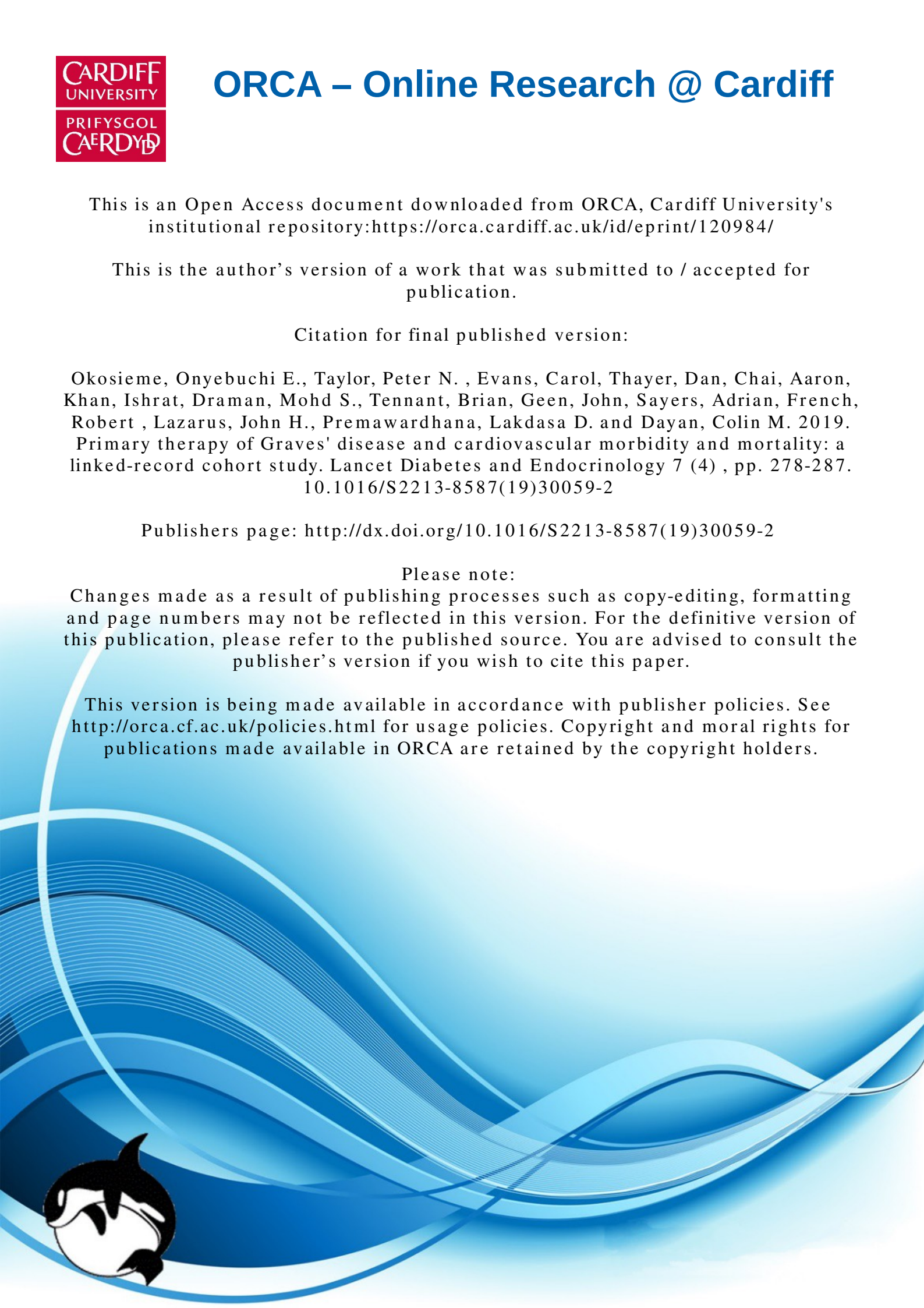


Primary therapy of Graves' disease and cardiovascular morbidity and mortality: a linked-record cohort study

Onyebuchi E Okosieme[1,2], Peter N Taylor[1], Carol Evans[3], Dan Thayer[4], Aaron Chai[1], Ishrat Khan[1], Mohd S Draman[1], Brian Tennant[5], John Geen[5,6], Adrian Sayers[7], Robert French[1], John H Lazarus [1], Lakdasa D Premawardhana[1,8], and Colin M Dayan[1]

1. Thyroid Research Group, School of Medicine, Cardiff University
2. Diabetes Department, Prince Charles Hospital, Cwm Taf University Health Board
3. Department of Medical Biochemistry & Immunology, University Hospital of Wales, Cardiff
4. SAIL Databank, School of Medicine, Swansea University
5. Clinical Biochemistry Department, Cwm Taf University Health Board
6. Faculty of Life Sciences and Education, University of South Wales
7. Department of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, United Kingdom.
8. Section of Endocrinology, Department of Medicine, YYF Hospital, Aneurin Bevan University Health Board

Key words: Graves' disease, hyperthyroidism, radioiodine therapy, thyroidectomy, mortality, major adverse cardiovascular events

Word count: 3542
Tables: **2**
Figures: **5**
References: **33**

Correspondence: Dr Onyebuchi E Okosieme, Diabetes Department, Prince Charles Hospital, Cwm Taf University Health Board, Gurnos Estate, Merthyr Tydfil, CF47 9DT, Telephone 01685728353, Fax 01685728448

ABSTRACT

BACKGROUND: Graves' disease is routinely treated with antithyroid drugs (ATD), radioactive-iodine (RAI), or surgery but it is uncertain whether the choice of initial therapy influences long-term outcomes. We evaluated cardiovascular morbidity and mortality according to the modality and effectiveness of primary therapy in Graves' disease.

METHODS: The study was conducted through linked datasets within the All-Wales Secure Anonymised Information Linkage (SAIL) Databank. Graves' disease patients were identified from a regional TSH-receptor-antibody test register, (n=4,189, female 82%, 1998–2013) and matched by age and sex to a control population in SAIL (n=16,756). Patients were grouped by treatment within one-year of diagnosis as: (1) ATD (n=3587), (2) RAI with resolved hyperthyroidism (RAI-Group-A, n=250), and (3) RAI with unresolved hyperthyroidism (RAI-Group-B, n=182). One-year landmark Kaplan-Meier and Cox regression models were used to analyse the association of treatment with all-cause mortality and major adverse cardiovascular events (MACE, myocardial infarction, heart failure, ischaemic stroke, or death). The relationship between FT4 concentration and outcomes was analysed using restricted cubic-spline regression models.

FINDINGS: Overall, patients had increased mortality compared to controls (HR 1·23, 95%CI 1·06, 1·42). Compared to ATD-treated patients, mortality was reduced in RAI-Group-A (HR 0·50, 95%CI 0·29, 0·85) but not RAI-Group-B (HR 1·51, 95%CI 0·96, 2·37). Persistently low-TSH at 1-year was associated with increased mortality independent of treatment modality (HR 1·55, 95%CI 1·08-2·24). Spline-regressions demonstrated a positive non-linear relationship between 1-year-FT4 and outcomes.

INTERPRETATION: Regardless of therapy modality, early and effective control of Graves' hyperthyroidism is associated with improved survival compared to less effective control.

FUNDING: National Institute for Social Care and Health Research (NISCHR), Wales

RESEARCH IN CONTEXT

Evidence before this study

We searched PubMed for articles published up to October 2018, without language restrictions, that addressed mortality outcomes in relation to treatment modality in patients with hyperthyroidism. Our search terms were ("hyperthyroidism" OR "thyrotoxicosis" OR "Graves' disease") AND ("mortality" OR "survival") AND ("treatment" OR "antithyroid drugs" OR "thionamides" OR "radioactive iodine" OR "thyroidectomy"). Although several studies reported excess mortality in radioiodine-treated cohorts, these studies did not include patients treated with antithyroid drugs which currently represents the predominant therapy modality for Graves' disease worldwide. Only one study to date has compared hyperthyroidism mortality according to treatment with antithyroid drugs or radioiodine. This study showed survival advantages following radioiodine-induced hypothyroidism compared to during antithyroid drug treatment. However, the study comprised a heterogeneous hyperthyroid cohort that was underpowered to show treatment effects in the Graves' disease subset of 345 patients.

Added value of this study

To the best of our knowledge ours is the first study to show that early and complete resolution of hyperthyroidism in patients with Graves' disease is associated with improved survival compared to less effective control. We show that survival in patients with Graves' disease is associated with the rapid restoration of normal thyroid status rather than therapy modality *per se*. Using flexible spline regression modelling we demonstrate a positive association between FT4 levels after initial therapy and subsequent cardiovascular morbidity and mortality.

Implications of all the available evidence

Our findings emphasise the importance of prompt elimination of hyperthyroidism and maintenance of normal thyroid status in the care of Graves' disease patients. The choice of therapy should be driven by the prospects of successful resolution of hyperthyroidism and therefore, early definitive therapy should be offered to patients unlikely to achieve remission with antithyroid drugs alone.

INTRODUCTION

Graves' disease is the most common cause of hyperthyroidism in iodine-sufficient countries, affecting 0·5–1.0% of the population[1,2]. Hyperthyroidism in Graves' disease results from unregulated stimulation of the thyrotropin (TSH) receptor by autoreactive TSH-receptor antibodies (TRAbs) which are detectable in virtually every patient with the disease[3]. Patients with uncorrected hyperthyroidism have an increased mortality risk and suffer significant cardiovascular morbidity including cardiac arrhythmias, strokes and heart failure[4,5]. Three treatment modalities for Graves' disease, namely thionamide drugs, radioactive iodine, and thyroidectomy have been available for decades, all with established safety and efficacy records. However, the approach to therapy is highly variable and largely dictated by regional preferences[2,6]. Primary thionamide therapy is favoured by the European Thyroid Association[7], while the American Thyroid Association endorses any of the modalities depending on individual patient characteristics[8]. In practice radioiodine is traditionally preferred in North America[9] although antithyroid drugs are increasingly used and are now the most common treatment for Graves' disease in the United States [10]. In contrast primary radioiodine therapy is less popular in Europe and Asia and remains restricted to patients who relapse following 12-18 months of thionamides[112].

