Title: Reference intervals in the diagnosis of thyroid dysfunction - time to treat patients and not numbers

Abstract: Although the assignment of a diagnosis of thyroid dysfunction appears quite simple, this is often not the case. Issues that lead to complexities regarding whether thyroid function is, in fact, normal include transient changes in thyroid parameters, inter- and intra-individual differences in thyroid parameters, age-related differences, and ethnic variations. Superimposed upon these considerations, is the understanding that a statistically calculated distribution of thyroid analytes does not necessarily coincide with intervals or cut offs that have predictive value for beneficial or adverse health outcomes. Based on current trial data, it is unclear whether certain adults would still benefit from levothyroxine, (such as those with TSH values > 10 mIU/L), since a limited number of these persons were included in randomised trials. Even if therapy is initiated for abnormal thyroid function, not all treated individuals are maintained at the desired treatment target, and therefore still may be subject to risks. The consequence of this is that not only does each patient's thyroid function need to be assessed on an individual basis with the entire clinical picture in mind, but also monitoring needs to be vigilant, and the targets for treatment re-assessed on an ongoing basis.
Reference intervals in the diagnosis of thyroid dysfunction
– time to treat patients and not numbers.

Jacqueline Jonklaas¹ and Salman Razvi²

¹Division of Endocrinology Georgetown University, Washington, DC, USA and ²Department of Endocrinology, University of Newcastle, Newcastle, UK

Key words: Hypothyroidism, hyperthyroidism, thyroid hormones, TSH, diagnosis, reference intervals

Disclosures: J.J. and S.R. have no relevant disclosures

Manuscript word count = 5727 words

References 110

Figures 7, tables 2.

Search strategy and selection criteria: References for this review were identified through searches of PubMed for articles published in English from January, 1971, to June, 2018, by use of the MeSH terms “hypothyroidism”, “hyperthyroidism”, “subclinical hypothyroidism”, “subclinical hyperthyroidism”, “thyroid disease”, “thyroid stimulating hormone”, “thyroid hormone”, “thyroxine”, “free thyroxine”, “triiodothyronine” in combination with the term “reference value”. Relevant articles were also identified through searches in the authors’ personal files. Articles resulting from these searches and relevant references cited in those articles were reviewed. Preference was given to more recent articles.
Although the assignment of a diagnosis of thyroid dysfunction appears quite simple, this is often not the case. Issues that lead to complexities regarding whether thyroid function is, in fact, normal include transient changes in thyroid parameters, inter- and intra-individual differences in thyroid parameters, age-related differences, and ethnic variations. Superimposed upon these considerations, is the understanding that a statistically calculated distribution of thyroid analytes does not necessarily coincide with intervals or cut offs that have predictive value for beneficial or adverse health outcomes. Based on current trial data, it is unclear whether certain adults would still benefit from levothyroxine, (such as those with TSH values > 10 mIU/L), since a limited number of these persons were included in randomised trials. Even if therapy is initiated for abnormal thyroid function, not all treated individuals are maintained at the desired treatment target, and therefore still may be subject to risks. The consequence of this is that not only does each patient’s thyroid function need to be assessed on an individual basis with the entire clinical picture in mind, but also monitoring needs to be vigilant, and the targets for treatment re-assessed on an ongoing basis.

Introduction

Thyroid dysfunction is one of the common conditions encountered by medical practitioners. Hypothyroidism, including the subclinical form, affects between 5 – 15% of adults, whereas hyperthyroidism, including the subclinical version, is seen in 0.5 – 2% of the population. Symptoms of thyroid dysfunction are non-specific and therefore thyroid function testing is one of the commonest tests to be requested. Clinical evaluation of patients with suspected thyroid dysfunction by assessing symptoms and examining for signs, though useful, is not very sensitive in confirming the disease. Measurement of thyroid function by quantifying circulating thyroid hormones and thyroid-stimulating hormone or thyrotropin (TSH) concentrations is deemed to be the most accurate and reliable approach to diagnose hypo and hyperfunction. The diagnosis of thyroid dysfunction has important implications for the patient as both hypo and hyperthyroidism are long-term conditions that may warrant prolonged therapy (life-long in the case of hypothyroidism and several months or years in case of hyperthyroidism).
The prevalence and incidence of the common thyroid disorder hypothyroidism is increasing. It is likely that this is, at least in part, due to a fall in the threshold level of TSH at which thyroid hormone therapy for hypothyroidism is being initiated (1). In this study approximately 30% of the individuals who were prescribed therapy were treated for subclinical hypothyroidism (SC Hypo) based on the documentation of their thyroid hormone levels (1). Therefore, more patients with borderline disease are being commenced upon treatment although the evidence of benefit is not present for quality of life, symptoms, or cognitive function (2), is unclear with respect to cardiovascular outcomes, and therefore the overall risk of harm is uncertain. Inappropriate and unnecessary treatment may lead to increased health and economic burden – both at personal and societal levels.

Understanding how reference intervals are determined and being aware of common pitfalls in the interpretation of thyroid function test results is crucial to provide safe and effective care for patients with suspected thyroid disease. This review aims to provide an insight into the variation in TSH, thyroxine (T4), and triiodothyronine (T3) in serum, impact on their respective reference intervals, the determinants of biologic variation, and its impact on interpretation of test results.

**Historical perspective of diagnosis of thyroid function**

Before the advent of thyroid function assays, the diagnosis of thyroid dysfunction was based purely on clinical grounds and, therefore, milder and less clinically apparent forms of the disease were most likely missed. Measuring clinical end-organ targets of thyroid function such as pulse rate, serum cholesterol, body temperature, ankle jerk reflex time, or basal metabolic rate were utilised instead. In the 1950s, protein bound iodine technique was used as an indirect estimate of the serum total T4 concentration. Later, from 1965 onwards, development of first generation radioimmunoassays led to the estimation of serum TSH, although these assays initially had limited functional sensitivity. Since then, TSH tests have evolved progressively and are able to detect lower TSH levels with improved sensitivity, which has led to the progressive reduction in the lower limit of the TSH level to approximately 0.4 mIU/L (3). At the same time, the upper limit of the TSH reference range has also decreased from \( \approx 10 \) to 4.5 mIU/L (4).

**Prevalence of thyroid dysfunction and prescriptions for treatment**
First and foremost, the prevalence of thyroid dysfunction is affected by the assay methods used to measure TSH and thyroid hormones (5, 6) and the statistical methods utilized to derive the reference intervals (7). With respect to statistical methods a recent study illustrated the different lower and upper limits of normal for TSH of 0.65-3.81 mIU/L and 0.27-4.87 mIU/L for the Hoffman and Tukey methods respectively (7). Efforts to harmonise both TSH and thyroid hormone assays are ongoing (8). Additionally, the prevalence of thyroid dysfunction is affected by the definitions of dysfunction employed and also the population being considered. As will be discussed, thyroid analytes falling outside a specific reference interval are often considered synonymous with the existence of disease, although there may not necessarily be a disease, or even a need for treatment. The decision as to whether disease is present may be better judged based on outcomes from epidemiologic studies (9). When defining reference intervals, narrower reference intervals will yield higher prevalence of thyroid dysfunction, and vice versa. The particular population under consideration also affects the prevalence of thyroid dysfunction. For example, populations with more older individuals and geographic regions affected by iodine deficiency harbour different burdens of thyroid dysfunction. Another factor affecting prevalence is the extent of screening for disease, with more comprehensive screening producing more abnormal thyroid analytes.

Hypothyroidism is the more common disorder, with overt disease affecting between 0.20-5.3% of the European population (10). Hyperthyroidism, on the other hand, has a prevalence of 0.75% in Europe (11). There is only one available treatment for hypothyroidism, namely thyroid hormone. Prescription of thyroid hormone in the United States has increased by 1.3% between 1999-2000 and 2011-2012, with the increase remaining significant after adjusting for the age of the population (12). Part of this increase could potentially be contributed to by the increased diagnosis of thyroid cancer, with resultant post-surgical hypothyroidism. Other factors associated with increased prescribing include less cigarette smoking and higher rates of obesity (13). The prevalence of treated hypothyroidism is highest in older age groups (14). It is possible that the trend for increased prescribing of thyroid hormone over time is partly due to increased treatment of mild degrees of hypothyroidism, including SC Hypo (1). The decreased prevalence of untreated SC Hypo in Norway between the two time periods 1995-1997 and 2006-2008 is certainly suggestive of increased treatment of mild hypothyroidism over time (15). Trends in prescribing antithyroidal medications for hyperthyroidism show more use of these agents and less use of radioiodine therapy and surgery over time, but
generally are not suggestive of altered thresholds for treatment of hyperthyroidism (16). However, algorithms used to decide whether treatment for subclinical hyperthyroidism is indicated incorporate both age and cardiovascular disease, so that rates of treatment may be greater in aging populations.

**Reference intervals and how they are determined**

Normality is relative and biochemical measurement of analytes of interest is frequently interpreted by comparison with earlier measurements taken from an apparently ‘healthy population’. The reference range for thyroid function tests, similar to many other biochemical analytes, is usually determined as the interval into which 95% of apparently euthyroid individuals fall. The distribution of both thyroid hormones (T4 and T3) is Gaussian and therefore the reference interval ranges two standard deviations below and above the mean. Serum TSH, however, has a non-Gaussian distribution, so the reference interval is obtained by including values that fall between 2.5\(^{th}\) and the 97.5\(^{th}\) ranked percentiles (Figure 1). This non-Gaussian distribution is common to other analytes such as, for example, prostate specific antigen. Reasons for the shift to the right in the curve could include inclusion of some individuals with autoimmune thyroid disease, altered biological activity of TSH especially in the elderly, and mutations in the TSH receptor. This method of defining reference intervals means that one in twenty (5%) apparently euthyroid individuals will have values outside this range. Similarly, the presence of thyroid disease is not excluded by results that fall within the reference range, without taking the complete clinical picture into account. The ‘normal’ TSH or thyroid hormone concentrations, for example, are influenced by a number of non-disease-related factors including age, co-existent illness, type of assay used, iodine status, time of sampling, ethnicity, amongst others – which are discussed later in the review. Examples of assay-related variations in FT4 and TSH reference intervals are illustrated in table 1 (6).

Another approach to determining reference intervals has been advocated (17). This suggestion alters the focus from a statistically calculated reference interval to defining an optimal range for thyroid function. This concept is similar to the one used to define optimal ranges for cardiovascular risk factors such as fasting glucose or total cholesterol levels – where the interval or cut-off is not based on its distribution but on its associated long-term risk to health. However, this strategy is not straightforward in defining reference intervals for
thyroid parameters as both hypo and hyperthyroidism are related to a higher risk of adverse outcomes.

**Relationship between TSH and thyroid hormone**

As behaves a sensitive signalling system, there is an inverse logarithmic relationship between serum TSH and thyroid hormone concentrations (Figure 2). The central component of this relationship includes TSH-free T4 (FT4) data pairs that are contributed by those individuals that have no thyroid abnormalities, defined by absence of thyroid antibodies and sonographic abnormalities (18), whereas the data pairs at either end of the range indicate the pituitary response to insufficient or excessive production of thyroid hormones when primary thyroid disease is present. The definition of which data points fall outside of the normal range depend on whether modelling incorporates TSH, FT4, and free triiodothyronine (FT3) individually, or as a composite. Use of composite methods yields a lower percentage of abnormal results compared use of a single analyte alone (19).

The linearity of the relationship between the logarithmically-transformed TSH and FT4 is fairly well-maintained at a population level. However, more detailed analyses show that this relationship is actually more complex. Factors affecting this relationship include age, sex, and whether an individual is being treated with thyroid hormone or not. With respect to age, when examining FT4 values within the normal range, TSH values are higher in older individuals. However, when examining low FT4 values, corresponding TSH values are lower in older individuals (20). This results in the relationship between TSH and FT4 being described best by two overlapping sigmoidal curves. In a study from Australia TSH values are higher for men than women across a range of FT4 values that fell within the normal range (20). Relationships between TSH and FT4 and TSH and FT3 may also differ depending on whether an individual is being treated with levothyroxine (LT4) or not, with displacement of the curves to the right and left respectively with treatment (21). Moreover, there is a narrower range of within-individual TSH-FT4 values than the range seen in a population (22).

