Current utility of first-line FT4 and TSH in screening for central hypothyroidism

Carol Evans1, Jiya Jacob2, Annabel Rodham1, Manjot Gill3, Laura Parry4, Alan Dodd5, Nadia El-Farhan4, Angharad Shore3, Andrew Lansdown6, Aled Rees7, Onyebuchi E. Okosieme2

1Department of Medical Biochemistry, Immunology and Toxicology, Cardiff and Vale University Health Board, University Hospital of Wales, Cardiff, UK
2Department of Diabetes, Prince Charles Hospital, Cwm Taf University Health Board, Merthyr Tydfil, UK
3Laboratory Medicine Service, Swansea Bay University Health Board, Morriston Hospital, Heol Maes Eglwys, Swansea, UK
4Department of Medical Biochemistry, Royal Gwent Hospital, Newport, UK
5Department of Biochemistry, Prince Charles Hospital, Cwm Taf Morgannwg University Health Board, Merthyr Tydfil, UK
6Department of Medicine, Cardiff and Vale University Health Board, University Hospital of Wales, Cardiff, UK
7Neuroscience and Mental Health Innovation Institute, School of Medicine, Cardiff University, Cardiff, UK

Correspondence
Carol Evans, Department of Medical Biochemistry, Immunology and Toxicology, Cardiff and Vale University Health Board, University Hospital of Wales, Cardiff CF14 4XW, UK.
Email: carol.evans9@wales.nhs.uk

Abstract

Background: Thyroid testing strategies vary across laboratories. First-line combined thyroid stimulating hormone (TSH) and freeT4 (FT4) have historically been preferred by many laboratories as this detects individuals with undiagnosed central hypothyroidism who can be missed with a first-line TSH-only strategy. However, an up-to-date evaluation of the utility of this approach is lacking.

Objectives: We investigated the clinical utility of first-line TSH and FT4 in the detection of central hypothyroidism in current day practice.

Design, Patients, and Measurements: The All-Wales laboratory information system was queried to identify thyroid function tests in patients aged ≥16 years with decreased FT4 and inappropriate TSH (low-FT4). The 1-year incidence of low-FT4 was determined using mid-year population data. Clinical information of patients with low-FT4 was reviewed to determine causes of low-FT4 and the incidence of central hypothyroidism.

Results: The incidence of low-FT4 varied according to FT4 assay method (range: 98–301 cases/100,000 population/year). Fifteen new cases of central hypothyroidism were detected in two health boards, equivalent to 2 cases/100,000 population/year. Positive predictive value of low-FT4 for central hypothyroidism was 2%–4%. In a cross-section of primary care patients, low-FT4 was detected in 0.5% of all thyroid tests with assay-related differences in detection rates.

Conclusions: Although low-FT4 is a common laboratory finding, the incidence of central hypothyroidism remains rare. With the currently increased rates of thyroid testing and increased use of medications that decrease FT4, low-FT4 has a much lower predictive value for central hypothyroidism than previously reported. Thyroid screening strategies will need to balance the yield from first line TSH and FT4 testing with the cost of investigating individuals with non-pathological laboratory abnormalities.

KEYWORDS
central hypothyroidism, FT4, thyroid function testing, TSH
Thyroid dysfunction affects 2%-5% of the global population and is readily diagnosed with modern sensitive laboratory tests of thyroid function. Query of the national laboratory information system (LIMS) revealed one million thyroid function tests (TFTs) comprising FreeT4 (FT4) and TSH were undertaken in our laboratories in Wales in 2019, a country with an approximate population of 3 million people. At the time, the testing strategy for the diagnosis of thyroid dysfunction in Wales was to undertake first-line free thyroxine (FT4) and thyroid stimulating hormone (TSH) measurements, with reflex addition of free triiodothyronine (FT3) if TSH was <0.1 mU/L. This approach was based on the 2006 UK Guidelines for the Use of Thyroid Function Tests which was viewed as best practice when the LIMS was implemented in 2009. More recently, the National Institute of Health and Care Excellence have recommended a first-line TSH strategy, with combined FT4 and TSH reserved for children, pregnant women, patients with hyperthyroidism, or adults with suspected central hypothyroidism.

