

Shorter versus Standard Antithyroid Drug Therapy in Graves' Disease with Lower Baseline TRAb Levels: Relapse Outcomes from a Propensity-Weighted Analysis

Salman Razvi, MD^{1,2}; Mia Holley, PhD³; Hannah Wheeler, MSc⁴; Jonathan Vernazza, MSc⁴; Kathryn Stewart, MD²; Simon Pearce, MD^{1,5}; Kilimangalam Narayanan, MD²; Vasileios Tsalidis, MD².

¹ Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK

² Department of Endocrinology, Queen Elizabeth Hospital, Gateshead, UK

³ School of Medicine, Faculty of Health Sciences and Wellbeing, University of Sunderland, Sunderland, UK

⁴ Department of Biochemistry, Queen Elizabeth Hospital, Gateshead, UK

⁵ Endocrine Unit, Royal Victoria Infirmary, Newcastle upon Tyne, UK

Email addresses for all authors:

Salman.razvi@newcastle.ac.uk

Mia.Holley@sunderland.ac.uk

Hannah.Wheeler3@nhs.net

Jonathan.Vernazza@nhs.net

Kathryn.Stewart1@nhs.net

Simon.pearce@newcastle.ac.uk

k.narayanan1@nhs.net

V.Tsalidis@nhs.net

Abstract

Background: The optimal duration of antithyroid drug (ATD) therapy in Graves' disease is uncertain. Although guidelines recommend 12–18 months, shorter courses may be sufficient in selected patients. We evaluated whether ATD duration <12 months is non-inferior to 12–18 months for relapse among patients with Graves' disease and low thyrotropin receptor antibody (TRAb) levels (>1.8 and <10.0 U/L) at diagnosis.

Methods: We analysed real-world data from adults with Graves' disease and baseline TRAb <10 U/L treated with ATDs at a single centre. The primary endpoint was relapse within 12 months of ATD cessation. A prespecified risk-difference non-inferiority framework was used, with a primary margin of +5% and an exploratory margin of +10%. Treatment effects were estimated using unadjusted analyses, inverse-probability-of-treatment weighting (IPTW; n=369), and 1:1 propensity-score matching (PSM; 104 matched pairs). Secondary outcomes included long-term relapse (median follow-up 58 months), TRAb levels at cessation, and restricted mean survival time (RMST) over 60 months.

Findings: At 12 months, relapse occurred in 23/134 (17.2%) patients treated <12 months and 48/235 (20.4%) treated for 12–18 months. The IPTW-adjusted risk difference was –1.5% (95% CI –11.0 to +8.0), meeting the 10% but not the 5% non-inferiority margin. The unadjusted estimate (–3.2%) met both margins, whereas the PSM estimate (+2.9%, 95% CI –7.3 to +13.1) met neither because of reduced precision. Long-term outcomes were similar between groups (IPTW Cox HR 0.90 [95% CI 0.64–1.28]; RMST difference +3.2 months over 60 months [–2.1 to +8.4]). TRAb levels at cessation were comparable.

Conclusions: Among patients with Graves' disease and low TRAb levels, shorter ATD courses (<12 months) produced outcomes broadly comparable to 12–18 months. Although strict non-inferiority was not confirmed at a 5% margin, findings meeting an exploratory 10% threshold support TRAb-guided individualisation of ATD duration and justify prospective randomised evaluation.

Introduction

Graves' disease is the most common cause of hyperthyroidism worldwide and imposes a substantial clinical, biochemical, psychological, and quality-of-life burden.¹ Antithyroid drugs (ATDs) are the recommended first-line treatment in most adults because they are non-destructive, avoid permanent hypothyroidism, and may induce durable immunological remission.^{2,3} Despite their long-standing use, however, the optimal duration of ATD therapy remains uncertain.⁴

Most international guidelines and expert consensus statements recommend a standard 12–18-month course of ATD therapy for adults with Graves' disease, largely irrespective of baseline disease characteristics (**Online Supplementary Table 1**).^{5–7} This approach implicitly treats Graves' disease as a homogeneous condition and does not account for marked inter-individual differences in relapse risk or immunological activity. As a result, patients with low relapse risk are commonly exposed to the same treatment duration as those with more aggressive disease.