**Physiologic Factors impacting on TSH and thyroid hormone levels**

In health, serum thyroid function parameters demonstrate significant inter-individual variation, whereas intra-individual variability is much narrower – suggesting that the
hypothalamic-pituitary-thyroid (HPT) axis set-point is unique for each individual (23). The
HPT axis is mainly determined by demographic, genetic and certain environmental factors, as
described below (and also listed in Table 2).

**Age:** Serum TSH levels rise with age and the increase is particularly pronounced after the age
of 70 years (24-26) (Figure 3), although this finding has not been confirmed in another study
(27). T4 levels have been shown to either remain stable (26) or increase (25, 27) with age,
whereas T3 levels decrease (25) Despite some studies suggesting that a marginally raised
TSH may not be detrimental (28-30) and international guidelines recommending a
conservative approach in this older group of individuals, who may have a frail phenotype and
comorbidities such as heart failure, (31, 32), there is no consensus whether age-specific
thyroid function reference ranges should be utilized. However, it is acknowledged that there
is a knowledge gap with respect to whether age-specific ranges may be needed. Use of age-
specific TSH reference ranges, especially in persons over 70 years, may result in the
reclassification of some "abnormal" results to "normal", and thus preventing overestimation
of subclinical hypothyroidism (33) Such a stance may prevent significant over-estimation of
subclinical thyroid disease (32). However, the use of age-specific TSH ranges has been
shown to have minimal impact on the reclassification of a person's thyroid status, except in
participants over 85 years in one study (34).

**Gender:** Women as a group tend to have higher TSH levels but this increase is accounted for
by the presence of positive thyroid peroxidase antibody (TPOAb) status, such that in a study
from the United States women have higher TSH values than men when those with risk factors
for thyroid disease are included, but similar values when those with risk factors are excluded
(24). Furthermore, the age-related increase in serum TSH that is observed is similar in both
genders (25, 26).

**Ethnicity:** Individuals of African-Caribbean descent have lower serum TSH levels than
Caucasians. The median and 97.5th percentile (signifying the upper limit of the reference
range) for TSH is 1.43 and 4.18 mIU/L in Caucasians. In African-Caribbeans, the
corresponding values are 1.19 and 3.63 mIU/L, respectively (24). Uniform reference ranges
that combine data across all racial groups risk misclassifying 8 and 4% with low TSH, and 2
and 6% with high TSH, in African-Caribbeans and Caucasians, respectively (35).
Genetics: Classical twin studies estimate the heritability of serum TSH and T4 levels between 39 – 65% (36, 37). A meta-analysis of thousands of euthyroid individuals of European descent reported significant independent associations for 26 single nucleotide polymorphisms (20 loci for TSH and 6 for FT4), that explain 5.6% and 2.3% of variation, respectively (38).

Pregnancy: Thyroid function is subjected to major changes during pregnancy related to increase in thyroxine-binding globulin, increased degradation of circulating thyroid hormones by placental deiodinase 3 and stimulation of the thyroid gland by human chorionic gonadotrophin. The net effect is a reduction in serum TSH levels and an increase in total thyroid hormones (39). Trimester-specific reference ranges for thyroid analytes have been suggested (40).

Body mass index: Individuals with higher body mass index have higher TSH levels (41-43) and lower or unchanged T4 concentrations. It seems likely that the increased body mass leads to these hormonal changes rather than vice versa as TSH levels reduce after weight loss (44). Moreover, the weight loss associated with treatment of overt hypothyroidism appears to be due to loss of body water (45).

Circadian and seasonal variation: Circadian variation in serum TSH and, to a lesser extent, in T3 is well described. Lowest levels of TSH are found in the afternoon and a more than 100% rise is observed just after midnight (46). In addition, T3 levels are generally slightly higher in winter while the changes in TSH and T4 are not consistent. Part of the reason for seasonal changes in thyroid function in certain countries could be attributed to variation in iodine intake (46). The circadian rhythm of TSH increases the upper limit of its reference interval with age mainly due to an increased in the amplitude of the nocturnal TSH surge in older individuals (47).

Pathologic Factors impacting on TSH and thyroid hormone levels

TPOAb status: Individuals with positive TPOAb levels tend to have higher serum TSH levels and lower T4 concentrations (27).
Smoking: Tobacco smoking is associated with lower TSH concentrations and higher T4 levels in a dose-related fashion (48) and smoking cessation leads to a rise in TSH and reduction in FT4 levels suggesting that the effects are reversible.

Illness: Thyroid function is impacted by illness and is termed nonthyroidal illness or sick euthyroid syndrome. The typical alterations observed in the acute phase are reduction in serum T3 (both total and free) levels, low or normal circulating T4 (both total and free), while TSH levels remain normal (49). The protracted phase of illness – especially if it is critical – is associated with low levels of all three thyroid-related hormones.

Drugs: Several drugs can affect thyroid function by nature of their iodine content, inhibitory effect on the hypothalamic-pituitary axis, destructive thyroiditis or immune-reactivating mechanisms (see Table 2).

Iodine status and iodine intake including contrast media: Both iodine deficiency and excess can affect TSH and thyroid hormone levels. Severe iodine deficiency can cause hypothyroidism and goitre while mild-moderate deficit can lead to hyperthyroidism due to an increased prevalence of toxic multinodular goitre. Iodine supplementation in an area that was previously iodine deficient can lead to an increase in TSH levels (50). Areas with high iodine intake have higher average and upper-limit of TSH levels (51) – reflecting the importance of iodine status on thyroid function. Iodine containing contrast media as used for CT scans or angiography contains 13,500 mcg free iodine per scan and thus has approximately 100 times more iodine than the recommended daily allowance of 150 mcg (52). Despite this substantial high iodine load remarkably few individuals proceed to develop thyroid dysfunction. In some susceptible persons, however, features of thyroid dysfunction (both hypo- and hyperthyroidism) can occur (53).

Laboratory Factors impacting on TSH and thyroid hormone levels

Inter- and intra-assay differences: Most tests for estimating TSH and thyroid hormones are performed using immunoassays. Currently, most immunoassays for FT4 and free T3 (FT3) exhibit significant biases that exceed intra-individual variability (Table 1) (3). Interference by
proteins (immunoglobulins and paraproteins) are the usual cause of erroneous thyroid function results. Results affected by these interferences and others (Table 2) can only be suspected by requesting clinicians when they are discordant with the clinical scenario.

Decisions about diagnosis and treatment of thyroid dysfunction (2000 words)

i) Diagnosis and treatment of hypothyroidism
   a) Making decisions regarding whether hypothyroidism is present
   A diagnosis of hypothyroidism can be confirmed by biochemical testing once it is suspected on clinical grounds. Characteristically, serum TSH levels are elevated and T4 concentrations are low. Circulating T3 tends to be in the reference range until profound hypothyroidism develops. On occasions, TSH elevation – particularly when mild or moderate – can be transient and therefore levels should be retested after an appropriate interval (usually at least 6 weeks) before confirming a diagnosis of hypothyroidism. Transient rises of TSH are observed during the recovery phase after nonthyroidal illness or due to certain drugs. Minor elevations of TSH noted in those over age 55 years normalize in approximately 40% of people followed up over an average of 32 months (54).

   The other important question that needs to be considered is whether TSH reference ranges impact on the diagnosis of thyroid disease. As noted before, several factors including age, iodine status, race and body mass index impact on the TSH reference range. Using an age-specific TSH range had minimal impact (<2%) in reclassifying a person’s thyroid status, except in the oldest old above the age of 85 years (<5%) in one study (34). Areas that are mild-moderately iodine deficient tend to have lower TSH ranges and thus prevalence rates for hypothyroidism are also lower. For example, in the Study of Health in Pomerania (SHIP), a population-based study of adults residing in North-East Germany which was till recently iodine deficient, the TSH reference range was 0.3 – 3.0 mIU/L (55). Correspondingly, the prevalence of hypothyroidism (both overt and subclinical) was 1.2% whereas low TSH was observed in a much higher proportion (11.3%). These prevalence rates are in stark contrast to thyroid disease prevalence data from areas which are considered iodine sufficient where hypothyroidism is more predominant.

   b) Subclinical hypothyroidism
The clinical significance of overt and subclinical hypothyroidism (SC Hypo) depends on the degree of TSH elevation. The debate regarding the long-term health implications of minor TSH elevations has continued for over 4 decades (56). Observational studies differ in their conclusions as to the significance of SC Hypo with cardiovascular outcomes or impairment in quality of life. A meta-analysis of 11 studies with individual data on more than 55,000 participants showed that only TSH elevations above 10 mIU/L were associated with cardiovascular disease (57). One randomised controlled trial in older (>65 years) individuals with persistent SC Hypo (TSH levels between 4.6 and 19.9 mIU/L) concluded that low-dose T4 replacement therapy provides no benefit in improving QoL or thyroid-related symptoms (58). However, one caveat of this study was that the mean pre-treatment TSH values in the experimental and control groups were relatively low at 6.41 and 6.38 mIU/L respectively. Another trial of T4 in older (>65 years) participants with SC Hypo (TSH > 5.5 mIU/L) showed no benefit of treatment on measures of cognitive function (59). These data suggest that T4 treatment of SC Hypo is of no benefit in improving symptoms, quality of life or cognitive function, at least in older individuals. There are no randomised controlled trials of SC Hypo that have evaluated incident cardiovascular events or mortality. However, a retrospective analysis of a General Practitioner database concluded that treatment of persistent SC Hypo (TSH levels between 5.0 and 10.0 mIU/L) was associated with lower incident fatal and non-fatal cardiovascular events only in people aged between 40 and 70 years but a similar benefit was not observed in those who were older than 70 years (60). Based on these data, in part, the European Thyroid Association formulated guidelines that recommend treatment based on TSH levels (greater than 10 mIU/L across all age groups) and the individual’s age (less than 70 years) (31). Potential factors to consider when deciding whether to observe or treat SC Hypo are shown in Figure 4a and 4b. The factors on each seesaw are not weighted according to their relative position on the seesaw. In managing older patients with SC Hypo a geriatric multidimensional assessment might also be helpful in decision-making.

c) Overt Hypothyroidism

Overt hypothyroidism is diagnosed when TSH levels are high (usually >10 mIU/L) and T4 levels are low. It is widely accepted that overt hypothyroidism should be treated (61). The only available treatment is thyroid hormone therapy. There is no evidence-based role for other therapies such as nutritional supplements including iodine as the sole management...
strategy in hypothyroidism (61). The aim of therapy is to normalize thyroid function and
ameliorate symptoms, if any. This can be achieved by T4 administration (usually in the form
of the synthetic LT4). However, approximately 10% of patients on LT4 therapy for
hypothyroidism are dissatisfied with their treatment. Combination therapy of T4 and T3 has
not been shown to provide any benefit over T4 monotherapy alone in a number of
randomised controlled trials (62-65). It is possible that there might be a certain group of
patients that could benefit from T4/T3 combination therapy (66), although this needs to be
confirmed in more trials conducted prospectively. Current guidelines recommend that T3
treatment should only be provided in a trial environment, and discontinued if no benefit is
perceived (67). Debates about how to define euthyroidism, and whether TSH is the best
indicator of euthyroidism continue (68).

ii) Diagnosis and treatment of hyperthyroidism

a) Making decisions regarding whether hyperthyroidism is present

As is the case for TSH values that are above the upper limit of the reference interval, many
TSH values that are below the lower limit of a generic reference interval may either be
normal for that particular individual or may be only transiently abnormal. For example TSH
values tend to be lower in African Americans (69). A TSH calculator has been proposed to
calculate reference intervals in individuals of different race and age, and this can be used to
compute a lower TSH reference interval for African Americans (70). Smokers tend to have
low TSH values than non-smokers (71). With respect to transient TSH abnormalities, studies
have examined the impact of both the degree of TSH suppression and the trend over time.
Among individuals over 60 years of age with TSH values that were less than 0.5 mIU/L but
greater than 0.05 mIU/L, as many as 76% of these had a normal TSH upon follow up after 12
months (72) (see Figure 5). In another study, also examining the natural history of untreated
endogenous subclinical hyperthyroidism (SC Hyper), an increasing percentage of patients
reverted to normal TSH values over time, such that 38% had normalized their TSH values
after 7 years of follow up (73) (see Figure 5).

