

Thyroid function tests—time for a reassessment

Denis StJ O'Reilly

Department of
Clinical
Biochemistry and
Clinic for Thyroid
Diseases, Royal
Infirmary, Glasgow
G4 0SF

Denis StJ O'Reilly,
*consultant clinical
biochemist*

doreilly@clinmed.
gla.ac.uk

BMJ 2000;320:1332-4

In 1999, 890 000 measurements of thyroid stimulating hormone were performed by Scottish hospital laboratories—approximately one test for every six of Scotland's 5.1 million people.¹ This number does not include tests performed in the non-NHS laboratories or as part of the screening programme for congenital hypothyroidism. Although laboratory statistics are not collected nationally in England and Wales, the market in the United Kingdom (population 59 million) for thyroid stimulating hormone diagnostic tests is currently estimated at 9-10 million each year.

A remarkable downgrading of the clinical aspects of hypothyroidism and hyperthyroidism has paralleled the inexorable increase in the number of thyroid function tests performed over the past 20 years. This has led to chaos in the diagnosis of hypothyroidism. It has been stated that a diagnosis of clinical hypothyroidism can be made on the basis of biochemical measurements alone and that signs and symptoms are unnecessary.² Other authors protest, and maintain that biochemical tests can be misleading and that the diagnosis can be made on clinical grounds alone.³ In hyperthyroidism, a suppressed thyroid stimulating hormone concentration is currently the cornerstone of biochemical diagnosis. No numerical value has been assigned to the serum concentration of thyroid stimulating hormone below which suppression is considered to occur. This value varies from centre to centre depending on the sensitivity of the local assay. Thus, to many non-specialists the diagnosis of hyperthyroidism is also confusing.

Methods

This review is based on my 20 years' postgraduate experience in providing biochemical thyroid function tests and treating patients with thyroid disorders. I have selected and highlighted some of the publications that have influenced my practice and call into question the increasing reliance on biochemical thyroid function tests in making a diagnosis.

Historical setting

The treatments currently used for hyperthyroidism and hypothyroidism were established by the beginning of the 1970s. Though the symptoms and signs of these disorders had been analysed and clinical scoring indices had been developed and validated in the 1960s, clinical diagnosis remained problematic.⁴⁻⁸ The clinical diagnostic schemes for hypothyroidism were similar,⁴⁻⁶ but there were considerable differences between diagnostic schemes for hyperthyroidism. For example, atrial fibrillation was considered by Wayne and Crooks to be one of the most powerful discriminating signs,^{6,7} but it was not included by Gurney et al.⁸ Age, on the other hand, was a major diagnostic factor according to Gurney et al.,⁸ but was not mentioned by Wayne or Crooks.^{6,7} From knowledge of the pathophysiology of the hypothalamic-pituitary-thyroid axis available at that time, it was believed that measuring the

Summary points

There are no data on the relative importance of biochemical thyroid function tests and clinical symptoms and signs in assessing thyroid dysfunction

Secretion of thyroid stimulating hormone is influenced by many factors other than the negative feedback inhibition by thyroxine or triiodothyronine

Changes in thyroid stimulating hormone, thyroxine, and triiodothyronine concentrations during systemic illness are poorly understood

Thyroid function tests cannot be interpreted in patients with systemic illness

Since thyroid stimulating hormone concentrations are distributed logarithmically in the population, minor changes are unlikely to be clinically important

The possibility of false positive and false negative results should be considered in interpreting thyroid stimulating hormone concentrations

concentration of serum thyroid stimulating hormone would simplify the diagnosis.

Hypothyroidism

The publication of a reliable and practical assay for thyroid stimulating hormone was a landmark.⁹ A normal range of < 0.5-4.2 mU/l was established, based on measurements from 29 control subjects. One of the first applications of the assay was in patients who had undergone subtotal thyroidectomy for Graves' disease.¹⁰ In 28 "unequivocally euthyroid" patients followed for three to 21 years, the mean concentration was 8.2 mU/l (range 1.3-34.0 mU/l). In four patients followed up for four to 12 years and in whom a therapeutic trial of thyroxine had shown no benefit, the thyroid stimulating hormone concentration range was 10.5-21.5 mU/l. These patients were considered to be unequivocally euthyroid by a group who had validated clinical indices for the diagnosis of hypoparathyroidism and hyperthyroidism.^{5,7} They were used to show the superiority of thyroid stimulating hormone measurements in detecting hypothyroidism, and no suggestion was made that the normal range could be widened.

