CLINICAL STUDY

Subclinical thyroid dysfunction and mortality: an estimate of relative and absolute excess all-cause mortality based on time-to-event data from cohort studies

Patrick Haentjens1,2, Alain Van Meerhaeghe1, Kris Poppe4,5 and Brigitte Velkeniers4,5

1Center for Outcomes Research and Laboratory for Experimental Surgery, Universitair Ziekenhuis Brussel and Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium, 2CEBAM, Centre for Evidence Based Medicine, Belgian Branch of the Cochrane Collaboration, B-3000 Leuven, Belgium, 3Service de Pneumologie ISPPC CHU A. Vésale, Montigny-le-Tillé, Belgium and 4Department of Endocrinology and 5Department of General Internal Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, B-1090 Brussels, Belgium

(Correspondence should be addressed to P Haentjens; Email: patrick.haentjens@uzbrussel.be)

Abstract

Objectives: To what extent persons with subclinical hyper- or hypothyroidism are more (or less) likely to die than euthyroid control subjects remains a matter of controversy.

Methods: We searched electronic reference databases up to July 31, 2007. Three reviewers independently assessed eligibility. Cohort studies published in full that reported data on the hazard ratio (HR) for mortality from all causes in persons with subclinical thyroid dysfunction versus euthyroid controls were included.

Results: Based on seven cohorts including 290 participants with subclinical hyperthyroidism, random-effects models estimated that the pooled HR for all-cause mortality was 1.41 (95% confidence interval (CI), 1.12–1.79; P = 0.004). Using the pooled HR and standard life-table methods applied to a US reference population, we estimated that a white US woman, when diagnosed with subclinical hyperthyroidism at age of 70, has an excess mortality of 1.5, 4.0, and 8.7% at 2, 5, and 10 years respectively after diagnosis. Likewise, a white US man has an excess mortality of 2.3, 5.7, and 10.7%. For the nine cohorts including 1580 participants with subclinical hypothyroidism, observed heterogeneity (Q test P = 0.006; I² = 63%) disappeared after pooling cohorts in predefined subgroups according to the presence or absence of a comorbid condition. In doing so, the pooled HR for all-cause mortality was 1.03 (95% CI, 0.78–1.35; P = 0.83) in cohorts from the community and 1.76 (95% CI, 1.36–2.30; P < 0.001) in cohorts of participants with comorbidities (P = 0.014 for heterogeneity among study groups).

Conclusions: Individuals with subclinical hyperthyroidism demonstrate a 41% increase in relative mortality from all causes versus euthyroid control subjects. Mathematical modeling suggests that absolute excess mortality after diagnosis might depend on age, with an increase beyond the age of 60, especially in aging men. For patients with subclinical hypothyroidism, the relative risk of all-cause mortality is increased only in patients with comorbid conditions.

European Journal of Endocrinology 159 329–341

Introduction

Mild thyroid dysfunction, defined as serum thyrotropin (TSH) levels outside the normal reference range with normal levels of thyroid hormones, is a frequent finding, with estimates of the prevalence of subclinical hyper- and hypothyroidism in the general population varying from 1.5 to 5.9% and 2.9 to 16% respectively (1, 2). The prevalence of both subclinical hyper- and hypo-thyroidism is greater in women and increases with age (2).

Subclinical thyroid dysfunction has been associated with various adverse clinical outcomes, including altered serum cholesterol levels, heart rhythm and rate, ventricular function, and risk of coronary artery disease (3–7). To what extent subclinical thyroid dysfunction may increase mortality when compared with euthyroid controls remains an unsolved issue (8). Recently several meta-analyses have quantified the effect of subclinical hyper- and hypothyroidism in the general population on different endpoints, including cardiovascular morbidity, mortality, and finally, overall (all-cause) mortality (9–12). The currently available evidence for a relationship of subclinical thyroid dysfunction with mortality is weak and inconclusive. Overall, the interpretation of time-to-event (survival) data is difficult, particularly if individual studies present the effects of thyroid dysfunction on mortality in a number of different ways and no further analyses are undertaken by a systematic reviewer to make estimates comparable and easier to interpret (13–15). Furthermore, even when individual

© 2008 European Society of Endocrinology

DOI: 10.1530/EJE-08-0110

Online version via www.eje-online.org
papers or systematic reviews provide powerful estimates of relative risk, they usually give no information about the absolute risk of death; that is, how many individuals will actually sustain the event of interest (death) (16, 17). Absolute risk figures are increasingly being considered as the optimal basis for individual clinical treatment decisions and public health policy strategies (16–18). Robust data in this regard are essential given the ongoing debate on whether or not to treat these subclinical thyroid dysfunctions (19–23).

In the present systematic reviews and meta-analyses, we quantify the impact of mild thyroid dysfunction, defined on the basis of TSH and free thyroxine (fT4) serum values, on overall mortality in different clinical situations, i.e., population-based cohorts and populations with associated comorbid conditions. In doing so, we include only cohort studies among adults with subclinical thyroid dysfunction not directly resulting from the treatment of the underlying comorbid condition. To do so with sufficient power, we select only the cohort studies that provide sufficient information to estimate a hazard ratio (HR), which takes into account the number and timing of events, and the time until last follow-up for each patient who has not experienced an event, i.e., has been censored. A further aim is to provide estimates of absolute mortality risks, year-by-year, up to 10 years after the diagnosis of subclinical thyroid dysfunction, both in women and men.

Methods

Data sources

We searched for English and non-English articles using Embase and Medline (Ovid and PubMed), with the last computerized search undertaken on July 31, 2007. To avoid missing any relevant study, we used broadly defined medical subject heading terms and text words, including the following: ‘hypothyroidism,’ ‘hyperthyroidism,’ or ‘thyroid hormones’ and ‘fatal outcome,’ ‘mortality,’ ‘survival,’ or ‘death’. The Medline search was limited to humans, and all adults 19 + years old. The computerized search was supplemented by a manual search of the bibliographies of all retrieved articles. Potentially relevant articles were assessed for inclusion against pre-specified eligibility and exclusion criteria. Searching was performed by three independent reviewers (B V, A V M, and P H). Differences were resolved by consensus of the three reviewers.