Another key question, in addition to whether a particular TSH value is actually abnormal or
not, is once a laboratory-defined abnormality has been identified, does that “abnormality”
have any clinical significance. Several scenarios are possible. An important disease may be
identified and treated, or a risk for a disease may be identified and monitored. Alternatively, the affected individual may actually have no disease at all, but nevertheless there may be personal costs, such as the anxiety and inconvenience associated with additional testing, and there may be financial costs. In a study of the effect of different low TSH cut-offs for reflex testing of FT4, using lower TSH values would have missed only a small percentage of FT4 values above the reference interval and many of these values were only marginally elevated (74). For example, triggering reflex testing of FT4 with a TSH value of 0.3 mIU/l, rather than 0.4 mIU/L in a cohort of 120,403 individuals would have avoided testing in 176 individuals, 76% of whom had minimal FT4 elevations. Whether these cases best represent individuals with disease whose identification has been missed, or avoidance of an unnecessary burden by not identifying patients without real disease, depends upon longitudinal studies of outcomes.

Being cognizant of even mild thyroid status abnormalities is supported by the finding that even lower and higher TSH values that are within the reference range predict future hyperthyroidism and hypothyroidism respectively, although the association was stronger for hypothyroidism (75). Similarly, there is a growing body of evidence that even TSH values that fall within the lower and upper parts of the normal range in euthyroid individuals may be associated with adverse outcomes. TSH values within the lower end of the normal range have been associated with diverse outcomes such as increased hip fractures in older women (76), lower serum cholesterol levels (77), lower body weight (78), and less metabolic syndrome. All-cause mortality has been shown to be either increased (30, 79), decreased (80), or in individuals over 85 years of age, not affected (29) by having low-normal TSH values. Combining multiple datasets of healthy individuals with normal TSH values shows that lower TSH values within the normal range are associated with higher risk of osteoporosis and vertebral fractures, compared with higher TSH values within the normal range (81). Risk of adverse metabolic and cardiovascular outcome may be decreased compared with individuals with TSH values in the upper end of the normal range. TSH values in the upper part of the reference range are associated with lower risk of stroke (82).

Despite the fact that the data about mortality is not consistent, these types of studies illustrate that thyroid function/dysfunction is associated with a continuum of risk and that assignment of specific reference intervals may be an oversimplification of the underlying physiology. Though such studies of endogenous thyroid function are interesting, it would be an
extrapolation to infer that these data also apply to those being treated for hypothyroidism with LT4, as the T<sub>4</sub>/T<sub>3</sub> ratios also differ between treated and untreated individuals. Therefore, such data also creates a quandary for the physician treating thyroid disease with respect to what particular TSH target is best for each individual patient. Moreover, if a narrower TSH goal such as the midportion of the normal range was desirable, this would represent an even greater challenge with respect to avoiding iatrogenic thyroid disease.

b) Subclinical hyperthyroidism

SC Hyper, by definition, exists when a TSH value below the lower limit of the reference interval is accompanied by a FT4 value within the reference interval. The prevalence of SC Hyper in Europe was 2.91 % (95% CI 2.63-3.21%) in a recent metanalysis (11).

Exogenous SC Hyper

Exogenous hyperthyroidism exists when the cause is treatment with thyroid hormone. Other than intentional suppression of serum TSH in an individual with moderate to high risk thyroid cancer, exogenous SC Hyper represents failure of a treatment program. The percentage of patients with SC Hyper is as high as 13-28% in some studies of LT4-treated patients (72, 83, 84) (see Figure 6), including a study of those over age 65 years (83). The different prevalences noted are likely due to different methodology and populations (e.g. health fair attendees versus longitudinal cohort) but nevertheless illustrate the magnitude of the issue. This is a serious problem, as it is conceivable that the risk of iatrogenic SC Hyper is greater than the risk of the hypothyroidism for which the thyroid hormone was initiated.

Risks of suppressed TSH caused by exogenous thyroid hormone extend to dysrhythmias and fractures (85).

Endogenous SC Hyper

Endogenous SC Hyper is most often due to mild Graves’ disease or toxic nodular disease. The latter is less likely to remit than the former (86). Observational studies show multiple risks of SC Hyper including atrial fibrillation (87, 88), hip fractures (89, 90), and dementia (91). Increased cardiovascular mortality (92) and all-cause mortality (93) have been shown in some, but not all studies. Assessment of the risks of treatment compared with monitoring without treatment are generally inferred from these observational studies.
There are few randomised or controlled trials comparing outcomes in treated and untreated patients with SC Hyper, and those that exist are small. For example, in a study of 28 postmenopausal women with toxic nodular goitres with serum TSH values of <0.2mIU/L, 12 were followed without treatment and 16 were followed after they received radioiodine therapy and achieved a mean TSH of 0.39-0.83 mIU/L. Bone mineral density continued to decline in those not treated, but stabilized in the treated group (94). In another study patients with toxic nodular disease were randomised to methimazole or observation. Ventricular premature beats decreased and a marker of bone density increased in those taking methimazole (95). On the other hand, treatment of SC Hyper in pre-menopausal women with either Graves’ disease or toxic nodular disease did not change the trend for loss of bone density over a 6-month period, compared with untreated controls. TSH values were 0.21-0.23 mIU/L prior to treatment and reached 3.8 mIU/L in the treated group (96). These studies were too small to perform analyses adjusting for such important factors as baseline TSH, aetiology of SC Hyper, and menopausal status. Only larger studies with the ability to stratify patients based on relevant risk factors such as age, menopausal status, functional status etc will answer the question of whether a personalized “reference interval” or “risk-adapted reference interval” could be used to make decisions about the care of individual patients. A small study of individuals with TSH values ranging from 0.005-0.02 mIU/L showed an improvement in mid-thigh muscle strength 6-9 months after their SC Hyper was treated (97). Thus, a relevant piece of information might be whether a patient is at risk for falls or whether their occupation requires physical activity. Another consideration, for example, in an individual who is at risk of falling or sustaining a fracture might be whether they have a normal serum sodium, with treatment being more strongly considered in a hyponatraemic patient who is already at extra risk of sustaining a fracture (98). Therefore, a rational decision about whether a particular patient would derive benefit from treatment of their SC Hyper is complex and might involve use of a dynamic algorithm that incorporated multiple patient characteristics. Not only would these characteristics be weighted differently in different patients, but also their importance would change over the life span of the patient. Examples of some patient characteristics or other factors favouring treatment and monitoring respectively are shown in Figures 7a and 7b. The factors on each see saw are not weighted according to their relative position on the see saw. Another factor to be considered would be
incorporating the patient in the decision making (99). This could potentially lead to greater
patient satisfaction with their choice and better adherence to therapy (100).

All such considerations of the risks and benefits of treatment of SC Hyper assume that
treatment will render the patient “euthyroid”. While it is clear that both iatrogenic
hypothyroidism and iatrogenic hyperthyroidism while treating hypothyroidism is associated
with increased mortality (101), similar data about overtreatment or undertreatment of SC
Hyper is not available. It is also not known whether targeting a particular TSH value within
the normal range while treating SC Hyper is desirable for optimum control of blood pressure
(102) and dyslipidaemia (103), and reduction of metabolic and cardiovascular risk (81).

When treating patients with hypothyroidism, targeting a specific TSH does not seem to affect
body composition, lipid profile, or energy expenditure (104-106), mood and cognition (107),
or satisfaction (108). Although, interestingly, lowering the TSH values of individuals with
hypercholesterolemia from the upper to the lower part of the normal range also lowered their
cholesterol (109). Studies of targeting particular TSH goals in patients with SC Hyper have
not been reported.

c) Overt hyperthyroidism

Overt hyperthyroidism, by definition, exists when a TSH value below the lower limit of the
reference interval is accompanied by a FT4 value above the upper limit of the reference
interval. In some cases, the low TSH may be accompanied by an elevated T3 (so-called T3
thyrotoxicosis), with the FT4 remaining normal. The benefits of treating overt
hyperthyroidism are undisputed (110). The goal of treatment is to normalize the patient’s
serum TSH and relieve symptoms of hyperthyroidism. The preceding discussion about what
is an ideal target TSH when treating SC Hyper also applies to treating overt hyperthyroidism.

Conclusions

Thyroid hormone has actions in all the cells of the body. Significant perturbations in thyroid
function can have manifestations in any organ system, and these effects can be well-
compensated and difficult to recognise due to their non-specificity, or can be obvious and
classic in their presentation. However, the rationale for treating overt thyroid disease is
undisputed. Mild thyroid disease forms a continuum with normal thyroid function and when
present can be associated with only subtle symptoms. The boundaries between normal thyroid function and thyroid disease are generally defined based upon statistically-derived reference intervals. The consequence of this is that thyroid parameters falling outside these ranges, but not proven to have any health consequences may merge with values that may be associated with adverse health outcomes. Moreover, the cut-off between these two classifications may be different for different health outcomes and for different population subgroups or at an individual level. This quandary highlights the need for further research that incorporates randomised controlled trials of the multiple health outcomes of treating or not treating various degrees of thyroid dysfunction. Such trials would have to provide sufficient power to stratify results by such parameters as age and co-existent chronic disease.

Such large trials can only be accomplished with multinational efforts. Perhaps groundwork in the form of education and advocacy can pave the way for the funding and commitment to such trials being performed by multiple centres with clinical trial expertise in the future. An alternative approach is for future trials to focus on high risk individuals, such as those at high risk of cardiovascular disease.

References


35. Boucai L, Surks MI 2009 Reference limits of serum TSH and free T4 are significantly influenced by race and age in an urban outpatient medical practice. Clin Endocrinol (Oxf) 70:788-793.


McAninch EA, Rajan KB, Miller CH, Bianco AC 2018 Systemic Thyroid Hormone Status During Levothyroxine Therapy In Hypothyroidism: A Systematic Review and Meta-Analysis. J Clin Endocrinol Metab.


Leader A, Ayzenfeld RH, Lishner M, Cohen E, Segev D, Hermoni D 2014 Thyrotropin levels within the lower normal range are associated with an increased risk of hip fractures in euthyroid women, but not men, over the age of 65 years. J Clin Endocrinol Metab 99:2665-2673.


907 102. Asvold BO, Bjoro T, Nilsen TI, Vatten LJ 2007 Association between blood pressure and serum thyroid-stimulating hormone concentration within the reference range: a


Reference intervals in the diagnosis of thyroid dysfunction
– time to treat patients and not numbers.

Jacqueline Jonklaas¹ and Salman Razvi²

¹Division of Endocrinology Georgetown University, Washington, DC, USA and ²Department of Endocrinology, University of Newcastle, Newcastle, UK

Key words: Hypothyroidism, hyperthyroidism, thyroid hormones, TSH, diagnosis, reference intervals

Disclosures: J.J. and S.R. have no relevant disclosures

Manuscript word count = 5727 words

References 110

Figures 7, tables 2.

Search strategy and selection criteria: References for this review were identified through searches of PubMed for articles published in English from January, 1971, to June, 2018, by use of the MeSH terms “hypothyroidism”, “hyperthyroidism”, “subclinical hypothyroidism”, “subclinical hyperthyroidism”, “thyroid disease”, “thyroid stimulating hormone”, “thyroid hormone”, “thyroxine”, “free thyroxine”, “triiodothyronine” in combination with the term “reference value”. Relevant articles were also identified through searches in the authors’ personal files. Articles resulting from these searches and relevant references cited in those articles were reviewed. Preference was given to more recent articles.
Summary

Although the assignment of a diagnosis of thyroid dysfunction appears quite simple, this is often not the case. Issues that lead to complexities regarding whether thyroid function is, in fact, normal include transient changes in thyroid parameters, inter- and intra-individual differences in thyroid parameters, age-related differences, and ethnic variations. Superimposed upon these considerations, is the understanding that a statistically calculated distribution of thyroid analytes does not necessarily coincide with intervals or cut offs that have predictive value for beneficial or adverse health outcomes. Based on current trial data, it is unclear whether certain adults would still benefit from levothyroxine, (such as those with TSH values > 10 mIU/L), since a limited number of these persons were included in randomised trials. Even if therapy is initiated for abnormal thyroid function, not all treated individuals are maintained at the desired treatment target, and therefore still may be subject to risks. The consequence of this is that not only does each patient’s thyroid function need to be assessed on an individual basis with the entire clinical picture in mind, but also monitoring needs to be vigilant, and the targets for treatment re-assessed on an ongoing basis.

