

## Guidance in Subclinical Hyperthyroidism and Subclinical Hypothyroidism: Are We Making Progress?

Wilmar M. Wiersinga

Department of Endocrinology and Metabolism, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

The ETA guidelines on subclinical hyperthyroidism (SHyper) in the present issue of *European Thyroid Journal* [1], together with the previously published ETA guidelines on subclinical hypothyroidism (SHypo) [2, 3], offer up-to-date recommendations on the management of subjects with subclinical thyroid dysfunction. Guidance in this field is most welcome because of continuing uncertainty whether or not therapeutic intervention will improve health outcomes. Although the evidence of associations between SHyper or SHypo and adverse health outcomes has become much stronger in the last decade, evidence is lacking that restoration of the euthyroid state reverses the risk of adverse health outcomes. There are no long-term randomized clinical trials demonstrating that treatment will do more good than harm [4]. Against this background, one may wonder whether the grades of evidence attached to some of the recommendations are not overrated. Nevertheless, the guidelines could be very helpful in making treatment decisions. In this editorial, however, I would like to explore the question if we are really making progress in our thoughts about SHyper and SHypo. In other words, which topics have not been addressed by the present guidelines? Are there less prominent but still clinically relevant issues?

### 'Subclinical' Is a Misnomer and Should Be Replaced by a Grading System

SHyper and SHypo are misnomers because the term 'subclinical' suggests the absence of symptoms and signs of thyroid hormone excess or deficiency, respectively. Such symptoms and signs, however, can sometimes be present, e.g. atrial fibrillation is a well-known manifestation of thyrotoxicosis and its prevalence is increased in SHyper [5]. Also, subjects with SHypo score slightly higher than controls on a clinical scale for hypothyroidism [6]. The confusing term 'subclinical' should thus better be avoided [7], and a more accurate terminology is required. SHyper and SHypo are defined exclusively by biochemical criteria (TSH outside but FT<sub>4</sub> and FT<sub>3</sub> within their respective reference ranges). Evered et al. [8] proposed 40 years ago to grade hypothyroidism along biochemical criteria. They distinguished between grade I (subclinical), grade II (mild), and grade III (overt) hypothyroidism (table 1) [8]. TSH becomes progressively higher and FT<sub>4</sub> progressively lower in the transition from grade I to grade III. The grading system of Evered et al. has not been adopted by the medical community, but in my view has lost none of its attractiveness. It might

**Table 1.** Proposed grading of hypothyroidism and hyperthyroidism

	TSH	FT <sub>4</sub>	FT <sub>3</sub>
Hypothyroidism grade IA	increased, >4.0 to <10 mU/l	normal	normal
Hypothyroidism grade IB	increased, ≥10 mU/l	normal	normal
Hypothyroidism grade II	increased	decreased	normal
Hypothyroidism grade III	increased	decreased	decreased
Hyperthyroidism grade IA	decreased, >0.1 to <4.0 mU/l	normal	normal
Hyperthyroidism grade IB	decreased, ≤0.1 mU/l	normal	normal
Hyperthyroidism grade II	decreased	normal	increased
Hyperthyroidism grade III	decreased	increased	increased

Normal, increased, decreased: values within, above, below respective reference intervals.

have become even more relevant in view of current guideline recommendations to prescribe levothyroxine in SHypo with TSH values of  $\geq 10$  mU/l (at least in subjects  $\leq 70$  years), but to be more conservative at TSH values between 4 and 10 mU/l [2]. My proposal would be to subdivide grade I (SHypo) into grade IA (TSH  $>4.0$  to  $<10$  mU/l) and grade IB ( $\geq 10$  mU/l). The same grading system may be applied in hyperthyroidism: grade III would indicate overt hyperthyroidism, grade II would indicate mild hyperthyroidism or T<sub>3</sub> toxicosis, and grade I would indicate SHyper (table 1). Again, in view of the higher risk of adverse health outcomes at TSH values  $\leq 0.1$  mU/l in comparison with TSH values  $>0.1$  to  $<0.4$  mU/l and the current recommendation to intervene when TSH is  $\leq 0.1$  mU/l [1], grade I could be subdivided into grade IA (TSH  $>0.1$  to  $<0.4$  mU/l) and grade IB (TSH  $\leq 0.1$  mU/l). The authors of the present SHyper guidelines already apply a grading system, which – although slightly different from my proposal – is welcome as the grading is more comprehensible than terms of ‘low’ or ‘suppressed’ values.