Study eligibility

We included longitudinal (cohort) studies published in full that reported long-term data on mortality from all causes in participants with subclinical hyper- and/or hypothyroidism versus euthyroid controls. Studies were required to define thyroid status as a subclinical hyper- or hypothyroidism based on TSH and fT4 levels prior to inclusion, and to report (or provide information to compute) a HR for death from all causes with a 95% confidence interval (CI). If a particular participant population was reported in more than one publication, we included only the article that provided the most complete data set. Reviews, case–control studies, uncontrolled studies, retrospective studies, studies in which mortality from all causes were not reported as a separate outcome, and studies in which participants had thyroid dysfunction other than subclinical hyper- or hypothyroidism were excluded. More specifically, studies enrolling patients with overt hyper- or hypothyroidism were not included.

Data extraction

The outcome of primary interest was the HR for mortality from all causes in participants with subclinical hyper- or hypothyroidism compared with euthyroid control individuals. Data on mortality from all causes according to the length of follow-up after the diagnosis of thyroid dysfunction were considered supportive.

The following data were abstracted: the first author’s name; the publication year; the country of origin; the number, mean age, and sex of the participants; the definition of thyroid dysfunction, based on the information as provided in the primary studies; the study design details, including source population (whether the cohort was recruited from the general population or from a group of participants with a specific comorbidity), starting year of study, and study duration; whether the reported HR was adjusted for age, gender, or other potential confounders; and losses to follow-up. Data were independently extracted by three of us (P H, A V M, and B V) and checked for accuracy in a second review. Differences in assessments were resolved by consensus of the three reviewers.

Statistical analyses

Two meta-analyses were conducted: one compared subclinical hyperthyroid participants versus euthyroid controls; and the other compared subclinical hypothyroid participants versus euthyroid controls. If sufficient information was available, we also explored the risk of death according to the length of follow-up after the diagnosis of subclinical thyroid dysfunction.

A reliable approach to perform a meta-analysis of survival data is to summarize each contributing study by a single number, along with its standard error, and then combine the summary statistic from each study using standard methods of meta-analysis. The (log) HR has been specifically designed for comparing two survival curves; it is the only summary statistic which allows for both censoring and time to the occurrence of an event (13–15).
From the eligible studies, the (log) HR and its standard error were retrieved as reported in each study, or were calculated from the published data. Whenever reported in the original paper, the adjusted (rather than unadjusted) HR was included in our analyses. If not reported in the original paper, a HR was calculated as described by Sutton et al., Parmar et al. and Tierney et al. (13–15). The log HR and its standard error were calculated directly, if the observed and the expected number of events were available for the group with thyroid dysfunction and the control group; indirectly, if the P values for the logrank, Mantel–Haenszel, or χ2 test were reported; or graphically, based on the published survival curves, if insufficient information was available for direct or indirect estimation. To track potential changes in the risk of death over time, we partitioned the time axis of the published survival curves into 1-year intervals and computed a (log) HR and its standard error for each 1-year interval, whenever possible (13–15, 24). The rationale for also focusing the analysis to studies with graphic displays is particularly relevant when assessing the effects of thyroid dysfunction on mortality according to the time of diagnosis (inclusion), because the risk of dying might only be higher during (or after) a specific time after diagnosis, and our interest is in establishing uniform time periods for use in making comparisons in trends across participant cohorts.

Pooled estimates of the mean effect of subclinical hyper- and hypothyroidism on mortality (pooled HR) and the corresponding 95% CIs were determined using the inverse variance fixed-effect model and the DerSimonian and Laird random-effects model (14, 25). Because the resultant point estimates were essentially similar to random- and fixed-effects models, we present only the random-effects analyses that incorporate both between- and within-study variation (providing the more conservative estimates).

The results were examined for heterogeneity by visually examining forest plots and using formal statistical tests for heterogeneity and trial inconsistency (14, 25). Between-study heterogeneity was assessed using the Cochrane Q test. P<0.10 indicating significance (14, 25), and formally quantified by the I2 statistic, with values less than 25% indicating low, 25–50% indicating moderate, and greater than 50% indicating high heterogeneity (26). To explain anticipated heterogeneity among study findings potential sources of heterogeneity were identified \textit{a priori}. Because all-cause mortality might vary according to participants’ and study characteristics, we plotted the effect size (HR) of each study against, in turn, mean age at entry (years), sex (percentage females), cohort size, publication year, starting year of study (calendar year), and the total duration of the study (years). We also performed formal random-effects meta-regression analyses. We further postulated that the findings of the studies would be affected by the following subgroup characteristics: geographic region defined according to the categories for the global burden of disease 2000 World Health Organization member states project (27), community based versus cohort recruited from a group of participants with a specific comorbidity, single baseline assessment versus repeated testing of thyroid function, and whether the HR was adjusted for age, gender, or other potential confounders.

To evaluate the effect of each selected study on the overall results of the meta-analysis, we performed a one-way sensitivity analysis, also defined \textit{a priori}.

Potential publication bias was explored visually by the funnel plot method of Sterne and Egger (28), and by Egger’s regression intercept test (14, 25). The potential implications for our results were assessed by Duval & Tweedie’s trim-and-fill method (29, 30).

