

RESEARCH

Clinical diagnosis of Graves' or non-Graves' hyperthyroidism compared to TSH receptor antibody test

Lauren Bell^{1,2}, Ann Louise Hunter^{1,2}, Angelos Kyriacou^{2,3}, Annice Mukherjee^{1,2} and Akheel A Syed^{1,2}

¹Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

²Endocrinology & Diabetes, Salford Royal NHS Foundation Trust, Salford, UK

³CEDM Centre of Endocrinology, Diabetes & Metabolism, Limassol, Cyprus

Correspondence should be addressed to A L Hunter: louise.hunter@manchester.ac.uk

Abstract

Background: TSH receptor antibody (TRAb) is considered the gold standard diagnostic test for the autoimmunity of Graves' disease (GD), which is commonly diagnosed clinically.

Aim: To evaluate the true positive (sensitivity) and true negative (specificity) rates of clinical diagnosis of GD or non-GD hyperthyroidism compared to the TRAb test.

Setting: University teaching hospital in North West England.

Participants: Patients in the Endocrinology service who had a TRAb measurement between December 2009 and October 2015.

Methods: Electronic patient records were studied retrospectively for a pre-TRAb clinical diagnosis of GD or non-GD hyperthyroidism. We examined descriptive statistics and binary classification tests; Fisher exact test was used to analyse contingency tables.

Results: We identified 316 patients with a mean age of 45 (range, 17–89) years; 247 (78%) were women. Compared to the TRAb result, clinical diagnosis had a sensitivity of 88%, specificity 66%, positive predictive value 72%, negative predictive value 84%, false negative rate 12%, false positive rate 34%, positive likelihood ratio 2.6 and negative likelihood ratio 0.2 ($P < 0.0001$).

Conclusions: Clinicians were liable to both over- and under-diagnose GD. The TRAb test can help reduce the number of incorrect or unknown diagnoses in the initial clinical assessment of patients presenting with hyperthyroidism.

Key Words

- ▶ Graves' disease
- ▶ thyroid-stimulating immunoglobulins
- ▶ hyperthyroidism
- ▶ sensitivity
- ▶ specificity

Endocrine Connections
(2018) **7**, 504–510

Introduction

Thyrotoxicosis, a clinical state resulting from inappropriately high thyroid hormone levels, is a condition with multiple aetiologies (1). It is commonly caused by Graves' disease (GD), toxic multinodular goitre (TMNG) or toxic adenoma and less commonly by thyroiditis, administration of iodinated contrast (2), immune checkpoint inhibitors (3), and extra-thyroidal causes such as struma ovarii, factitious thyrotoxicosis, trophoblastic tumours producing human chorionic gonadotrophin (hCG) (4), and TSH-producing pituitary adenomas.

Graves' disease, the commonest cause of hyperthyroidism, has an annual incidence of 20–50 per 100,000 population, a peak incidence between 30 and 50 years of age, and a lifetime risk of 3% for women and 0.5% for men (5). The diagnosis of GD is made on the basis of typical clinical features of hyperthyroidism such as weight loss, fatigue, heat intolerance, tremor, palpitations and diffuse thyroid enlargement, plus specific clinical features of GD including orbitopathy, thyroid dermopathy (pretibial myxoedema) and thyroid acropachy. Serum analyses

<http://www.endocrineconnections.org>
<https://doi.org/10.1530/EC-18-0082>

© 2018 The authors
Published by Bioscientifica Ltd



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

typically show suppressed thyroid-stimulating hormone (TSH; thyrotropin) and elevated thyroid hormones, tetraiodothyronine (T₄; thyroxine) and triiodothyronine (T₃) (6). Additional diagnostic tests can include imaging (commonly ultrasound and radioisotope uptake study) and thyroid autoantibodies, which can help to distinguish GD from other causes of thyrotoxicosis.