Classifying both hyperthyroidism and hypothyroidism in grades (IA, IB, II, III) provides a rather accurate estimate of the severity of the condition. The grading system follows the natural history of both conditions (starting with grade IA and often ending with grade III), and the reverse sequence of grading occurs after initiating treatment to restore euthyroidism (with normalization first of FT<sub>4</sub>, then of T<sub>3</sub>, and lastly of TSH in case of hyperthyroidism, and with normalization first of T<sub>3</sub>, then of FT<sub>4</sub>, and lastly of TSH in case of hypothyroidism) [9]. The four international thyroid associations may consider installing a committee to examine whether the grading system has sufficient advantages to adopt it universally.

### The Presence of Thyroid Disease Should Be an Additional Criterion for the Diagnosis of SHyper or SHypo

The current definition of SHyper and SHypo is by biochemical criteria only and does not stipulate the biochemical abnormality should be related to thyroid disease. Although the vast majority of TSH values outside the reference range in the presence of normal FT<sub>4</sub> and T<sub>3</sub> values are caused by thyroid disease and the sequela of its treatment, the same biochemical constellation may occur in conditions not related to thyroid pathology. Examples are interference in TSH assays by heterophilic antibodies, glucocorticoid excess or deficiency, nonthyroidal illness, and obesity. Thus, what we call subclinical thyroid dysfunction is sometimes caused by altered regulation of the hypothalamus-pituitary-thyroid axis and is not related to diseases of the thyroid gland itself. For example, a slightly elevated TSH in obese subjects will normalize upon weight reduction, and there is no evidence that under these circumstances levothyroxine will be helpful [10]. I would abstain from treating SHyper with radioactive iodine or antithyroid drugs if there is no positive proof the low or suppressed TSH is caused by thyroid pathology. Similarly, I would start levothyroxine treatment of SHypo with more confidence if I know thyroid pathology is present. Autoimmune thyroiditis is the most common cause of SHypo, but TPO and/or Tg antibodies are not detectable in about 20% of cases. In such patients, thyroid ultrasonography may provide early evidence for thyroid autoimmunity [2]. Ideally, thyroid disease should be demonstrated, but nonthyroid-related causes should be excluded in all cases. One may thus consider adding the presence of thyroid pathology as another criterion for the diagnosis: SHyper and SHypo are defined by an abnormal TSH (in the presence

of normal free thyroid hormones) which is related to thyroid disease. It would facilitate treatment decisions because a subject with an abnormal TSH that is not related to thyroid pathology is unlikely to benefit from treatment directed against thyroid hormone deficiency or excess.

### Limitations of TSH Reference Ranges

Application of reference ranges to determine whether or not a given TSH value is abnormal is not as straightforward as it looks. 'What is normal?' is almost a philosophical question, and its answer (that what is not abnormal) is fraught with difficulties. My favourite quote on this issue is from Benson [11]: 'The normal range has a vague but comforting role in laboratory medicine. It looms on the horizon of our consciousness, perfectly symmetrical like a Mount Fujiyama, somewhat misty in its meanings, yet gratefully revered and acknowledged. Far from being pure and simple, however, like a cherished illusion of childhood, on close examination it proves to be maddeningly complex and is indeed one of the most stubborn and difficult problems limiting the usefulness of clinical laboratory data.' The population-based NHANES III Survey in the USA has been a hallmark study for establishing reliable TSH reference ranges [12]. In their so-called 'reference population', the median TSH was 1.39 mU/l with a reference interval of 0.45–4.12 mU/l (P2.5–P97.5). However, a clear age-dependent effect on TSH values was observed: median TSH values and their reference ranges in the age groups 20–29, 60–69, 70–79, and  $\geq 80$  years are 1.26 (0.40–3.56), 1.67 (0.49–4.33), 1.76 (0.45–5.90), and 1.90 (0.33–7.50) mU/l, respectively [12]. In view of the higher TSH values with advancing age, the prevalence of SHypo may thus be significantly overestimated unless an age-specific range for TSH is used [13]. The guidelines do not propose age-specific reference ranges (which I would have found quite logical), but recommend a very conservative attitude in prescribing levothyroxine in subjects with SHypo of 70 years and older, but based on other considerations than the upper normal limit of TSH that increases with age [2].