In the final step of the analysis, statistically significant HRs were translated to estimates of the absolute survival difference(s) between thyroid dysfunction and control groups, year-by-year, up to 10 years after the diagnosis of thyroid dysfunction, using the formula:

\[
\text{absolute survival difference} = \exp(\ln P_c \times \text{HR}) - P_c
\]

where HR is the HR as estimated by the current meta-analyses and P_c is the probability of survival in the control group, year-by-year, up to 10 years after the diagnosis of thyroid dysfunction (16, 18, 31). The year-by-year probability of survival in the control group, P_c, was calculated using standard life-table methods, applied to population-based data on age- and sex-specific mortality (31, 32). We used this method to estimate the year-by-year survival probability P_c, up to 10 years after the diagnosis of thyroid dysfunction, for imaginary cohorts of women for initial age at the diagnosis of thyroid dysfunction ranging from 20 to 80 years, by decade (31, 32). Similar models were applied to hypothetical cohorts of men 20–80 years of age, again performing analyses by decade (16, 18). The upper and lower 95% CIs of the pooled HR for all-cause mortality were used to compute the corresponding upper and lower 95% CIs of absolute mortality for each age group at several time intervals. Point estimates are presented with corresponding 95% CIs for each age group at several time intervals, providing information on magnitude and precision (uncertainty) of absolute excess mortality, as well as the role that chance may play in the results.

The option to model excess mortality only for statistically significant estimates of pooled HRs was decided \textit{a priori}.

\section*{Results}

\subsection*{Study characteristics}

Our initial search identified 759 unique publications that were narrowed by preliminary review of the abstracts to 30 potentially relevant original papers (Fig. 1) (33). Out of these 30 potentially relevant papers, 12 were excluded...
because they related to overt thyroid dysfunction and/or treatment of thyroid dysfunction at inclusion (34–45). 2 papers included thyroid dysfunction, directly related to the treatment of the underlying comorbid condition (46, 47), 6 papers did not provide sufficient information to calculate a HR for all-cause mortality (48–53), and 1 paper reported data on the same participants as another included paper (54). Thus, nine papers, all published in English, were included in the present review (55–63). Each paper properly addressed almost all items of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for reports of observational studies (64). Specific pieces of information pertaining to differential study characteristics and thyroid dysfunction are listed in Table 1. The nine papers included seven cohorts that provided survival data for participants with subclinical hyperthyroidism and nine cohorts that provided survival data for participants with subclinical hypothyroidism. Considered together, the cohorts included 290 participants with subclinical hyperthyroidism and 1580 participants with subclinical hypothyroidism, and 13 039 euthyroid control subjects. Survival was documented during an observation period that ranged from 2 to 20 years after the diagnosis of thyroid dysfunction.

Quantitative data synthesis: subclinical hyperthyroidism

Seven cohorts compared mortality from all causes among participants with subclinical hyperthyroidism (Fig. 2) (55, 56, 58, 60–63). According to the data in the individual cohorts, the HR for all-cause mortality ranged from 0.84 to 2.22. In six cohorts, the HR was greater than unity, with 95% CIs excluding unity in only one of these cohorts (55). In this cohort, 1 out of 71 participants with suppressed TSH had a high fT₄ level, and 3 participants developed overt hyperthyroidism 2, 3, and 4 years after inclusion, none of whom died (55). Using a DerSimonian and Laird random-effects model, the pooled summary HR for all-cause mortality was 1.41 (95% CI, 1.12–1.79; P=0.004).

No important heterogeneity was demonstrated between the studies (Q test P=0.39; I²=4.6%). Accordingly, none of the predefined categorical and meta-regression sensitivity analyses was statistically significant (data not shown). Interestingly, the pooled HRs for studies using a single baseline assessment or repeated testing of thyroid function were 1.10 (95% CI, 0.76–1.58; three studies; I²=0%) (56, 57, 62) and 1.65 (95% CI 1.24–2.20; four studies; I²=0%) (55, 58–61, 63) respectively, with a P value of 0.09 for between group variation.

One-way sensitivity analyses demonstrated that the overall effect size and its statistical significance were consistent across the studies and did not depend on any single study (data not shown).

An exploratory analysis of the five cohorts, providing sufficient information to quantify changes in the risk of death over time, revealed no differences in mortality during the first year after the diagnosis of subclinical hyperthyroidism (55, 56, 58, 61, 63). Beyond the first year, the 95% CIs of the HRs of all cohort studies excluded unity (P values 0.017 to <0.001). When compared with the main analysis the magnitude of the HRs was slightly higher, with pooled HRs of 1.75–2.37.

The visual inspection of a funnel plot of effect size versus precision suggested that publication bias may have occurred due to the absence of or inability to find at least one small positive study on the right of the summary estimate (Fig. 3). The Egger’s regression intercept test provided no evidence of publication bias (intercept, −0.351; 90% CI, −2.037 to 1.335; P=0.71). However, a non-significant Egger’s test may be due to low statistical power and cannot be taken as evidence that bias is absent. Duval & Tweedie developed a method that allows us to impute potentially missing studies. The method is known as ‘trim and fill’, as the method initially trims the asymmetric studies from the
Table 1  Characteristics of the cohort studies included in the meta-analyses.