In 1973, the data on which the concept of subclinical hypothyroidism was based were published.¹¹ The reference range for thyroid stimulating hormone, established from measurement in 29 subjects,¹⁰ was used to classify 22 euthyroid subjects as having

subclinical hypothyroidism. In six of the 22 subjects given a therapeutic trial of thyroxine, treatment showed no benefit, and 10 had originally been recruited as normal controls.

Whickham survey

The Whickham survey was a further landmark.¹² All Whickham residents with a serum thyroid hormone concentration >6 mU/l were diagnosed as being hypothyroid, irrespective of their clinical status. This reinforced the view that the serum thyroid stimulating hormone concentration defined hypothyroidism.

The 20 year follow up study of the Whickham survey has yielded invaluable data on the natural history of thyroid disorders.¹³ A main conclusion of the study, disseminated to most non-specialists in a review published in the *BMJ*, was that "thyroid stimulating hormone concentrations above 2 mU/l are associated with an increased risk of hypothyroidism."²² Half of the population (male and female) fall into this category.¹² This conclusion was based on the change in the slope of the line obtained when the log of the serum thyroid stimulating hormone concentration was related to the logit probability of developing hypothyroidism over a 20 year period in women (see box).¹³ The probability of a 40 year old woman with a thyroid stimulating hormone of 2.1 mU/l developing hypothyroidism is low—at 1 in 50 over 20 years. In men, the probability is so low that an equivalent equation could not be derived.¹³

Clinical features ignored

The review also highlighted the fact that in making a diagnosis of clinical or overt hypothyroidism "symptoms are not considered a criterion by some authorities."²² The review claimed great authority. It was pointed out that some of the data on which it was based had been collected for the consensus statement for good practice and audit measures in the management of hypothyroidism and hyperthyroidism published on behalf of the Royal College of Physicians of London and the Society for Endocrinology.¹⁴ This publication makes no reference to the clinical manifestations or clinical diagnosis of hypothyroidism. Thus, the clinical features of hypothyroidism seem to have been relegated to the status of historical curiosities.

Hyperthyroidism

Assays capable of defining the lower end of the statistically derived reference range became available in the early 1980s. One evaluation of such an assay reported that all of 110 hyperthyroid patients studied had a thyroid stimulating hormone concentration <0.07 mU/l,

Relation between concentration and risk

The equation to describe the relation between the probability of developing hypothyroidism and the serum thyroid stimulating hormone concentration is¹³:
 $\ln [P/(1-P)] = b_0 + b_1 \ln$ thyroid stimulating hormone + 0.027 age (+ 1.79 if antibody positive).
 $b_0 = -5.02, b_1 = 0.30$ if thyroid stimulating hormone <2 mU/l
 $b_0 = -6.38, b_1 = 1.97$ if thyroid stimulating hormone ≥ 2 mU/l

Citation frequency (in BIDS) of published papers on the clinical assessment of hypothyroidism and hyperthyroidism in relation to UK groups and worldwide, 1987-97

First author	Number of citations	
	Worldwide	UK groups
Murray ⁴	0	0
Billewicz ⁵	20	2
Wayne ⁶	9	1
Crooks ⁷	0	0
Gurney ⁸	6	0
Klein ¹⁹	28	0

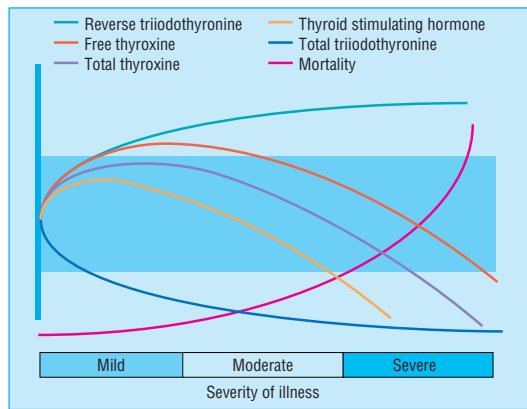
and all 62 euthyroid control subjects had concentrations >0.07 mU/l.¹⁵ However, some clinically euthyroid subjects with abnormally low thyroid stimulating hormone concentrations were classified as having subclinical hyperthyroidism.¹⁵ Assays can now detect thyroid stimulating hormone in serum at concentrations of 0.005 mU/l.¹⁶ At this low concentration, hyperthyroid patients were not distinguished from some euthyroid, though ill, patients.¹⁶ The range of thyroid stimulating hormone concentrations in patients whose condition stabilised on thyroxine replacement treatment was <0.005 to >10.00 mU/l.¹⁶ It is therefore clear that measurement of the thyroid stimulating hormone concentration has failed to deliver what was expected of it.