The autoimmune production of TSH receptor antibodies (TRAbs) is central to the pathogenesis of GD. TRAbs are heterogeneous and may either have a stimulating effect (TSH receptor stimulating antibody, TSAb) or an inhibitory effect (TSH receptor blocking antibody, TBAb) or rarely a neutral effect on the TSH receptor. TSAbs dominate in GD hyperthyroidism (7). TRAbs can be measured using two different molecular techniques: thyrotropin-binding inhibiting immunoglobulin (TBII) assays and bioassays. In clinical practice, TRAb is measured using third-generation TBII assays, which detect TRAb inhibition of TSH binding to its receptor, and are non-invasive, inexpensive and commercially available. The 3rd generation TBII assays have been found to have a sensitivity of over 97.2% and a specificity of over 98.3% (8). They are unable to distinguish between stimulatory or inhibitory TRAb; however, this information can usually be deduced from clinical and biochemical tests (7, 9). Furthermore, the first immunoassay method declared to measure serum TSAb concentration has recently been successfully developed in an automated commercial platform with a sensitivity of 100% and specificity of 99% (10).

UK guidelines, which are over 10 years old, recommend the use of the TRAb test to determine the aetiology of clinically ambiguous cases of hyperthyroidism (6). More recent American Thyroid Association (ATA) guidelines place greater emphasis on the use of TRAb, particularly in monitoring the course of Graves' disease, but similarly recommend that its place is in the diagnosis of GD in those patients whose aetiology of hyperthyroidism is unclear from their presentation and biochemistry (1). The value of TRAb measurement in the initial clinical assessment of all patients presenting with thyrotoxicosis remains a subject of debate (9). Thus, much reliance is placed on clinical judgement for making a diagnosis of GD; however, there is a dearth of data examining the accuracy of clinical diagnosis compared to objective tests.

The aim of our study was to assess the accuracy of the clinical diagnosis of Graves' or non-Graves' hyperthyroidism, made by a UK secondary care service, compared to TRAb measurement as the gold standard investigation.

Subjects and methods

We undertook a retrospective analysis of patients with a diagnosis of thyrotoxicosis to evaluate the accuracy of clinical diagnosis of GD and non-GD hyperthyroidism compared to TRAb results.

Setting and patients

We studied patients who presented with thyrotoxicosis to the Endocrinology outpatient department of a university teaching hospital in North West England between December 2009 and October 2015. The department was staffed by 10 consultants, two endocrine nurse specialists and three annually rotating specialist trainees, comprising a total of 30 individual specialists over the course of the study period. A total of 512 individual patients with TRAb measurements were identified from lab records in the study period, of which 316 (62%) were included in the study (Fig. 1), after excluding tests requested by departments outwith Endocrinology, tests performed at other laboratories, and patients with insufficient clinical information recorded. The project was approved by the Clinical Audit department of our institution and electronic patient records (EPR) were reviewed for demographic data and the pre-test clinical diagnosis of GD or non-GD hyperthyroidism. The latter included TMNG, thyroiditis, solitary toxic nodule, amiodarone-induced thyrotoxicosis, hyperemesis gravidarum and alemtuzumab-associated thyrotoxicosis. A post-test diagnosis, if applicable, was also recorded. Data on thyroid peroxidase (TPOAb) titres, where available, were also recorded.

TRAb assay

The TRAb assay used in the study period was a commercial third-generation TSH receptor autoantibody enzyme linked immunosorbent assay (ELISA) kit supplied by Thermo Scientific B.R.A.H.M.S (Hennigsdorf, Germany) and performed by the Department of Clinical Immunology, Northern General Hospital, Sheffield, UK, as per the manufacturer's instructions (personal communication). The presence of TRAb was detected based on the inhibition of binding of the biotin-labelled human monoclonal antibody M22 with immobilised TSH receptors in ELISA plates. Streptavidin peroxidase and tetramethylbenzidine were added to determine the amount of M22 bound to the plate. The absorbance of the mixture at 450 nm was read using an ELISA plate