Despite published guidelines the approach to thyroid testing may be driven by pragmatic considerations. While a first-line TSH strategy is adequate to screen for common causes of primary thyroid dysfunction, the suggested benefit of the combined FT4 (or total T4) plus TSH approach is the earlier identification of patients with central hypothyroidism or hypopituitarism. Central hypothyroidism is often subtle in clinical presentation with a biochemical picture comprising decreased FT4 and inappropriate TSH concentration. Thus, it has been argued that the diagnosis of central hypothyroidism may be missed with a first-line TSH approach. On the other hand, central hypothyroidism is rare, seldom presents without accompanying features of pituitary dysfunction, and combined TSH and FT4 screening incurs increased costs and unnecessary follow up. Ultimately, the key consideration underpinning a thyroid testing policy is the benefit of pituitary case-finding set against the cost of routine screening including costs from laboratory testing, endocrine and radiological evaluation for pituitary dysfunction, and the impact of treatment of individuals with laboratory test abnormalities without intrinsic pituitary disease.

To date, only a few studies have addressed the utility of TSH and FT4 or total T4 in the detection of central hypothyroidism. While these studies suggest that a first-line TSH approach would miss a portion of patients with central hypothyroidism, they were carried out over a decade or two ago. Since then requests for thyroid tests have increased substantially and all UK laboratories now measure FT4 instead of total T4. Recent years have also seen a growing use of medications that decrease FT4 concentrations including antipsychotic and antidepressant medications. Thus, it is unclear whether a first-line TSH and FT4 strategy remains efficient in the current thyroid testing landscape or whether this would incur excessive costs from laboratory tests and investigation of false-positive results without justifiable yield. The purpose of this study was thus to evaluate the current utility of TSH and FT4 in the detection of central hypothyroidism using routine clinical and laboratory data across several health board populations in Wales. Specifically, we determined the incidence of low FT4 and normal or low TSH, and the yield for cases of central hypothyroidism in patients with this biochemical finding.

**METHODS**

### 2.1 Data collection

LIMs in four Health Boards in Wales (Aneurin Bevan [AB], Cardiff and Vale [CV], Cwm Taf Morgannwg [CT] and Swansea Bay [SB]) were queried for TFTs on patients aged ≥16 years. The four health boards were chosen to represent the main laboratories serving the majority of the South Wales population. Data from two of the four participating health boards (CT and CV) were evaluated to determine the annual incidence rates of central hypothyroidism. In addition, cross-sectional primary care data were obtained from all four health boards to determine the causes of low-FT4 and the contribution of medications to low-FT4 in primary care patients. Clinical information was obtained from electronic or paper request forms and from information on the Welsh Clinical Portal which is a national digital patient record for Wales.

For the annual incident rates (AB and CV), data were collected for a 12-month period to identify incident cases in which FT4 was below the lower reference interval and TSH within the reference range (i.e., low-FT4). The 1-year data for CT covered incident cases from 1st January to 31st December 2019. In CV, clinical information from electronic test requesting was available from May 2022 onwards, and so the 1-year data in CV covered incident cases from 1st June 2022 to 31st May 2023. Of the cases of low-FT4, we excluded patients with pre-existing thyroid disease, known pituitary disease with central hypothyroidism, pregnant patients, or patients with non-thyroidal illness. After exclusions we determined the number of patients with a new diagnosis of central hypothyroidism due to pituitary disease and thereby estimated the annual population incidence of central hypothyroidism based on the population served by the health boards laboratories. To determine the aetiology of low-FT4 for those patients whose investigations were started but not concluded during the incidence year, clinical information for each patient was reviewed from the detection of low-FT4 till the end of the study follow-up period in December 2023.