These variations in treatment are in part driven by lack of evidence regarding mortality risk with respect to Graves' disease management. Randomised controlled trials on mortality outcomes in Graves' disease are lacking and are unlikely to be undertaken given the logistic implications. Although observational studies have addressed mortality end-points in patients with hyperthyroidism[4,12-18], only a few studies specifically distinguished Graves' disease from other causes of hyperthyroidism[17,19-23]. This distinction is relevant given that Graves' disease patients

are younger and express a different cardiovascular phenotype from those with thyroid nodules[2,5,20,22]. In addition, unlike patients with toxic nodules who require definitive treatment for cure, Graves' disease patients have a choice of all three treatments, underpinning the need for treatment-specific outcome data.

Although previous cohort studies have reported excess mortality in radioiodine-treated patients[15,17,18,23,24], these cohorts excluded patients treated with thionamides which currently represents the predominant therapy modality for Graves' disease worldwide[2]. Only one study to date has compared mortality according to thionamide versus radioiodine treatment[19]. The study by Boelaert *et al* observed reduced mortality after radioiodine induced hypothyroidism compared to during thionamide therapy. However, the study comprised a heterogeneous hyperthyroidism clinic cohort (n=1036) that was underpowered to show treatment effects in the Graves' disease subset (n=345)[19]. Thus, it remains unclear whether primary radioiodine therapy offers survival advantages over thionamides and hence whether current Graves' disease treatment strategies are optimal. Our objective in the present study was to evaluate cardiovascular morbidity and mortality according to the modality and effectiveness of primary therapy in a large cohort of patients with Graves' disease.

METHODS

Data Sources and study population

The study was conducted through linked datasets within the Secure Anonymised Information Linkage (SAIL) Databank. SAIL is a nationwide repository of routinely collected health and social care data in Wales run by the Health Informatics Research Unit of Swansea University. The databank contains over two billion anonymised person-based records linked to hospital admission and primary care data[25]. All National Health Service (NHS) users in Wales are identified by a unique

10-digit identifier and anonymity of data within the databank is ensured by multi-party encryption through a split-file data management system[26]. For this study, we created a dataset of patients with hyperthyroidism, diagnosed between January 1998 and December 2013 in clinics across three health boards serving an estimated 1·4 million residents in South Wales. We included all patients who could be linked to the Welsh Demographic Service, a population register of individuals registered with a general practitioner (GP) or receiving healthcare in Wales. Additional outcomes were obtained from primary care data held within SAIL, covering about 75% of the population, and the Patient Episode Database for Wales (PEDW), a dataset of all inpatient admissions in Wales.

Patients with Graves' disease were identified from electronic records of TRAb tests performed at the biochemistry department of the University Hospital of Wales, Cardiff, which served as the regional TRAb-assay laboratory during the period covered by the study. TRAbs are sensitive and specific markers of Graves' disease[27] and are routinely included in the diagnostic work-up of Graves' disease in our catchment clinics. Patients were defined as Graves' Hyperthyroidism if they had a positive TRAb test together with evidence of hyperthyroidism from either: (a) low-TSH in laboratory records, (b) record of hyperthyroidism in GP or hospital datasets (ICD-10 E05, read codes C02), or (c) record of thionamide prescription in GP data (carbimazole, read codes fa1 and propylthiouracil, fa2). The patient inclusion flow chart is presented in figure 1. For a background control population, we identified a sample of individuals in the Welsh Demographic Service, matched to cases on birth week and sex, at a ratio of 4 control subjects per patient (n=16,756). Individuals with hyperthyroidism on our loaded datasets were excluded from the controls.

Treatment and outcome data

Iodine-131 treatment data was extracted from radioiodine treatment logs in nuclear medicine departments serving the study population. Each department holds statutory records of all radioisotope treatment including isotope type, dose, and date of administration. Information on thyroidectomy was obtained from PEDW using the Office of Population censuses and Surveys version 4 (OPCS-4) classification code for thyroidectomy, B081-B091. Outcome events were recorded from PEDW using the relevant ICD-10 codes for atrial fibrillation (I48), acute myocardial infarction (I21-I23), acute coronary syndrome (I20), heart failure (I50), ischaemic stroke (I63, I65-I68), haemorrhagic stroke (I60-I62), diabetes mellitus (E10-E14), thyroid cancer (C73), and breast cancer (C50). Mortality information was obtained from the Office for National Statistics (ONS).

Laboratory assays

TRAbs were performed with second generation radioimmunoassays while thyroid hormones (FT4 and TSH) were performed using several different immunoassays. The laboratory assays, reference ranges and performance measures are detailed in the supplementary appendix and supplementary table 1.

Statistical analysis

Baseline data is summarised as means (standard deviation, SD) for normally distributed values or median (interquartile range, IQR) for non-normal data. Differences between groups were analysed with chi-squared test, 2-tailed t-test or Mann-Whitney tests as appropriate. The primary study outcome was all-cause mortality while secondary outcome was major adverse cardiovascular events (MACE), a composite of acute myocardial infarction, ischaemic stroke, heart failure, or death. Patients were grouped according to primary treatment using a landmark

approach to avoid immortal time bias. The landmark was set at one-year post-diagnosis which was considered sufficient time for primary therapy without incurring excessive data losses from pre-landmark censoring.

Treatment groups comprised: (1) ATD, patients who received only antithyroid drugs in the first year of diagnosis, (2) RAI-Grp-A, patients who received radioiodine and achieved hyperthyroidism control by one-year, and (3) RAI-Grp-B, patients who received radioiodine but had not achieved hyperthyroidism control by one-year. We excluded patients if they were censored before one-year (n=93) or had thyroidectomy in the first-year due to small numbers (78 patients, 2 deaths). Control of hyperthyroidism after radioiodine was defined as levothyroxine initiation, or TSH above the reference limit, or persistent euthyroidism (>6 months), each without antithyroid drug use. The diagnosis date was the date of the first positive TRAb or low-TSH, whichever was recorded earlier, and follow-up was from diagnosis until event or study end.

We analysed outcomes according to primary treatment using Kaplan-Meier curves and Cox regression models. First, we compared event rates in patients and population controls using the logrank test and an adjusted Cox regression model to account for baseline comorbidity. The analysis was undertaken in the whole cohort and stratified by primary treatment group, with the corresponding matched population for each group as controls. In a second approach, we used a Cox regression model to compare event risk between treatment groups with the antithyroid drug group as reference category. Models were adjusted for 5-year age bands, gender, diagnosis-year, baseline TRAb concentration, and comorbidity scores, using a modification of the Charlson Comorbidity Index (supplementary table 2). Stratified analysis was undertaken by age (>50 vs. $\leq$50 years) and co-morbidity

score (0 vs ≥1) with interaction terms for sub-groups. In sensitivity analysis, we sequentially excluded pregnant patients (n=63) and patients with borderline TRAb levels (n=266) at diagnosis. The validity of the proportional hazard assumption was examined by Schoenfeld residuals test and visual inspection of residuals plots against time.

As an alternative means of handling time bias, treatment and thyroid status were incorporated as time-dependent covariates in a multivariable Cox proportional hazard regression model in which patients contributed person-time as follows: (1) ATD, from diagnosis to definitive treatment with radioiodine or thyroidectomy; (2) RAI without hyperthyroidism control, from radioiodine therapy until control of hyperthyroidism, (3) RAI with hyperthyroidism control, following radioiodine control of hyperthyroidism; (4) Thyroidectomy, following thyroidectomy.

To directly evaluate the impact of thyroid status, we analysed outcomes according to thyroid function after one-year of treatment. We derived a one-year FT4 and TSH from time-weighted averages of all available tests at 9–12 months excluding patients censored before one-year. TSH was entered in a Cox regression model as low, normal, or high-TSH with the normal-TSH as reference category. Models were adjusted for age, gender, comorbidity, baseline TRAb, FT4 and TSH, and then stratified by primary treatment modality based on treatment received by the test date. For FT4, we used restricted cubic splines to model a possible non-linear relationship between one-year FT4 and outcomes. Four equally-spaced knots were set at percentiles 5, 35, 65 and 95 according to Harrell's recommendation[28]. To account for differences in assay methods, FT4 was converted to multiples of the upper limit of the assay reference range (ULN). Restricted cubic splines were also used to model

the longitudinal FT4 change in in the first 24-months of treatment with cubic knots set at 0, 3, 6, 9, 12, and 18 months.

Serial thyroid function tests were missing in 29% of patients. The characteristics of patients with thyroid tests were comparable to those with missing tests except that patients with tests were younger (mean age 47 vs 49 years, P<0.001), and missing tests were more frequent in patients treated with antithyroid drugs compared to radioiodine or surgery (37% vs 3% vs 16%, P<0.001) (supplementary appendix, supplementary table 3). Thus, we addressed missing tests in the one-year TSH analysis using multiple variable imputation by chained equations[29] under the missing at random assumption (supplementary appendix). We generated 50 imputed datasets with 500 iterations and fitted Cox proportional models within each dataset after which estimates were pooled according to Rubin's rules. Sensitivity analysis using a complete case analysis gave similar results, so we report results from the imputed datasets while the complete case analysis is presented in the supplementary appendix. Statistical analysis was performed via remote desktop with the SAIL Databank using Stata version 16 for Windows (Stata Corp., College Station, TXS, USA).

Regulatory Approvals

The study was registered with the Research and Development Office of Cwm Taf Health Board and approval for data linkage with individual health boards was obtained (reference CT/688/140049/16). Approval for access to anonymised data in the SAIL databank was granted by the SAIL Information Governance Review panel (project reference, 0233). Ethical permission is granted for conducting data analysis in the SAIL Databank and separate ethics committee review was not required for this study.

Role of the funding source

The funding source had no role in study design, data analysis, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

The final study cohort comprised 4,189 TRAb-positive patients with Graves' disease and 16,756 controls, with follow-up ranging from 6 months to 16·8 years in the SAIL databank. Baseline characteristics are presented in table 1. 82% of patients were females with peak age 30–60 years. Overall, antithyroid drug treatment alone was used in 74% of patients while 20% and 6% received radioiodine or thyroidectomy respectively. Of the whole cohort, 4,019 patients with at least one-year follow-up were eligible for inclusion in the landmark analysis including 3,587 patients treated with antithyroid drugs (ATD) alone in the first-year, 250 radioiodine-treated patients with resolved hyperthyroidism by one-year (RAI-Grp-A), and 182 radioiodine-treated patients with unresolved hyperthyroidism by one-year (RAI-Grp-B). The characteristics of the landmark cohort is presented in Supplementary table 4. As in the full cohort, radioiodine treated patients were older than medically-treated patients (mean age 53 vs 47 years) and had more severe disease at diagnosis in terms of higher baseline levels of FT4 and TRAbs than ATD patients. Although both radioiodine groups were well-balanced in baseline characteristics and iodine-131 dose (median 555 Mbq), RAI-Grp-A patients received earlier iodine-131 treatment than RAI-Grp-B (median time to treatment of 108 vs 188 days).