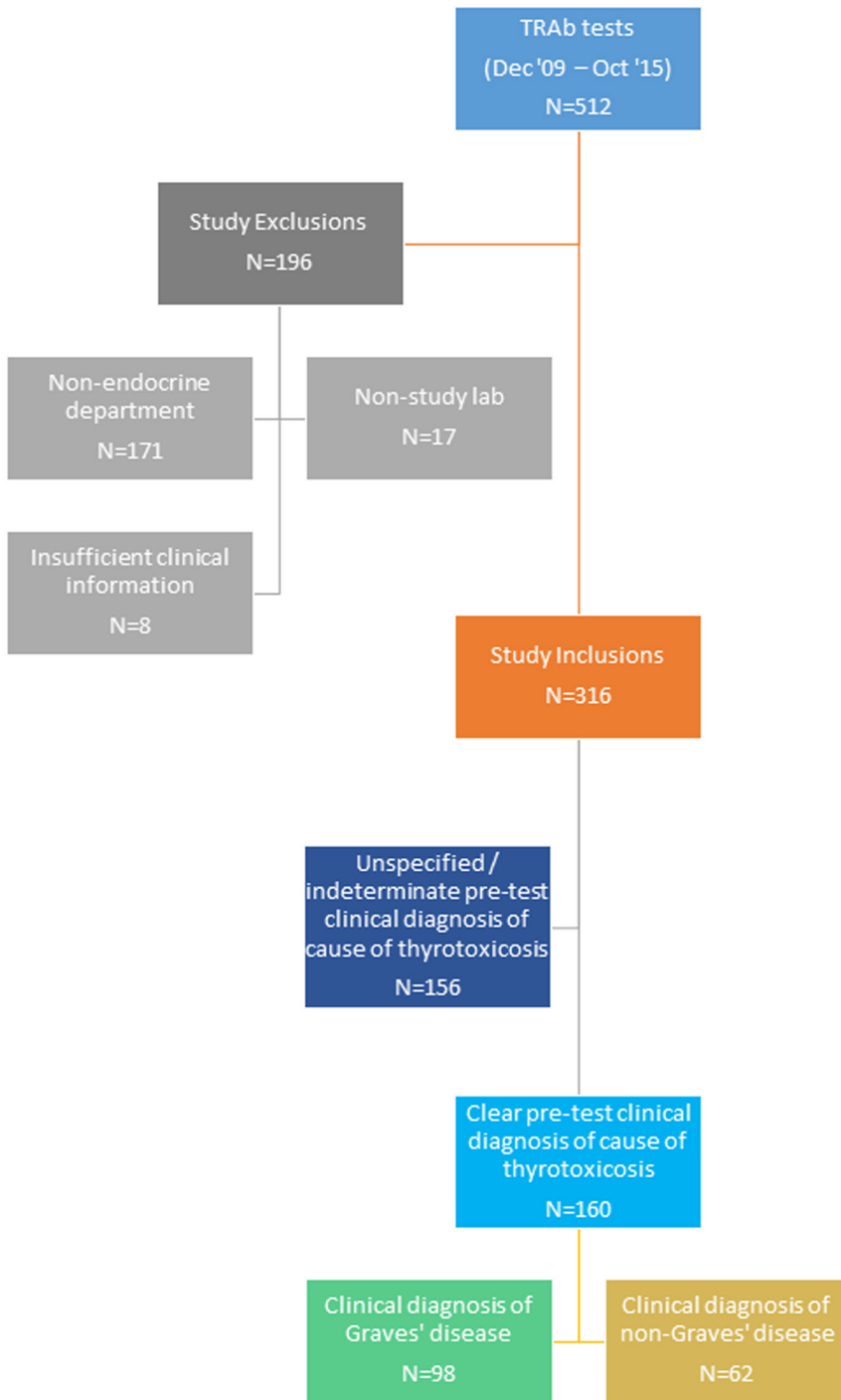


Figure 1
Flowchart of patient selection and classification.

reader. The intra-assay coefficient of variation (CV) was 13.57% at 2.28 IU/L and 8.35% at 35.8 IU/L. The manufacturer's lower boundary for a positive sample was ≥ 1.5 IU/L. The units per litre corresponded to the international standard for TRAb (90/672 from the National Institute for Biological Standardisation and Control, Potters Bar, UK). The manufacturer-reported sensitivity and specificity were 98.8% and 99.6%,

respectively (personal communication, Thermo Fisher Scientific).

