

Research paper

Variability in thyroid function test requests across general practices in south-west England

Bijay Vaidya

Department of Endocrinology, Royal Devon & Exeter Hospital, Exeter, UK

Obioha C Ukoumunne and Joanna Shuttleworth

Peninsula Collaboration for Leadership in Applied Health Research and Care, University of Exeter Medical School, University of Exeter, Exeter, UK

Alan Bromley

Department of Clinical Chemistry, Royal Cornwall Hospital, Truro, UK

Aled Lewis

Department of Biochemistry, Royal Devon & Exeter Hospital, Exeter, UK

Chris Hyde

Peninsula Technology Assessment Group, University of Exeter Medical School, University of Exeter, Exeter, UK

Anthea Patterson and Simon Fleming

Department of Clinical Chemistry, Royal Cornwall Hospital, Truro, UK

Julie Tomlinson

Department of Public Health, Cornwall & Isles of Scilly Primary Care Trust & University of Plymouth, Truro, UK

ABSTRACT

Background The number of thyroid function tests (TFTs) performed in the UK and other countries has increased considerably in recent years. Inconsistent clinical practice associated with inappropriate requests for tests is thought to be an important cause for this increase.

Aim To study the extent of variability in requests for TFTs from general practices.

Methods We analysed routine data on all TFTs on patients aged 16 years and over carried out by two hospitals in south-west England (Royal Cornwall Hospital and Royal Devon & Exeter Hospital) during 2010 at the request of 107 general practices.

Results A total of 195 309 TFT requests were made for 148 412 patients (63% female). The total requests included 192 108 tests for thyroid-stimulating hormone (TSH), 43 069 for free thyroxine (FT4) and

1972 for free tri-iodothyronine (FT3). The number of TSH tests per 1000 list size varied widely across the practices, ranging from 84 to 482. Most of the variation was due to heterogeneity across practices and only 24% of this was accounted for by prevalence of hypothyroidism and socio-economic deprivation.

Conclusions There is wide variation in TFT requests from general practice and scope to reduce both unnecessary TFTs and the variability in the clinical practice. Further studies are required to understand the causes for the variability in testing thyroid function.

Keywords: audit, clinical practice variation, general practice, laboratory tests, primary care, thyroid

How this fits in with quality in primary care

What do we know?

The number of TFTs performed in the UK and other countries has increased considerably in recent years, and inappropriate requests for tests are thought to be an important cause for this increase. National guidelines to help clinicians use TFTs appropriately have been published.

What does the paper add?

Despite the guidelines, there is wide variation in the rate of requests for TFTs from general practice and scope to reduce them.

Introduction

Thyroid disorders are common in the community, with hypothyroidism alone affecting about 2% of women.¹ Because symptoms of thyroid disorders are often non-specific and highly prevalent in the general population, biochemical tests are necessary for their diagnosis and monitoring. Thyroid function tests (TFTs) include serum thyroid-stimulating hormone (TSH), free thyroxine (FT4) and free tri-iodothyronine (FT3). In primary care, the first point of investigation for thyroid disorders, TSH is the first line test for most patients followed by an FT4 test if TSH is outside the reference range. FT3 testing is generally reserved for endocrinologists or done at the discretion of the laboratory.

The number of TFTs performed in the UK and other countries has increased considerably in recent years,² and it has been suggested that inappropriate requests for tests is an important cause.^{3,4} National guidelines to help clinicians use TFTs appropriately have been published.² Although it is difficult to measure the degree of 'appropriateness' of requests for TFTs at the population level, the presence of variation in clinical practice is an indicator.⁴ Our study aims to establish the extent of variability in TFT requesting across general practices referring to two hospitals in south-west England.

Methods

The analyses are based on routine data on all TFTs carried out by the biochemistry laboratories at the Royal Cornwall Hospitals NHS Trust (RCHT) and the Royal Devon & Exeter Hospital NHS Foundation Trust (RDE) and requested by general practices during 2010. The catchment areas of these hospital laboratories are the predominantly rural counties of Cornwall and Devon in the south-west of England, with a combined population of 1.7 million persons. Together, the two hospitals serve a catchment popu-

lation of approximately 800 000. The datasets included patient's gender and age; type (TSH, FT4 or FT3) and result of TFTs; and the name of the requesting general practice. Both laboratories use chemiluminescent immuno-assay (Roche Modular Analytics E170 analyser) for analysing TFTs.