<table>
<thead>
<tr>
<th>First author (year published)</th>
<th>Country</th>
<th>Source populationa</th>
<th>Starting year of studyb</th>
<th>Duration of study (years)</th>
<th>Definition of subclinical hyperthyroidism</th>
<th>Definition of subclinical hypothyroidism</th>
<th>Assessment of thyroid functionc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parle (2001) (55)</td>
<td>England and Wales</td>
<td>Community dwellers</td>
<td>1988</td>
<td>11</td>
<td>TSH &lt; 0.5 mU/l</td>
<td>TSH &lt; 0.5 mU/l</td>
<td>Repeated</td>
</tr>
<tr>
<td>Radacsi (2003) (56)</td>
<td>Germany</td>
<td>Patients recovering from stroke or hip fracture surgery</td>
<td>...</td>
<td>2</td>
<td>TSH &lt; 0.1 mU/l</td>
<td>TSH &lt; 13–27 pmol/l</td>
<td>Repeated</td>
</tr>
<tr>
<td>Imaizumi (2004) (57)</td>
<td>Japan</td>
<td>Atomic bomb survivors</td>
<td>1984</td>
<td>12</td>
<td>TSH &gt; 0.5 mU/l</td>
<td>TSH &gt; 0.1 mU/l</td>
<td>Single baseline</td>
</tr>
<tr>
<td>Rodondi (2005) (59)</td>
<td>USA</td>
<td>Community dwellers</td>
<td>1997</td>
<td>4</td>
<td>TSH &lt; 0.4 mU/l</td>
<td>TSH &gt; 13–23 pmol/l</td>
<td>Single baseline</td>
</tr>
<tr>
<td>Walsh (2005) (60)</td>
<td>Australia</td>
<td>Community dwellers</td>
<td>1984</td>
<td>13</td>
<td>TSH &lt; 0.4 mU/l</td>
<td>TSH &gt; 10.3–23.2 pmol/l</td>
<td>Single baseline</td>
</tr>
<tr>
<td>Cappola (2006) (61)</td>
<td>USA</td>
<td>Community dwellers</td>
<td>1989</td>
<td>2</td>
<td>TSH &lt; 0.44 mU/l</td>
<td>TSH &gt; 10.3–23.2 pmol/l</td>
<td>Single baseline</td>
</tr>
<tr>
<td>Chubb (2006) (62)</td>
<td>Australia</td>
<td>Diabetes type 2 patients</td>
<td>1993</td>
<td>9.3</td>
<td>TSH &lt; 0.34 mU/l</td>
<td>TSH &gt; 10.3–23.2 pmol/l</td>
<td>Single baseline</td>
</tr>
<tr>
<td>Iervasi (2007) (63)</td>
<td>Italy</td>
<td>Cardiac patients</td>
<td>2007</td>
<td>2.7</td>
<td>TSH &lt; 0.3 mU/l</td>
<td>TSH &gt; 9–23.8 pmol/l</td>
<td>Repeated</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First author (year published)</th>
<th>Number of patients with subclinical hyperthyroidism</th>
<th>Number of patients with subclinical hypothyroidism</th>
<th>Number of euthyroid controls</th>
<th>Prevalence of subclinical hyperthyroidism</th>
<th>Prevalence of subclinical hypothyroidism</th>
<th>Age at entry (years)</th>
<th>Proportion women (%)</th>
<th>Adjustment of HRd</th>
<th>Loss to follow-up (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parle (2001) (55)</td>
<td>71</td>
<td>94</td>
<td>1026</td>
<td>6.5%</td>
<td>8.4%</td>
<td>69</td>
<td>84%</td>
<td>No</td>
<td>0%</td>
</tr>
<tr>
<td>Radacsi (2003) (56)</td>
<td>11</td>
<td>5</td>
<td>64</td>
<td>14.7%</td>
<td>7.2%</td>
<td>77</td>
<td>65%</td>
<td>No</td>
<td>0%</td>
</tr>
<tr>
<td>Imaizumi (2004) (57)</td>
<td>...</td>
<td>257</td>
<td>2293</td>
<td>...</td>
<td>10.1%</td>
<td>61</td>
<td>61%</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Gussekloo (2004) (58)</td>
<td>17</td>
<td>30</td>
<td>472</td>
<td>3.5%</td>
<td>6.0%</td>
<td>85</td>
<td>66%</td>
<td>No</td>
<td>13%</td>
</tr>
<tr>
<td>Rodondi (2005) (59)</td>
<td>...</td>
<td>338</td>
<td>2392</td>
<td>...</td>
<td>12.4%</td>
<td>75</td>
<td>51%</td>
<td>Yes</td>
<td>...</td>
</tr>
<tr>
<td>Walsh (2005) (60)</td>
<td>39</td>
<td>119</td>
<td>1906</td>
<td>2.0%</td>
<td>5.9%</td>
<td>50</td>
<td>49%</td>
<td>Yes</td>
<td>5%</td>
</tr>
<tr>
<td>Cappola (2006) (61)</td>
<td>47</td>
<td>496</td>
<td>2639</td>
<td>1.7%</td>
<td>15.8%</td>
<td>73</td>
<td>60%</td>
<td>Yes</td>
<td>0%</td>
</tr>
<tr>
<td>Chubb (2006) (62)</td>
<td>7</td>
<td>33</td>
<td>342</td>
<td>1.8%</td>
<td>8.6%</td>
<td>64</td>
<td>100%</td>
<td>No</td>
<td>0%</td>
</tr>
<tr>
<td>Iervasi (2007) (63)</td>
<td>98</td>
<td>208</td>
<td>1905</td>
<td>4.9%</td>
<td>9.8%</td>
<td>61</td>
<td>33%</td>
<td>Yes</td>
<td>...</td>
</tr>
</tbody>
</table>

aCommunity setting versus medical patients.

bEllipses, no data.

cSingle baseline assessment versus repeated testing of thyroid function.

dHR, hazard ratio; no, calculated from the published data; yes, retrieved as reported in each study; the models were adjusted for age, sex, race, smoking status, diabetes mellitus, prevalent cardiovascular disease, poor or fair health, blood pressure, total cholesterol level, creatinine level, education, income, and use of thyroid hormone and angiotensin-converting enzyme inhibitors (59); age and gender (60); age, sex, clinical cardiovascular disease at baseline, atrial fibrillation at baseline, thyroid medication use during follow-up, race, smoking status, diabetes, low-density lipoprotein cholesterol, use of lipid-lowering medications, hypertension, body mass index, and C-reactive protein (61); and age, sex, ischemic and nonischemic heart disease, and levels of thyrotropin, free thyroxine, and free triiodothyronine (63).
right-hand side to locate the unbiased effect (in an iterative procedure), and then fills the plot by re-inserting the trimmed studies on the right as well as their imputed counterparts to the left the mean effect. When looking for missing studies based on both a fixed and random-effects models, the trim-and-fill method suggested that no studies were missing (Fig. 3). Moreover, the recalculated pooled risk estimate, imputed by the trim-and-fill method, was identical to our original HR, suggesting that publication bias is unlikely to cause a material change in our findings.