Statistical analysis

We performed descriptive statistics of demographic characteristics with parametric tests (or non-parametric tests for non-normative data), with measures of dispersion as

appropriate. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR), false negative rate (FNR) and likelihood ratio of clinical diagnosis were computed against a TRAb-positive or TRAb-negative result after excluding TRAb-borderline results. The Fisher exact test was used to analyse contingency tables of categorical variables. A two-sided $P < 0.05$ was considered statistically significant and 95% confidence intervals (95% CI) were reported as a measure of precision. Data were analysed with GraphPad Prism, version 7.00 (GraphPad Software) and IBM SPSS Statistics, version 23.0.0 (IBM).

Results

We studied 316 patients with a pre-TRAb test clinical diagnosis of GD or non-GD hyperthyroidism.

Patient demographics

The age of the patients at the time of the TRAb measurement was a mean \pm standard error of 45.2 ± 0.9 (range, 17–89) years; 247 (78%) were women; 279 (88%) patients were white, 16 (5%) Asian, 13 (4%) were black and the remaining eight (3%) were of other ethnicity.

Clinical diagnosis of GD or non-GD hyperthyroidism compared to TRAb result

A clear pre-test clinical diagnosis as made by the clinician was identified in 160 patients, which included GD in 98 (61.25%) and non-GD hyperthyroidism in 62 (38.75%) patients. The overall prevalence of TRAb-positive results was 50.62% (95% CI, 42.62–58.61%). Of the 98 patients with a pre-test clinical diagnosis of GD, 71 (72.45%) had a TRAb-positive result and 27 (27.55%) had a TRAb-negative result. Of the 62 patients with a pre-test clinical diagnosis of non-GD hyperthyroidism, 52 (83.87%) had a TRAb-negative result and 10 (16.13%) had a TRAb-positive result. Compared to the TRAb result, clinical diagnosis of GD had a sensitivity of 87.65% (78.74–93.15%) (Fig. 2), specificity of 65.82% (54.85–75.33%), PPV of 72.45% (62.88–80.32%), NPV of 83.87% (72.79–91.00%), FNR of 12.35% (6.85–21.26%), FPR of 34.18% (24.67–45.15%), a positive likelihood ratio of 2.57 (1.87–3.52) and a negative likelihood ratio of 0.19 (0.10–0.34) ($P < 0.0001$ for all). Given that the lifetime risk of GD is higher in females, sub-group analyses categorised by sex showed no statistically significant difference in sensitivity, specificity, PPV and NPV of clinical diagnosis in men compared to women

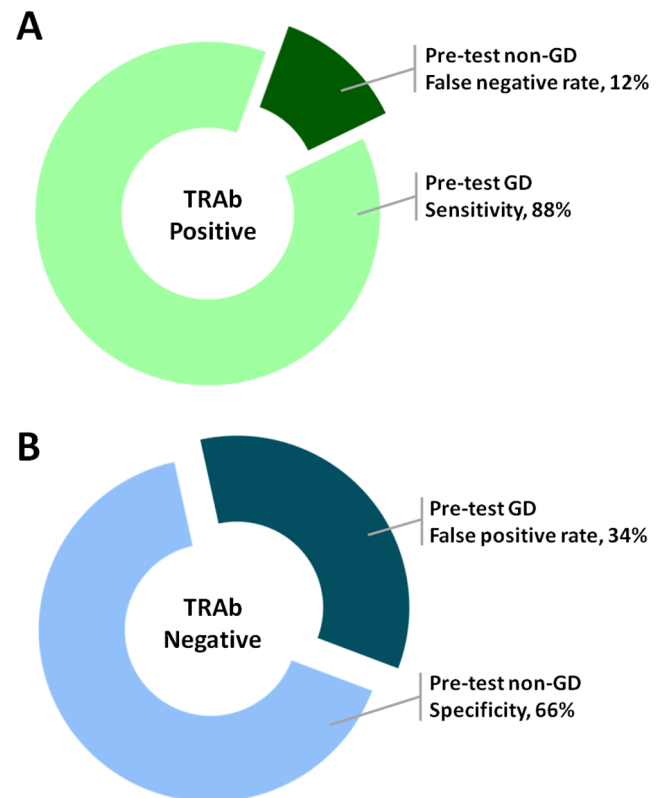


Figure 2

Sensitivity (A) and specificity (B) of pre-test clinical diagnosis of Graves' disease compared to TSH receptor antibody (TRAb) result.

