subclinical hypothyroidism was not associated with clear improvement in quality of life or measures of cognitive function (17, 21, 22, 27, 28). Findings about intermediate outcomes were also consistent with the 2004 USPSTF review. Trials found no clear benefits of treatment of subclinical hypothyroidism with greater TSH levels, other trials also tended to report a slight trend toward beneficial effects, although nonsignificant. However, differences were small (−0.7 to 0 mmol/L [−28 to 0 mg/dL] for total cholesterol levels and −0.6 to 0.1 mmol/L [−22 to 2 mg/dL] for LDL cholesterol levels). Treatment for subclinical hypothyroidism was not associated with beneficial effects on HDL cholesterol levels (−0.1 to 0.1 mmol/L [−4 to 4 mg/dL]) or triglyceride levels (−0.4 to 0.1 mmol/L [−32 to 11 mg/dL]).

Table 4—Continued

<table>
<thead>
<tr>
<th>Previous Report Findings</th>
<th>Studies Identified in Update</th>
<th>Limitations</th>
<th>Consistency</th>
<th>Applicability</th>
<th>Summary of Findings</th>
<th>Overall Quality†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changes in lipid levels</td>
<td>9 RCTs</td>
<td>Inconsistent</td>
<td>Study populations in United Kingdom, Brazil, Italy, Turkey, Norway, Kuwait, and Japan</td>
<td>3 of 8 good- and fair-quality trials found treatment associated with lower total cholesterol and LDL cholesterol levels; for total cholesterol levels, other trials also tended to report a slight trend toward beneficial effects, although nonsignificant. However, differences were small (−0.7 to 0 mmol/L [−28 to 0 mg/dL] for total cholesterol levels and −0.6 to 0.1 mmol/L [−22 to 2 mg/dL] for LDL cholesterol levels). Treatment for subclinical hypothyroidism was not associated with beneficial effects on HDL cholesterol levels (−0.1 to 0.1 mmol/L [−4 to 4 mg/dL]) or triglyceride levels (−0.4 to 0.1 mmol/L [−32 to 11 mg/dL]).</td>
<td>Fair</td>
<td></td>
</tr>
<tr>
<td>Incidental findings</td>
<td>5 RCTs for subclinical hypothyroidism</td>
<td>Only 1 trial directly compared harms between treated and nontreated adults; all other trials reported ad hoc adverse effects</td>
<td>Not able to assess</td>
<td>Study populations conducted in United Kingdom, Japan, Brazil, Italy</td>
<td>Only 1 trial in subclinical hypothyroidism patients directly compared harms between treated and nontreated adults and found no difference in withdrawals due to side effects; all other trials reported ad hoc adverse effects</td>
<td>Poor</td>
</tr>
<tr>
<td>BMI/weight changes</td>
<td>6 RCTs</td>
<td>Consistent</td>
<td>Levothyroxine associated with no effect on BMI (difference range, −1 to 1 kg/m²) or weight (difference of −1 kg in 1 study)</td>
<td>Fair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overt hypothyroidism‡</td>
<td>Not assessed</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

BMI = body mass index; HDL = high-density lipoprotein; KQ = key question; LDL = low-density lipoprotein; NA = not applicable; RCT = randomized, controlled trial; TSH = thyroid-stimulating hormone.

* Hyperthyroidism results are summarized in the full report (14).
† The aggregate internal validity rating of the body of the evidence based on study limitations, precision, consistency, and applicability.
‡ Asymptomatic or mildly symptomatic patients with biochemically overt thyroid disease.
levels, which may be associated with increased risk for adverse clinical outcomes.

As detailed in the full report (14), only 2 poor-quality trials evaluated effects of treatment of subclinical hyperthyroidism on intermediate outcomes (32, 33). Although a recent systematic review of observational studies found that subclinical hyperthyroidism may be associated with an increased risk for fractures, it only found trends, not statistically significant effects (40).

The harms of screening remain poorly studied and sparsely reported. Potential harms of screening for subclinical hypothyroidism include false-positive test results, anxiety related to test results, and harms of treatment (including overreplacement or overtreatment), but evidence is too limited to estimate effects on any of these outcomes. Two prospective cohort studies suggest that approximately 40% of persons with subclinical hypothyroidism were biochemically euthyroid after 3 years of watchful waiting, suggesting that overdiagnosis and subsequent overtreatment could be an issue (41, 42).