### 2.2 Analytical methods

In CT and SB health boards, FT4 and TSH were measured with Roche Elecsys analysers. In these health boards, reference ranges were 11.0–25.0 pmol/L for FT4 and 0.27–4.20 mU/L for TSH. In 2019, in two health boards, AB and CV, FT4 and TSH were measured with the Abbott Alinity (three sites in AB and one site in CV) and the Abbott Architect (one site in CV). Reference ranges were 9.1–19.0 pmol/L for FT4 and 0.30–4.40 mU/L for TSH in AB and CV. Between 2022
and 2023 at CV, FT4 and TSH were measured using Abbott Alinity at all laboratory sites. Reference ranges were 8.9–17.3 pmol/L for FT4 and 0.30–4.40 mU/L for TSH. The coefficient of variation was <8% for all assays at all levels of quality control in all Health Boards.

2.3 Statistical analysis

Data are summarised descriptively using means (standard deviation) for normally distributed data and medians (interquartile range) for non-normally distributed data. Data are compared across diagnostic groups, i.e., central hypothyroidism versus low-FT4 from other causes, using t-test for normally distributed data and Kruskal–Wallis for non-normal data. The annual incidence of central hypothyroidism per 100,000 population was estimated using the mid-year population for the health board obtained from the Office for National Statistics. Using this we calculated the positive predictive value (PPV) for a low-FT4 for the diagnosis of central hypothyroidism as the proportion of patients with a low-FT4 who had central hypothyroidism. Data was collected on an excel spreadsheet and analysed using Stata version 17.0 for Windows (Stata Corp.).

3 RESULTS

3.1 One-year incidence of low-FT4

Using data from two health boards (CT, CV), we identified 909 patients in CT and 461 patients in CV health boards with a low-FT4 over a 1-year period. The flow chart for exclusions is shown in Figure 1. After exclusion of patients with known thyroid disease, known pituitary disease with central hypothyroidism, pregnancy, non-thyroidal illness, and patients with tests that were not repeated or normal on repeat testing, 387 patients in CT and 247 patients in CV had persistent low-FT4. In addition, we excluded patients who were not evaluated further, comprising 351 patients in CT and 191 patients in CV. These were predominantly patients with requests from primary or community care centres with insufficient clinical details on their electronic request forms to determine the aetiology of low-FT4 or the reasons why further evaluation was not done. After all exclusions, 36 patients in CT and 56 patients in CV were investigated further with endocrine evaluation and MRI. Using the mid-year population estimates for the population served by the laboratories, the overall incidence of Low-FT4 was 301.4 (95% confidence interval [CI]: 300.3, 302.5) per 100,000 population for CT while for CV the incidence was 92.2 (95% CI: 91.9, 92.5) per 100,000. The incidence of Low-FT4 was significantly higher in CT where the Roche assay was used than in CV where the Abbott assay was used. Incidence rate ratio for CT compared to CV was 3.28 (95% CI: 2.93, 3.68 p < .001).

3.2 One-year incidence of central hypothyroidism

Six patients in CT and nine patients in CV were confirmed to have central hypothyroidism with evidence of pituitary disease either secondary to structural abnormalities on MRI or other pituitary hormone deficiencies. Based on the number of patients with persistent low-FT4 after excluding patients with pre-existing thyroid or pituitary disease, pregnancy, and non-thyroidal illness (CT, n = 387, and CV, n = 247, Figure 1) the PPV for a low-FT4 for the diagnosis of central hypothyroidism was 0.02 for CT and 0.04 for CV. The causes of central hypothyroidism in the affected 15 patients were pituitary macroadenoma (n = 6), post-trans-

---

**FIGURE 1** One-year incidence of Low-FT4 and central hypothyroidism. aTests done in an acute illness setting or normalisation on repeat testing. bRequests from primary care practices with unavailable clinical information, unknown indications for test, and endocrine evaluation not undertaken.
sphenoidal surgery \((n = 3)\), empty sella syndrome \((n = 1)\), hypophysitis \((n = 2)\), immunotherapy related \((n = 1)\), and idiopathic pituitary stalk thickening \((n = 2)\). All patients with central hypothyroidism had additional pituitary hormone axes dysfunction which either preceded or was diagnosed concurrently with central hypothyroidism. Thus a pituitary aetiology was apparent at or around the time of diagnosis of central hypothyroidism in all of these cases. Based on the background population, the incidence of diagnosed central hypothyroidism, was 1.99 (95% CI: 1.98, 2.00) cases per 100,000 population for CT and 1.80 (95% CI: 1.79, 1.81) cases per 100,000 for CV. These rates were not significantly different between the two sites, incidence rate ratio of 1.11 (95% CI: 0.32, 3.48, \(p = .84\)) for CT compared to CV.