However, Graves' disease is highly heterogeneous and relapse rates following ATD withdrawal range from approximately 30% to 70%.⁶ One of the strongest and most consistent predictors of relapse is the concentration of thyrotropin receptor antibodies (TRAb) at diagnosis and at treatment cessation.⁸ Higher TRAb titres reflect greater autoimmune activity and are associated with increased relapse risk, whereas patients with lower baseline TRAb levels generally have a lower probability of relapse than those with higher antibody titres.⁹ Nevertheless, disease severity and immunological profile are not currently used to guide treatment duration, raising concerns that a fixed 12–18-month strategy may lead to overtreatment in selected patients.

Although ATDs are generally safe, prolonged therapy is not without cost. Serious adverse effects such as agranulocytosis, hepatotoxicity, and pancreatitis are uncommon, while many adverse effects are dose- and time-dependent and occur predominantly during the early phase of treatment.^{1,10} Nevertheless, clinically important adverse events may occur unpredictably and can necessitate treatment discontinuation. Longer treatment courses also increase cumulative patient burden through repeated clinic visits, laboratory

monitoring, and disruption to work and daily life. From a health-system perspective, extended therapy increases resource use without clear evidence of incremental benefit for patients at lower risk of relapse.

Growing evidence suggests that immunological remission often occurs earlier than the traditionally recommended treatment duration, particularly in patients with lower baseline TRAb levels. Several studies demonstrate that TRAb normalisation commonly occurs within the first 6–12 months of ATD therapy in such patients, and contemporary cohorts indicate that relapse risk is highest in the early months following cessation, with many ~~low-risk~~ patients remaining stably euthyroid thereafter.^{8,11-13} These observations challenge the rationale for uniformly prolonged therapy and support the hypothesis that some patients may safely discontinue ATDs earlier.

Despite this, robust evidence to support shorter ATD courses in well-defined populations is limited. Few studies have directly compared short versus standard treatment durations, and most historical trials did not stratify by TRAb or disease severity. Modern observational datasets are often heterogeneous and have rarely applied analytic strategies to mitigate confounding by indication or treatment-duration bias, limiting causal interpretation (**Online Supplementary Table 2**).

To address this evidence gap, we conducted a retrospective observational study of adults with Graves' disease and lower baseline TRAb levels at diagnosis. Because relapse risk is strongly related to baseline TRAb concentration, we focused on patients with lower baseline TRAb levels, a subgroup in whom earlier ATD withdrawal may be most biologically plausible. We compared relapse outcomes between patients treated with short-duration (<12 months) ATD therapy and those treated for the guideline-recommended 12–18 months. Using propensity-score-based methods, including inverse-probability-of-treatment weighting and matching, we aimed to reduce confounding and approximate balanced comparisons between treatment strategies.

The objective of this study was to determine whether, in patients with lower TRAb Graves' disease, a shortened course of ATD therapy is non-inferior (NI) to standard 12 – 18 months with respect to relapse risk. A clinically conservative non-inferiority framework was

prespecified, with primary and exploratory thresholds to reflect both clinical relevance and the expected imprecision of real-world observational data. Using complementary adjusted analyses, we evaluated whether shorter ATD courses could represent a reasonable, individualised alternative for selected patients. We hypothesised that, among patients with Graves' disease and lower baseline TRAb levels, a shortened course of ATD therapy (<12 months) would not be associated with a clinically meaningful increase in relapse risk compared with standard-duration therapy (12–18 months).

Methods

Study Design and Population

This was a proof-of-concept, exploratory, retrospective observational cohort study of adults with newly diagnosed Graves' disease treated with ATDs at a single UK secondary-care endocrine centre, using data from a prospectively maintained clinical database between October 2007 and August 2024. All adult patients (≥ 18 years) with the first episode of newly diagnosed Graves' hyperthyroidism, confirmed by suppressed TSH, elevated free thyroid hormones, and positive TRAb levels, were eligible. For this analysis, we restricted inclusion to patients with lower TRAb levels (>1.8 and <10 U/L). This threshold was selected *a priori* based on previous studies demonstrating a graded relationship between baseline TRAb levels and relapse risk, with more favourable outcomes observed at lower antibody titres.^{13,14} The aim was to identify a subgroup with lower immunological activity while retaining sufficient sample size for meaningful comparative analysis. Patients with TRAb-negative Graves' disease were not included because this represents a distinct clinical subgroup and was not the focus of the present study. Patients were required to have complete baseline thyroid function and TRAb data, documented dates of ATD initiation and cessation, and a minimum of 12 months of post-treatment follow-up to allow ascertainment of relapse. Individuals who had received prior radioiodine therapy, thyroid surgery, lacked essential exposure or outcome data, or those that became pregnant during ATD therapy or during the 12-month post-ATD cessation period were excluded. This study was conducted as a service evaluation of routinely collected clinical data. Formal Research Ethics Committee review was not required under UK Health Research Authority guidance. The project was