We also obtained data on the following practice characteristics from the Network of Public Health Observatories⁵ for the year 2010: list size, percentage of patients with hypothyroidism recorded on practice disease registers, Index of Multiple Deprivation (IMD) score as a measure of socio-economic deprivation, and percentage of patients aged over 65 years.

Eligible records were TFT requests made by NHS general practices for patients aged 16 and over. We excluded records relating to practices outside the usual catchment areas of the hospital laboratories. Data for two separate practices in the RDE dataset were excluded because their records were combined under a common name and there was insufficient information to disaggregate them. In addition, records were excluded for a practice for which characteristics were not available. We analysed records from 107 practices (RCHT, 57; RDE, 50).

Statistical analysis

We summarised the distribution of tests per 1000 list size at general practice level within each hospital site and overall. Both means (with standard deviations) and medians (with interquartile ranges) are reported as the former reflects the volume of tests and the latter is appropriate for quantifying the average of skewed distributions. Random effects Poisson regression models were fitted to TSH test request rate (outcome) in order to quantify the extent to which prevalence of hypothyroidism, deprivation score and percentage of patients aged over 65 years (predictors) account for variability in practice test request behaviour. This model explicitly recognises the variation across clusters beyond chance. Analyses were carried out using Stata Statistical Software (Release 12.1, 2011; Stata Corporation, College Station, TX, USA).

Results

A total of 195 309 test requests for 148 412 patients [mean (SD) age 60 years (19); age range 16–105 years; 63.1% female] were made to the two hospital laboratories (Table 1). Of these, 46 897 (24%) were subsequent tests to patients who had already been tested earlier in the year. The total requests included 192 108 tests for TSH, 43 069 for FT4 and 1972 for FT3; 15.5% of the TSH results were outside the laboratory reference range (0.35–4.5 mIU/L).

The number of TSH tests per 1000 list size varied widely across practices, ranging from 84 to 482 (Table 2 and Figure 1). Using formula from Hayes and Bennett,⁶ the standard deviation of the test rate across practices that would be expected given the overall test rate and the list sizes is 7. The observed standard deviation of 83 was considerably greater than this,

indicating that most of the variation across practices (over 99%) is due to between-practice heterogeneity as opposed to mere sampling variability. Marked variation was also seen in the rates for FT4 and FT3 (Table 2). The inclusion of the prevalence of hypothyroidism ($P < 0.001$) and deprivation score ($P = 0.02$) as predictors in the random effects Poisson regression model together accounted for only 23.8% of the between-cluster variance component, indicating that they account for relatively little of the marked differences in TSH test rate. The percentage of patients aged over 65 years did not explain any extra variability. When the model was further adjusted for the proportion of TSH tests that were outside the laboratory range, it accounted for 53.8% (an extra 30%) of the between-cluster variance component. The higher the proportion of tests in the normal range, a crude proxy for unnecessary testing, the greater the TSH test rate ($P < 0.001$).

Table 1 Demographic characteristics of patients for whom thyroid function tests were requested

Features*	Royal Cornwall Hospital	Royal Devon & Exeter Hospital	Both sites
Total number of patients	89 498	58 914	148 412
Female, %	61.3	65.9	63.1
Age in years, mean (SD; range)	60 (18; 16–104)	59 (19; 16–105)	60 (19; 16–105)
Total number of tests requests	119 897	75 412	195 309
Female, %	62.2	67.2	64.1
Age in years, mean (SD; range)	61 (18; 16–104)	60 (19; 16–105)	61 (18; 16–105)
Total number of first test requests**	89 498	58 914	148 412
Female, %	61.3	65.9	63.1
Age in years, mean (SD; range)	60 (18; 16–104)	59 (19; 16–105)	60 (19; 16–105)
Total number of subsequent tests requests**	30 399	16 498	46 897
Female, %	65.1	71.8	67.5
Age in years, mean (SD; range)	64 (17; 16–102)	63 (18; 16–103)	64 (17; 16–103)
Total number of TSH tests	118 374	73 734	192 108
% of TSH results that are abnormal***	15.1	16.2	15.5
Total number of FT4 tests	19 648	23 421	43 069
% of FT4 results that are abnormal***	12.8	16.6	14.6
Total number of FT3 tests	472	1500	1972
% of FT3 results that are abnormal***	34.9	38.8	35.9

* For a given test request any combination of TSH, FT4 and FT3 may be carried out, so the sum of the individual tests types is greater than the number of test requests. ** 'First' tests include the first made for patients in 2010 and 'subsequent' tests are those for patients who had already been tested earlier in the year. *** Laboratory reference ranges: 0.35–4.5 mIU/L for TSH, 11–24 pmol/L for FT4, and 4.0–6.8 pmol/L for FT3.