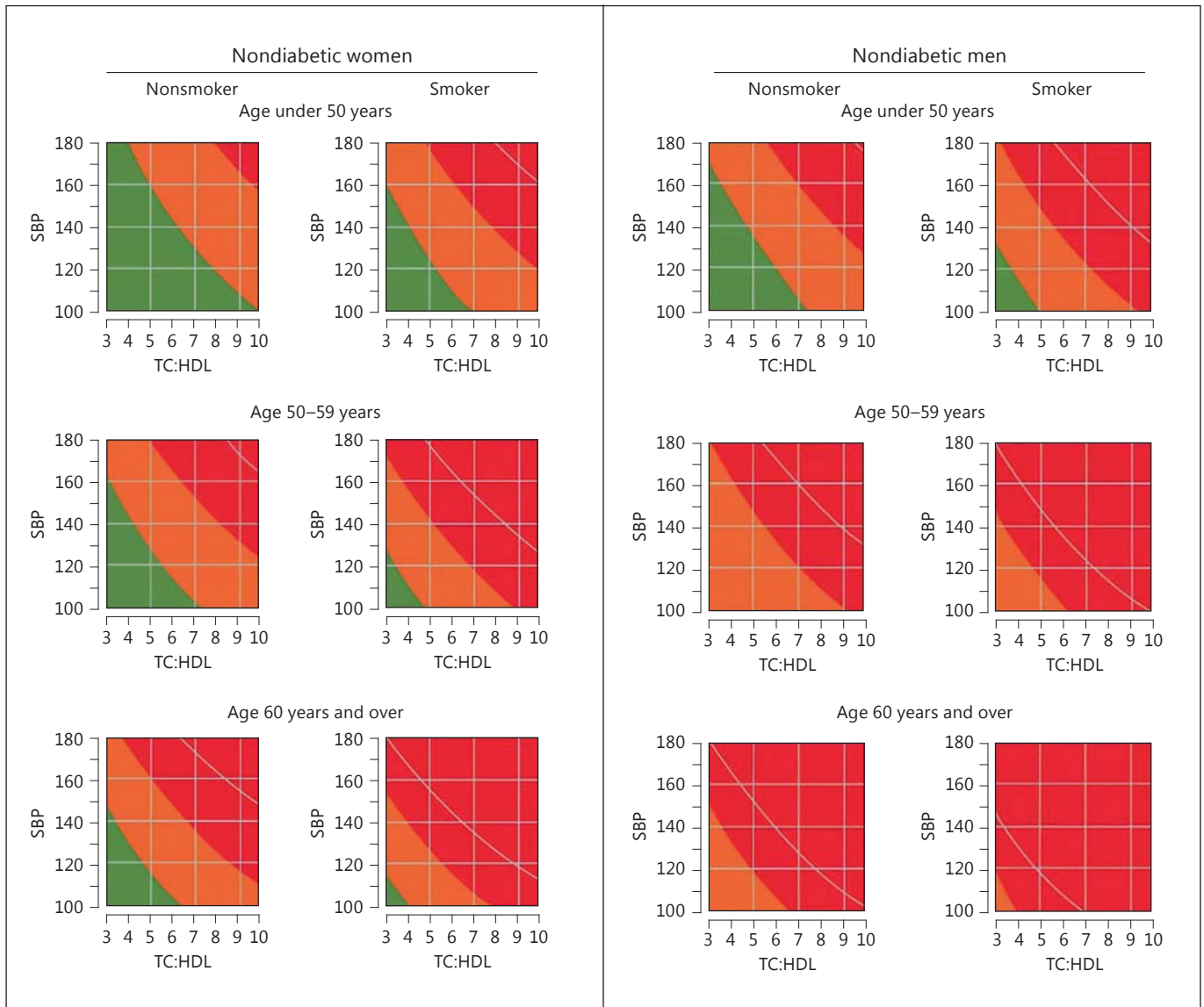
In contrast, the need for trimester-specific TSH reference ranges in pregnant women is widely recognized [3, 14]. The problem, however, is that most laboratories have not established their own trimester-specific TSH ranges among women residing in their own region and applying their own customary TSH assay. In that case the guidelines suggest using the so-called 'international trimester-specific reference ranges', with upper normal TSH limits

of 2.5, 3.0, and 3.5 in the 1st, 2nd, and 3rd trimesters, respectively. Several papers have now been published demonstrating that the upper normal TSH limits established in the local region are almost without exception much higher than the international ones [15, 16]. As compared to regional reference ranges, application of the international reference ranges results in significantly higher prevalence rates of SHypo in each trimester. One must conclude that following the recommendation of previous guidelines to treat SHypo in pregnancy under application of the international TSH reference ranges, many women might have been treated with levothyroxine, which would have been unnecessary if regional reference ranges had been applied.