Overall, 228 deaths (5·4%) were recorded in patients compared to 765 (4·6%) in controls. Crude and adjusted Cox regression models for mortality and other events are presented in table 2. Patients had increased mortality (HR 1·23, 95%CI 1·06,

1·42) and MACE (HR 2·47, 95%CI 2·16, 2·81) compared to controls and also had increased risks for atrial fibrillation, congestive cardiac failure, and ischaemic stroke (table 2). Figure 2 shows survival curves in patients versus controls, presented for the whole cohort and stratified by primary treatment groups. The adjusted Cox regression model showed an increase in mortality and MACE in all treatment groups apart from RAI-Grp-A patients (HR 0·65, 95%CI 0·38, 1·13). Logrank tests gave consistent results with the adjusted model except for mortality in RAI-Grp-B (figure 2d) and MACE in RAI-Grp-A (figure 2g) which failed to reach significance by logrank but were significant in the adjusted model. The results remained unchanged when all treatment groups and the control population were included in an adjusted Cox regression model with a shared frailty term for case-control clusters.

Between-group comparisons are shown in figure 3. RAI-Grp-A patients had reduced risk of mortality (HR 0·50, 95% CI 0·29, 0·85) and MACE (HR 0·59, 95% CI 0·38, 0·92) compared to the ATD group. In contrast, RAI-Grp-B patients had an increase in MACE risk (HR 1·52, 95% CI 1·01, 2·28) with no survival advantage over ATD patients. Similar results were seen in sub-groups without comorbidity but not those with comorbidity (P for interaction 0.039). There was no interaction with age (>50 vs $\leq$50 years) and results remained unchanged after excluding pregnant patients or patients with borderline TRAbs (supplementary tables 5 and 6).

The results of the time-dependent multivariable analysis are presented in supplementary tables 7 and 8. The models incorporated serial thyroid hormones and treatment modality as time-dependent variables. Mortality and MACE risk increased with age, male gender, and co-morbidity but reduced with increasing year of diagnosis possibly reflecting temporal improvements in overall cardiovascular survival. Increased risk of mortality and MACE was observed with increasing levels

of serial FT4 but not serial TSH. With respect to treatment, reduced mortality and MACE risk was seen after radioiodine induced control of hyperthyroidism compared to during antithyroid drug therapy. In contrast reduced mortality was not seen after radioiodine therapy that had not controlled hyperthyroidism or after thyroidectomy although there was low power due to fewer thyroidectomies (supplementary table 7–8).

To evaluate the influence of thyroid status on outcomes we analysed the relationship of TSH at one-year with mortality and MACE using a landmark Cox regression model (figure 4). Of the patients with a one-year TSH, 898 (39%) had low-TSH while 429 (19%) had high-TSH. The fully adjusted model showed increased mortality and MACE risk with Low-TSH but not high-TSH (figure 4). This association was observed in the whole cohort as well as in the antithyroid drug group with no significant interaction with treatment (P value for interaction of 0.325 and 0.565 for mortality and MACE respectively, figure 4). Similar results were obtained from the imputation model (figure 3) and the complete case analysis (supplementary figure 1).

The relationship between FT4 and outcomes was modelled with restricted cubic splines which showed increasing hazard with FT4 concentrations above the reference range (figure 5). This relationship was seen in the whole cohort and in drug-treated patients but not radioiodine-treated patients (figure 5). Cubic splines were also used to model the longitudinal change in FT4 after diagnosis. All treatment groups showed a rapid drop in FT4 after diagnosis which was most pronounced and sustained in the successfully treated radioiodine group (supplementary figure 2).

DISCUSSION

We have used linked health records to evaluate all-cause mortality and cardiovascular event risk in patients with Graves' disease and compared outcomes

according to the modality and effectiveness of primary treatment. Compared to population controls, Graves' disease patients had a 23% increase in all-cause mortality and over twice the risk of a major cardiovascular event. Furthermore, early and successful treatment with radioiodine was associated with a 50% reduction in mortality compared to primary treatment with antithyroid drugs. Notably, survival advantages were not seen in radioiodine treated patients who failed to achieve control of hyperthyroidism. Patients with persistently low-TSH after one-year of treatment had a 55% increase in mortality even in patients treated with antithyroid drugs alone. Thus, for the first time, we show that mortality in patients with Graves' disease is reduced by early and intensive correction of hyperthyroidism regardless of therapy modality.

Our findings confirm the increased mortality risk previously reported in some[19,20,22] but not all[17,30] hyperthyroidism outcome studies. However, only a few studies to date have specifically addressed mortality risk in relation to hyperthyroidism treatment. Previous large-scale cohort-studies suggested an excess of cardiovascular mortality in radioiodine-treated patients compared to the background population[15,17,18,31]. A 1980s United States thyroid clinic study observed no difference in mortality between radioiodine and thyroidectomy-treated women[32] whereas more recent registry studies from Finland [24] and Sweden[23] have reported survival advantages of thyroidectomy over radioiodine treatment. The interpretation of these data is clearly constrained by study designs that excluded thionamide-treated patients[15,17,18,23,24,31,32], failed to account for thyroid status[23,31,32], and did not control for bias from uneven time-intervals preceding definitive therapy[15,17,18,31,32]. In their Birmingham clinic study, Boelaert *et al* applied a time-dependent regression model that incorporated thyroid status to show survival benefits relative to thionamides

following radioiodine-induced hypothyroidism[19]. However, their study was underpowered to show benefits in the Graves' disease subset[19].

In the present study, we have employed a study design that distinguished treatment effects from the effects of hyperthyroidism to show that the predictor of cardiovascular morbidity and mortality is the effectiveness of primary treatment rather than therapy modality *per se*. Our one-year TSH analysis shows that the excess mortality in Graves' disease is driven by exposure to uncontrolled hyperthyroidism and that elimination of the hyperthyroid state whether with antithyroid drugs or radioiodine yields survival benefits. This is biologically plausible given the established association of hyperthyroidism with cardiovascular disease[4] and is consistent with a recent Danish register-based study that reported excess mortality in association with cumulative periods of low-TSH in patients with treated and untreated hyperthyroidism[14]. Subsequent analyses of the Danish register also showed increased mortality with cumulative hypothyroidism exposure[33] highlighting the need for strict surveillance of therapy to avoid periods of uncontrolled hypothyroidism or hyperthyroidism. As our data shows, persistent thyroid dysfunction is common during Graves' disease therapy, with low or high TSH recorded in 58% of patients at one-year.

The choice of primary therapy in Graves' disease is currently driven by regional preferences[7,8]. Our findings indicate that a greater value should now be placed on the rapid elimination of hyperthyroidism with maintenance of a normal-TSH as standard of care. Thus, patients with unfavourable characteristics for thionamide-induced relapse such as severe disease, or high TRAb levels should be offered primary radioiodine therapy. The use of radioiodine should be governed by treatment protocols designed to optimise treatment success through efficient patient

preparation and administration of ablative doses of iodine-131. Primary definitive treatment will not be a priority for patients with mild disease and in such cases thionamides should be used at doses and duration that promote rapid and sustained control of hyperthyroidism.

The strength of our study is the unique dataset of 4,189 TRAb-positive Graves' disease patients with long-term follow-up and robust and innovative data linkage. Ours is the first large-scale hyperthyroidism cohort study to exploit TRAbs as a specific Graves' disease marker, thereby excluding patients with toxic nodules, thyroiditis, or non-thyroidal illness. By employing a landmark regression approach, we eliminated treatment time bias and additionally provided a clinically relevant time-point from which to evaluate primary therapy. Our study limitation is that the retrospective design could have introduced bias from undocumented confounders such as smoking, orbitopathy, physician treatment preference, patient compliance with thionamides, and unrecorded exposure to thyroid hormone excess. However, our adjustments for comorbidity and biochemical characteristics would to some extent have adjusted for these confounders. In keeping with conventional practice, patients with more severe baseline disease received radioiodine, but it is unlikely that this influenced our results given that both radioiodine-treated groups had similar baseline features yet strikingly divergent outcomes. Lastly, thyroid tests were missing in a proportion of patients. However, we minimised potential bias from this using principled imputation methods and robust sensitivity checks. Ultimately, potential confounders will only be fully addressed in well-designed, randomised controlled trials which are yet to be undertaken.

In conclusion, early and intensive control of hyperthyroidism in Graves' disease is associated with improvements in long-term cardiovascular morbidity and mortality

regardless of therapy modality. Greater emphasis should now be placed on ensuring prompt elimination of hyperthyroidism and maintenance of a normal thyroid status in the management of Graves' disease.

Contributors

CMD and OO conceived and designed the study. OO performed the statistical analysis with contributions from CMD, RF, PNT, DT, and AS. OO wrote the first draft of the manuscript and all authors contributed to subsequent drafts and approved the manuscript before submission for publication. OO had full access to the study data and vouches for the integrity and accuracy of the data analysis.

Declaration of interests

OO reports grants from the National Institute for Social Care and Health Research, Wales, during the conduct of the study, and grants from Concordia International, outside the submitted work. DT reports grants from Bristol-Myers Squibb, grants from Pfizer, grants from Janssen-Cilag Ltd., and grants from Medical Ethics Pty., outside the submitted work. CMD reports advisory board membership of Apitope, advisory board membership of Sanofi-Genzyme, and lecture fees from Sanofi-Genzyme, outside the submitted work.

Acknowledgements

OO was supported for the study by an Academic Health Science Collaboration Fellowship of the National Institute of Social Care and Health Research Wales.

RESULTS

TABLE 1: Baseline characteristics of the whole cohort

TABLE 2: Hazard ratios for mortality, MACE, and other events

FIGURE 1: Study Flow chart

SUPPLEMENTARY FIGURES AND TABLES

REFERENCES

1. Nystrom HF, Jansson S, Berg G. Incidence rate and clinical features of hyperthyroidism in a long-term iodine sufficient area of Sweden (Gothenburg) 2003-2005. *Clin Endocrinol (Oxf)* 2013; **78**(5): 768-76.

2. Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol* 2018; **14**(5): 301-16.

3. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet* 2016; **388**(10047): 906-18.

4. Brandt F, Green A, Hegedüs L, Brix TH. A critical review and meta-analysis of the association between overt hyperthyroidism and mortality. *Eur J Endocrinol* 2011; **165**(4): 491-7.

5. Biondi B, Kahaly GJ. Cardiovascular involvement in patients with different causes of hyperthyroidism. *Nat Rev Endocrinol* 2010; **6**(8): 431-43.