3.3 Demographic and biochemical characteristics of patients with low-FT4

The characteristics of the patients with low-FT4 are presented according to aetiology and health board in Table 1. The majority of the patients with low-FT4 were female although in CV more males had central hypothyroidism compared to other causes of low-FT4. There were no significant age differences with average age ranging from 48 to 53 years across the groups. Due to the difference in assay bias, the initial or index FT4 differed across health boards with lower concentrations in CV (Abbott assay) than CT (Roche assay). Patients with central hypothyroidism had lower FT4 levels than patients with low-FT4 from other causes. The FT4 results were not normally distributed for either health board and this difference was statistically significant for CV (Abbott assay) but not CT (Roche assay). TSH was also lower for central hypothyroidism than for other causes of low-FT4.

| TABLE 1 | Patient characteristics according to aetiology of Low-FT4 and normal TSH. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | CT Low-FT4: other causes | Central hypothyroidism | p-value | CV Low-FT4: other causes | Central hypothyroidism | p-value |
| Number         | 903              | 6                |        | 452              | 9                |        |
| Sex            |                  |                  |        |                  |                  |        |
| Female         | 589 (65.2%)      | 5 (83.3%)        | .35    | 294 (65.8%)      | 3 (33.3%)        | .043   |
| Male           | 314 (34.8%)      | 1 (16.7%)        |        | 153 (34.2%)      | 6 (66.7%)        |        |
| Age, years     |                  |                  |        |                  |                  |        |
| Mean (SD)      | 50.1 (17.3)      | 48.8 (20.4)      | .86    | 51.9 (18.6)      | 53.4 (21.2)      | .80    |
| Median (IQR)   | 50 (37, 62)      | 52 (27, 65)      | .91    | 51 (38, 65)      | 58 (37, 66)      | .84    |
| Index FT4, pmol/L |                  |                  |        |                  |                  |        |
| Mean (SD)      | 10.2 (1.0)       | 8.3 (2.5)        | <.001  | 8.0 (0.9)        | 6.9 (1.1)        | <.001  |
| Median (IQR)   | 10.50 (10.00, 10.70) | 8.85 (6.00, 10.60) | .072  | 8.40 (7.90, 8.70) | 6.80 (6.30, 7.30) | <.001  |
| Index TSH, mU/L |                  |                  |        |                  |                  |        |
| Mean (SD)      | 2.04 (0.98)      | 1.42 (1.18)      | .12    | 1.73 (1.16)      | 1.19 (1.03)      | .17    |
| Median (IQR)   | 2.00 (1.33, 2.74) | 1.41 (0.41, 2.54) | .23    | 1.58 (0.88, 2.43) | 1.06 (0.49, 1.58) | .15    |

Abbreviations: CT, Cwm Taf Morgannwg University Health Board; CV, Cardiff and Vale University Health Board; IQR, interquartile range; SD, standard deviation.

3.4 Review of Low-FT4 results from primary care patients

To determine the aetiologies and detailed medication histories for a cross section of primary care patients, we reviewed low-FT4 tests requested from primary care centres using data from 4 health boards collected over periods ranging from 3 to 20 weeks. Out of 100,057 sequential TFTs, we identified 463 episodes (0.5% of all tests) with low-FT4, i.e., a low FT4 with inappropriate TSH (Table 2). The rates of low-FT4 varied across laboratories with an average of 0.3% of TFTs in the Abbott laboratories and 1.2% of TFTs in the Roche laboratories (Table 2). After excluding duplicate tests, 396 unique patients had a low-FT4 of which 33 patients underwent endocrine evaluation. Of these, 3 patients were subsequently confirmed to be new cases of secondary hypothyroidism with pituitary disease. Two patients had pituitary macroadenomas while one had an empty sella syndrome.
confirmed on a subsequent sample. Reasons for our low PPV compared to this earlier study may include the laboratory reference ranges employed, an increase in the volume of thyroid function testing, precision of the FT4 assay around the lower reference interval, increased use of medications that can cause this biochemical finding, or a combination of these factors.