Another limitation of reference ranges is the narrow intra-individual variation in serum TSH [17]. TSH values of an individual subject apparently are always located in a narrow area somewhere within the much wider reference range; thus, one subject might have TSH values at the lower end of the reference range, and another always at the higher end. With regard to variance of TSH assays, analytical, intra-individual, and inter-individual variation coefficients of TSH results have been reported as 7.5, 16.2, and 31.7%, respectively [18]. The ratio of intra-individual to inter-individual variation can be used for the reliability of reference ranges. If the ratio is  $>1.4$ , the population-based reference range works as intended. If the ratio is  $<0.6$ , the population-based reference range is an insensitive measure in the majority of subjects. Actually, the ratio is  $<0.6$  for all thyroid function tests. For TSH, the literature mentions ratios of 0.36 [19], 0.49 [17], and 0.50 [18]. The substantial intra-individual variation in TSH may explain why among subjects with SHypo and the same TSH and FT<sub>4</sub> values, some have symptoms and others not: the original TSH value of the ones who have symptoms might have been much lower than of those without symptoms, and they must have travelled a longer distance along the TSH/FT<sub>4</sub> regression line to arrive at the same TSH value. Consequently, their fall in FT<sub>4</sub> concentrations within the reference range is greater, enhancing the possibility to have symptoms [9].

### Associations with Adverse Health Outcomes Continues for TSH Values within the Reference Range

The association between abnormal thyroid function tests and adverse health outcomes also holds true for test results within the reference range. For instance, the high-



**Fig. 1.** Joint British Societies' cardiovascular disease risk prediction chart for nondiabetic women (left) and men (right) [25]. SBP = Systolic blood pressure; TC:HDL = serum total cholesterol to HDL cholesterol ratio; green, orange, and red area depict cardiovascular

disease risk <10, 10–20, and >20% over the next 10 years, respectively; white curve denotes cardiovascular disease risk of 30% over the next 10 years.

er the  $FT_4$  value is within the reference range of 10–22 pmol/, the higher the prevalence of atrial fibrillation [20], and the lower the TSH value is within the reference range of 0.5–5.0 mU/l, the higher the prevalence of osteoporosis in healthy postmenopausal women [21]. The odds of adverse outcomes for higher TSH levels within the reference range compared to lower TSH levels within the reference range is significant for combined cardiovascular outcomes (OR 1.21, 95% CI: 1.15–1.27), for combined meta-

bolic outcomes (OR 1.37, 95% CI: 1.27–1.48), and for combined bone outcomes (OR 0.55, 95% CI: 0.41–0.72) [22]. It should come as no surprise that the risk of adverse health outcomes expands to values within the reference range, as there is no good physiological reason why the risk should stop at the borders of the (arbitrarily defined) reference range. I do agree with the comment that 'the continuum of effects across the reference range of thyroid function suggest that it might be more appropriate to



consider thyroid hormone levels as “risk factors” for disease (similar to blood pressure or cholesterol in cardiovascular disease) rather than consider a particular level to be normal or abnormal’ [22]. Along similar lines, a study among US community-dwelling subjects  $\geq 65$  years of age reports that higher TSH and lower FT<sub>4</sub> concentration within the euthyroid range (TSH: 0.45–4.5 mU/l) are associated with lower risk of multiple adverse events, including mortality [23]. It suggests tolerance for lower thyroid hormone levels in older people, in agreement with the guidelines to be rather conservative in prescribing levothyroxine to elderly subjects with SHypo [2]. The data support shifting the upper normal limit of the TSH reference range upward in older people [13]. More recently, an individual participant data analysis of 14 cohorts observed no association of TSH levels within the reference range (0.45–4.5 mU/l) with risk of coronary heart disease events or mortality, but found a U-shaped association with FT<sub>4</sub> levels within the reference range [24]; according to the authors, chance findings cannot be completely excluded.