The 2-, 5-, and 10-year excess mortality data after the diagnosis of subclinical hyperthyroidism in the various age and sex groups are shown in Table 2. The same estimates are presented graphically in Fig. 4. At the age of 70, the differences in absolute risk of death are 1.5, 4.0, and 8.7% and 2.2, 5.7, and 10.7% in women and men respectively. Beyond the age of 80, these excess risks start to decline because of the competing probabilities of death attributable to thyroid dysfunction and baseline probability of death in the general population.

Overall, the excess mortality after the diagnosis of subclinical hyperthyroidism depends on age, with very low excess mortality until the age of 50, both among white US women and men with subclinical hyperthyroidism. Beyond the age of 60, excess mortality increases, especially in aging men. At any given age, excess mortality after subclinical hyperthyroidism is always higher among men than among women. Similar findings were observed when applying our models to a UK and a Belgian reference population (unpublished observations).

### Quantitative data synthesis: subclinical hypothyroidism

Nine cohorts involving 1580 patients were included in the primary analysis of mortality from all causes among participants with subclinical hypothyroidism.

#### Figure 2 Random-effects meta-analysis for the primary outcome of interest (hazard ratio for mortality from all causes) for patients with subclinical hyperthyroidism

<table>
<thead>
<tr>
<th>Study</th>
<th>Hazard ratio (95% confidence interval)</th>
<th>Weight (%)</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parle, 2001</td>
<td>1.86 (1.33 to 2.60)</td>
<td>41.1</td>
<td>1.86 (1.33 to 2.60)</td>
</tr>
<tr>
<td>Radaci, 2003</td>
<td>2.22 (0.61 to 8.16)</td>
<td>3.2</td>
<td>2.22 (0.61 to 8.16)</td>
</tr>
<tr>
<td>Gussekloo, 2004</td>
<td>1.12 (0.46 to 2.68)</td>
<td>7.0</td>
<td>1.12 (0.46 to 2.68)</td>
</tr>
<tr>
<td>Walsh, 2005</td>
<td>0.84 (0.28 to 2.50)</td>
<td>4.6</td>
<td>0.84 (0.28 to 2.50)</td>
</tr>
<tr>
<td>Cappola, 2006</td>
<td>1.08 (0.72 to 1.62)</td>
<td>29.8</td>
<td>1.08 (0.72 to 1.62)</td>
</tr>
<tr>
<td>Chubb, 2006</td>
<td>2.01 (0.50 to 8.10)</td>
<td>2.8</td>
<td>2.01 (0.50 to 8.10)</td>
</tr>
<tr>
<td>Iervasi, 2007</td>
<td>1.22 (0.68 to 2.40)</td>
<td>11.5</td>
<td>1.22 (0.68 to 2.40)</td>
</tr>
<tr>
<td>Pooled (95 % CI)</td>
<td>1.41 (1.12 to 1.79)</td>
<td>100.0</td>
<td>1.41 (1.12 to 1.79)</td>
</tr>
</tbody>
</table>

Tests for heterogeneity $\chi^2 = 6.29$, df = 6, $P = 0.39$, $I^2 = 4.6$
Tests for overall effect $Z = -2.89$, $P = 0.004$

#### Figure 3 Publication bias and its potential impact for subclinical hyperthyroidism

The blue circles represent the observed individual studies, the curved lines represent the funnel plot, and the blue diamond is the hazard ratio (HR) and 95% confidence interval for the meta-analysis. The red diamond is the HR and 95% confidence interval for the meta-analysis, after adjusting for publication bias according to the trim-and-fill methodology. When looking for missing studies based on both a fixed- and random-effects models, the trim-and-fill method suggests that no studies are missing: there are no red circles representing imputed (missing) studies, the red curved line representing an adjusted funnel plot are identical to the curves of the funnel plot of the observed studies, and the imputed hazard ratio (HR) and 95% confidence interval (red diamond) are identical to the calculated hazard ratio (HR) and 95% confidence interval (blue diamond).
Table 2 Excess mortality from all causes in US patients with subclinical hyperthyroidism, by gender and age at the time of diagnosis.

<table>
<thead>
<tr>
<th>Age at diagnosis (years)</th>
<th>White US women with subclinical hyperthyroidism</th>
<th>White US men with subclinical hyperthyroidism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage excess mortality (95% confidence interval) after diagnosis</td>
<td>Percentage excess mortality (95% confidence interval) after diagnosis</td>
</tr>
<tr>
<td></td>
<td>2-years</td>
<td>5-years</td>
</tr>
<tr>
<td>20</td>
<td>0.04 (0.01–0.07)</td>
<td>0.10 (0.03–0.18)</td>
</tr>
<tr>
<td>30</td>
<td>0.05 (0.01–0.09)</td>
<td>0.13 (0.04–0.25)</td>
</tr>
<tr>
<td>40</td>
<td>0.06 (0.03–0.22)</td>
<td>0.33 (0.09–0.62)</td>
</tr>
<tr>
<td>50</td>
<td>0.25 (0.07–0.47)</td>
<td>0.70 (0.20–1.33)</td>
</tr>
<tr>
<td>60</td>
<td>0.62 (0.18–1.16)</td>
<td>1.75 (0.50–3.30)</td>
</tr>
<tr>
<td>70</td>
<td>1.50 (0.43–2.84)</td>
<td>4.00 (1.16–3.97)</td>
</tr>
<tr>
<td>80</td>
<td>3.77 (1.09–7.05)</td>
<td>8.77 (2.61–15.08)</td>
</tr>
</tbody>
</table>

These data are based on a hypothetical approach: all point estimates and corresponding 95% confidence intervals are generated using mathematical modeling, i.e., these data are not based on individual patient data, but derive from aggregated data reported in original cohort studies and life-tables. Point estimates are presented with corresponding 95% confidence intervals for each age group at several time intervals, providing information to the reader on magnitude and the precision (uncertainty) of absolute excess mortality, as well as the role that chance may play in the results. For further details on the assumptions made, see materials and methods section.