Thyroid testing has doubled over the last two decades from 1 test for every 6 of the population in 1999 to 1 in every 3 in 2019. We have shown that the number of cases of central hypothyroidism in the population has not significantly changed over the last two decades. Thus, doubling the population screened will have the effect of halving the prevalence, which is the fraction of subjects that have the disease in the population screened. Prevalence is known to influence the performance characteristics of a test as the prevalence decreases the PPV decreases. In addition, the high volume of testing will magnify any imperfections in assay performance e.g., precision at the lower reference interval. This makes it very important to choose the reference range carefully. In the health board that used a Welsh laboratories pathology harmony Roche reference range, in which the lower reference interval of FT4 was slightly below the manufacturers reported 2.5 centile, the incidence of low-FT4 was just over three times that in a neighbouring health board using a 99 centile FT4 reference range.

In the review of Low-FT4 results from primary care patients across four health boards, of those who had a repeat TFT (281 patients), a striking finding is the number of patients in whom the subsequent TFT was normal 184/281 (65%). Preiss et al. previously reported normalisation of TFTs in 67/266 (25%) of patients, 48 on repeating the original sample and a further 19 on a repeat sample from the patient. These findings were attributed to the FT4 assay precision and stability of calibration. European Thyroid Association Guidelines on the Diagnosis and Management of Central Hypothyroidism recommend that the diagnosis of central hypothyroidism should be confirmed by the combined findings of serum FT4 concentrations below the lower limit of the normal range and inappropriately low/normal TSH concentrations on at least two separate determinations. Our data and that of Priess supports this recommendation.

FT4 was lower in patients with central hypothyroidism compared to other causes of low FT4. This difference was significant for CV but not for CT (p = .07) perhaps due to the smaller number of patients.

### 3.5 Medication history

The medication history of 334 patients was reviewed for potential culprit medications that may have caused the biochemical finding of low-FT4. These included selective serotonin reuptake inhibitors (SSRIs, 98 patients), antiepileptics (69 patients), antipsychotics (36 patients), antidepressants (50 patients) and glucocorticoids (21 patients), with some patients taking more than one of these medication types. We were surprised to find that 20 patients had commenced levothyroxine therapy in primary care without endocrine evaluation.

### 4 DISCUSSION

The approach to thyroid screening varies across laboratories. Historically, many laboratories have preferred a first-line combined FT4 and TSH strategy on the premise that this approach detects cases of central hypothyroidism which would be missed with a TSH-only approach. Here, we have undertaken a reappraisal of this strategy in the current practice of increased thyroid testing. We found that central hypothyroidism remains rare. We observed a 1-year incidence of 2 cases per 100,000 of the adult population. This is roughly similar to the incidence of 8–10 new adult cases of hypopituitarism per million people every year reported in 1998. However, a finding of low FT4 with inappropriate TSH is now relatively common. Over the course of 1 year, we found an incidence of low-FT4 ranging from 92 per 100,000 (0.1%) to 301 per 100,000 (0.3%) depending on the FT4 assay and reference range employed. In a cross-section of primary care patients, low-FT4 was detected in 0.5% of all thyroid tests with assay-related differences in detection rates. Thus, while assays may differ in their detection rates for low FT4, the incidence of central hypothyroidism remains consistently rare.

In our 1-year cohort we identified 15 new cases of secondary hypothyroidism from a total of 1370 TFTs with low FT4 and inappropriate TSH. Using the number of patients with persistent low-FT4 after exclusions as denominator the PPV for a low-FT4 for the diagnosis of central hypothyroidism was 2.4%. This is much lower than the PPV derived from data presented by Preiss et al. in 2008 of 29% on the first TFT and 55% if the laboratory findings were
with central hypothyroidism in the latter sites. Furthermore, patients with low-FT4 from other causes, i.e., without pituitary disease may represent the lower tail of individuals with normal thyroid function. In our primary care cohort, we identified a significant number of medications: anti-epileptics, anti-psychotics, anti-depressants including SSRIs and steroids which can all cause a decrease in measured FT4 without clinical significance to the patient.\(^9\)–\(^11\) Prescriptions for antidepressants in England and Wales have almost doubled in the past decade so that in 2017/2018 11% of individuals were taking \(\geq 1\) antidepressants (mainly SSRIs) on any day.\(^15\) This is likely to be another cause of the poor PPV for identification of secondary hypothyroidism that we have observed.