### **Risk Stratification according to Individual TSH Values and Comorbidities**

The observed associations between SHyper and SHypo and adverse health outcomes do not constitute evidence for a causal relationship. Causality is, however, likely in view of its biologic plausibility (e.g. hyperthyroidism is a definite risk factor for atrial fibrillation and bone loss) and the presence of a dose-response relationship. The latter provides the rationale for the recommendation to intervene in hyperthyroidism grade IB (SHyper TSH  $\leq 0.1$  mU/l) and in hypothyroidism grade IB (SHypo TSH  $\geq 10$  mU/l). Guidelines agree that in grade IA other risk factors should be considered in the decision to intervene or to abstain from treatment. In hyperthyroidism grade IA (SHyper TSH  $>0.1$  to  $<0.4$  mU/l) the presence of old age ( $>65$  years), postmenopause, osteoporosis, and cardiovascular risk factors should tip the balance towards intervention, which hopefully will do more benefit than harm; however, in hypothyroidism grade IA (SHypo TSH  $>4.0$  to  $<10$  mU/l) the presence of age  $<70$  years, symptoms, pregnancy (desire), and cardiovascular risk factors would favour intervention with levothyroxine. In these recommendations TSH values are considered as just another risk factor for a particular disease, and the decision to treat or not to treat depends on the context of the subject with respect to age and oth-

er risk factors. It reminds me of clinical decision making as to when to start with antihypertensive drugs or statins in apparently healthy people at risk of cardiovascular disease. High blood pressure and high cholesterol are associated with adverse health outcomes, but the decision to treat may depend on other risk factors as well (e.g. age, sex, smoking, diabetes). To tackle this problem, charts have been outlined which immediately visualize when treatment is warranted: definitely if the cardiovascular risk over the next 10 years is  $>20\%$  (fig. 1) [25]. One wonders if similar charts could be constructed to evaluate the utility of treatment of a particular TSH level in conjunction with other risk factors.

### **How to Proceed Further: Urgent Need for Randomized Clinical Trials**

The last decade has seen much progress, especially thanks to the Thyroid Studies Collaboration. This consortium combined data from 11 international prospective cohort studies, allowing individual participant data analysis from about 55,000 individuals with about 543,000 person-years of follow-up. Individual participant data analyses is generally considered the highest level of nonrandomized evidence, and they have provided reliable quantitative estimates of involved risks. However, whether or not preventive intervention will do more good than harm does require randomized clinical trials with a large sample size and a long follow-up. To do such trials is not easy. A trial on SHyper comparing <sup>131</sup>I treatment with placebo has been discontinued because recruitment was very low. Fortunately, the European Commission has recently funded the TRUST trial, a multicentre double-blind placebo-controlled randomized trial in 3,000 adults age 65 and older with persistent SHypo (NCT01660126). Still, we will get no evidence-based guidance on how to manage younger subjects. Financing such studies is a problem. However, the numbers of subjects using thyroid hormone medication are huge (in the order of 2.5% of the population), and have increased significantly over the last decade: levothyroxine sodium prescriptions in the period 2006–2010 increased by 33% in the Netherlands, 37% in the UK, and 42% in the USA [26, 27]. It looks like the main reason for the increase in thyroid hormone users is the treatment of SHypo: median TSH at the initiation of levothyroxine fell from 8.7–7.9 mU/l between 2001 and 2009, with a population-adjusted OR of 1.30 (95% CI: 1.19–1.42) for prescribing L-T<sub>4</sub> at TSH  $\leq 10$  mU/l [27]. It thus seems

not unreasonable to ask for financial assistance to set up the logistics for data acquisition and storage of such large trials, which could be conducted by individual endocrinologists in their own practice in various European countries. Would the ETA feel up to the challenge to initiate such studies?

## Disclosure Statement

The author declares that he has no conflict of interest.