Introduction

Thyroid dysfunction is one of the common conditions encountered by medical practitioners. Hypothyroidism, including the subclinical form, affects between 5 – 15% of adults, whereas hyperthyroidism, including the subclinical version, is seen in 0.5 – 2% of the population. Symptoms of thyroid dysfunction are non-specific and therefore thyroid function testing is one of the commonest tests to be requested. Clinical evaluation of patients with suspected thyroid dysfunction by assessing symptoms and examining for signs, though useful, is not very sensitive in confirming the disease. Measurement of thyroid function by quantifying circulating thyroid hormones and thyroid-stimulating hormone or thyrotropin (TSH) concentrations is deemed to be the most accurate and reliable approach to diagnose hypo and hyperfunction. The diagnosis of thyroid dysfunction has important implications for the patient as both hypo and hyperthyroidism are long-term conditions that may warrant prolonged therapy (life-long in the case of hypothyroidism and several months or years in case of hyperthyroidism).
The prevalence and incidence of the common thyroid disorder hypothyroidism is increasing. It is likely that this is, at least in part, due to a fall in the threshold level of TSH at which thyroid hormone therapy for hypothyroidism is being initiated (1). In this study approximately 30% of the individuals who were prescribed therapy were treated for subclinical hypothyroidism (SC Hypo) based on the documentation of their thyroid hormone levels (1). Therefore, more patients with borderline disease are being commenced upon treatment although the evidence of benefit is not present for quality of life, symptoms, or cognitive function (2), is unclear with respect to cardiovascular outcomes, and therefore the overall risk of harm is uncertain. Inappropriate and unnecessary treatment may lead to increased health and economic burden – both at personal and societal levels.

Understanding how reference intervals are determined and being aware of common pitfalls in the interpretation of thyroid function test results is crucial to provide safe and effective care for patients with suspected thyroid disease. This review aims to provide an insight into the variation in TSH, thyroxine (T4), and triiodothyronine (T3) in serum, impact on their respective reference intervals, the determinants of biologic variation, and its impact on interpretation of test results.

**Historical perspective of diagnosis of thyroid function**

Before the advent of thyroid function assays, the diagnosis of thyroid dysfunction was based purely on clinical grounds and, therefore, milder and less clinically apparent forms of the disease were most likely missed. Measuring clinical end-organ targets of thyroid function such as pulse rate, serum cholesterol, body temperature, ankle jerk reflex time, or basal metabolic rate were utilised instead. In the 1950s, protein bound iodine technique was used as an indirect estimate of the serum total T4 concentration. Later, from 1965 onwards, development of first generation radioimmunoassays led to the estimation of serum TSH, although these assays initially had limited functional sensitivity. Since then, TSH tests have evolved progressively and are able to detect lower TSH levels with improved sensitivity, which has led to the progressive reduction in the lower limit of the TSH level to approximately 0.4 mIU/L (3). At the same time, the upper limit of the TSH reference range has also decreased from ≈ 10 to 4.5 mIU/L (4).

**Prevalence of thyroid dysfunction and prescriptions for treatment**
First and foremost, the prevalence of thyroid dysfunction is affected by the assay methods used to measure TSH and thyroid hormones (5, 6) and the statistical methods utilized to derive the reference intervals (7). With respect to statistical methods a recent study illustrated the different lower and upper limits of normal for TSH of 0.65-3.81 mIU/L and 0.27-4.87 mIU/L for the Hoffman and Tukey methods respectively (7). Efforts to harmonize both TSH and thyroid hormone assays are ongoing (8). Additionally, the prevalence of thyroid dysfunction is affected by the definitions of dysfunction employed and also the population being considered. As will be discussed, thyroid analytes falling outside a specific reference interval are often considered synonymous with the existence of disease, although there may not necessarily be a disease, or even a need for treatment. The decision as to whether disease is present may be better judged based on outcomes from epidemiologic studies (9). When defining reference intervals, narrower reference intervals will yield higher prevalence of thyroid dysfunction, and vice versa. The particular population under consideration also affects the prevalence of thyroid dysfunction. For example, populations with more older individuals and geographic regions affected by iodine deficiency harbour different burdens of thyroid dysfunction. Another factor affecting prevalence is the extent of screening for disease, with more comprehensive screening producing more abnormal thyroid analytes.

Hypothyroidism is the more common disorder, with overt disease affecting between 0.20-5.3% of the European population (10). Hyperthyroidism, on the other hand, has a prevalence of 0.75% in Europe (11). There is only one available treatment for hypothyroidism, namely thyroid hormone. Prescription of thyroid hormone in the United States has increased by 1.3% between 1999-2000 and 2011-2012, with the increase remaining significant after adjusting for the age of the population (12). Part of this increase could potentially be contributed to by the increased diagnosis of thyroid cancer, with resultant post-surgical hypothyroidism. Other factors associated with increased prescribing include less cigarette smoking and higher rates of obesity (13). The prevalence of treated hypothyroidism is highest in older age groups (14). It is possible that the trend for increased prescribing of thyroid hormone over time is partly due to increased treatment of mild degrees of hypothyroidism, including SC Hypo (1). The decreased prevalence of untreated SC Hypo in Norway between the two time periods 1995-1997 and 2006-2008 is certainly suggestive of increased treatment of mild hypothyroidism over time (15). Trends in prescribing antithyroidal medications for hyperthyroidism show more use of these agents and less use of radioiodine therapy and surgery over time, but
generally are not suggestive of altered thresholds for treatment of hyperthyroidism (16).

However, algorithms used to decide whether treatment for subclinical hyperthyroidism is indicated incorporate both age and cardiovascular disease, so that rates of treatment may be greater in aging populations.

Reference intervals and how they are determined

Normality is relative and biochemical measurement of analytes of interest is frequently interpreted by comparison with earlier measurements taken from an apparently ‘healthy population’. The reference range for thyroid function tests, similar to many other biochemical analytes, is usually determined as the interval into which 95% of apparently euthyroid individuals fall. The distribution of both thyroid hormones (T4 and T3) is Gaussian and therefore the reference interval ranges two standard deviations below and above the mean. Serum TSH, however, has a non-Gaussian distribution, so the reference interval is obtained by including values that fall between 2.5th and the 97.5th ranked percentiles (Figure 1). This non-Gaussian distribution is common to other analytes such as, for example, prostate specific antigen. Reasons for the shift to the right in the curve could include inclusion of some individuals with autoimmune thyroid disease, altered biological activity of TSH especially in the elderly, and mutations in the TSH receptor. This method of defining reference intervals means that one in twenty (5%) apparently euthyroid individuals will have values outside this range. Similarly, the presence of thyroid disease is not excluded by results that fall within the reference range, without taking the complete clinical picture into account. The ‘normal’ TSH or thyroid hormone concentrations, for example, are influenced by a number of non-disease-related factors including age, co-existent illness, type of assay used, iodine status, time of sampling, ethnicity, amongst others – which are discussed later in the review. Examples of assay-related variations in FT4 and TSH reference intervals are illustrated in table 1 (6).

Another approach to determining reference intervals has been advocated (17). This suggestion alters the focus from a statistically calculated reference interval to defining an optimal range for thyroid function. This concept is similar to the one used to define optimal ranges for cardiovascular risk factors such as fasting glucose or total cholesterol levels – where the interval or cut-off is not based on its distribution but on its associated long-term risk to health. However, this strategy is not straightforward in defining reference intervals for
thyroid parameters as both hypo and hyperthyroidism are related to a higher risk of adverse outcomes.

**Relationship between TSH and thyroid hormone**

As behaves a sensitive signalling system, there is an inverse logarithmic relationship between serum TSH and thyroid hormone concentrations (*Figure 2*). The central component of this relationship includes TSH-free T4 (FT4) data pairs that are contributed by those individuals that have no thyroid abnormalities, defined by absence of thyroid antibodies and sonographic abnormalities (18), whereas the data pairs at either end of the range indicate the pituitary response to insufficient or excessive production of thyroid hormones when primary thyroid disease is present. The definition of which data points fall outside of the normal range depend on whether modelling incorporates TSH, FT4, and free triiodothyronine (FT3) individually, or as a composite. Use of composite methods yields a lower percentage of abnormal results compared use of a single analyte alone (19).

The linearity of the relationship between the logarithmically-transformed TSH and FT4 is fairly well-maintained at a population level. However, more detailed analyses show that this relationship is actually more complex. Factors affecting this relationship include age, sex, and whether an individual is being treated with thyroid hormone or not. With respect to age, when examining FT4 values within the normal range, TSH values are higher in older individuals. However, when examining low FT4 values, corresponding TSH values are lower in older individuals (20). This results in the relationship between TSH and FT4 being described best by two overlapping sigmoidal curves. *In a study from Australia* TSH values are higher for men than women across a range of FT4 values that fell within the normal range (20).

Relationships between TSH and FT4 and TSH and FT3 may also differ depending on whether an individual is being treated with levothyroxine (LT4) or not, with displacement of the curves to the right and left respectively with treatment (21). Moreover, there is a narrower range of within-individual TSH-FT4 values than the range seen in a population (22).

**Physiologic Factors impacting on TSH and thyroid hormone levels**

In health, serum thyroid function parameters demonstrate significant inter-individual variation, whereas intra-individual variability is much narrower – suggesting that the
hypothalamic-pituitary-thyroid (HPT) axis set-point is unique for each individual (23). The HPT axis is mainly determined by demographic, genetic and certain environmental factors, as described below (and also listed in Table 2).

Age: Serum TSH levels rise with age and the increase is particularly pronounced after the age of 70 years (24-26) (Figure 3), although this finding has not been confirmed in another study (27). T4 levels have been shown to either remain stable (26) or increase (25, 27) with age, whereas T3 levels decrease (25) Despite some studies suggesting that a marginally raised TSH may not be detrimental (28-30) and international guidelines recommending a conservative approach in this older group of individuals, who may have a frail phenotype and comorbidities such as heart failure, (31, 32), there is no consensus whether age-specific thyroid function reference ranges should be utilized. However, it is acknowledged that there is a knowledge gap with respect to whether age-specific ranges may be needed. Use of age-specific TSH reference ranges, especially in persons over 70 years, may result in the reclassification of some "abnormal" results to "normal", and thus preventing overestimation of subclinical hypothyroidism (33) Such a stance may prevent significant over-estimation of subclinical thyroid disease (32). However, the use of age-specific TSH ranges has been shown to have minimal impact on the reclassification of a person's thyroid status, except in participants over 85 years in one study (34).

Gender: Women as a group tend to have higher TSH levels but this increase is accounted for by the presence of positive thyroid peroxidase antibody (TPOAb) status, such that in a study from the United States women have higher TSH values than men when those with risk factors for thyroid disease are included, but similar values when those with risk factors are excluded (24). Furthermore, the age-related increase in serum TSH that is observed is similar in both genders (25, 26).

Ethnicity: Individuals of African-Caribbean descent have lower serum TSH levels than Caucasians. The median and 97.5th percentile (signifying the upper limit of the reference range) for TSH is 1.43 and 4.18 mIU/L in Caucasians. In African-Caribbeans, the corresponding values are 1.19 and 3.63 mIU/L, respectively (24). Uniform reference ranges that combine data across all racial groups risk misclassifying 8 and 4% with low TSH, and 2 and 6% with high TSH, in African-Caribbeans and Caucasians, respectively (35).
Genetics: Classical twin studies estimate the heritability of serum TSH and T4 levels between 39 – 65% (36, 37). A meta-analysis of thousands of euthyroid individuals of European descent reported significant independent associations for 26 single nucleotide polymorphisms (20 loci for TSH and 6 for FT4), that explain 5.6% and 2.3% of variation, respectively (38).

Pregnancy: Thyroid function is subjected to major changes during pregnancy related to increase in thyroxine-binding globulin, increased degradation of circulating thyroid hormones by placental deiodinase 3 and stimulation of the thyroid gland by human chorionic gonadotrophin. The net effect is a reduction in serum TSH levels and an increase in total thyroid hormones (39). Trimester-specific reference ranges for thyroid analytes have been suggested (40).

Body mass index: Individuals with higher body mass index have higher TSH levels (41-43) and lower or unchanged T4 concentrations. It seems likely that the increased body mass leads to these hormonal changes rather than vice versa as TSH levels reduce after weight loss (44). Moreover, the weight loss associated with treatment of overt hypothyroidism appears to be due to loss of body water (45).