(Fig. 5) (55–63). According to the data in the individual cohorts, the HR for all-cause mortality ranged from 0.49 to 2.01. The HR was statistically significant in three cohort studies only. The point estimates of these three cohort studies were located on the opposite sides of the ‘no difference’ line, with one cohort study (58) reporting a significantly decreased HR of 0.49 (95% CI, 0.24–0.99; P=0.048) and two cohorts (57, 63) reporting a significantly increased HR of 1.77 (95% CI, 1.18–2.65; P=0.006) and 2.01 (95% CI, 1.33–3.04; P<0.001) respectively. Overall, the pooled summary estimate did not show an increased risk of death among participants with subclinical hypothyroidism. The pooled summary HR for all-cause mortality was 1.22 (95% CI, 0.95–1.57; P=0.12) using a DerSimonian and Laird random-effects model.

Significant heterogeneity was demonstrated between the studies (Q test P=0.006; I²=63%). A predefined categorical meta-analysis indicated that the source population explained a significant proportion of the variation in risk estimates (R²=44%, P=0.014 for between study variation). The pooled HR for all-cause mortality was 1.03 in cohorts recruited from the community (55, 58–61) (95% CI, 0.78–1.35; P=0.83) and 1.76 in cohorts recruited from a group of participants with a specific comorbidity (95% CI, 1.36–2.30; P<0.001), i.e., atomic bomb survivors (57), diabetes type 2 (62), cardiac (63), stroke (56), or hip fracture patients (56). The number of patients for the analysis of cohorts recruited from the community was 1258 vs 295 for those with a specific comorbidity. No other a priori variable could explain the heterogeneity between studies (all P values >0.10). In particular, there were no differences in the summary estimates between studies using a single baseline assessment or repeated testing of thyroid function, with pooled HRs of 1.35 (95% CI, 0.98–1.87; five studies) (57, 59–62) and 1.02 (95% CI 0.66–1.57; four studies) (55, 56, 58, 63) respectively and a P value of 0.31 for between-group variation.

The two cohorts recruited from a group of participants with a specific comorbidity that provided sufficient information to quantify changes in the risk of death over time (57, 63), revealed no differences in mortality during the first year after the diagnosis of subclinical hypothyroidism. Beyond the second year, the 95% CIs of the HRs all excluded unity (P values 0.022 to <0.0001).

One-way sensitivity analyses demonstrated that the results for the overall pooled summary estimate did not depend on any single study (unpublished observations).
A funnel plot of effect size versus precision was not perfectly symmetrical, but when we formally tested for publication bias using the Egger’s regression intercept test and the trim-and-fill method, we found none (figure and data available from the authors).

We did not calculate estimates of absolute excess mortality, given the non-significant overall pooled HR for all-cause mortality in patients with subclinical hypothyroidism.

Discussion

In the two meta-analyses reported here, participants with subclinical hyperthyroidism demonstrated a significant 1.41-fold increase in relative likelihood of death from all causes versus euthyroid control subjects. The absolute risk of death and the corresponding excess all-cause mortality attributable to subclinical hyperthyroidism were dependent on age and, to a lesser extent, on male gender. At any given age, excess mortality was always higher in men. By contrast, participants with subclinical hypothyroidism, the overall finding of no difference in mortality reflects the results of six out of the nine relevant cohorts, with mortality from all causes being increased only in participants with a specific comorbidity.

Our systematic reviews and meta-analyses have a number of novel aspects compared with previous meta-analyses on mortality in patients with subclinical thyroid dysfunction (10, 12). Because we summarized the survival data of each relevant study by a single number that takes into account data censoring (HR), we were able to produce pooled survival estimates less prone to bias, and to formally explore potential sources of heterogeneity. Another key feature is that we translated relative mortality risks into absolute risks of death, allowing quantification of excess mortality attributable to subclinical hyperthyroidism. Our approach differs from the meta-analyses previously published by Singh et al. (10) and Völzke et al. (12). The primary aim of the Singh et al. meta-analysis (10) was to evaluate the impact of subclinical hypothyroidism on the risk of coronary artery disease and mortality from cardiovascular causes. Völzke et al. analyzed the impact of (clinical and subclinical) thyroid dysfunction on all-cause mortality, included only prospective cohort studies within the general population and excluded selected populations such as patients with comorbid conditions (12). When compared with the two previously published meta-analyses (10, 12), we adopted less strict inclusion criteria. We reasoned that the inclusion of a larger panel of a priori defined observational cohort studies might contribute to highlight the causes of heterogeneity. This might generate new hypotheses and finally help define those populations that might benefit from randomized interventional studies.

We provide data indicating that after the diagnosis of subclinical hyperthyroidism mortality from all causes is significantly increased, with a pooled HR of 1.41 with a 95% CI of 1.12–1.79. Our results differ from the conclusions on the impact of subclinical hyperthyroidism on all-cause mortality in two previously published meta-analyses (10, 12). Unlike these meta-analyses (10, 12), we
decided to include the Parle et al. study (55), mainly because none of the patients described in the Parle et al. study were treated with l-thyroxine at recruitment. We do acknowledge that in the Parle et al. study 1 out of 71 participants with suppressed TSH had a high fT4 level and that 3 participants developed overt hyperthyroidism. 2, 3, and 4 years after inclusion, none of whom died (55). The contribution of these cases to overall mortality in patients with suppressed TSH is therefore very unlikely. Although suppressed TSH probably indicates mild thyroid excess in the absence of thyroid hormone treatment it might also reflect other factors such as non-thyroidal illness. The normal thyroxine values measured in the individual studies make confounding by euthyroid sick syndrome less likely, but not impossible.