Clinical aspects

During this period the clinical aspects of hyperthyroidism have also been downgraded. Most current undergraduate textbooks treat the clinical diagnosis of thyroid dysfunction by referring the student to lists. In the current edition of the *Oxford Textbook of Medicine*, this matter is dismissed in less than a line, and the reader is referred to unweighted lists of the symptoms and signs.¹⁷ In the popular postgraduate textbook of *Clinical Endocrinology*, the biochemical diagnosis and assessment of hyperthyroidism are given before the clinical features.¹⁸ Medical journals are now effectively devoid of references to the clinical features of hyperthyroidism. Though a symptom rating scale for the diagnosis of hyperthyroidism was described in 1988,¹⁹ the clinical scoring systems for assessing hypothyroidism and hyperthyroidism are now rarely cited (table).

Non-thyroidal illness syndrome

We have recently become aware of the complexity of the effects of non-thyroidal illness on the hypothalamic-pituitary-thyroid axis and thyroid hormone metabolism. Figures like the one shown (taken from a recent review²⁰) are frequently used to illustrate the nature of the changes that occur in serum thyroid hormone concentrations in the non-thyroidal illness syndrome. These figures have never been published with a numerical scale or error bars. The problem of interpreting free thyroxine was summarised by the author: "It is common to find that a sample obtained from a patient with non-thyroidal illness syndrome may have a raised free thyroxine by one method but a normal or low free thyroxine by another."²⁰ The equilibrium dialysis reference method used to profile free thyroxine in the figure is technically demanding and currently not established in the United Kingdom. As the original legend to the figure explains:



The effects of illness on the concentrations of thyroid hormones: the shaded area represents the reference range for each method (reproduced with permission from Beckett and Wilkinson²³)

The profile for free thyroxine is that obtained using equilibrium dialysis and low sample dilution. The level of free thyroxine found using commercial methods will be heavily method dependent. A profile of free triiodothyronine is not included as some ultrafiltration methods suggest that normal or raised free triiodothyronine may be found in illness whilst equilibrium dialysis methods usually show diminished or normal concentrations.²⁰

What free thyroxine and free triiodothyronine assays actually measure is controversial.²¹ However, what is clear is that we cannot interpret thyroid function tests in systemically ill patients.

Current status of thyroid function tests

Our understanding of the complexity of the cerebral-hypothalamic-pituitary-thyroid axis and the mechanism of thyroid hormone action has grown enormously. Current knowledge indicates that the cardiac effects of thyroid hormones, which are clinically very important, are mediated via the α_1 thyroid hormone receptor independent of the β receptors, which are the dominant regulators of thyroid stimulating hormone secretion.²²

False positive and negative results

Overlap between the statistically derived normal and abnormal ranges is accepted in diagnostic tests, giving rise to false positive and false negative results. These concepts have not been applied to measurements of thyroid stimulating hormone. Rather than accepting that the test can be fallible, we transfer the problem to the patient. In patients with systemic disease, the non-thyroidal illness syndrome is invoked to explain the anomalous results, and healthy subjects are diagnosed as having subclinical hypothyroidism or hyperthyroidism.¹¹⁻¹⁵ The distribution of the serum thyroid stimulating hormone concentration in the population is logarithmic.¹³ Thus, minor deviations from the statistically derived reference range are unlikely to be clinically meaningful.¹¹

Confusion

Studies in 1580 inpatients²³ and in 630 patients admitted as medical emergencies²⁴ found that thyroid function tests performed as screening tests yielded abnormal results in 33% and 20% of patients respectively. In both studies, the biochemical tests suggested thyroid disease

incorrectly (that is, they gave false positive results) in nine cases out of 10. Thus, indiscriminate use of thyroid function tests is more likely to confuse than to help.

We do not know how important the thyroid function tests are for making a diagnosis of thyroid dysfunction. It is a matter of personal judgment. Experience has shown that thyroid function tests, like all the signs and symptoms associated with hypothyroidism and hyperthyroidism, are not totally reliable. As it becomes clear that biochemical assessments cannot deliver the diagnostic accuracy expected of them, the fact that the clinical aspects of assessing thyroid dysfunction are being sidelined is a cause for concern. Doing more biochemical tests will lead to further confusion, not the hoped for clarity. The information obtained from thyroid function tests, despite its quantitative numerical appearances, is "soft." How soft has yet to be established.

I thank Dr David Lyon for mathematical help, Dr Ann Wales for obtaining the citation data given in the table, and Drs G H Beall and H G Gray for constructive comments and discussion.

Competing interests: None declared.