Circadian and seasonal variation: Circadian variation in serum TSH and, to a lesser extent, in T3 is well described. Lowest levels of TSH are found in the afternoon and a more than 100% rise is observed just after midnight (46). In addition, T3 levels are generally slightly higher in winter while the changes in TSH and T4 are not consistent. Part of the reason for seasonal changes in thyroid function in certain countries could be attributed to variation in iodine intake (46). The circadian rhythm of TSH increases the upper limit of its reference interval with age mainly due to an increased in the amplitude of the nocturnal TSH surge in older individuals (47).

Pathologic Factors impacting on TSH and thyroid hormone levels

TPOAb status: Individuals with positive TPOAb levels tend to have higher serum TSH levels and lower T4 concentrations (27).
Smoking: Tobacco smoking is associated with lower TSH concentrations and higher T4 levels in a dose-related fashion (48) and smoking cessation leads to a rise in TSH and reduction in FT4 levels suggesting that the effects are reversible.

Illness: Thyroid function is impacted by illness and is termed nonthyroidal illness or sick euthyroid syndrome. The typical alterations observed in the acute phase are reduction in serum T3 (both total and free) levels, low or normal circulating T4 (both total and free), while TSH levels remain normal (49). The protracted phase of illness – especially if it is critical – is associated with low levels of all three thyroid-related hormones.

Drugs: Several drugs can affect thyroid function by nature of their iodine content, inhibitory effect on the hypothalamic-pituitary axis, destructive thyroiditis or immune-reactivating mechanisms (see Table 2).

Iodine status and iodine intake including contrast media: Both iodine deficiency and excess can affect TSH and thyroid hormone levels. Severe iodine deficiency can cause hypothyroidism and goitre while mild-moderate deficit can lead to hyperthyroidism due to an increased prevalence of toxic multinodular goitre. Iodine supplementation in an area that was previously iodine deficient can lead to an increase in TSH levels (50). Areas with high iodine intake have higher average and upper-limit of TSH levels (51) – reflecting the importance of iodine status on thyroid function. Iodine containing contrast media as used for CT scans or angiography contains 13,500 mcg free iodine per scan and thus has approximately 100 times more iodine than the recommended daily allowance of 150 mcg (52). Despite this substantial high iodine load remarkably few individuals proceed to develop thyroid dysfunction. In some susceptible persons, however, features of thyroid dysfunction (both hypo- and hyperthyroidism) can occur (53).

Laboratory Factors impacting on TSH and thyroid hormone levels

Inter- and intra-assay differences: Most tests for estimating TSH and thyroid hormones are performed using immunoassays. Currently, most immunoassays for FT4 and free T3 (FT3) exhibit significant biases that exceed intra-individual variability (Table 1) (3). Interference by
proteins (immunoglobulins and paraproteins) are the usual cause of erroneous thyroid function results. Results affected by these interferences and others (Table 2) can only be suspected by requesting clinicians when they are discordant with the clinical scenario.

Decisions about diagnosis and treatment of thyroid dysfunction (2000 words)

i) Diagnosis and treatment of hypothyroidism

a) Making decisions regarding whether hypothyroidism is present

A diagnosis of hypothyroidism can be confirmed by biochemical testing once it is suspected on clinical grounds. Characteristically, serum TSH levels are elevated and T4 concentrations are low. Circulating T3 tends to be in the reference range until profound hypothyroidism develops. On occasions, TSH elevation – particularly when mild or moderate – can be transient and therefore levels should be retested after an appropriate interval (usually at least 6 weeks) before confirming a diagnosis of hypothyroidism. Transient rises of TSH are observed during the recovery phase after nonthyroidal illness or due to certain drugs. Minor elevations of TSH noted in those over age 55 years normalize in approximately 40% of people followed up over an average of 32 months (54).

The other important question that needs to be considered is whether TSH reference ranges impact on the diagnosis of thyroid disease. As noted before, several factors including age, iodine status, race and body mass index impact on the TSH reference range. Using an age-specific TSH range had minimal impact (<2%) in reclassifying a person’s thyroid status, except in the oldest old above the age of 85 years (<5%) in one study (34). Areas that are mild-moderately iodine deficient tend to have lower TSH ranges and thus prevalence rates for hypothyroidism are also lower. For example, in the Study of Health in Pomerania (SHIP), a population-based study of adults residing in North-East Germany which was till recently iodine deficient, the TSH reference range was 0.3 – 3.0 mIU/L (55). Correspondingly, the prevalence of hypothyroidism (both overt and subclinical) was 1.2% whereas low TSH was observed in a much higher proportion (11.3%). These prevalence rates are in stark contrast to thyroid disease prevalence data from areas which are considered iodine sufficient where hypothyroidism is more predominant.

b) Subclinical hypothyroidism
The clinical significance of overt and subclinical hypothyroidism (SC Hypo) depends on the degree of TSH elevation. The debate regarding the long-term health implications of minor TSH elevations has continued for over 4 decades (56). Observational studies differ in their conclusions as to the significance of SC Hypo with cardiovascular outcomes or impairment in quality of life. A meta-analysis of 11 studies with individual data on more than 55,000 participants showed that only TSH elevations above 10 mIU/L were associated with cardiovascular disease (57). One randomised controlled trial in older (>65 years) individuals with persistent SC Hypo (TSH levels between 4.6 and 19.9 mIU/L) concluded that low-dose T4 replacement therapy provides no benefit in improving QoL or thyroid-related symptoms (58). However, one caveat of this study was that the mean pre-treatment TSH values in the experimental and control groups were relatively low at 6.41 and 6.38 mIU/L respectively. Another trial of T4 in older (>65 years) participants with SC Hypo (TSH > 5.5 mIU/L) showed no benefit of treatment on measures of cognitive function (59). These data suggest that T4 treatment of SC Hypo is of no benefit in improving symptoms, quality of life or cognitive function, at least in older individuals. There are no randomised controlled trials of SC Hypo that have evaluated incident cardiovascular events or mortality. However, a retrospective analysis of a General Practitioner database concluded that treatment of persistent SC Hypo (TSH levels between 5.0 and 10.0 mIU/L) was associated with lower incident fatal and non-fatal cardiovascular events only in people aged between 40 and 70 years but a similar benefit was not observed in those who were older than 70 years (60).

Based on these data, in part, the European Thyroid Association formulated guidelines that recommend treatment based on TSH levels (greater than 10 mIU/L across all age groups) and the individual’s age (less than 70 years) (31). Potential factors to consider when deciding whether to observe or treat SC Hypo are shown in Figure 4a and 4b. The factors on each see saw are not weighted according to their relative position on the see saw. In managing older patients with SC Hypo a geriatric multidimensional assessment might also be helpful in decision-making.

c) Overt Hypothyroidism

Overt hypothyroidism is diagnosed when TSH levels are high (usually >10 mIU/L) and T4 levels are low. It is widely accepted that overt hypothyroidism should be treated (61). The only available treatment is thyroid hormone therapy. There is no evidence-based role for other therapies such as nutritional supplements including iodine as the sole management.
strategy in hypothyroidism (61). The aim of therapy is to normalize thyroid function and ameliorate symptoms, if any. This can be achieved by T4 administration (usually in the form of the synthetic LT4). However, approximately 10% of patients on LT4 therapy for hypothyroidism are dissatisfied with their treatment. Combination therapy of T4 and T3 has not been shown to provide any benefit over T4 monotherapy alone in a number of randomised controlled trials (62-65). It is possible that there might be a certain group of patients that could benefit from T4/T3 combination therapy (66), although this needs to be confirmed in more trials conducted prospectively. Current guidelines recommend that T3 treatment should only be provided in a trial environment, and discontinued if no benefit is perceived (67). Debates about how to define euthyroidism, and whether TSH is the best indicator of euthyroidism continue (68).

ii) Diagnosis and treatment of hyperthyroidism

a) Making decisions regarding whether hyperthyroidism is present

As is the case for TSH values that are above the upper limit of the reference interval, many TSH values that are below the lower limit of a generic reference interval may either be normal for that particular individual or may be only transiently abnormal. For, example TSH values tend to be lower in African Americans (69). A TSH calculator has been proposed to calculate reference intervals in individuals of different race and age, and this can be used to compute a lower TSH reference interval for African Americans (70). Smokers tend to have low TSH values than non-smokers (71). With respect to transient TSH abnormalities, studies have examined the impact of both the degree of TSH suppression and the trend over time. Among individuals over 60 years of age with TSH values that were less than 0.5 mIU/L but greater than 0.05 mIU/L, as many as 76% of these had a normal TSH upon follow up after 12 months (72) (see Figure 5). In another study, also examining the natural history of untreated endogenous subclinical hyperthyroidism (SC Hyper), an increasing percentage of patients reverted to normal TSH values over time, such that 38% had normalized their TSH values after 7 years of follow up (73) (see Figure 5).

Another key question, in addition to whether a particular TSH value is actually abnormal or not, is once a laboratory-defined abnormality has been identified, does that “abnormality” have any clinical significance. Several scenarios are possible. An important disease may be
identified and treated, or a risk for a disease may be identified and monitored. Alternatively, the affected individual may actually have no disease at all, but nevertheless there may be personal costs, such as the anxiety and inconvenience associated with additional testing, and there may be financial costs. In a study of the effect of different low TSH cut-offs for reflex testing of FT4, using lower TSH values would have missed only a small percentage of FT4 values above the reference interval and many of these values were only marginally elevated (74). For example, triggering reflex testing of FT4 with a TSH value of 0.3 mIU/l, rather than 0.4 mIU/L in a cohort of 120,403 individuals would have avoided testing in 176 individuals, 76% of whom had minimal FT4 elevations. Whether these cases best represent individuals with disease whose identification has been missed, or avoidance of an unnecessary burden by not identifying patients without real disease, depends upon longitudinal studies of outcomes.

Being cognizant of even mild thyroid status abnormalities is supported by the finding that even lower and higher TSH values that are within the reference range predict future hyperthyroidism and hypothyroidism respectively, although the association was stronger for hypothyroidism (75). Similarly, there is a growing body of evidence that even TSH values that fall within the lower and upper parts of the normal range in euthyroid individuals may be associated with adverse outcomes. TSH values within the lower end of the normal range have been associated with diverse outcomes such as increased hip fractures in older women (76), lower serum cholesterol levels (77), lower body weight (78), and less metabolic syndrome. All-cause mortality has been shown to be either increased (30, 79), decreased (80), or in individuals over 85 years of age, not affected (29) by having low-normal TSH values. Combining multiple datasets of healthy individuals with normal TSH values shows that lower TSH values within the normal range are associated with higher risk of osteoporosis and vertebral fractures, compared with higher TSH values within the normal range (81). Risk of adverse metabolic and cardiovascular outcome may be decreased compared with individuals with TSH values in the upper end of the normal range. TSH values in the upper part of the reference range are associated with lower risk of stroke (82).

Despite the fact that the data about mortality is not consistent, these types of studies illustrate that thyroid function/dysfunction is associated with a continuum of risk and that assignment of specific reference intervals may be an oversimplification of the underlying physiology.

Though such studies of endogenous thyroid function are interesting, it would be an
extrapolation to infer that these data also apply to those being treated for hypothyroidism with LT4, as the T\textsubscript{4}/T\textsubscript{3} ratios also differ between treated and untreated individuals. Therefore, such data also creates a quandary for the physician treating thyroid disease with respect to what particular TSH target is best for each individual patient. Moreover, if a narrower TSH goal such as the midportion of the normal range was desirable, this would represent an even greater challenge with respect to avoiding iatrogenic thyroid disease.

b) Subclinical hyperthyroidism

SC Hyper, by definition, exists when a TSH value below the lower limit of the reference interval is accompanied by a FT4 value within the reference interval. The prevalence of SC Hyper in Europe was 2.91 % (95% CI 2.63-3.21%) in a recent metanalysis (11).