Table 2 Distribution of number of tests per 1000 list size across practices in 2010

Test type	Royal Cornwall Hospital (57 practices)			Royal Devon & Exeter Hospital (50 practices)			All 107 practices		
	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range	Mean (SD)	Median (IQR)	Range
All TSH	275 (81)	269 (215–329)	125–453	213 (72)	200 (170–245)	84–482	246 (83)	229 (187–302)	84–482
First TSH*	206 (56)	204 (167–242)	64–340	168 (51)	159 (138–189)	61–333	188 (57)	178 (147–231)	61–340
Subsequent TSH*	69 (29)	63 (48–88)	24–155	45 (24)	41 (31–51)	14–149	58 (29)	50 (38–71)	14–155
All FT4	47 (9)	47 (40–53)	20–65	69 (20)	68 (58–82)	17–109	57 (19)	54 (43–67)	17–109
First FT4*	27 (5)	27 (25–30)	9–40	43 (11)	42 (38–50)	12–63	34 (11)	31 (26–41)	9–63
Subsequent FT4*	20 (5)	19 (15–22)	9–33	26 (10)	25 (21–32)	2–60	23 (9)	21 (17–26)	2–60
All FT3	1.1 (0.5)	1.0 (0.8–1.4)	0–2.5	5.0 (2.8)	4.8 (3.2–5.7)	0–15.9	2.9 (2.8)	1.7 (0.9–4.7)	0–15.9
First FT3*	0.6 (0.3)	0.6 (0.4–0.9)	0–1.2	2.5 (1.5)	2.4 (1.6–3.2)	0–8.3	1.5 (1.4)	1.0 (0.6–2.3)	0–8.3
Subsequent FT3*	0.4 (0.3)	0.4 (0.3–0.5)	0–1.3	2.4 (1.6)	2.3 (1.3–2.9)	0–7.6	1.4 (1.5)	0.8 (0.4–2.1)	0–7.6

* 'First' tests include the first made for patients in 2010 and 'subsequent' tests are those for patients who had already been tested earlier in the year. IQR, interquartile range. SD, standard deviation.

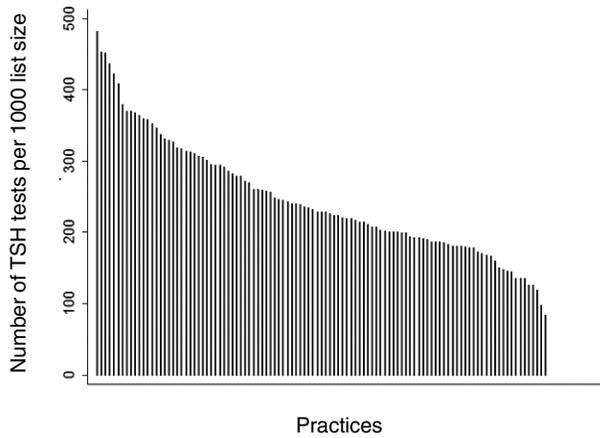


Figure 1 Number of TSH tests requested per 1000 list size in 2010 (all practices; $n = 107$)

Discussion

This study shows wide variation in the TFT test rate across general practices in south-west England. Three quarters of this variability was not explained by practice-level prevalence of hypothyroidism, socio-economic deprivation and percentage of patients aged over 65 years. O’Kane and colleagues also found a wide variation in the rates of different biochemistry tests requests, including TFTs, from general practices in Northern Ireland.⁷

The recent increases in many tests ordered, unmatched by obvious change in level of disease, alongside wide variation in practice strongly suggest that some test ordering is inappropriate. Our result, that high levels of test ordering are related to high proportions of results in the normal range for a practice, provides support for this view. There is thus potential for both health gain and savings to be made in NHS expenditure. The latter will be important as the UK Department of Health has set a target of £20 billion efficiency savings to be made through the Quality, Innovation, Productivity and Prevention (QIPP) programme.⁸ Our data show that if the TSH test rates in the practices were all reduced to 229 per 1000 list size (the current median) there would be a 7% reduction in tests ordered. In Cornwall and Exeter this would equate to approximately 13 000 TSH tests per annum, costing £91 000 if priced at £7 per test. Furthermore, inappropriate tests may also harm individual patients by increasing the risk of false-positive results leading to increased anxiety, a cascade of further investigations and unnecessary treatments.