Our review has several limitations. We did not include non-English-language articles, and we could not assess for publication bias using graphical or statistical methods because of small numbers of studies (43). Limitations of the evidence include methodological shortcomings in most studies, the lack of studies conducted in the United States, small sample sizes, relatively brief duration of follow-up (4 to 12 months), and variability in criteria used to define thyroid dysfunction and in the treatment regimens. The applicability of the evidence on treatment is also uncertain because few trials clearly evaluated screen-detected populations or described how patients were identified (17, 20, 21, 27).

Additional research may help clarify the benefits and harms of thyroid screening. To better understand potential benefits and harms of thyroid screening, research is needed on the prevalence of unrecognized overt thyroid disease and on effects of treatment in such patients. Trials that evaluate clinical outcomes associated with thyroid screening versus no screening would provide the most direct evidence but may require large samples and long duration of follow-up to evaluate cardiovascular outcomes and mortality rates. Therefore, it may be prudent to first conduct well-designed trials of treatment of subclinical hypothyroidism versus placebo or no treatment in screen-detected populations because determining treatment efficiency is a prerequisite for effective screening interventions. Observational studies could help provide important information on effects of screening and treatment but should be conducted in well-defined populations and account for important confounders (such as patient demographic characteristics, medical and psychiatric co-morbid conditions, risk factors for cardiovascular disease, and concomitant medication use) in their design and analysis. Another emerging area with implications for screening is research suggesting that subclinical hypothyroidism may be protective in older persons (41, 44, 45) and that the reference ranges for TSH should be adjusted upward in older adults (46, 47).

search to clarify criteria for abnormal thyroid function would have important implications for defining the target populations and understanding the effect of screening.

In conclusion, screening can identify patients with subclinical thyroid dysfunction or undiagnosed overt thyroid disease, but direct evidence on benefits and harms of screening versus no screening remains unavailable. Trials of treatment of subclinical hypothyroidism suggest potential beneficial effects on total cholesterol and LDL cholesterol levels, but results were inconsistent and the magnitude of effect was of uncertain clinical significance. The only study showing a beneficial effect of treatment on cardiovascular events was observational and susceptible to residual confounding. Trials on the effects of treatment of subclinical hypothyroidism on other clinical and intermediate outcomes showed no clear beneficial effects, and data on harms were poor. More research is needed to understand effects of treatment of subclinical thyroid dysfunction and screen-detected, undiagnosed overt thyroid disease.

From Oregon Health & Science University, Portland, Oregon.

Disclaimer: The authors of this article are responsible for its content. Statements in the article should not be construed as endorsement by the Agency for Healthcare Research and Quality of the U.S. Department of Health and Human Services.

Acknowledgment: The authors thank Agency for Healthcare Research and Quality Medical Officer Jennifer Croswell, MD, MPH. They also thank the U.S. Preventive Services Task Force Lead Work Group, including Jessica Herzstein, MD, MPH; Wanda Nicholson, MD, MPH, MBA; Timothy Wilt, MD, MPH, and Virginia A. Moyer, MD, MPH. In addition, the authors thank Raj Sehgal, MD; Paul N. Gorman, MD; Howard Balshem, MS; and Rose Relevo, MLS, who were coauthors of a comparative effectiveness review on this same topic, on which this article builds.


Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M14-1456.

Requests for Single Reprints: J. Bruin Rugge, MD, MPH, Oregon Health & Science University, 3181 Southwest Sam Jackson Park Road, Portland, OR 97239; e-mail, ruggeb@ohsu.edu.

Current author addresses and author contributions are available at www.annals.org.

References

Screening and Treatment of Thyroid Dysfunction


47. Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the U.S. population: implications for the prevalence of subclinical hypothyroidism. J Clin Endocrinol Metab. 2007;92:4575-82. [PMID: 17911171]
Current Author Addresses: Dr. Rugge: Oregon Health & Science University, Mail Code FM, 3181 Southwest Sam Jackson Park Road, Portland, OR 97239.
Ms. Bougatsos and Dr. Chou: Oregon Health & Science University, Mail Code BICC, 3181 Southwest Sam Jackson Park Road, Portland, OR 97239.

Author Contributions: Conception and design: J.B. Rugge, C. Bougatsos, R. Chou.
Analysis and interpretation of the data: J.B. Rugge, C. Bougatsos, R. Chou.
Drafting of the article: J.B. Rugge, C. Bougatsos, R. Chou.
Critical revision of the article for important intellectual content: J.B. Rugge, R. Chou.
Final approval of the article: J.B. Rugge, C. Bougatsos, R. Chou.
Statistical expertise: R. Chou.
Obtaining of funding: R. Chou.
Administrative, technical, or logistic support: C. Bougatsos.
Collection and assembly of data: J.B. Rugge, C. Bougatsos, R. Chou.