Exogenous SC Hyper

Exogenous hyperthyroidism exists when the cause is treatment with thyroid hormone. Other than intentional suppression of serum TSH in an individual with moderate to high risk thyroid cancer, exogenous SC Hyper represents failure of a treatment program. The percentage of patients with SC Hyper is as high as 13-28% in some studies of LT4-treated patients (72, 83, 84) (see Figure 6), including a study of those over age 65 years (83). The different prevalences noted are likely due to different methodology and populations (e.g. health fair attendees versus longitudinal cohort) but nevertheless illustrate the magnitude of the issue. This is a serious problem, as it is conceivable that the risk of iatrogenic SC Hyper is greater than the risk of the hypothyroidism for which the thyroid hormone was initiated. Risks of suppressed TSH caused by exogenous thyroid hormone extend to dysrhythmias and fractures (85).

Endogenous SC Hyper

Endogenous SC Hyper is most often due to mild Graves’ disease or toxic nodular disease. The latter is less likely to remit than the former (86). Observational studies show multiple risks of SC Hyper including atrial fibrillation (87, 88), hip fractures (89, 90), and dementia (91). Increased cardiovascular mortality (92) and all-cause mortality (93) have been shown in some, but not all studies. Assessment of the risks of treatment compared with monitoring without treatment are generally inferred from these observational studies.
There are few randomised or controlled trials comparing outcomes in treated and untreated patients with SC Hyper, and those that exist are small. For example, in a study of 28 postmenopausal women with toxic nodular goitres with serum TSH values of <0.2 mIU/L, 12 were followed without treatment and 16 were followed after they received radioiodine therapy and achieved a mean TSH of 0.39-0.83 mIU/L. Bone mineral density continued to decline in those not treated, but stabilized in the treated group (94). In another study patients with toxic nodular disease were randomised to methimazole or observation. Ventricular premature beats decreased and a marker of bone density increased in those taking methimazole (95). On the other hand, treatment of SC Hyper in pre-menopausal women with either Graves’ disease or toxic nodular disease did not change the trend for loss of bone density over a 6-month period, compared with untreated controls. TSH values were 0.21-0.23 mIU/L prior to treatment and reached 3.8 mIU/L in the treated group (96).

These studies were too small to perform analyses adjusting for such important factors as baseline TSH, aetiology of SC Hyper, and menopausal status. Only larger studies with the ability to stratify patients based on relevant risk factors such as age, menopausal status, functional status etc will answer the question of whether a personalized “reference interval” or “risk-adapted reference interval” could be used to make decisions about the care of individual patients. A small study of individuals with TSH values ranging from 0.005-0.02 mIU/L showed an improvement in mid-thigh muscle strength 6-9 months after their SC Hyper was treated (97). Thus, a relevant piece of information might be whether a patient is at risk for falls or whether their occupation requires physical activity. Another consideration, for example, in an individual who is at risk of falling or sustaining a fracture might be whether they have a normal serum sodium, with treatment being more strongly considered in a hyponatraemic patient who is already at extra risk of sustaining a fracture (98). Therefore, a rational decision about whether a particular patient would derive benefit from treatment of their SC Hyper is complex and might involve use of a dynamic algorithm that incorporated multiple patient characteristics. Not only would these characteristics be weighted differently in different patients, but also their importance would change over the life span of the patient. Examples of some patient characteristics or other factors favouring treatment and monitoring respectively are shown in Figures 7a and 7b. The factors on each see saw are not weighted according to their relative position on the see saw. Another factor to be considered would be
incorporating the patient in the decision making (99). This could potentially lead to greater patient satisfaction with their choice and better adherence to therapy (100).

All such considerations of the risks and benefits of treatment of SC Hyper assume that treatment will render the patient “euthyroid”. While it is clear that both iatrogenic hypothyroidism and iatrogenic hyperthyroidism while treating hypothyroidism is associated with increased mortality (101), similar data about overtreatment or undertreatment of SC Hyper is not available. It is also not known whether targeting a particular TSH value within the normal range while treating SC Hyper is desirable for optimum control of blood pressure (102) and dyslipidaemia (103), and reduction of metabolic and cardiovascular risk (81). When treating patients with hypothyroidism, targeting a specific TSH does not seem to affect body composition, lipid profile, or energy expenditure (104-106), mood and cognition (107), or satisfaction (108). Although, interestingly, lowering the TSH values of individuals with hypercholesterolemia from the upper to the lower part of the normal range also lowered their cholesterol (109). Studies of targeting particular TSH goals in patients with SC Hyper have not been reported.

c) Overt hyperthyroidism

Overt hyperthyroidism, by definition, exists when a TSH value below the lower limit of the reference interval is accompanied by a FT4 value above the upper limit of the reference interval. In some cases, the low TSH may be accompanied by an elevated T3 (so-called T3 thyrotoxicosis), with the FT4 remaining normal. The benefits of treating overt hyperthyroidism are undisputed (110). The goal of treatment is to normalize the patient’s serum TSH and relieve symptoms of hyperthyroidism. The preceding discussion about what is an ideal target TSH when treating SC Hyper also applies to treating overt hyperthyroidism.

Conclusions

Thyroid hormone has actions in all the cells of the body. Significant perturbations in thyroid function can have manifestations in any organ system, and these effects can be well-compensated and difficult to recognise due to their non-specificity, or can be obvious and classic in their presentation. However, the rationale for treating overt thyroid disease is undisputed. Mild thyroid disease forms a continuum with normal thyroid function and when
present can be associated with only subtle symptoms. The boundaries between normal thyroid function and thyroid disease are generally defined based upon statistically-derived reference intervals. The consequence of this is that thyroid parameters falling outside these ranges, but not proven to have any health consequences may merge with values that may be associated with adverse health outcomes. Moreover, the cut-off between these two classifications may be different for different health outcomes and for different population subgroups or at an individual level. This quandary highlights the need for further research that incorporates randomised controlled trials of the multiple health outcomes of treating or not treating various degrees of thyroid dysfunction. Such trials would have to provide sufficient power to stratify results by such parameters as age and co-existent chronic disease. Such large trials can only be accomplished with multinational efforts. Perhaps groundwork in the form of education and advocacy can pave the way for the funding and commitment to such trials being performed by multiple centres with clinical trial expertise in the future. An alternative approach is for future trials to focus on high risk individuals, such as those at high risk of cardiovascular disease.

References


35. Boucai L, Surks MI 2009 Reference limits of serum TSH and free T4 are significantly influenced by race and age in an urban outpatient medical practice. Clin Endocrinol (Oxf) 70:788-793.


McAninch EA, Rajan KB, Miller CH, Bianco AC 2018 Systemic Thyroid Hormone Status During Levothyroxine Therapy In Hypothyroidism: A Systematic Review and Meta-Analysis. J Clin Endocrinol Metab.


76. Leader A, Ayzenfeld RH, Lishner M, Cohen E, Segev D, Hermoni D 2014 Thyrotropin levels within the lower normal range are associated with an increased risk of hip fractures in euthyroid women, but not men, over the age of 65 years. J Clin Endocrinol Metab 99:2665-2673.


Asvold BO, Bjoro T, Nilsen TI, Vatten LJ 2007 Association between blood pressure and serum thyroid-stimulating hormone concentration within the reference range: a


Table 1. Effect of use of different assays on the reference intervals for FT4 and TSH

<table>
<thead>
<tr>
<th>Assay</th>
<th>Analyte and Units</th>
<th>Lower limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott Architect</td>
<td>FT4 (pmol/L)</td>
<td>10.64</td>
<td>15.53</td>
</tr>
<tr>
<td>Beckman Unicel</td>
<td></td>
<td>7.93</td>
<td>14.00</td>
</tr>
<tr>
<td>Roche Cobas</td>
<td></td>
<td>12.54</td>
<td>19.64</td>
</tr>
<tr>
<td>Siemens Advia Centaur</td>
<td></td>
<td>11.80</td>
<td>18.97</td>
</tr>
<tr>
<td>Abbott Architect</td>
<td>TSH (mIU/L)</td>
<td>0.51</td>
<td>3.67</td>
</tr>
<tr>
<td>Beckman Unicel</td>
<td></td>
<td>0.57</td>
<td>3.60</td>
</tr>
<tr>
<td>Roche Cobas</td>
<td></td>
<td>0.60</td>
<td>4.31</td>
</tr>
<tr>
<td>Siemens Advia Centaur</td>
<td></td>
<td>0.63</td>
<td>4.27</td>
</tr>
</tbody>
</table>

*From Barth et al (5)*
Table 2. Main factors impacting on thyroid function tests in euthyroid individuals

<table>
<thead>
<tr>
<th>Factor</th>
<th>Impact on TSH</th>
<th>Impact on Thyroid hormones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>↑ particularly after birth and after 70 years</td>
<td>→ T4; ↓ T3</td>
</tr>
<tr>
<td>Iodine</td>
<td>Deficiency can cause ↓</td>
<td>→ T4; → or ↑ T3</td>
</tr>
<tr>
<td></td>
<td>Excess can lead to ↑ or ↓</td>
<td>↓ or ↑ T4 and T3</td>
</tr>
<tr>
<td>Smoking</td>
<td>↓</td>
<td>↑ T4 and T3</td>
</tr>
<tr>
<td>Time of sampling</td>
<td>↑ morning; ↓ late afternoon</td>
<td>→ T4; T3 levels mirror TSH</td>
</tr>
<tr>
<td><strong>Assay interference</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacroTSH</td>
<td>↑</td>
<td>→ T4 and T3</td>
</tr>
<tr>
<td>Biotin</td>
<td>↓</td>
<td>↑FT4, ↑FT3</td>
</tr>
<tr>
<td>Heterophile antibodies</td>
<td>↓ or ↑</td>
<td>↓ or ↑ T4 and T3</td>
</tr>
<tr>
<td>Thyroid hormone autoantibodies</td>
<td>→</td>
<td>↓ or ↑ T4 and T3</td>
</tr>
<tr>
<td>Non-thyroidal illness</td>
<td>↓ early phase; ↑ recovery phase</td>
<td>↓ T4 and T3 in acute phase</td>
</tr>
<tr>
<td>Non-thyroidal drugs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amiodarone</td>
<td>↑ common but occasionally ↓</td>
<td>↓ or ↑ T4; ↓ T3</td>
</tr>
<tr>
<td>Lithium</td>
<td>↑ common but occasionally ↓</td>
<td>↓T4</td>
</tr>
<tr>
<td>Glucocorticosteroids</td>
<td>↓</td>
<td>→T4 ; ↓ T3</td>
</tr>
<tr>
<td>Rifampin</td>
<td>↑ or →</td>
<td>↓T4</td>
</tr>
<tr>
<td>Dopamine/agonists</td>
<td>↓</td>
<td>→ or ↓ T4 and T3</td>
</tr>
<tr>
<td>Tyrosine kinase inhibitors</td>
<td>↑</td>
<td>↓ T4 and T3</td>
</tr>
<tr>
<td>Immune modulators</td>
<td>↓ (example, Graves’ disease by Alemtuzumab) or ↑ (example, Hashimoto’s thyroiditis by interferon alfa)</td>
<td>↑ T4 and T3 (Graves’ disease) and ↓ T4 and T3 (Hashimoto’s thyroiditis)</td>
</tr>
<tr>
<td>Obesity</td>
<td>↑</td>
<td>→ or ↓ T4; ↑ T3</td>
</tr>
<tr>
<td>TPOAb +ve status</td>
<td>↑</td>
<td>↓ T4, → or ↓ T3</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>↓</td>
<td>↑TT4, ↑TT3, ↓FT4, ↓FT3</td>
</tr>
</tbody>
</table>

→ = unchanged, ↑ = increased, ↓ = decreased

T4/T3 denote both total and free thyroid hormones unless otherwise specified, TT4 = total T4, TT3 = total T3, FT4 = free T4, FT3 = free T3
Euthyroid 95%

Subclinical hyperthyroid 2.5%

TSH reference range

Subclinical hypothyroid 2.5%

With use of the 95% reference interval definition, 5% of individuals will fall outside of the “euthyroid range”
Figure 2. TSH – thyroid hormone relationship
Figure 3. TSH reference interval by various age groups

The reference interval is also affected by co-existent illness, type of assay used, iodine status, time of sampling, ethnicity.
Figure 4. Factors based on expert opinion favouring (a) observation and (b) treatment of subclinical hypothyroidism. All factors shown favour observation and are not weighted according to their position on the seesaw. All factors shown favour treatment and are not weighted according to their position on the seesaw.
Figure 5. The natural history of untreated hyperthyroidism

**Effects of time:**
Patients with endogenous SC hyper and TSH values 0.1-0.4 mIU/L (based on Vadiveloo et al., 2011)

- TSH normalized at 2 year follow up in 17%
- TSH normalized at 5 year follow up in 32%
- TSH normalized at 7 year follow up in 38%

**Effects of initial TSH value:**
One-year follow up testing in patients over 55 yrs with low TSH values (based on Parle et al., 1991)

- Of those with TSH values of 0.05-0.5 mIU/L, 76% normalized
- Of those with TSH values less than 0.05 mIU/L, 13% normalized
- Of all patients with low TSH values, 61% normalized
Figure 6. Prevalence of iatrogenic thyroid disease among those treated for hypothyroidism

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Somwaru</th>
<th>Canaris</th>
<th>Parle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperthyroid</td>
<td>9.7</td>
<td>0.9</td>
<td>7.0</td>
</tr>
<tr>
<td>SC Hyperthyroid</td>
<td>27.8</td>
<td>20.7</td>
<td>13.0</td>
</tr>
<tr>
<td>Euthyroid</td>
<td>46.0</td>
<td>60.1</td>
<td>53.0</td>
</tr>
<tr>
<td>Hypothyroid</td>
<td>2.1</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>SC Hypothyroid</td>
<td>14.4</td>
<td>17.6</td>
<td>26.0</td>
</tr>
</tbody>
</table>
Figure 7. Factors based on expert opinion favouring (a) observation and (b) treatment of subclinical hyperthyroidism.