6. Bartalena L. Diagnosis and management of Graves disease: a global overview. *Nat Rev Endocrinol* 2013; **9**(12): 724-34.

7. Kahaly GJ, Bartalena L, Hegedüs L, Leenhardt L, Poppe K, Pearce SH. 2018 European Thyroid Association Guideline for the Management of Graves' Hyperthyroidism. *European Thyroid Journal* 2018; **7**(4): 167-86.

8. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. *Thyroid* 2016; **26**(10): 1343-421.

9. Burch HB, Burman KD, Cooper DS. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab* 2012; **97**(12): 4549-58.

10. Brito JP, Schilz S, Singh Ospina N, et al. Antithyroid Drugs-The Most Common Treatment for Graves' Disease in the United States: A Nationwide Population-Based Study. *Thyroid* 2016; **26**(8): 1144-5.

11. Bartalena L, Burch HB, Burman KD, Kahaly GJ. A 2013 European survey of clinical practice patterns in the management of Graves' disease. *Clin Endocrinol (Oxf)* 2016; **84**(1): 115-20.

12. Osman F, Gammage MD, Franklyn JA. Hyperthyroidism and cardiovascular morbidity and mortality. *Thyroid* 2002; **12**(6): 483-7.

13. Flynn RW, Macdonald TM, Jung RT, Morris AD, Leese GP. Mortality and vascular outcomes in patients treated for thyroid dysfunction. *J Clin Endocrinol Metab* 2006; **91**(6): 2159-64.

14. Lillevang-Johansen M, Abrahamsen B, Jorgensen HL, Brix TH, Hegedus L. Excess mortality in treated and untreated hyperthyroidism is related to cumulative periods of low serum TSH. *J Clin Endocrinol Metab* 2017.

15. Franklyn JA, Sheppard MC, Maisonneuve P. Thyroid function and mortality in patients treated for hyperthyroidism. *Jama* 2005; **294**(1): 71-80.

16. Bauer DC, Rodondi N, Stone KL, Hillier TA. Thyroid hormone use, hyperthyroidism and mortality in older women. *The American journal of medicine* 2007; **120**(4): 343-9.

17. Metso S, Jaatinen P, Huhtala H, Auvinen A, Oksala H, Salmi J. Increased cardiovascular and cancer mortality after radioiodine treatment for hyperthyroidism. *J Clin Endocrinol Metab* 2007; **92**(6): 2190-6.

18. Franklyn JA, Maisonneuve P, Sheppard MC, Betteridge J, Boyle P. Mortality after the treatment of hyperthyroidism with radioactive iodine. *N Engl J Med* 1998; **338**(11): 712-8.

19. Boelaert K, Maisonneuve P, Torlinska B, Franklyn JA. Comparison of mortality in hyperthyroidism during periods of treatment with thionamides and after radioiodine. *J Clin Endocrinol Metab* 2013; **98**(5): 1869-82.

20. Brandt F, Thvilum M, Almind D, et al. Graves' disease and toxic nodular goiter are both associated with increased mortality but differ with respect to the cause of death: a Danish population-based register study. *Thyroid* 2013; **23**(4): 408-13.

21. Schwensen CF, Brandt F, Hegedus L, Brix TH. Mortality in Graves' orbitopathy is increased and influenced by gender, age and pre-existing morbidity: a nationwide Danish register study. *Eur J Endocrinol* 2017; **176**(6): 669-76.

22. Giesecke P, Rosenqvist M, Frykman V, et al. Increased Cardiovascular Mortality and Morbidity in Patients Treated for Toxic Nodular Goiter Compared to Graves' Disease and Nontoxic Goiter. *Thyroid* 2017.

23. Giesecke P, Frykman V, Wallin G, et al. All-cause and cardiovascular mortality risk after surgery versus radioiodine treatment for hyperthyroidism. *Br J Surg* 2017.

24. Essi R, Saara M, Heini H, Matti V, Anssi A, Pia J. Cardiovascular Morbidity and Mortality After Treatment of Hyperthyroidism with Either Radioactive Iodine or Thyroidectomy. *Thyroid* 2018; **28**(9): 1111-20.

25. Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; **9**: 157.

26. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009; **9**: 3.

27. Autilio C, Morelli R, Locantore P, Pontecorvi A, Zuppi C, Carrozza C. Stimulating TSH receptor autoantibodies immunoassay: analytical evaluation and clinical performance in Graves' disease. *Ann Clin Biochem* 2018; **55**(1): 172-7.

28. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988; **80**(15): 1198-202.

29. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; **30**(4): 377-99.

30. Nyirenda MJ, Clark DN, Finlayson AR, et al. Thyroid disease and increased cardiovascular risk. *Thyroid* 2005; **15**(7): 718-24.

31. Goldman MB, Maloof F, Monson RR, Aschengrau A, Cooper DS, Ridgway EC. Radioactive iodine therapy and breast cancer. A follow-up study of hyperthyroid women. *Am J Epidemiol* 1988; **127**(5): 969-80.

32. Hoffman DA, McConahey WM, Diamond EL, Kurland LT. Mortality in women treated for hyperthyroidism. *Am J Epidemiol* 1982; **115**(2): 243-54.

33. Lillevang-Johansen M, Abrahamsen B, Jorgensen HL, Brix TH, Hegedus L. Over- and Under-Treatment of Hypothyroidism Is Associated with Excess Mortality: A Register-Based Cohort Study. *Thyroid* 2018; **28**(5): 566-74.

Primary therapy of Graves' disease and cardiovascular morbidity and mortality: a linked-record cohort study

Supplementary Appendix

Supplementary Methods

Modified Charlson Comorbidity Scores

Comorbidity scores were derived using a modification of the Charlson Comorbidity Index (CCI) [1]. Individual scores were calculated as the sum of weighted scores for each of 17 disease conditions identified from hospital admission data using ICD-10 codes for secondary diagnosis fields (DIAG2-DIAG14) [2,3] (supplementary table 1).

Laboratory assays

Between January 1998 and April 2004 TSH receptor antibodies (TRAbs) were measured with a 2nd generation radioimmunoassay (RIA) based on inhibition of labelled TSH binding to TSH receptor coated tubes, RSR Limited, Cardiff [4,5]. The assay reference ranges were: <1.0, negative, 1.0-1.5 borderline >1.5, positive with analytical range of 1—40 U/L, and lower detection limit of 0.33 U/L. The sensitivity of the assay was 92.0% with specificity of 100%. Between April 2004 and December 2013, TRAbs were measured with a 2nd generation RIA which was also based on inhibition of radio-labelled TSH binding to TSH receptor coated tubes, B.R.A.H.M.S. TRAK human, Thermo Scientific [6]. The assay reference ranges were: <1.0, negative, 1.0-1.5 borderline, and >1.5, positive, with analytical range of 1—40 U/L, functional sensitivity of 1.0 +/- 0.2 U/L and analytical sensitivity of 0.3 U/L. The sensitivity of the assay was 98.8% with specificity of 99.6%. The assay principles, performance, and reference ranges for thyroid function tests (FT4 and TSH) are presented in supplementary table 2. To account for differences in assay methods, FT4 was transformed to multiples of the upper limit of the assay reference range (ULN) and TSH was categorised as low, normal or high-TSH according to the assay reference range.

Restricted Cubic Spline Regression

We modelled a potential non-linear relationship between the FT4 concentration at one year and mortality or MACE using restricted cubic spline regressions. Cubic splines allow flexible smooth transformations of the relationship between a quantitative covariate and an outcome [7]. We used the *mkspline* co-mmand in Stata

to set 4 equally-spaced knots at percentiles 5, 35, 65 and 95 according to the recommendation by Harrell [8]. Varying the positions of the knots did not significantly influence our estimates. The reference values were set at the FT4 assay upper limit (x 1.0). Predicted hazard ratios (HR) were derived from Cox regression models adjusted for age, gender, year of diagnosis, baseline TRAb concentration, and comorbidity. We used the *xblc* post-estimation package in Stata to plot the regression between FT4 and log HR for mortality/MACE. Models were fitted for the whole cohort and then stratified by treatment. P values for non-linearity were obtained using likelihood ratio tests. We also used cubic splines to model longitudinal change in FT4 concentration in the first year of treatment with cubic knots set at 0, 3, 6, 9, and 12 months.

Missing data imputation

Serial thyroid function tests were available for patients with Graves' disease but not for the control population. Thyroid function tests were missing in 29% of patients. One-year TSH was available in 80% of patients with serial thyroid function tests equivalent to 57% of the eligible landmark cohort. All other variables had complete data. Missing thyroid tests were due to lack of electronic access and linkage of laboratory data to the SAIL databank from some laboratories at various study time segments. Where laboratory linkage occurred, thyroid tests were available in >90% of patients in clinics served by the laboratory, reducing the likelihood of confounding by data not missing at random.

The characteristics of patients with missing and non-missing thyroid tests is shown in supplementary table 3. Patients with missing tests were older (49 vs 47 years) and missing tests were more frequent in patients treated with antithyroid drugs compared to radioiodine or surgery (37% vs 3% vs 16%, P<0.001). Patients with missing tests also had lower TRAb levels (median 5.8 vs 7.8 IU/L) but in logistic regression, the only factors associated with thyroid test missingness were older age (OR 1.01, 95%CI 1.00, 1.02 per year increase in age) and treatment with antithyroid drugs (OR 2.28, 95%CI 1.61, 3.23, antithyroid drugs vs radioiodine/Surgery). Missingness was not associated with sex, comorbidity, TRAB concentration, mortality or MACE. Thus, we assumed the data were not 'Missing Completely at Random, (MCAR)' but were 'Missing at Random, (MAR)'.

In the one-year TSH analysis we addressed missing tests using multiple variable imputation by chained equations [9]. The imputation model comprised all predictor and outcome variables in the analysis. We generated 50 imputed datasets and fitted Cox proportional models within each dataset after which estimates were pooled according to Rubin's rules [10]. In sensitivity analysis we repeated the analysis using a complete case analysis which showed identical estimates to the imputation model (Supplementary figure 1). To exclude the possibility that missingness at one-year was associated with stable or worse disease control, we modelled two sensitivity scenarios in which a normal-TSH and a low-TSH at one-year were assumed for patients with serial TSH but missing one-year TSH. These scenarios gave similar results to the imputation model (data not shown).

References

1. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases* 1987; **40**(5): 373-83.

2. http://www.drfosterhealth.co.uk/hospital-guide/methodology/. 2018.

3. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004; **57**(12): 1288-94.

4. Sanders J, Evans M, Betterle C, et al. A human monoclonal autoantibody to the thyrotropin receptor with thyroid-stimulating blocking activity. *Thyroid* 2008; **18**(7): 735-46.

5. Smith BR, Bolton J, Young S, et al. A new assay for thyrotropin receptor autoantibodies. *Thyroid* 2004; **14**(10): 830-5.

6. Costagliola S, Morgenthaler NG, Hoermann R, et al. Second generation assay for thyrotropin receptor antibodies has superior diagnostic sensitivity for Graves' disease. *J Clin Endocrinol Metab* 1999; **84**(1): 90-7.

7. Orsini N, Greenland S. A procedure to tabulate and plot results after flexible modeling of a quantitative covariate. *Stata Journal* 2011; **11**(1): 1-29.

8. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988; **80**(15): 1198-202.

9. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; **30**(4): 377-99.

10. Rubin D, B. Multiple Imputation for Nonresponse in Surveys. Hoboken, New Jersey, USA: John Wiley and Sons; 2004.

			Treatment groups			
	Population controls n=16756	**All Patients n=4189**	**Antithyroid drugs n=3094**	**Iodine-131 n=829**	**Thyroidectomy n=266**	**P for trend [c]**
Age, yrs						
Mean ± SD	48 ± 16	48 ± 16	48 ± 16	50 ± 16	39 ± 13	<0·001
<20	204 (1·2)	51 (1·2)	37 (1·2)	6 (0·7)	8 (3·0)	<0·001
20-29	2176 (13·0)	544 (13·0)	402 (13)	83 (10·0)	59 (22·2)	
30-39	3420 (20·4)	855 (20·4)	626 (20·2)	138 (16·6)	91 (34·2)	
40-49	3688 (22·0)	922 (22·0)	667 (21·6)	195 (23·5)	60 (22·6)	
50-59	3240 (19·3)	810 (19·3)	601 (19·4)	184 (22·2)	25 (9·4)	
60-69	2340 (14·0)	585 (14·0)	443 (14·3)	124 (15·0)	18 (6·8)	
≥70	1688 (10·1)	422 (10·1)	318 (10·3)	99 (11·9)	5 (1·9)	
Sex						
Female	13656 (81·5)	3414 (81·5)	2527(81·7)	656 (79·1)	231 (86·8)	0·017
Male	3100 (18·5)	775 (18·5)	567 (18·3)	173 (20·9)	35 (13·2)	
Comorbidity						
Absent	15453 (92·2)	3599 (85·9)	2647 (85·6)	715 (86·2)	237 (89·1)	0·263*
Present	1303 (7·8)	590 (14·1)	447 (14·4)	114 (13·8)	29 (10·9)	
TRAb IU/L	—	6·9 (3·1, 16·8)	6·0 (2·8, 13·6)	11·3 (4·1, 26·4)	15·9 (5·8, 36·3)	<0·001
Thyroid status						
FT4, pmol/L[a]	—	31·8 ± 22·0	29·5 ± 19·7	36·2 ± 25·4	36·5 ± 25·6	<0·001
FT4 >40 pmol/L[a]	—	719 (24)	390 (20)	255 (32)	74 (33)	<0·001
Person years follow up	103374	25266	16481	6878	1907	

TABLE 1: Baseline characteristics of the whole cohort

Data represents numbers (%), mean ± SD, or median (interquartile range). a, data available for 2962 patients, TRAb, TSH receptor antibody; c, P values are for trend across treatment categories. Patients and controls were matched for age· *Comorbidity was increased in patients compared to controls (P<0·001)

Outcome	Number of cases (%) Graves' disease N=4189	Controls [a] N=16756	Crude HR (95% CI)	P value	Adjusted HR (95 CI) [b]	P value
All-cause Mortality	228 (5·4)	765 (4·6)	1·22 (1·05, 1·42)	0·008	1·23 (1·06, 1·42)	0·033
MACE[c]	329 (7·9)	651 (3·9)	2·18 (1·91, 2·48)	<0·001	2·47 (2·16, 2·81)	<0·001
Atrial Fibrillation	240 (5·7)	379 (2·3)	2·62 (2·23, 3·08)	<0·001	2·67 (2·25, 3·15)	<0·001
Acute Myocardial Infarction	38 (0·9)	122 (0·7)	1·26 (0·87, 1·81)	0·220	1·22 (0·84, 1·76)	0·287
Acute coronary syndrome	32 (0·8)	74 (0·4)	1·75 (1·15, 2·65)	0·008	1·65 (1·09, 2·51)	0·018
Congestive Cardiac Failure	109 (2·6)	240 (1·4)	1·85 (1·47, 2·31)	<0·001	1·78 (1·42, 2·23)	<0·001
Ischaemic Stroke	41 (1·0)	98 (0·6)	1·70 (1·17, 2·43)	0·005	1·60 (1·11, 2·30)	0·012
Haemorrhagic Stroke	14 (0·3)	37 (0·2)	1·53 (0·83, 2·83)	0·175	1·50 (0·81, 2·78)	0·196
Type 1 Diabetes	29 (0·7)	42 (0·3)	2·71 (1·68, 4·37)	<0·001	2·53 (1·56, 4·08)	<0·001
Type 2 Diabetes	170 (4·1)	474 (2·8)	1·46 (1·23, 1·74)	<0·001	1·39 (1·16, 1·66)	<0·001
All Diabetes	199 (4·8)	516 (3·1)	1·57 (1·33, 1·85)	<0·001	1·46 (1·24, 1·71)	<0·001
Thyroid Cancer	8 (0·2)	5 (0·0)	6·52 (2·13, 19·93)	<0·001	6·67 (2·18, 20·41)	0·001
Breast Cancer	38 (0·9)	143 (0·9)	1·08 (0·76, 1·55)	0·442	1·06 (0·74, 1·52)	0·798

TABLE 2: Hazard ratios for mortality, MACE, and other events

[a] Controls are matched for age and sex [b] Hazard ratios are derived from Cox proportional models adjusted for age, sex and comorbidity, [c] MACE, major adverse cardiovascular event comprising acute myocardial infarction, congestive cardiac failure, ischaemic cerebrovascular accident, or death

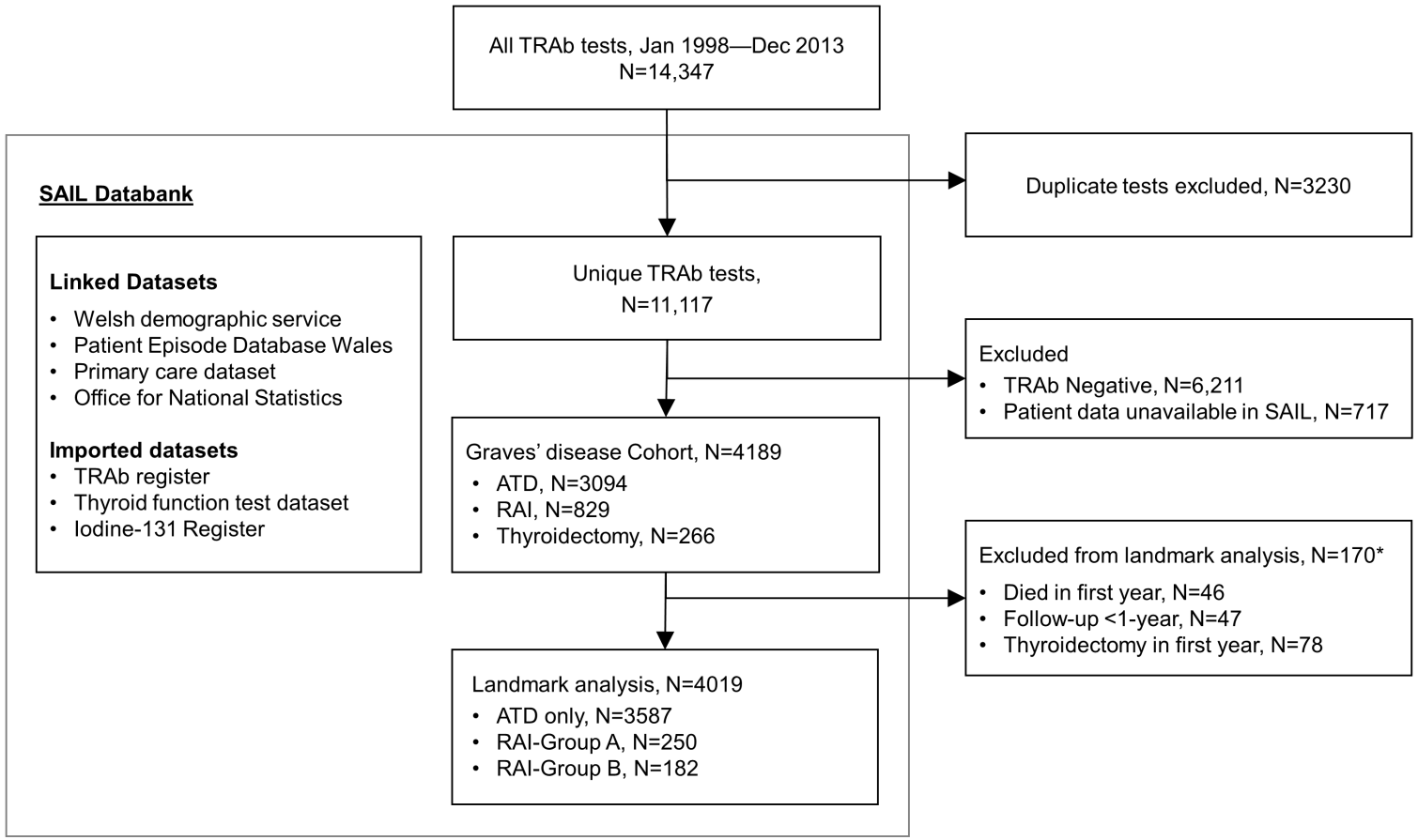


FIGURE 1: Study flow chart

TRAb, TSH receptor antibody; ATD, Antithyroid Drugs; RAI, radioactive iodine; RAI-Group-A, radioiodine with hyperthyroidism control by one-year, and RAI-Group-B, radioiodine without hyperthyroidism control by one-year; *Numbers add up to 171 for 170 patients because one patient had thyroidectomy and died in the first year.