(Table 1). As the incidence of GD is lower in patients over the age of 50 years, sub-group analyses by age showed that clinical diagnosis had significantly greater specificity and NPV in older patients, but no significant difference in sensitivity or PPV.

Post-TRAb test clinical diagnosis

A false positive clinical diagnosis of GD was recorded in 27 patients, of which five (18.5%) patients' diagnoses were subsequently amended by clinicians in follow-up appointments. Of the remaining patients, three (11.1%) kept their diagnosis of GD at the clinician's discretion. Additionally, 14 (51.9%) patients kept their diagnosis of GD as the clinician had requested the TRAb measurement to monitor disease relapse or (in pregnant patients), to predict risk to the foetus of developing thyroid dysfunction. Three (11.1%) patients did not have their TRAb result acknowledged by the clinician and of the remaining two (7.4%), one had a corrected diagnosis to thyrotoxicosis of an indeterminate cause and one was lost to follow-up. Ten patients with an incorrect pre-test clinical diagnosis of non-GD were recorded, of

Table 1 Sensitivity and specificity of the clinical diagnosis of Graves' or non-Graves' hyperthyroidism compared to TSH receptor antibody result in all patients and in subgroups.

	All (N=160)	Female (N=131)	Male (N=29)	P	Age ≤ 50 (N=109)	Age > 50 (N=51)	P
Sensitivity (%)	87.65	86.76	92.31	ns	88.71	84.21	ns
Specificity (%)	65.82	63.49	75.00	ns	53.19	84.38	<0.0001
PPV (%)	72.45	71.95	75.00	ns	71.43	76.19	ns
NPV (%)	83.87	81.63	92.31	ns	78.13	90.00	0.0327

NPV, negative predictive value; ns, non-significant; PPV, positive predictive value.

which the diagnosis of nine (90.0%) were subsequently corrected by clinicians in follow-up appointments to GD hyperthyroidism; one (10.0%) patient did not have the TRAb result acknowledged by the clinician.

Indeterminate/unspecified pre-test clinical diagnosis compared to TRAb result

Of 156 patients with a differential/indeterminate/unspecified pre-test clinical diagnosis of thyrotoxicosis, the TRAb test was positive in 72 (46.2%) and negative in 84 (53.8%) patients. Consequently, 128 (82.1%) patients were given a diagnosis by a clinician in follow-up appointments. Of the remaining 28 patients, 14 (9.0%) did not have their TRAb test acknowledged and 14 (9.0%) kept their diagnosis of thyrotoxicosis of an indeterminate cause.

TRAb titres in patients with true positive vs false negative (or absent) pre-test clinical diagnosis of GD

In patients with a pre-test clinical diagnosis of GD who had confirmatory raised TRAb titres (true positive group, $n=71$), the mean \pm s.e.m. TRAb titre was 11.48 ± 1.46 (range, 1.9–68.7) IU/L. In patients with a pre-test clinical diagnosis of non-GD who had raised TRAb titres (false negative group, $n=10$), the mean \pm s.e.m. TRAb titre was 6.43 ± 1.85 (range, 1.8–21.7) IU/L. There was no statistical difference in TRAb titres between the true positive vs false negative groups ($P>0.99$).

In patients with absent (indeterminate/unspecified) pre-test clinical diagnosis but a subsequent positive TRAb result and subsequent post-test diagnosis of GD (undiagnosed group, $n=82$), the mean \pm s.e.m. TRAb value was 11.50 ± 1.33 (range, 1.6–70.2) IU/L. There was no statistical difference in TRAb titres between the true positive vs undiagnosed groups ($P>0.99$).