(a) Graves disease?
- Low or mid-range FT4 and T3?
- No Symptoms
- TSH concentration > 0.1 mIU/L
- Pre-menopausal or taking estrogen
- Age < 65 yrs

Already taking a beta adrenergic blocker?
- Normal bone density
- No heart disease
- No cognitive impairment

Favour Observation

(b) Cognitive impairment
- Heart disease
- Decreased bone density
- Toxic nodular disease
- High normal FT4 or T3 levels?
- Symptoms
- Post-menopausal or estrogen deficient
- Age > 65 yrs

Favour Treatment

All factors shown favour observation and are not weighted according to their position on the see saw.

All factors shown favour treatment and are not weighted according to their position on the see saw.
The authors would like to thank the authors for providing their time and expertise. We believe that the suggested changes have improved the manuscript. The changes made to the manuscript are indicated in tracking changes.

Reviewer #1:
Comment 2. I would like to suggest that the authors consider a table giving examples of the variation in analytes as measured by different commercial assays.

Response 2
Lines 157-158
A table, referred to in lines 157-158, has been added as table 1 showing the variations in FT4 and TSH according to selected commercial assays.

Comment 3. Table 1. Readers should know whether impact on thyroid hormones is on total T4 and FT4 or just one of these, especially as total T4 is used more in USA and FT4 in Europe.

Response 3
Table 2: The table has been amended to indicate which thyroid hormone metabolite (total hormone or free fraction) is affected by the variables in question. In addition, a section on laboratory interferences impacting thyroid function has been added.

Comment 4. Fig 2 The intervals will also depend on the TSH assay

Response 4
Lines 155-158
Some of the other factors affecting the reference intervals (lines 155-158) as discussed in the manuscript text have been added to the figure legend also.

Comment 5. Figs 3 and 6. These could be confusing. eg in fig 3 because the items are on the right hand side of the see saw does normal lipid profile suggest that observation should not be favoured? I do not think so but for non thyroidologists this should be clarified.

Response 5
Lines 355-358 and 495-496
A legend has been added to figures 4 and 7 (old figures 3 and 6) to explain that the factors illustrated are not weighted according to their position on the see saw.

Comment 6. Fig 5 There should be discussion of the variation between authors in the prevalence.

Response 6.
Lines 455
Some examples of reasons for the variation in prevalence have been added to the text.

Comment 7. line 75 their

Response 7
Line 80-81
The “their” is intended to refer to the respective reference intervals for TSH, T4 and T3. The word respective has been added to clarify.

Comment 9.line 220 yes there is a huge amount of iodine in contrast agents but the dissociation constant of I from these substances means that there is much less iodine delivered to the system than the concentration in the agent

Response 9.
Line 282-292
The free iodine content per CT scan has been clarified.

Comment 10. line 231 circulating thyroid...

Response 10
Line 241
The word circulating has been added for clarity
Comment 11. line 246 please indicate exactly which moieties eg FT4 and or total T4 etc

Response 11

The table (now table 2) has been modified to indicate which moieties are affected. In addition, the section in the Main text (now line 271-275) has clarification added to which thyroid hormone measurements are affected by non-thyroidal illness.

Comment 12. line 315 et seq. The sentence is confusing eg treatment based on TSH levels less than or greater than 10. Please explain more clearly.

Response 12.

These sentences have been made clearer with respect to the General Practitioner Database study as well as the European Thyroid Association guidelines based on age and TSH levels.

Comment 13. line 470 In T3 toxicosis the T4 concentrations are within the reference intervals.

Response 13.

Mention of T3 thyrotoxicosis has been added.

Comment 14. line 485 are generally ...

Response 14

The sentence has been corrected to state “are” instead of “is”.

Comment 15. In the conclusions the authors indicate the need for further trials but perhaps also allude to the difficulties in conducting these trials particularly in relation to their required power. They should therefore discuss other strategies (eg advocacy etc) to further their recommendations.

Response 15.

Some potential strategies are now mentioned.

Reviewer #2:

Comment 1. Terminology - subclinical vs. overt disease

1a) Throughout the manuscript the authors loosely use the term “mild” (pg 4, line 118; pg 4 line 120; pg 15, line 483) or “borderline” thyroid disease (pg 3, line 68). Since subclinical hyperthyroidism and subclinical hypothyroidism later on receive their own dedicated sections within the manuscript, it is unclear whether “mild degrees of hypothyroidism” refers to mild forms of overt disease, subclinical disease, or in fact, is taken to mean the lower limit of euthyroid. The authors should specify when it is overt/subclinical/euthyroid throughout the text, to avoid confusion. For example, (pg 2, line 65) "the prevalence and incidence of ...hypothyroidism is increasing, ...due to a fall in the threshold level of TSH at which thyroid hormone therapy for hypothyroidism is being initiated". This sentence may need rephrasing. Does the statement refer to prevalence of overt hypothyroidism, or is it in fact subclinical hypothyroidism that is being increasingly treated?

Response 1a

It is not always possible for the authors to be as specific as they or the reviewer would like. Many of the studies cited do not have information about the FT4 level accompanying the TSH value. For example in the study by Taylor et al (reference 1), FT4 levels were only available for 66% of the individuals and of those 31% had subclinical hypothyroidism. This is acknowledged in lines 70-71, and line 128.

1b) Furthermore, initiation of therapy does not always equate to diagnosis of disease and a true increase in disease prevalence within the population. Especially since some newly identified subclinical hypothyroid persons are started
on levothyroxine, based purely on blood results and without convincing disease-specific symptoms. This issue needs to be clearly addressed otherwise overdiagnosis and overtherapy cannot be excluded, especially given the negative findings (for both primary and secondary outcomes) of the largest trial of levothyroxine benefit to date (Stott et al., NEJM 2017)1.

Response 1b
Lines 71-76
The authors agree and have made this point in line 71-76.

Comment 2. No benefit of thyroid hormone therapy
I disagree with the statement "the evidence of benefit is unclear" for "patients with borderline disease" (page 3, line 68). Both the TRUST RCT (NEJM, 2017)1 and a recent meta-analysis including all 21 RCTs (JAMA, 2018)2 have demonstrated that levothyroxine administration is of no benefit in improving Quality of Life or thyroid-related symptoms in subclinical hypothyroidism, as well as for several other outcomes. In the manuscript the authors acknowledge the results of the TRUST RCT (page 10, line 302), but offer a caveat for the finding of no levothyroxine benefit - by noting that the mean pre-levothyroxine TSH values in both treatment and placebo groups were only mildly deranged from the onset (6.41 and 6.38 mIU/L respectively). However, these are the most common TSH values in adults with subclinical hypothyroidism, and in the JAMA meta-analysis the baseline mean TSH values ranged from 4.4 to 12.8 mIU/L prior to levothyroxine administration, with two studies (99 participants) who had a mean baseline TSH >10mIU/L. Therefore, both papers showed no benefit of levothyroxine on the outcomes of Quality of Life or thyroid-related symptoms. I suggest major revision of the use of "unclear benefit of levothyroxine", with a complete description of current data.

Response 2
Line 74
The recent metanalysis by Feller et al has now been added.

Comment 3. TSH limits
(Pg 3, line 88): "reduction in the lower limit of the TSH level to 0.3-0.4 mIU/L". Studies published in high impact journals use 0.45 mIU/L (JAMA3, Circulation4) and 0.40 mIU/L (NEJM1, US Preventive Services Task Force5) as the lower limit cut-off for euthyroidism. I suggest removing 0.3 mIU/L, as this is not based on guidelines, nor on large cohorts which clearly show increased risk of clinical outcomes with 0.3mIU/L.

Response 3
Line 95
This has been changed.

Comment 4. Assessing patients on an individual case-by-case basis
This review stresses the importance of assessing patients on an individual case-by-case basis and the need to regularly re-assess the clinical and laboratory picture, in order to tailor treatment and target-levels accordingly. Although this seems wise advice, this statement is not evidence based. In the largest trial (Stott et al., NEJM1), no benefit of levothyroxine therapy could be identified for any of the subgroups. This statement should be removed from the abstract or at least updated; for example "based on current trial data, it is unclear whether certain adults would still benefit from levothyroxine, (such as those with TSH > 10 mIU/L), since a limited number of these persons were included in RCTs". Similarly in Fig 3, it should be clearly acknowledged that several of the factors listed as favouring treatment are not based on RCTs (e.g positive antibodies, hyperlipidaemia, age <65 years, TSH 7-10mIU/L), but are based on expert opinion - at least until an RCT demonstrates otherwise, that subgroups would gain more clinical benefits.

Response 4
Lines 42-45
The authors have added the suggested statement to the abstract and the acknowledgement of the factors being employed in figure 4 (formerly figure 3) to be based on expert opinion.

Comment 5. Age-specific thyroid reference ranges
(Pg 6, line 194): It is correct that "there is no consensus whether age-specific thyroid reference ranges should be utilized", but it should be more carefully discussed. The US Preventive Services Task Force in their published final recommendations statement for thyroid dysfunction screening do indeed identify a knowledge gap and reinforce that age-specific ranges may be needed5. (Pg 9, line 285) referenced Kahapola-Arachchige et al., (Clin Endocrinol, 2012) which demonstrated that in their large sample (N = 148,938), increasing age was associated with an increase in median and 97.5th percentile TSH levels. However, using age-specific TSH ranges had minimal impact in reclassifying a person's thyroid status, except in the oldest old participants >85years. I would like to draw the authors'
attention to the TEARS Study published afterwards in 2013; featuring a large population-based cohort (N = 153,127). In keeping with previous published data from large cohorts in the United States (NHANES III, 7, 8), and Australia (Busselton Health Survey), the TEARS Study demonstrated a significant increase in median TSH with increasing age. It demonstrated that the use of age-specific TSH reference ranges, especially in persons >70 years, would result in the reclassification of many "abnormal" results to "normal," and thus preventing significant overestimation of subclinical hypothyroidism.

A more careful discussion on age- (as well as race-) specific reference ranges for TSH would add to this article, in order to prevent significant over- and under-estimation of subclinical thyroid disease.

Response 5

Lines 211-218

The authors have added some discussion as requested. The references from Boucai and Surks have already been incorporated.

Comment 6.  Sex and TSH distribution

The authors should clarify their viewpoint on effects of sex on TSH levels:

(Pg 6, line 173): "TSH values are higher for men than women across a range of FT4 values".

(Pg 6, line 196): "women tend to have higher TSH levels".

The latter excerpt is based on the NHANES III study. However, the more recent and larger population-based TEARS study contradicts this and reports significantly higher median TSH levels for males compared to females, though the difference is minimal (median TSH = males 1.72 vs females 1.70 mIU/L; p < 0.001)*

Response 6

Lines 220 and 221-223

The authors have attempted clarification as follows. For a specific FT4, men tend to have higher TSH values in a study from Australia. However in a study from the United States, as a group women tend to have higher TSH values because of autoimmune. Both these observations are clarified in the manuscript text.