The formal analysis of the published survival curves available from five cohorts indicates that the increased likelihood of death is not present immediately after the diagnosis of subclinical hyperthyroidism, but becomes apparent during the second year, and then continues for up to 10 years after the diagnosis of subclinical hyperthyroidism. These findings are underpinned by a strong biological rationale (4, 6–8). Subclinical hyperthyroidism is increasingly recognized as having significant consequences on the cardiovascular, skeletal, and central nervous systems (2). When compared with age-matched control groups of euthyroid control persons, patients with subclinical hyperthyroidism have an up to threefold increased risk for atrial fibrillation, a higher prevalence of arterial hypertension, a significantly higher heart rate, a significant increase in left ventricular mass, a significantly decreased bone mineral density, a threefold increased risk for hip fracture, a fourfold increased risk for vertebral fracture, and a more than threefold increased risk of dementia and Alzheimer’s disease (4, 6–8, 65–67). All these cardiac, bone, and nervous abnormalities might play a role in determining the increased mortality in persons with subclinical hyperthyroidism, given the evidence from large population-based cohort studies that each aforementioned condition by itself has a strong association with increased risk of death (68–73). With regard to specific causes of death resulting in increased mortality among persons with subclinical hyperthyroidism, only two out of the seven studies included in our meta-analysis found that the increase in all-cause mortality was related to a higher cardiovascular mortality (55, 63). A likely explanation for increased mortality becoming apparent only after the first year of diagnosis is that the severity of the adverse effects resulting from subclinical thyroid dysfunction, left untreated, depends on the duration of the dysfunction.

For participants with subclinical hypothyroidism, on the other hand, the overall pooled relative risk of mortality from all causes was not increased. These findings might be confounded because not all studies excluded and/or corrected for treatment with thyroid hormones during follow-up of the participants. In addition, excess mortality related to subclinical hypothyroidism might become apparent only after a very long duration of thyroid dysfunction or during follow-up periods much longer than those available from the studies published at the time being. Significant heterogeneity was observed, and a categorical meta-analysis indicated that all-cause mortality was higher in participants with a specific comorbidity, i.e., atomic bomb survivors (57), diabetes type 2 (62), cardiac (63), stroke (56), or hip fracture patients (56). The adverse effects of subclinical hypothyroidism on mortality in participants with comorbid conditions remain largely speculative. Subclinical hypothyroidism has been linked to increased cardiovascular risk factors (dyslipidemia, hypercoagulation, and decreased fibrinolysis), and an alteration in endothelium-mediated vasodilatation with increased cardiovascular vulnerability (2–10). Moreover, thyroid hormones regulate cardiac function through transcriptional actions and non-genomic effects involving substrate metabolism. Mild hypothyroidism might therefore contribute to a higher mortality in patients with stroke, heart disease, and diabetes mellitus. Finally, muscle metabolism and exercise tolerance are altered in subclinical hypothyroidism (74–76). In patients with poor physical fitness, subclinical hypothyroidism might account for an additional neuromuscular deficit compromising recovery after stroke or hip fracture.

A recent review has indicated that subclinical hypothyroidism is associated with an increased risk of coronary heart disease (9). The summary odds ratio for coronary heart disease was 1.81 (95% CI, 1.38–2.39) in studies that adjusted or matched for demographic characteristics, and 2.38 (95% CI, 1.53–3.69) in studies that adjusted for most cardiovascular risk factors. This odds ratio was not significant when pooling only cohort studies (9). Individual studies also suggest that a significant difference in the prevalence of coronary heart disease exists according to the severity of subclinical hypothyroidism (TSH > 10 mU/l), whereas the increase is of borderline significance with mild subclinical hypothyroidism (60). Although it would seem important to assess the extent of TSH elevation as it may relate to mortality, evidence toward a higher all-cause mortality rate with increasing levels of TSH is provided by only one study (59), precluding any formal meta-analysis in this regard. Nevertheless, our meta-analysis showed that there is an increase in hazard for mortality associated with subclinical hypothyroidism in patients with comorbid conditions. Whether this extra mortality associated with subclinical hypothyroidism may serve as a base to characterize the risk of individual patients and as a guide to treatment decision needs to be confirmed by prospective data from randomized controlled trials. At the time being, it remains to be proven that treatment is indeed beneficial in the prevention of excess mortality.
The findings presented in this analysis may have clinical implications. They highlight the predictive risk of subclinical hyperthyroidism for all-cause mortality according to age at the time of diagnosis, and also the presence of gender-related differences in the association between subclinical hyperthyroidism and all-cause mortality. Our results favor targeting interventions at both men and women aged 60 years and older to potentially reduce the burden of mortality associated with subclinical hyperthyroidism in old age. If deaths are causally related to subclinical hyperthyroidism, then it might be expected that interventions that normalize TSH levels would increase survival (23). Experts have suggested that only patients with TSH < 0.1 mU/l require intervention, however, there is minimal and only indirect data that support this recommendation (23). Our findings do not allow further conclusions on the impact of low versus undetectable TSH on all-cause mortality. Even though two studies indicate that treatment improves bone mineral density, there are no empirical data to indicate that there is a survival advantage associated with the treatment of subclinical hyperthyroidism. For the youngest age groups, this question is unlikely to be resolved from empirical data, even if obtained from well-designed randomized placebo-controlled clinical trials. On the basis of the information presented in this paper, consider, for example, that women diagnosed with subclinical hyperthyroidism at the age of 30 years are enrolled into a randomized placebo-controlled trial for 5 years, and that the intervention reduces excess mortality to zero, that is the probability of dying for the actively treated group would be reduced to the baseline probability of death for healthy euthyroid women. In order to achieve a power of 80% at a significance level of 0.05 with a two-tailed test, the sample size in each group would need to be 12,350 at this age. For the older age groups, however, the sample size in each group would need to be 24,800, 9,722, and 4,124 at the age of 50, 60, and 70 respectively.