Discussion

We evaluated the accuracy of clinical diagnosis of GD or non-GD hyperthyroidism compared to TRAb

measurement in a UK university teaching hospital. Whereas previous observational studies involving thyroid autoimmunity in clinical practice have evaluated the sensitivity and specificity of TRAb tests against the premise of definitiveness of clinical diagnosis, we approached our study from the opposite direction: that the TRAb result was the definitive diagnosis against which we evaluated the sensitivity and specificity of clinical diagnosis. We non-judgementally accepted the pre-test clinical diagnosis as given to patients by specialist clinicians in the Endocrine service and report that pre-test clinical diagnosis had an overall sensitivity of 88% and specificity of 66%. The specificity of clinical diagnosis was remarkably weaker in younger patients (53% specificity), despite hyperthyroidism being more common than in older individuals (84% specificity). In patients with an absent (indeterminate/unspecified) pre-test clinical diagnosis, the TRAb test could have distinguished between GD (46%) and non-GD hyperthyroidism (54%). We did not find any difference in TRAb titres between GD patients who were diagnosed on their clinical features alone (pre-test diagnosis of GD) and those patients who did not receive a diagnosis of GD until after the TRAb test had been performed (post-test diagnosis of GD). As biochemical thyrotoxicosis is diagnosed earlier and earlier in the natural history of hyperthyroidism, the classical clinical features of GD may not be evident in many patients at presentation, making clinical diagnosis more challenging and inherently weaker. The TRAb test would make diagnosis more secure in these situations.

The performance of a TBII-based 3rd generation TRAb assay has previously been evaluated against a clinical diagnosis of GD in a retrospective and prospective cohort of patients in a UK clinic (11). The assay was found to be a reliable tool to determine the aetiology of the hyperthyroidism with a sensitivity and specificity of 95% and 92%, respectively. However, the clinical diagnosis of GD may be inherently flawed. An earlier study had found that 8% of patients clinically diagnosed to have GD had non-GD hyperthyroidism when re-evaluated with thyroid uptake scintigraphy; after taking this into account,

the sensitivity of TBII assays was 98.7%, leading the authors to conclude that TRAb-negative Graves' disease was extremely rare (12).

In the past, there has been controversy surrounding the value of a TRAb measurement in the initial clinical assessment of hyperthyroid patients. One reason for this apprehension was a lack of confidence in the early TBII assays (7, 13). The sensitivity and specificity of modern 3rd generation TRAb assays has significantly improved. A recent systematic review and meta-analysis reported that the overall pooled sensitivity and specificity of the 2nd and 3rd generation TRAb assays was 97.1% and 97.4%, and 98.3% and 99.2%, respectively, with little difference between the types of immunoassay methods employed (human or porcine receptor, manual or automated procedure) (8); the likelihood of a TRAb-positive individual to have GD was >1000 to >3000 fold greater (depending upon the type of assay) compared to a TRAb-negative person. Another study that compared the performance characteristics of different TBII and bioassays for TRAbs reported 100% specificity for all the assays (14). This study is in the minority in reporting poorer sensitivity for some of the TBII assays, and has had its methodology criticised (15). Furthermore, in untreated hyperthyroid patients the sensitivity and specificity of 3rd generation TBII assays is of the order of 98% and 99%, respectively (8). Therefore, in the appropriate clinical setting (i.e. a patient with hyperthyroidism), the choice of a bioassay vs a binding assay seems to have little importance (16). It is important to note that the variability between one clinician's opinion and another is likely to be greater than that between one assay and another. One opinion is that perhaps the guidelines and clinical practice have yet to acknowledge the technical advances of the past decade in this field (16).