Comment 7.  BMI and TSH levels - direction of causality relationship

Hypothyroidism is known to be associated with weight gain. However, the direction of the causality arrow is controversial. (Pg 8, line 238): In keeping with the manuscript, I am aware that novel view suggests that BMI/obesity (cause) may lead to an increase in TSH level (effect), potentially via increasing leptin levels and increasing susceptibility to thyroid autoimmunity. However, I would like to point out that the literature is conflicting and that many studies exist still supporting the traditional stance where the causality arrow is reversed and it is indeed changes in the thyroid gland that drives BMI changes. For example, administering levothyroxine to an overtly hypothyroid patient can induce weight loss. Similarly, treating overt hyperthyroidism leads to significant weight gain. Since the literature is conflicting and there is no clear consensus on mechanism of action, I suggest that the authors remain neutral and rephrase "it is likely that…" (pg 8, line 239).

Response 7

Lines 247-251

This statement has been rephrased and an additional reference has been added.

Specific comments - Text

Comment 8*  (Pg 5, line 134): For "non-Gaussian distributions, so the reference interval is obtained by including values that fall between the 2.5th and the 97.5th ranked percentiles. This method of defining reference intervals means that 1 in 20 (5%) apparently euthyroid individuals will have values outside this range". The authors then go on to later mention another approach where optimal ranges are determined according to health risk, which seems more relevant to patients. (However, as the authors already point out, this method is difficult to use since both subclinical hypothyroidism and subclinical hyperthyroidism are related to an increased risk of adverse outcomes). Consequently, it is unclear what point is being made about the effect of non-Gaussian distributions on reference intervals, and this should be clarified.

Response 8

Lines 147-151

The impact of potential factors such as unrecognized autoimmune thyroid disease on the shift to the right in the TSH distribution range has now been added.

Comment 9*  (Pg 7, line 228): consider adding that trimester-specific reference ranges for TSH during pregnancy and the post-partum interval have been suggested.

Response 9
This has been added as requested.

Comment 10*
(Pg 8, line 256): "seasonal changes in thyroid function could be attributed to variation in iodine intake". Iodine intake is known to influence the secretion of thyroid hormones. However, in this day and age of iodine fortification of food, the impact of any seasonal variation in dietary iodine intake is attenuated. (World Health Organization access to iodized salt is a global public health strategy. Cow's milk is rich in iodine due to dairy-farming practices - iodine fortified concentrates given to cattle particularly in the winter months14). Nevertheless, I acknowledge that the UK has no national iodine fortification programme as in other countries, and so is heavily reliant on dairy products for iodine intake - posing a problem for those unable to drink cow's milk. I suggest revising this statement, since geographical or socioeconomic factors may play a greater role in variations in iodine intake than seasonal changes.

Response 10
Line 257
The study cited refers to Scandinavia and the statement has been modified to state that this relationship may only be seen in some countries.

Specific comments - Tables & Figures
Comment 11*
Table 1: BMI - arrows for increased TSH, increased T3, equivocal T4. Firstly, should it not instead depict increased TSH with lower T4 & lower T3?

Response 11
Line 247-251
The table (now Table 2) has been amended. Body mass index has been changed to obesity to better convey the effects of increased body mass on thyroid function. As far as the effects of obesity on the various thyroid function parameters is concerned, the data is convincing for a direct positive relationship between BMI and TSH as well as T3 (both FT3 and TT3). However, the relationship between T4 and BMI is conflicting, with the majority of studies, however, suggesting no effect or a slight reduction. We have made this clearer in the section on body mass index (line 247) as well as in Table 2.

Comment 12*
Figure 2: Typo. Please correct the units on your scale bar for TSH reference range from "mU/L" to "mIU/L"

Response 12
The units on the scale bar for TSH have been corrected.

Comment 13*
Figure 6: Consistency: American vs British spelling. Throughout the manuscript (text & figures) American English spellings have been employed. However, the Figure 6 title uses "favoring" (American spelling), but the graphic uses "favour observation" (British spelling).

Response 13
Thank you. The manuscript has been corrected to use British spelling, in keeping with Lancet D&E requirements

Reviewer #3:

Comment 1. Lines 81 and 82: you might mention measurement of serum cholesterol and the ankle jerk reflex time.

Response 1
Line 88
These additional parameters have been added to the description of end organ parameters

Comment 2. Line 11 and 118: the trend for increased prescribing thyroid hormone can also be related to on increased thyroid surgery due to the well known "thyroid cancer epidemic"

Response 2
Line 123-126
Increasing cases of post-surgical hypothyroidism have now been mentioned

Comment 3. Lines 133ff: yes, it is true that TSH distribution is non-Gaussian. You might mention other analytes that are also nongaussian in their distribution, including prostate specific antigen. Also, since you say in line 136 that one
in 20 "apparently euthyroid individuals" will have a serum TSH outside the reference range, widens figure 1 label people in these ends of the curve as having "subclinical" thyroid disease? Also, it might be worthwhile for you to enumerate potential reasons for the nongaussian distribution, especially the shift to the right in the curve. Potential explanations, as you know, include mutations in the TSH receptor, altered biological activity of TSH, subclinical autoimmune thyroid disease, etc. Also, you do not mention macro TSH as a cause of falsely elevated serum TSH, or biotin as a cause of laboratory interference in certain TSH assays.

Response 3

Lines 147-151

The consequence of use of the 95% definition has been added to the legend of figure 1. Further discussion of the non-Gaussian distribution has been added to the manuscript text as requested in lines 147-151.

Comment 4. Section "Relationship between TSH and thyroid hormones": consider a figure to illustrate the relationship between TSH and free thyroid hormones for easier reader comprehension.

Response 4

Line 172

This has been added as a new figure 2.

Comment 5. In line 173, you state "TSH values are higher for men than women across a range of free T4 values", but in line 196 you state that women have higher TSH levels than men. Which is it?

Response 5

Lines 220 and 221-223

Both these statements are correct. For a specific FT4, men tend to have higher TSH values in a study from Australia. However in a study from the United States, as a group women tend to have higher TSH values because of autoimmunity. Both these observations are clarified in the manuscript text.

Comment 6. Line 225 TPOAb status: elevated TSH is a consequence of chronic autoimmune thyroiditis (revealed by the presence of TPOAbs) and should not be considered as other physiologic conditions described in this section of the review (e.g. age, gender, circadian and seasonal variations). Consider listing the "Factors impacting on TSH and thyroid hormone" in subgroups: e.g. physiologic conditions, pathologic conditions unrelated to true thyroid disease, laboratory factors.

As for the latter group, consider introducing information on possible interfering agents/conditions.

A more clinical section on how to deal with possible interference (e.g. change assay, use serial dilution) or possible confounding conditions (e.g. for obese patients should TSH testing be repeated if slightly elevated values are found?) could be of interest.

Response 6.

Line 196-301

This section has been re-orders into i) physiologic, ii) pathologic, and iii) laboratory factors as requested by the reviewer. A section on dealing with laboratory interferences is precluded by the word limit requirement of the review.

Comment 7. In line 252ff you talk about seasonal variation in TSH, being higher in the winter than in the summer. You then say that this observation is "not consistent". Why do you say that?

Response 7

Line 255-256

T3 levels are generally higher in the winter, but the reports of TSH and FT4 changes with different seasons have been different in different studies.

Comment 8. In this paper, you place all of the focus on serum TSH values as a reason to either treat or not to treat patients. It would be reasonable, I think, to include data showing that TSH does not correlate with a variety of clinical outcomes, whereas levels of free T4 and/or free T3 do: for example, Zulewski et al (JCEM 1997) found no correlation of hypothyroid symptoms to TSH levels but did find a correlation with free T4 and free T3. Also, normalizing serum TSH values does not necessarily normalize thyroid hormone-sensitive serum analytes: for example, SHBG (Alevizaki et al. Wein Klin Woch. 2005) or carotid intimal thickness (Clausen et al. Clinical Endocrinology 2009;70:932).

Response 8

Lines 374-375

The reviewer brings up some complex issues regarding how to define euthyroidism which space constraints do not allow us to address. A reference that addresses this issue somewhat has been alluded to in lines 374-375.
Comment 9. Line 278: reference 48 applies only to people over age 55

Response 9

The age of individuals in the cited study has been added.

Comment 10. The discussions of overt hyperthyroidism and hypothyroidism are really not relevant for this paper.

Response 10

The authors believe these topics are relevant, as it is the reference intervals in use that will determine whether overt or subclinical disease is present. For example, with a FT4 assay that has a lower upper limit, a case of subclinical hyperthyroidism may become overt hyperthyroidism.

Comment 11. In line 370ff you discuss the fact that many studies have shown morbidities associated with serum TSH levels within the reference range, and how this might relate to TSH targets in patients treated with levothyroxine. Yet, you provide little guidance on how this information might be used clinically. For example, some experts have suggested that TSH targets should be higher in elderly patients, as you know. Also, what is your recommendation for managing a patient with, let’s say a FT4 of 1.6 ng/dl and a TSH of 4? Or a patient with a FT4 of 0.8 ng/dl and a TSH of 0.4? Which is more important, the FT4 or the TSH?

Response 11

These data about TSH values and their association with outcomes are based on epidemiologic studies in individuals not treated with thyroid hormone. The authors did not intend to infer that these data could be directly translated into treatment goals for treating individuals with thyroid disease. Clarification has been added in lines 431-433.

Comment 12. Line 463: it’s one thing to say “treat patients, not numbers” but to cite a study showing improved “cognition” in persons with normal serum TSH values after being treated with methimazole is a bit irresponsible, I think. As you know, the study was not randomized or blinded, the 7 subjects that received methimazole were younger than those who did not, and the treatment group had a shorter follow-up period. Also, the results were not statistically significantly different. I do not believe that you think will want to propose that any worrisome or bothersome symptom or sign could potentially be treated with thyroid hormone or methimazole, to change the TSH in a direction that seems favorable.

Response 12

The authors agree and have removed the reference to this study.

Comment 13. Table 1: it is unclear why you have "antiepileptic agents" listed as a cause for increased serum TSH and low "thyroid hormones". It was my impression that these drugs cause a falsely low free T4 and most assays including equilibrium dialysis using diluted serum (Surks et al. JAMA 1996;275:1495). You also show in Table 1 that "pregnancy" has an increased impact on "thyroid hormones". This is ambiguous: you mean it has an impact on thyroid hormones in general (increased or decreased), or the thyroid hormone levels go up? The use of the arrows makes things confusing.

Response 13

Rifampin has been used as an example instead. The impact of pregnancy on T4 and FT4 has been clarified. An explanation of the meaning of the arrows has been added as a legend to table 2 (previous table 1).

Comment 14. Figure 4: again, the Parle paper (your reference 66 I believe) was only in “the elderly”.

Response 14

The age of the patients studied has been added to the figure (now figure 5).

Comment 15. Figure 5: Please state more clearly that each flow chart refers to an original article (for example by introducing “et al.” and the year of publication.

Response 15

The figure has been modified to state that the data is based upon the study cited and the year of publication has been added to the figure.
Comment 16
Figures 3 and 6: The meaning of the text on each side of the balance is not immediately clear. It looks like each side of the figure is a “scale” when, in fact, the “scale” is really between the two panels in the figure.

Response 16
The reviewer is correct that the comparison is between the two different panels. A legend has been added to the figures (now figures 4 and 7) to clarify that the position of the parameters on each see saw does not indicate their relative weight.

Reviewer #4:

Comment 1. Lines 121-124: in recent years treatment of subclinical hyperthyroidism seems to be more frequently prescribed, especially in older patients with cardiovascular disease.

Response 1
Lines 134-136
Acknowledgement of considerations for older individuals with cardiovascular disease has been added.

Comment 2. Lines 188-194: beside serum TSH cut-off level the presence of a frail phenotype and comorbidities such as heart failure should be taken into account in the clinical workup.

Response 2
Lines 209-210
The frail phenotype has been acknowledged.

Comment 3. Figure 3a and 3b: In managing older patients with subclinical hypothyroidism (and, perhaps, with subclinical hyperthyroidism) a geriatric multidimensional assessment should be useful.

Response 3
Lines 357-358
A comment about such an assessment has been added.

Comment 4. Line 488: “Morevover” should be Moreover.

Response 4.
Line 535
This mis-spelling has been corrected.