## References

- 1 Biondi B, Bartalena L, Cooper SD, Hegedüs L, Laurberg P, Kahaly GJ: The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. *Eur Thyroid J* 2015;4:149–163.
- 2 Pearce SHS, Brabant G, Duntas LH, Monzani F, Peeters RP, Razvi S, Wemeau J-L: 2013 ETA Guideline: Management of subclinical hypothyroidism. *Eur Thyroid J* 2013;2:215–228.
- 3 Lazarus J, Brown RS, Daumerie C, Hubalewska-Dydejczyk A, Negro R, Vaidya B: 2014 ETA Guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J* 2014;3:76–92.
- 4 Stott DJ, Bauer DC, Ford I, et al: The dilemma of treating subclinical hypothyroidism: risk that current guidelines do more harm than good. *Eur Thyroid J* 2014;3:137–138.
- 5 Sawin CT, Geller A, Wolf PA, et al: Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N Engl J Med* 1994;331:1249–1252.
- 6 Zulewski H, Müller B, Exer P, et al: Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. *J Clin Endocrinol Metab* 1997;82:771–776.
- 7 Duntas LH: Subclinical hypothyroidism: a misnomer in search of a new name. *Thyroid* 2001;11:361–362.
- 8 Evered DC, Ormston BJ, Smith PA, et al: Grades of hypothyroidism. *Br Med J* 1973;1:657–662.
- 9 Wiersinga WM: Hypothyroidism and myxedema coma; in Jameson JL, DeGroot LJ, et al (eds): *Endocrinology: Adult and Pediatric*. Philadelphia, Elsevier Saunders, 2016, pp 1540–1556.
- 10 Laurberg P, Knudsen N, Andersen S, Carle A, Pedersen IB, Karmisholt J: Thyroid function and obesity. *Eur Thyroid J* 2012;1:159–167.
- 11 Benson ES: The concept of the normal range. *Hum Pathol* 1972;3:152–155.
- 12 Hollowell JG, Staehling NW, Flanders WD, et al: Serum TSH, T<sub>4</sub>, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab* 2002;87:489–499.
- 13 Surks MI, Hollowell JG: Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. *J Clin Endocrinol Metab* 2007;92:4575–4582.
- 14 Stagnaro-Green A, Abalovich M, Alexander E, et al: Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *J Clin Endocrinol Metab* 2011;21:1–45.
- 15 Amouzegar A, Ainy E, Khazan M, Mehran L, Hedayati M, Azizi F: Local versus international recommended TSH references in the assessment of thyroid function during pregnancy. *Horm Metab Res* 2014;46:206–210.
- 16 Medici M, Korevaar TI, Visser WE, Visser TJ, Peeters RP: Thyroid function in pregnancy: what is normal? *Clin Chem* 2015;61:704–713.
- 17 Andersen S, Pedersen KM, Bruun NH, Laurberg P: Narrow individual variations in serum T<sub>4</sub> and T<sub>3</sub> in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab* 2002;87:1068–1072.
- 18 Browning MC, Ford RP, Callaghan SJ, Fraser CG: Intra- and interindividual biological variation of five analytes used in assessing thyroid function: implications for necessary standards of performance and the interpretation of results. *Clin Chem* 1986;32:962–966.
- 19 Nagayama I, Yamamoto K, Saito K, Kuzuya T, Saito T: Subject-based reference values in thyroid function tests. *Endocr J* 1993;40:557–562.
- 20 Gammage MD, Parle JV, Holder RL, et al: Association between serum free thyroxine concentration and atrial fibrillation. *Arch Intern Med* 2007;167:928–934.
- 21 Kim DJ, Khang YH, Koh JM, Shong YK, Kim GS: Low normal TSH levels are associated with low bone mineral density in healthy postmenopausal women. *Clin Endocrinol* 2006;64:86–90.
- 22 Taylor PN, Razvi S, Pearce SH, Dayan CM: A review of the clinical consequences of variation in thyroid function within the reference range. *J Clin Endocrinol Metab* 2013;98:3562–3571.
- 23 Cappola AR, Arnold AM, Wulczyn K, Carlson M, Robbins J, Psaty BM: Thyroid function in the euthyroid range and adverse outcomes in older adults. *J Clin Endocrinol Metab* 2015;100:1088–1096.
- 24 Asvold BO, Vatten LJ, Bjoro T, et al: Thyroid function within the normal range and risk of coronary heart disease: an individual participant data analysis of 14 cohorts. *JAMA Intern Med* 2015;175:1037–1047.
- 25 British Cardiac Society, British Hypertension Society, Diabetes UK, HEART UK, Primary Care Cardiovascular Society, The Stroke Association: JBS2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005;91(suppl V):1–52.
- 26 de Jong NW, Baljet GM: Use of T<sub>4</sub>, T<sub>4</sub>+T<sub>3</sub>, and T<sub>3</sub> in the Dutch population. *Eur Thyroid J* 2012;1:135–136.
- 27 Taylor PN, Iqbal A, Minassian C, et al: Failing threshold for treatment of borderline elevated thyrotropin levels – balancing benefits and risks: evidence from a large community-based study. *JAMA Intern Med* 2014;174:32–39.