Questions have also been raised over the cost-effectiveness of the assay in routine clinical practice. The results from our study indicate that for a significant number of patients the correct diagnosis of GD or non-GD hyperthyroidism cannot be made on clinical presentation alone. A consequence of incorrect diagnosis may include unnecessary investigations, treatment and follow-up. The cost of a TRAb test in a UK clinic is £14.83 (personal communication, Department of Clinical Immunology, Northern General Hospital, Sheffield). In contrast, a thyroid uptake scintigraphy scan costs £165 (personal communication, Department of Nuclear Imaging, Salford Royal Hospital, Salford) and the average cost to the NHS of an outpatient attendance was £117 in 2015–16 (17). It has been suggested that a TRAb measurement at the

initial clinical assessment of hyperthyroid patients would result in a cost saving, achieved through a reduction in hospital clinic appointments and number of thyroid uptake scintigraphy scans requested (11), and may be a small price to pay to increase the clinician's confidence in the diagnosis and management plan.

In our study, 90% of patients with an incorrect pre-test clinical diagnosis of non-GD hyperthyroidism subsequently had their diagnosis corrected by clinicians in follow-up appointments. These findings suggest clinicians utilise a positive TRAb measurement over their clinical judgement to confirm a diagnosis of GD. A TRAb measurement in the initial clinical assessment of all patients could potentially prevent an incorrect diagnosis being made in the first place. However, only a quarter of the patients with a clinical diagnosis of GD but negative TRAb result had their diagnosis subsequently corrected by clinicians in follow-up appointments. Further research into clinicians' apparent lack of confidence in a negative TRAb result is needed.

The current UK guidelines and the more recent ATA guidelines recommend the use of TRAb in clinically ambiguous cases to determine the aetiology of the hyperthyroidism (1, 6). The results from our study confirm the value of TRAb in this clinical scenario. In patients with an indeterminate/unspecified pre-test clinical diagnosis, the TRAb test was able to distinguish between GD and non-GD hyperthyroidism. Post-TRAb test, 82% of patients had the aetiology of their thyrotoxicosis confirmed by clinicians in follow-up appointments. Therefore, universal TRAb measurement in the initial clinical assessment of all patients with hyperthyroidism would lessen unspecified/indeterminate diagnoses as well as false positive and false negative diagnoses. As a consequence, this would lead to more certainty of the diagnosis and more timely delivery of definitive management for patients.

A limitation of our study was that it was a retrospective analysis. A prospective analysis would allow for a better understanding of clinical reasoning and judgement in the assessment of hyperthyroid patients and assessing any reduction in the number of incorrect or indeterminate diagnoses and any cost savings achieved. Also, patients studied were seen at one hospital site only – namely a university teaching hospital site – and thus this study may not be representative of the whole population of thyrotoxic patients and their clinicians. However, records studied encompassed a six-year period, and in addition to providing a regional tertiary service in some disease areas, the hospital's endocrinology department provides a secondary care

service to its local population. Therefore, we suggest that these patients were not untypical of those presenting to UK endocrinology secondary care, nor were the clinicians involved untypical of UK practitioners.

Conclusion

We report that clinical diagnosis of GD or non-GD hyperthyroidism has remarkably poorer sensitivity and specificity compared to the TRAb test. We conclude that the clinical diagnosis of GD or non-GD hyperthyroidism may not reliably identify patients with or without GD, respectively, and that testing for TRAb is of value in the initial clinical assessment of all patients presenting with hyperthyroidism. We encourage our colleagues to consider the accuracy of their own diagnostic practices.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Author contribution statement

L B and A A S conceived and designed the study. L B performed data collection. L B, A L H and A A S analysed data. A K and A M contributed to data interpretation. All authors contributed to the writing of the manuscript.

Acknowledgements

The authors thank the Department of Clinical Immunology, Northern General Hospital, Sheffield, UK and Thermo Fisher Scientific for particulars of the TRAb assay and its performance characteristics.