A variety of educational and administrative strategies have been trialled to reduce unnecessary TFTs,^{4,9–11} but the results are sometimes disappointing. This may be because the interventions were chosen as they were standard approaches thought to be worthy of evaluation, rather than interventions appropriate to the

context and the particular behaviours needing to be tackled to achieve reduced testing. Current behaviour-change science indicates that the barriers to change must be clearly identified.¹²

A limitation of our study is the lack of data on why the tests were requested as many of the requests had incomplete or unclear indication for the tests. This precluded us from further exploring the potential causes for the variability in TFTs requests seen in the study. Although we have gained some insights into the reasons for variability in test ordering in primary care,^{9,13,14} this knowledge is incomplete. We believe that understanding the causes of variation between high- and low-ordering practices using mixed methods research is the next step. As well as informing our ultimate behaviour change strategy, this will also lead to better definition of the most appropriate level of test ordering; any assumption that this is the median rate needs to be explored.

ACKNOWLEDGEMENTS

This report presents independent research funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health in England.

REFERENCES

- 1 Vanderpump MP, Tunbridge WM, French JM *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clinical Endocrinology (Oxf)* 1995;43:55–68.
- 2 UK Guidelines for the Use of Thyroid Function Tests. www.british-thyroid-association.org/info-for-patients/Docs/TFT_guideline_final_version_July_2006.pdf (accessed 25/11/12).

- 3 Roti E, Gardini E, Magotti MG *et al*. Are thyroid function tests too frequently and inappropriately requested? *Journal of Endocrinological Investigation* 1999;22:184–90.
- 4 Rao GG, Crook M and Tillyer ML. Pathology tests: is the time for demand management ripe at last? *Journal of Clinical Pathology* 2003;56:243–8.
- 5 *National General Practice Profiles*. www.apho.org.uk/PRACPROF (accessed 01/09/12).
- 6 Hayes RJ and Bennett S. Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology* 1999;28:319–26.
- 7 O’Kane MJ, Casey L, Lynch PL, McGowan N and Corey J. Clinical outcome indicators, disease prevalence and test request variability in primary care. *Annals of Clinical Biochemistry* 2011;48:155–8.
- 8 QIPP – Quality, Innovation, Productivity and Prevention. www.dh.gov.uk/health/category/policy-areas/nhs/quality/qipp (accessed 25/11/12).
- 9 Schectman JM, Elinsky EG and Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Archives of Internal Medicine* 1991;151:2163–6.
- 10 Baker R, Falconer Smith J and Lambert PC. Randomised controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scandinavian Journal of Primary Health Care* 2003;21:219–23.
- 11 Verstappen WH, van der Weijden T, Sijbrandij J *et al*. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. *The Journal of the American Medical Association* 2003;289:2407–12.
- 12 Grol R and Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. *Medical Journal of Australia* 2004;180:S57–60.
- 13 Dupuits FM, Pop P, Hasman A and Schouten HJ. A rational request behavior: the development of prediction instruments regarding thyroid function tests in primary care. *Methods of Information in Medicine* 1999;38:31–6.
- 14 Valenstein P and Schifman RB. Duplicate laboratory orders: a College of American Pathologists Q-Probes study of thyrotropin requests in 502 institutions. *Archives of Pathology & Laboratory Medicine* 1996;120:917–21.

ETHICAL APPROVAL

As this was an audit of the routine clinical practice, a formal ethical approval was not necessary.

CONTRIBUTORS

The research question originated from the Peninsula CLAHRC question generation workshop in Cornwall led by JT and SF. SF, CH, AP, BV, JS and OU were involved in the development and design of the study. AB, AL and JS acquired the data, and OU did the statistical analysis. BV, JT, OU and JS prepared the draft of the manuscript. All authors were involved in the critical revision of the draft and approved the final version of the manuscript. BV is guarantor for the study.

PEER REVIEW

Not commissioned; externally peer reviewed.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Dr Bijay Vaidya
 Department of Endocrinology
 Royal Devon & Exeter Hospital
 Barrack Road
 Exeter EX2 5DW
 UK
 Email: b.vaidya@exeter.ac.uk

Received 29 January 2013

Accepted 13 March 2013