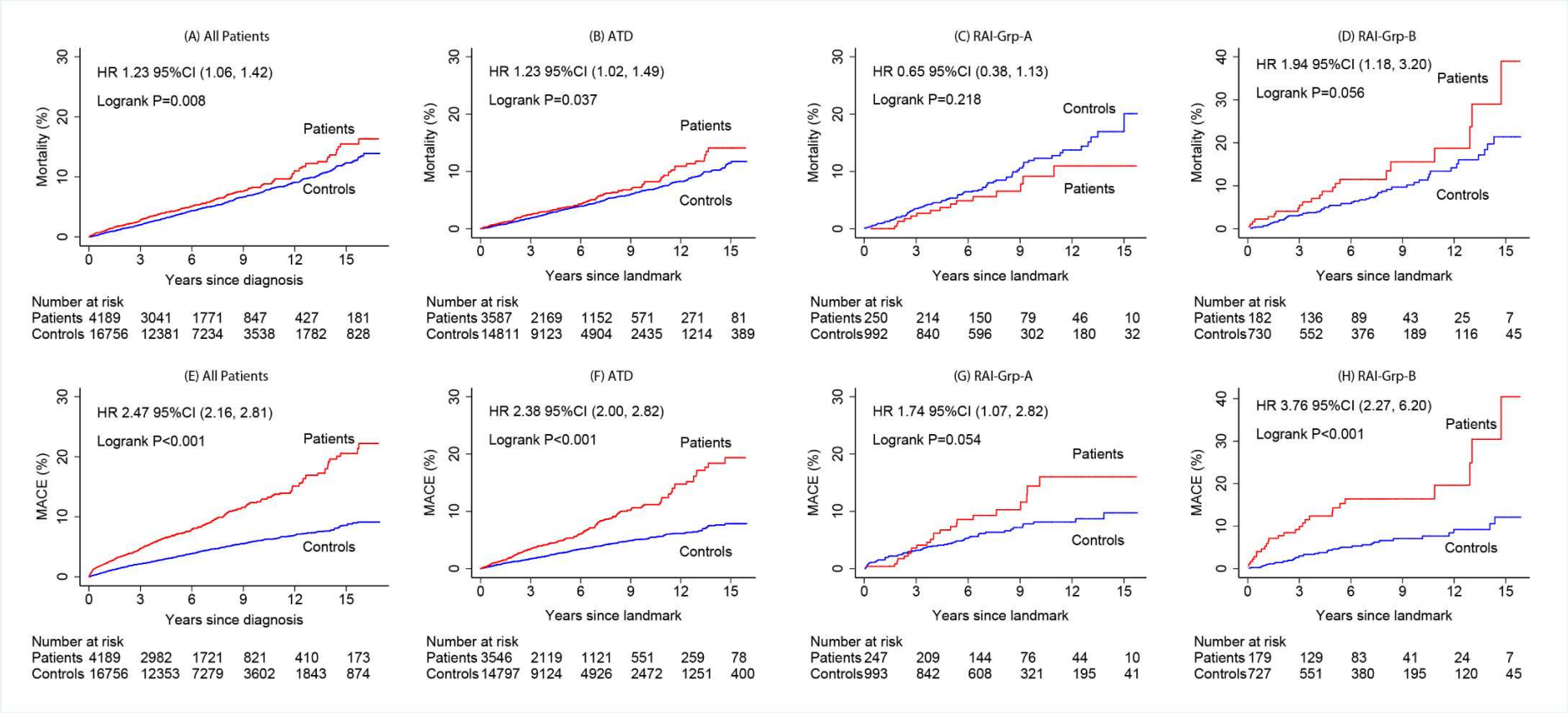
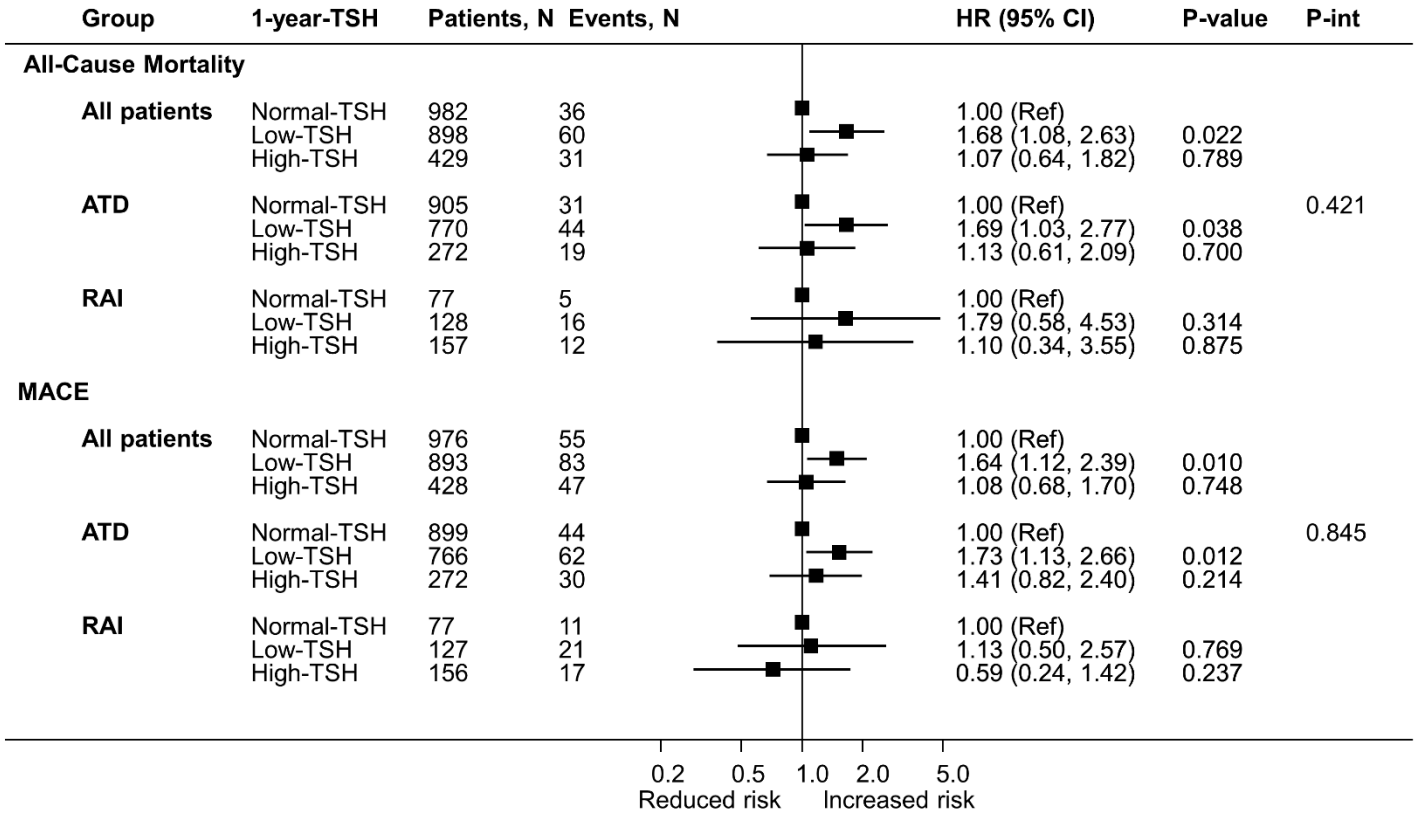


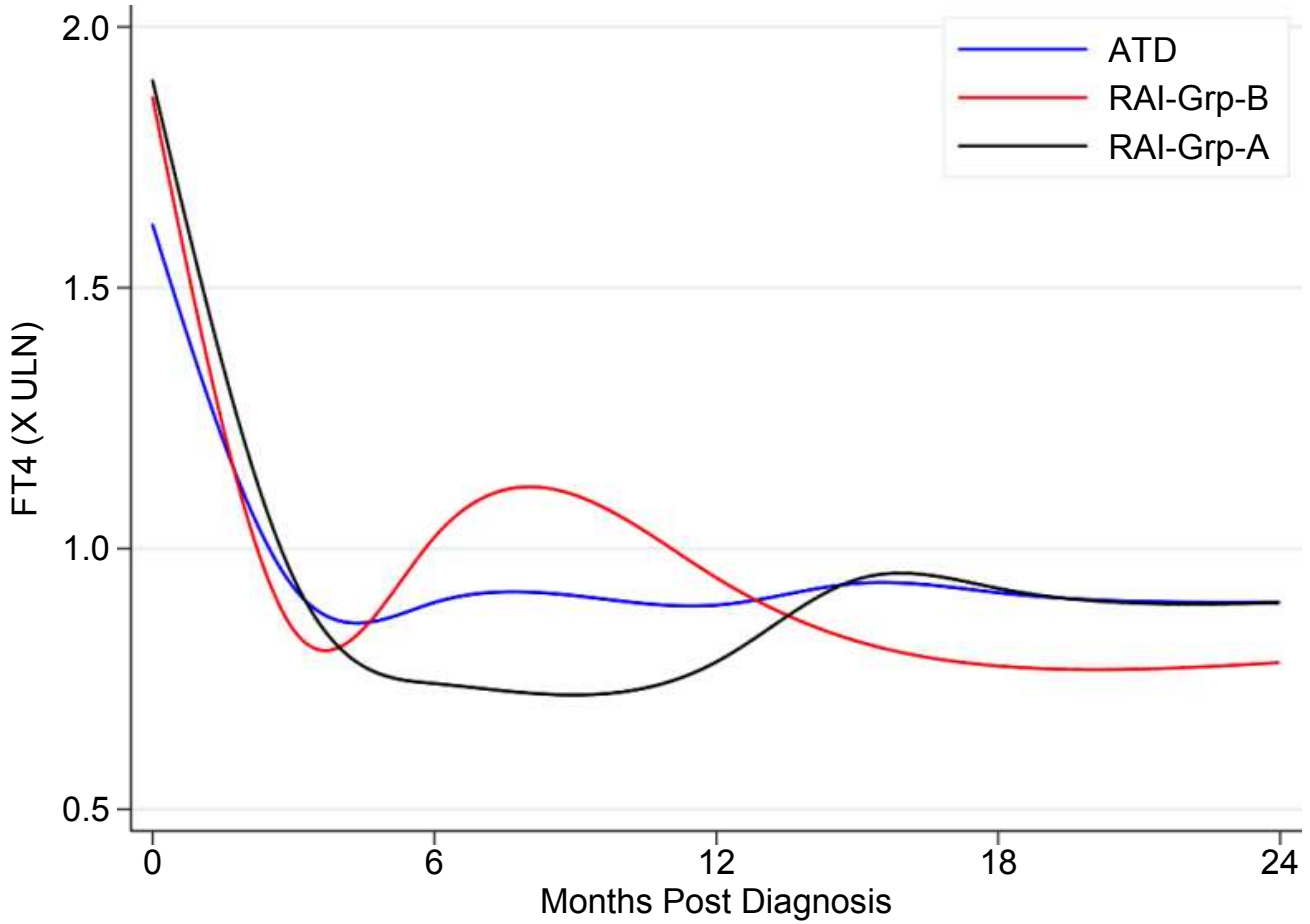
FIGURE 2: Kaplan Meier curves for mortality or MACE

Kaplan Meier curves for cumulative mortality (panels A-D) or MACE (panels E-H) in the whole cohort and by primary treatment. ATD, patients who received antithyroid drugs alone in the first one-year, RAI-Group-A, radioiodine with hyperthyroidism control by one-year, and RAI-Group-B, radioiodine without hyperthyroidism control by one-year. HR, Hazard ratios derived from Cox regression models adjusted for age, sex, and comorbidity.