We acknowledge several limitations of our study which are inevitable in a retrospective real-world multi-centre study such as this. The 1-year incidence data in the two health boards (CV and CT) was collected over different calendar years due to the unavailability of electronic data before 2022 in one of the health boards. Accordingly, patients with incident low-FT4 were followed up for varying durations after detection. Thus, the incidence rates reported in each site represents annual incidence rates for different calendar years. Interestingly, a proportion of individuals with low-FT4 did not have repeat tests and some with persistent abnormal results were not worked up further. In addition, for reasons that are unclear, some patients with low-FT4 were started on Levothyroxine without endocrine evaluation. While these decisions may have been mitigated by factors not captured in our review, it is important that clinicians are educated on a rational and safe approach to patients with low-FT4. These will include confirming abnormal results with repeat tests, considering factors such as pregnancy, non-thyroidal illness, and potential culprit medications, and setting rational thresholds for seeking endocrine evaluation. In addition, routine thyroid testing should be avoided in the acute illness setting except where there are clinical indications to suspect undiagnosed thyroid dysfunction. Guidance on the approach to evaluating patients with low-FT4 should be developed jointly by endocrinologists and clinical biochemists and will depend on local pathways and resource arrangements.

Case finding for secondary hypothyroidism may not be the only reason for choosing a first line TSH and FT4 TFT strategy. Mena and Kabadi\(^16\) have investigated use of FT4 or TSH individually or FT4 and TSH combined to identify the probable diagnostic category in hospital patients. Their study was not designed to identify new cases of central hypothyroidism and their cohort was not limited to cases where the test was used to diagnose disease and included patients being monitored for treatment of known thyroid disease. They concluded that both tests together were superior to either one alone to correctly identify the category of thyroid disorder. This study was based on samples rather than individual patients. Of the 3022 TFTs reviewed 51/3022 showed a pattern of decreased FT4 and normal TSH, an incidence of 2%. This is higher than our incidence but their study was limited to hospital patients unlike primary care and hospital patients in our study.

Although low-FT4 is a common laboratory finding, central hypothyroidism remains a rare cause of thyroid dysfunction. Low FT4 has a poor PPV for central hypothyroidism and unnecessary repeat testing and further investigation of patients without pituitary pathology is a potential disadvantage of a first line TSH and FT4 thyroid testing strategy. There may however be other reasons for laboratories to choose such a testing strategy. Both tests together have been found to be superior in hospital patients.\(^16\) Furthermore, it may not be possible for laboratories or electronic requesting systems to easily identify those patients (children, pregnant women, patients with hyperthyroidism, or adults with suspected central hypothyroidism)\(^3\) that require both FT4 and TSH measurement front line. So, in some cases the approach to thyroid testing is likely to remain driven by pragmatic considerations. Lastly, if a first line TSH and FT4 screening strategy is chosen, guidance should be provided to clinicians on the common causes of low FT4 without central hypothyroidism, the need to repeat tests with low-FT4, and the thresholds for seeking specialist input.

**ACKNOWLEDGEMENTS**

This service review was undertaken on behalf of the ACB Wales Standardisation Committee. We thank all members of the committee for their contribution to discussions on the thyroid function testing strategies. We thank all participating Health Board laboratories for supporting this service review.

**ORCID**

Carol Evans http://orcid.org/0000-0002-9401-394X

Aled Rees https://orcid.org/0000-0002-1165-9092

Onyebuchi E. Okosime http://orcid.org/0000-0003-2221-992X

**REFERENCES**


8. Price A. Screening for central hypothyroidism is unjustified. BMJ. 2001;322:798.