References

- Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, Rivkees SA, Samuels M, Sosa JA, Stan MN, *et al.* 2016 American Thyroid Association Guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* 2016 **26** 1343–1421. (<https://doi.org/10.1089/thy.2016.0229>)
- Dunne P, Kaimal N, MacDonald J & Syed AA. Iodinated contrast-induced thyrotoxicosis. *CMAJ* 2013 **185** 144–147. (<https://doi.org/10.1503/cmaj.120734>)
- Corsello SM, Barnabei A, Marchetti P, De Vecchis L, Salvatori R & Torino F. Endocrine side effects induced by immune checkpoint inhibitors. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 1361–1375. (<https://doi.org/10.1210/jc.2012-4075>)
- Tingi E, Syed AA, Kyriacou A, Mastorakos G & Kyriacou A. Benign thyroid disease in pregnancy: a state of the art review. *Journal of Clinical and Translational Endocrinology* 2016 **6** 37–49.
- Smith TJ & Hegedus L. Graves' disease. *New England Journal of Medicine* 2016 **375** 1552–1565. (<https://doi.org/10.1056/NEJMra1510030>)
- Beastall G, Beckett G, Franklyn J, Fraser W, Hickey J, John R, Kendall-Taylor P, Nevens B & Vanderpump M. UK Guidelines for the Use of Thyroid Function Tests. London, UK: Association for Clinical Biochemistry, British Thyroid Association & British Thyroid Foundation, 2006. (available at: http://www.british-thyroid-association.org/sandbox/bta2016/uk_guidelines_for_the_use_of_thyroid_function_tests.pdf)
- Matthews DC & Syed AA. The role of TSH receptor antibodies in the management of Graves' disease. *European Journal of Internal Medicine* 2011 **22** 213–216. (<https://doi.org/10.1016/j.ejim.2011.02.006>)
- Tozzoli R, Bagnasco M, Giavarina D & Bizzaro N. TSH receptor autoantibody immunoassay in patients with Graves' disease: improvement of diagnostic accuracy over different generations of methods. Systematic review and meta-analysis. *Autoimmunity Reviews* 2012 **12** 107–113. (<https://doi.org/10.1016/j.autrev.2012.07.003>)
- Ajjan RA & Weetman AP. Techniques to quantify TSH receptor autoantibodies. *Nature Clinical Practice Endocrinology and Metabolism* 2008 **4** 461–468. (<https://doi.org/10.1038/ncpendmet0886>)
- Tozzoli R, D'Aurizio F, Villalta D & Giovannella L. Evaluation of the first fully automated immunoassay method for the measurement of stimulating TSH receptor autoantibodies in Graves' disease. *Clinical Chemistry and Laboratory Medicine* 2017 **55** 58–64. (<https://doi.org/10.1515/cclm-2016-0197>)
- Theodoraki A, Jones G, Parker J, Woolman E, Martin N, Perera S, Thomas M, Bunn C, Khoo B, Bouloux PM, *et al.* Performance of a third-generation TSH-receptor antibody in a UK clinic. *Clinical Endocrinology* 2011 **75** 127–133. (<https://doi.org/10.1111/j.1365-2265.2011.04022.x>)
- Paunkovic J & Paunkovic N. Does autoantibody-negative Graves' disease exist? A second evaluation of the clinical diagnosis. *Hormone and Metabolic Research* 2006 **38** 53–56. (<https://doi.org/10.1055/s-2006-924979>)
- Davies TF, Roti E, Braverman LE & DeGroot LJ. Thyroid controversy – stimulating antibodies. *Journal of Clinical Endocrinology and Metabolism* 1998 **83** 3777–3785. (<https://doi.org/10.1210/jcem.83.11.5056-1>)
- Diana T, Wuster C, Kanitz M & Kahaly GJ. Highly variable sensitivity of five binding and two bio-assays for TSH-receptor antibodies. *Journal of Endocrinological Investigation* 2016 **39** 1159–1165. (<https://doi.org/10.1007/s40618-016-0478-9>)
- Bitcon V, Donnelly J & Kiaei D. Sensitivity of assays for TSH-receptor antibodies. *Journal of Endocrinological Investigation* 2016 **39** 1195–1196. (<https://doi.org/10.1007/s40618-016-0520-y>)
- Barbesino G & Tomer Y. Clinical review: clinical utility of TSH receptor antibodies. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 2247–2255. (<https://doi.org/10.1210/jc.2012-4309>)
- Department of Health. Reference costs 2015–16. London, UK: Department of Health, 2016. (available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/577083/Reference_Costs_2015-16.pdf)

Received in final form 22 February 2018

Accepted 12 March 2018

Accepted Preprint published online 12 March 2018