SUPPLEMENTARY FIGURE 1: Hazard ratios for mortality or MACE by one-year TSH, complete case analysis

HR (95% CI) are obtained from multivariable Cox proportional hazard regression models. Analysis was conducted in all patients and in treatment subgroups according to treatment in the first-year with antithyroid drugs (ATD) or radioactive iodine (RAI). P-int, P values for interaction with treatment group. Results are obtained from a complete case analysis



SUPPLEMENTARY FIGURE 2: FT4 profile post-diagnosis

Longitudinal profile of FT4 concentrations after diagnosis according to treatment groups. FT4 is in multiples of the upper limit of the reference range (X ULN); ATD, antithyroid drugs, RAI-Grp-A, radioiodine with control of hyperthyroidism by 12 months, RAI-Grp-B, radioiodine without control of hyperthyroidism by 12 months

		FT4, pmol/L		TSH, mU/L	
Laboratory	**Assay method, platform**	**Reference Interval**	**CV (%)**	**Reference Interval**	**CV (%)**
Cardiff and Vale					
Jan 1998 - May 2011	CCIA, ADVIA Centaur, Bayer[1]	9·8–23·1	2·31 at 13·9 pmol/L	0·35 - 5·50	2·44 at 5·65 mU/L
May 2011 - Jan 2013	CMIA, Abbott ARCHITECT[2]	9·0 –19·1*	<10 at all levels	0·30 - 4·40	<10 at all levels
Cwm Taf (1)					
Jan 1998 - March 2007	CCIA, IMMULITE 2000, DPC[3]	10·3–24·5	<10 at all levels	0·4 - 4·0	<5·0 at all levels
March 2007 - Dec 2013	ECLIA, Roche E-170[4]	10·3–24·5	8·2 at 8·2 pmol/L	0·4 - 4·0	2·4 at 6·67 mU/L
Cwm Taf (2)					
Jan 1995 – April 2011	CCIA, ADVIA Centaur, Bayer[1]	10·0-25·0	<10 at all levels	0·35-5·50**	<10 at all levels
April 2011-Sept 2012	CCIA, ADVIA Centaur, Bayer[1]	10·0-25·0	<10 at all levels	0·35-5·50	<10 at all levels
Sept 2012-Dec 2013	ECLIA, Roche E-170[4]	10·3–24·5	8·2 at 8·2 pmol/L	0·4 - 4·0	2·4 at 6·67 mU/L
Aneurin Bevan (1)					
1998-2006	CCIA, IMMULITE 2000, DPC[3]	11·5–22·7	<10 at all levels	0·4 - 4·0	<5 at all levels
2006-2013	CMIA, Abbott ARCHITECT[2]	9·0–19·1	4·8% at 23·3 pmol/L	0·30 - 4·40	4·1% at 26·8 mU/L
Aneurin Bevan (2)					
1998—2000	MEIA, Abbott Axsym[2]	9·0–19·1	4·6 at 15·4 pmol/L	0·30 - 4·40	4·9 at 2·8 mU/L
2000—2006	CCIA, ADVIA Centaur, Bayer[1]	9·8–23·1	<10 at all levels	0·35 - 5·50	<10 at all levels
2007—2013	CMIA, Abbott ARCHITECT[2]	9·0–19·1	<10 at all levels	0·30 - 4·40	<10 at all levels

Supplementary Table 1: Thyroid hormone assays

CV, coefficient of variation; CCIA, competitive chemiluminescent immunoassay, CMIA, chemiluminescent microparticle immunoassay; ECLIA, electrochemiluminescence immunoassay; MEIA, Microparticle Enzyme Immunoassay; 1. Bayer Diagnostics, Newbury, UK; 2. Abbott Diagnostics, Maidenhead, Berks, UK; 3. Diagnostics Product Corporation, Llanberis, Wales; 4. Roche Diagnostics GmbH, Mannheim, Germany, *from January to December 2013 the reference range for this assay was changed to 9.2—21.0 pmol/L, **from September 2009 to April 2011, reference range for this assay was 0.3-6.0 mU/L

Condition	ICD-10 diagnosis codes	Score
Acute myocardial infarction	I21, I22, I252	1
Congestive heart failure	I50	1
Peripheral vascular disease	I71, I790, I739, R02, Z958, Z959	1
Cerebral vascular disease	I60, I61, I62, I63, I65, I66, G450, G451, G452, G458, G459, G46, I64, G454, I670, I671, I672, I674, I675, I676, I677, I678, I679, I681, I682, I688, I69	1
Dementia	F00, F01, F02, F051	1
Pulmonary disease	J40, J41, J42, J44, J43, J45, J46, J47, J67, J44, J60, J61, J62, J63, J66, J64, J65	1
Connective tissue disease	M32, M34, M332, M053, M058, M059, M060, M063, M069, M050, M052, M051, M353	1
Peptic ulcer disease	K25, K26, K27, K28	1
Liver disease	K702, K703, K73, K717, K740, K742, K746, K743, K744, K745	1
Diabetes	E109, E119, E139, E149, E101, E111, E131, E141, E105, E115, E135, E145	1
Diabetes with complications	E102, E112, E132, E142, E103, E113, E133, E143, E104, E114, E134, E144	2
Hemiplegia or paraplegia	G81, G041, G820, G821, G822	2
Renal disease	N03, N052, N053, N054, N055, N056, N072, N073, N074, N01, N18, N19, N25	2
Cancer	C0, C1, C2, C3, C40, C41, C43, C45, C46, C47, C48, C49, C5, C6, C70, C71, C72, C73, C74, C75, C76, C81, C82, C83, C84, C85, C883, C887, C889, C900, C901, C91, C92, C93, C940, C941, C942, C943, C945, C947, C95, C96	2
Metastatic cancer	C77, C78, C79, C80	3
Severe liver disease	K729, K766, K767, K721	3
HIV	B20, B21, B22, B23, B24	6

Supplementary Table 2: Modified Charlson Comorbidity Scores

Reference [2,3]

	Patients with available thyroid tests					Patients with missing thyroid tests	
	All patients n=2969	Medical therapy n=1950	Iodine-131 n=796	Thyroidectomy n=223	P for trend, treatment groups [a]	All patients n=1220	P value, missing vs non-missing [b]
Age, yrs							
Mean ± SD	47 ± 16	46 ± 16	50 ± 16	39 ± 13	<0·001	49 ± 16	<0.001
<20	42 (1·4)	29 (1·5)	6 (0·8)	7 (3·1)	<0·001	9 (0.7)	
20-29	409 (13·8)	278 (14·3)	81 (10·2)	50 (22·4)		135 (11.1)	
30-39	630 (21·2)	424 (21·7)	133 (16·7)	73 (32·7)		225 (18.4)	
40-49	669 (22·5)	426 (21·9)	191 (24·0)	52 (23·3)		253 (20.7)	
50-59	550 (18·5)	360 (18·5)	170 (21·4)	20 (9·0)		260 (21.3)	
60-69	390 (13·1)	252 (12·9)	120 (15·1)	18 (8·1)		195 (16.0)	
≥70	279 (9·4)	181 (9·3)	95 (11·9)	3 (1·4)		143 (11.7)	
Sex							
Female	2422 (81·6)	1596 (81·9)	633 (79·5)	193 (86·5)	0·050	992 (81.3)	0.841
Male	547 (18·4)	354 (18·2)	163 (20·5)	30 (13·5)		228 (18.7)	
Comorbidity							
Absent	2564 (86·4)	1678 (86·1)	688 (86·4)	198 (88·8)	0·532	1035 (84.9)	0.198
Present	405 (13·6)	272 (13·9)	108 (13·6)	25 (11·2)		185 (15.2)	
TRAb IU/L	7·8 (3·2, 18·4)	6·4 (2·9, 14·0)	11·4 (4·2, 27·0)	16·4 (5·6, 37·5)	<0·001	5.8 (2.75, 12.9)	<0.001
Thyroid status							
FT4, pmol/L	31·8 ± 22·0	29·5 ± 19·7	36·2 ± 25·4	36·5 ± 25·6	<0·001	—	
FT4 >40 pmol/L	719 (24)	390 (20)	255 (32)	74 (33)	<0·001	—	
Missing tests	1220 (29)	1144 (37)	33 (4)	43 (16)	<0.001	—	

Supplementary Table 3: Baseline characteristics in patients with missing and non-missing thyroid function tests

Data represents numbers (%), mean ± SD, or median (interquartile range). TRAb, TSH receptor antibody; a, P values are for trend across treatment categories in patients with thyroid tests. b, P values are for difference between missing and non-missing tests.