Because our analyses support no adverse effect of subclinical hypothyroidism upon mortality in community-dwelling patients, treating these individuals may have limited benefit in this regard. This is, however, a very controversial subject that has not been settled. According to a 20-year follow-up survey on the natural history of subclinical hypothyroidism conducted in the UK, the annual risk for developing overt hypothyroidism after 20 years is 4.3% in women with increased TSH concentrations and thyroid antibodies and 2.6% in women with subclinical hypothyroidism without thyroid antibodies (77). In a prospective cohort study, analyzing the natural course and risk factors for the development of overt thyroid failure among Spanish patients with subclinical hypothyroidism and no previous history of thyroid disease, the mean time free of L-T4 therapy was 56 months (78). In this context, a watchful waiting approach with periodic evaluations of TSH levels, and the initiation of treatment only when related clinical manifestations become apparent, might be more appropriate in these individuals with no obvious comorbid conditions. Our categorical meta-analysis, on the other hand, provides further indirect evidence that treatment should be considered in patients with subclinical hypothyroidism and associated comorbidity.

There are several limitations that may affect the inferences derived from our calculations. Published survival curves and HRs are composite estimates that reflect the typical experience of patients during the years of observation. In most studies, the interval between the diagnosis of subclinical thyroid dysfunction and subsequent mortality ranged from a few years to a few decades. Moreover, thyroid dysfunction is not necessarily stable over time (79), and we have no information available on the duration of subclinical thyroid dysfunction prior to its diagnosis. To address these concerns, we performed predefined analyses exploring the potential impact of study duration, or single baseline assessment versus repeated testing of thyroid function. This commonly used approach may not detect and cannot solve inherent problems affecting the primary studies, for example, if specific information on the duration of subclinical thyroid dysfunction prior to its diagnosis is not gathered. Our results are subject to ecologic fallacy, an inherent limitation of any meta-analysis not based on individual patient data but derived from aggregated information on, for example, mean age and gender proportions in the original cohorts. For the reference US population (or any other reference population from a different country), available information on age- and gender-specific mortality rates is based on current cross-sectional data on patients of different ages. There is no guarantee that these rates will remain constant over the time span for which absolute risks and excess mortality are calculated. Life-table method-based analyses, as used in our study, assume a constant relative risk of death, which may be unrealistic. Although mortality trends are stable in many countries, it is not possible to determine how future relative risks of death would change over time with the data to hand. Finally, the generalizability (external validity) of our absolute risk and excess mortality estimates to other populations and its use in clinical risk prediction are limited by the fact that the data were restricted to a white US population. Similar findings were observed when applying our models to reference populations from the European Union. As a general rule, appropriate CIs are mentioned to show uncertainty around the estimates of any outcome measure and to allow cautious interpretation of the data.

The strengths of our study, on the other hand, include the use of standard meta-analytic procedures for retrieval, assessment of relevance, and statistical processing of the data. All estimates were based on data published in peer-reviewed journals and official government reports, providing access to survival data in well-defined cohorts of participants with subclinical thyroid dysfunction, and robust information on US age- and sex-specific all-cause mortality data. All cohort studies were published within
the past decade, minimizing any effect of secular trends and changes in medical practice. Meta-analysis of survival data requires specific techniques because of data censoring. If data censoring is ignored, this may bias the overall estimates. Potentially useful information about timing of events (deaths) and the shape of the published survival curves was not discarded in the current systematic review and meta-analysis, as we retrieved or computed a HR and its standard error for each contributing cohort study, and then combined them by using validated approaches for meta-analysis. In doing so, we always considered the group with euthyroid function as the reference group, an approach sometimes resulting in differences between our calculated estimates and HRs reported in the original papers (55). Another key feature of our systematic review is that we explored potential sources of heterogeneity and quantified the potential impact of publication bias on our findings. When we formally tested for publication bias, we found that our results or interpretations were not affected in a significant way. These tests, however, have limited power and are difficult to interpret when used for a small number of studies. Although limited by a quite small number of studies and a relatively small number of adverse events, our analyses provide evidence for statistically significant and clinically relevant differences in the magnitude of mortality between different patient groups with subclinical thyroid dysfunction. The different cohorts varied in size, but no single study was so large as to dominate the overall results. From a public health perspective, absolute risk calculations and estimates of excess mortality such as those presented here not only allow estimation of the consequences of various diseases and complications, but also a more appropriate allocation of resources for competing causes of mortality.

In summary, meta-analyses based on data from prospective cohort studies indicate that the relative risk of mortality from all causes is 41% higher for participants with subclinical hyperthyroidism than for euthyroid controls. Mathematical models suggest that among participants with subclinical hyperthyroidism the magnitude of excess mortality might depend on age and, to a lesser extent, on male gender. For patients with subclinical hypothyroidism, on the other hand, meta-analyses indicate that the relative risk of mortality from all causes is increased only among patients with associated comorbidity. This information should be taken into account when designing treatment strategies or other courses of action for both women and men diagnosed with subclinical thyroid dysfunction.

Acknowledgements

Author Contributions: B V had the idea for the study. P H, A V M, and B V designed the study, collected the data, had full access to all of the data, take responsibility for the integrity of the data and the accuracy of the data analysis, and are the guarantors for the paper. P H did the statistical analysis. K P provided advice on interpretation of the results. P H, B V and A V M drafted the paper. All authors revised the paper critically and approved the final manuscript. This work was supported by grants from the Willy Gepts Foundation, Academisch Ziekenhuis van de Vrije Universiteit Brussel, to B V. The authors declare no conflict of interest, ethical approval was not required for this study.

References


20 Cooper DS. Thyroid disease in the oldest old: the exception to the rule. *Journal of the American Medical Association* 2004 292 2651–2654.


23 Surks MI, Ortiz E, Daniels GH, Sawin CT, Cobin RH, Franklyn JA, Jermann M, Burman KD, Denke MA, Gorman C, Cooper RS & Weisman NJ. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *Journal of the American Medical Association* 2004 291 228–238.


www.eje-online.org


76 Harper ME & Seifert EL. Thyroid hormone effects on mitochondrial energetics. Thyroid 2008 18 145–156.