- 1 The NHS in Scotland. *Laboratory statistics 1999*. Edinburgh: Information and Statistics Division, The NHS in Scotland, 1999.
- 2 Weetman AP. Hypothyroidism: screening and subclinical disease. *BMJ* 1997;314:1175-8.
- 3 Skinner GRB, Thomas R, Taylor M, Sellarajah M, Bolt S, Krett S, et al. Thyroxine should be tried in clinically hypothyroid but biochemically euthyroid patients. *BMJ* 1997;314:1764.
- 4 Murray IPC. The clinical diagnosis of thyroid disease. *Med J Aust* 1964;1:827-31.
- 5 Billewicz WZ, Chapman RS, Crooks J, Day ME, Gossage J, Wayne E, et al. Statistical methods applied to the diagnosis of hypothyroidism. *Q J Med* 1969;38:255-66.
- 6 Wayne EJ. Clinical and metabolic studies in thyroid disease. *BMJ* 1960;i:78-90.
- 7 Crooks J, Murray IPC, Wayne EJ. Statistical methods applied to the clinical diagnosis of thyrotoxicosis. *Q J Med* 1959;28:211-34.
- 8 Gurney C, Owen SG, Hall R, Roth M, Harper M, Smart GA. Newcastle thyrotoxicosis index. *Lancet* 1970;ii:1275-8.
- 9 Hall R, Amos J, Ormston BJ. Radioimmunoassay of human serum thyrotrophin. *BMJ* 1971;i:582-5.
- 10 Hedley AJ, Hall R, Amos J, Michie W, Crooks J. Serum thyrotropin levels after subtotal thyroidectomy for Graves disease. *Lancet* 1971;i:455-8.
- 11 Evered DC, Ormston BJ, Smith PA, Hall R, Bird T. Grades of hypothyroidism. *BMJ* 1973;i:657-62.
- 12 Tunbridge WMG, Evered DC, Hall R, Appleton D, Brewis M, Clark F, et al. The spectrum of thyroid disease in a community: the Whickham survey. *Clin Endocrinol* 1977;7:481-93.
- 13 Vanderpump MPJ, Tunbridge WMG, French JM, Appleton D, Bates M, Clark F, et al. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol* 1995;43:55-68.
- 14 Vanderpump MPJ, Ahlquist JAO, Franklyn JA, Clayton RN. Consensus statement for good practice and audit measures in the management of hypothyroidism and hyperthyroidism. *BMJ* 1996;313:539-44.
- 15 Seth J, Kellert HA, Caldwell G, Sweeting VM, Beckett GJ, Gow SM, et al. A sensitive immunoradiometric assay for serum thyroid stimulating hormone: a replacement for the thyrotropin releasing hormone test? *BMJ* 1984;289:1334-6.
- 16 Wilkinson E, Rae PWH, Thomson KJT, Toft AD, Spencer CA, Beckett GJ. Chemiluminescent third generation assay (Amerlite TSH30) of thyroid stimulating hormone in serum or plasma assessed. *Clin Chem* 1993;39:2166-73.
- 17 McGregor AM. The thyroid gland and disorders of thyroid function. In: Weatherall DJ, Ledingham JGG, Warrell DA, eds. *Oxford textbook of medicine*. 3rd ed. Oxford: Oxford University Press, 1996:1603-21.
- 18 Hall R. Hyperthyroidism and Graves disease. In: Besser GM, Thorner MO, eds. *Clinical endocrinology*. 2nd ed. London: Mosby Wolfe, 1994:17: 1-24.
- 19 Klein I, Trzepacz PT, Roberts M, Levey GS. Symptom rating scale for assessing hyperthyroidism. *Arch Intern Med* 1988;148:387-90.
- 20 Beckett GJ, Wilkinson E. Thyroid hormone metabolism and thyroid function tests in non-thyroidal illness. *CPD Bull Clin Biochem* 1998;1:9-14.
- 21 Ekins R. Measurement of free hormones in blood. *Endocrin Rev* 1990;11:5-46.
- 22 Weiss RE, Murata Y, Cua K, Hayashi Y, Seo H, Refetoff S. Thyroid hormone action on liver, heart, and energy expenditure in thyroid hormone receptor β -deficient mice. *Endocrinology* 1998;139:4945-52.
- 23 Spencer C, Eigen A, Shen D, Duda M, Qualls S, Weiss S, et al. Specificity of sensitive assays of thyrotropin (TSH) used to screen for thyroid disease in hospitalised patients. *Clin Chem* 1987;33:1391-6.
- 24 Small M, Buchanan L, Evans R. Value of screening thyroid functions in acute medical admissions to hospital. *Clin Endocrinol* 1990;32:185-191. (Accepted 3 November 1999)