	ATD	RAI-Group-A	RAI-Group-B	P value
Number (%)	3587 (89%)	250 (6%)	182 (5%)	
Age				
Mean ± SD	46·9 ± 15·8	52·6 ± 15·0	52·9 ± 15·8	<0·001[a, b]
Range	18-92	19-88	20-92	
Gender				
Male	642 (17·9%)	55 (22·0%)	41 (22·5%)	0·089
Female	2945 (82·1%)	195 (78·0%)	141 (77·5%)	
Comorbidity				
Present	485 (13·5%)	35 (14·0%)	30 (16·5%)	0·728
Absent	3102 (86·5%)	215 (86·0%)	152 (83·5%)	
TRAbs, U/L	6·7 (3, 15·8)	11·2 (4·1, 26·6)	11·4 (4·4, 26·1)	<0·001[a, b]
Radioiodine therapy				
Dose, Mbq	-	555 (546, 563)	555 (544, 566)	0·984[d]
Days from diagnosis to treatment	-	108 (53, 178)	188 (88, 302)	<0·001 [d]
FT4, pmol/L				
Baseline	31·2 ± 21·8	35·2 ± 25·2	34·8 ± 21·7	<0·003[a, b]
12 months	17·4 ± 9·3	15·9 ± 7·3	21·5 ± 10·6	<0·001[a, b, c]
Baseline – 12 months	16·1 ± 23·0	21·1 ± 26·9	13·9 ± 21·3	<0·001[b, c]
TSH at 12 months, mU/L				
Low	770 (39·6%)	44 (18·9%)	84 (64·6%)	<0·001 [a, c]
Normal	905 (46·5%)	41 (17·7%)	36 (27·7%)	
High	272 (13·9%)	147 (63·4%)	10 (7·7%)	

Supplementary Table 4: Characteristics of patients in the one-year landmark cohort

a ATD vs RAI-Group-A, b ATD vs RAI-Group-B, c RAI-Group-A vs RAI-Group-B; ANOVA with pairwise comparisons with Tukey's test; d RAI-Group-A vs RAI-Group-B, Kruskal Wallis; data is in numbers (%), mean + standard deviation, or median (interquartile range). Excluded patients who had surgery in the first 12 months (n=78) or <12 months survival (n=89).

	All-Cause Mortality			MACE		
	Events/Total, N	HR (95% CI)	P value	Events/Total, N	HR (95% CI)	P value
Whole cohort						
ATD	145/3527	1·00 (Ref)		176/3384	1·00 (Ref)	
RAI-Group-A	13/236	0·47 (0·27, 0·83)	0·009	17/224	0·49 (0·30, 0·81)	0·005
RAI-Group-B	21/174	1·52 (0·96, 2·42)	0·083	21/159	1·49 (0·94, 2·35)	0·089
No comorbidities						
ATD	82/3047	1·00 (Ref)		118/2982	1·00 (Ref)	
RAI-Group-A	4/201	0·20 (0·07, 0·56)	0·002	7/195	0·25 (0·12, 0·54)	0·001
RAI-Group-B	13/146	1·45 (0·80, 2·63)	0·195	16/137	1·68 (0·99, 2·84)	0·052
Age≥50 yrs						
ATD	129/1502	1·00 (Ref)		152/1387	1·00 (Ref)	
RAI-Group-A	11/137	0·43 (0·23, 0·79)	0·007	15/127	0·46 (0·27, 0·79)	0·005
RAI-Group-B	19/96	1·52 (0·93, 2·48)	0·102	17/84	1·36 (0·82, 2·25)	0·236
Age<50 yrs						
ATD	16/2025	1·00 (Ref)		24/1997	1·00 (Ref)	
RAI-Group-A	2/99	1·04 (0·23, 4·79)	0·963	2/97	0·74 (1·70, 3·18)	0·680
RAI-Group-B	2/78	1·36 (0·31, 6·09)	0·691	4/75	2·32 (0·79, 6·78)	0·125

Supplementary Table 5: Mortality and MACE by primary treatment: excluding pregnant patients

HR (95% CI) are obtained from multivariable Cox proportional hazard regression models. Analysis was conducted in all patients and in subgroups according to treatment in the first-year. ATD, Antithyroid drugs, RAI-Group-A, radioiodine treatment with control of hyperthyroidism, RAI-Group-B, radioiodine treatment without control of hyperthyroidism

	All-Cause Mortality			MACE		
	Events/Total, N	**HR (95% CI)**	**P value**	**Events/Total, N**	**HR (95% CI)**	**P value**
Whole cohort						
ATD	142/3353	1·00 (Ref)		166/3215	1·00 (Ref)	
RAI-Group-A	13/229	0·46 (0·26, 0·82)	0·009	17/218	0·50 (0·30, 0·83)	0·007
RAI-Group-B	21/166	1·53 (0·96, 2·44)	0·071	19/151	1·38 (0·86, 2·33)	0·184
No comorbidities						
ATD	82/2903	1·00 (Ref)		114/2841	1·00 (Ref)	
RAI-Group-A	4/197	0·20 (0·07, 0·54)	0·002	7/192	0·25 (0·12, 0·54)	0·001
RAI-Group-B	13/139	1·44 (0·79, 2·60)	0·119	15/130	1·62 (0·94, 2·77)	0·082
Age≥50 yrs						
ATD	126/1408	1·00 (Ref)		142/1297	1·00 (Ref)	
RAI-Group-A	11/131	0·42 (0·22, 0·78)	0·006	15/121	0·48 (0·28, 0·82)	0·007
RAI-Group-B	19/89	1·53 (0·94, 2·50)	0·089	15/77	1·23 (0·72, 2·11)	0·443
Age<50 yrs						
ATD	16/1945	1·00 (Ref)		24/1918	1·00 (Ref)	
RAI-Group-A	2/98	1·02 (0·22, 4·69)	0·984	2/97	0·72 (1·67, 3·11)	0·660
RAI-Group-B	2/77	1·38 (0·31, 6·16)	0·673	4/74	2·31 (0·79, 6·74)	0·126

Supplementary Table 6: Mortality and MACE by primary treatment: excluding patients with borderline TRAbs

Patients with borderline TRAbs (1.0-1.5 IU/L) were excluded. HR (95% CI) are obtained from multivariable Cox proportional hazard regression models. Analysis was conducted in all patients and in subgroups according to treatment in the first-year. ATD, Antithyroid drugs, RAI-Group-A, radioiodine treatment with control of hyperthyroidism, RAI-Group-B, radioiodine treatment without control of hyperthyroidism

	Model 1		Model 2	
	HR (95% CI)	**P value**	**HR (95% CI)**	**P value**
Age, years [a]	1·09 (1·08, 1·10)	<0·001	1·09 (1·08, 1·10)	<0·001
Sex				
Male	1·00	Ref	1·00	Ref
Female	0·57 (0·43, 0·76)	<0·001	0·65 (0·47, 0·90)	0·010
Co-morbidity				
Absent	1·00	Ref	1·00	Ref
Present	3·07 (2·29, 4·13)	<0·001	3·30 (2·36, 4·63)	<0·001
Year of diagnosis [b]	0·68 (0·58, 0·79)	<0·001	0·63 (0·54, 0·79)	<0·001
Baseline TRAb [c]	1·15 (0·99, 1·32)	0·069	1·14 (0·96, 1·35)	0·144
Thyroid hormones				
Baseline FT4 [d]	—	—	0·99 (0·99, 1·00)	0·474
Baseline TSH	—	—		
TSH 0.1—0·4 mU/L			1·00	Ref
TSH <0·1 mU/L			0·99 (0·95, 1·02)	0·384
Serial FT4 [d]	—	—	1·40 (1·11, 1·77)	0·005
Serial TSH				
Normal/High	—	—	1·00	Ref
Low	—	—	1·00 (0·66, 1·54)	0.980
Treatment				
ATD	1·00	Ref	1·00	Ref
RAI without hyperthyroidism control	0·91 (0·67, 1·25)	0·571	0·97 (0·70, 1·36)	0·863
RAI with hyperthyroidism control	0·26 (0·08, 0·82)	0·022	0·27 (0·09, 0·87)	0·028
Thyroidectomy	0·60 (0·15, 2·44)	0·472	0·33 (0·05, 2·39)	0·272

Supplementary Table 7: Multivariable analysis for all-Cause mortality

HR (95% CI) are obtained from multivariable Cox proportional hazard regression models. Model 1, all patients, n=4189; Model 2, patients with thyroid function tests, n=2969; Serial thyroid hormones and treatment were set as time-dependent covariates. Patients contributed person-years of treatment from diagnosis to radioiodine treatment or thyroidectomy (ATD), following radioiodine to control of hyperthyroidism (RAI without hyperthyroidism control), following radioiodine and control of hyperthyroidism (RAI with hyperthyroidism control), and following thyroidectomy (thyroidectomy). a, HR per 5-year increase in age; b, HR per 5-year increase in diagnosis year, c, HR per 10 IU/L increase in TRAb concentration, d, HR per 10 pmol/L increase in FT4 concentration

	Model 1		Model 2	
	HR (95% CI)	**P value**	**HR (95% CI)**	**P value**
Age, years	1·08 (1·07, 1·09) [a]	<0·001	1·08 (1·07, 1·09) [a]	<0·001
Sex				
Male	1·00	Ref	1·00	Ref
Female	0·56 (0·44, 0·71)	<0·001	0·61 (1·07, 1·09)	<0·001
Co-morbidity				
Absent	1·00	Ref	1·00	Ref
Present	2·91 (2·28, 3·72)	<0·001	2·95 (2·22, 3·91)	<0·001
Year of diagnosis	0·75 (0·67, 0·85) [b]	<0·001	0·74 (0·64, 0·86) [b]	<0·001
Baseline TRAb	1·04 (0·93, 1·17) [c]	0·493	1·06 (0·92, 1·22) [c]	0·404
Thyroid hormones				
Baseline FT4	—	—	0·99 (0·99, 1·00) [d]	0·618
Baseline TSH	—	—		
TSH 0.1—0·4 mU/L			1·00	Ref
TSH <0·1 mU/L			0·99 (0·96, 1·01)	0·317
Serial FT4	—	—	1·52 (1·09, 1·94) [d]	0·007
Serial TSH				
Normal/High	—	—	1·00	Ref
Low	—	—	1·00 (0·89, 1·78)	0.990
Treatment				
ATD	1·00	Ref	1·00	Ref
RAI without hyperthyroidism control	1·02 (0·78, 1·34)	0·857	1·08 (·081, 1·43)	0·601
RAI with hyperthyroidism control	0·44 (0·20, 0·94)	0·033	0·45 (0·21, 0·96)	0·040
Thyroidectomy	0·35 (0·09, 1·43)	0·144	0·19 (0·03, 1·40)	0·103

Supplementary Table 8: Multivariable analysis for MACE

HR (95% CI) are obtained from multivariable Cox proportional hazard regression models. Model 1, all patients, n=4189; Model 2, patients with thyroid function tests, n=2969; Serial thyroid hormones and treatment were set as time-dependent covariates. Patients contributed person-years of treatment from diagnosis to radioiodine treatment or thyroidectomy (ATD), following radioiodine to control of hyperthyroidism (RAI without hyperthyroidism control), following radioiodine and control of hyperthyroidism (RAI with hyperthyroidism control), and following thyroidectomy (thyroidectomy). a, HR per 5-year increase in age; b, HR per 5-year increase in diagnosis year, c, HR per 10 IU/L increase in TRAb concentration, d, HR per 10 pmol/L increase in FT4 concentration