reviewed and approved by the Caldicott Guardian at Gateshead Health NHS Foundation Trust (reference no. 2627.0001). The study was conducted in accordance with the principles of the Declaration of Helsinki as revised in 2013. Data were anonymised prior to analysis, and the requirement for individual informed consent was waived in accordance with institutional policies.

Because decisions regarding ATD duration were made in routine clinical practice, and not predetermined by protocol, confounding by indication was expected. To address this, we used causal-inference methods to compare shorter versus standard ATD therapy in patients with Graves' disease and lower TRAb levels.

Procedures

At diagnosis, all patients underwent standard clinical and biochemical evaluation, including serum TSH, free T4 (FT4), free T3 (FT3), and TRAb measurement. Thyroid function parameters and TRAb were measured initially on the Roche Cobas e601 and subsequently on the Roche e802 platforms. The reference ranges changed slightly over the follow-up period; for TSH from 0.40–4.00 mU/L to 0.30–4.50 mU/L, FT4 from 9.0–25.0 pmol/L to 10.0–22.0 pmol/L, FT3 from 3.8–7.0 pmol/L to 3.1–6.8 pmol/L, and TRAb from <1.0 to <1.8 U/L. Clinical assessment included documentation of smoking status, goitre size based on WHO criteria,¹⁵ and the presence of Graves' orbitopathy.¹⁶ Thyroid function and TRAb levels were monitored at routine clinic visits during treatment and after ATD withdrawal.

ATD therapy (typically carbimazole or occasionally propylthiouracil) was prescribed and titrated according to clinician judgement. No patients received adjunctive therapies specifically intended to modify the course of Graves' disease. Treatment practices varied across endocrinologists: some routinely treated for the guideline-recommended 12–18 months regardless of immunological activity, whereas others discontinued therapy earlier once biochemical stability and TRAb normalisation were achieved. Thus, ATD duration reflected real-world clinical behaviour rather than a standardised protocol. ATD cessation was defined as the date therapy was intentionally stopped. TRAb values at cessation were recorded but were not used to determine exposure assignment.

Following ATD withdrawal, thyroid function testing was routinely scheduled at approximately 3, 6 and 12 months according to local clinical practice and clinician preference, although the exact timing varied between patients. Relapse was defined as recurrent overt hyperthyroidism, characterised by suppressed TSH with elevated FT4 or FT3, in conjunction with elevated TRAb levels and resulting in recommencement of ATD therapy. This definition was chosen to maximise specificity and ensure that relapse represented clinically meaningful recurrent Graves' disease requiring treatment, as isolated subclinical hyperthyroidism may be transient and does not necessarily warrant treatment. Long-term outcomes and follow-up durations were extracted from health records and manually verified.

Patients were categorised according to total ATD exposure:

- Short duration: <12 months
- Standard duration: 12–18 months (guideline-recommended reference group)

Outcomes

Primary outcome: Relapse within 12 months after ATD cessation, defined as recurrent overt hyperthyroidism requiring recommencement of ATD treatment.

Secondary outcomes:

1. TRAb level at ATD cessation, analysed as both continuous (log-transformed) and categorical (<1.8 vs ≥1.8 U/L).
2. Long-term relapse, assessed over the entire available follow-up period.

Sensitivity analyses of relapse outcomes

1. 12-month landmark survival analysis, restricted to patients who were event-free at 12 months. Landmark analysis evaluates subsequent outcomes among patients who remain event-free at a pre-defined time point, thereby reducing the influence of early events on longer-term comparisons.

2. Restricted mean survival time (RMST), representing the average relapse-free survival time accrued over a fixed follow-up period, was compared between the groups up to 60 months following ATD cessation.

Propensity Score and Weighting

A propensity score representing the probability of receiving short-duration ATD therapy was estimated using logistic regression including the following baseline variables: age, sex, smoking status, TRAb at diagnosis, FT4 at diagnosis, treating clinician, goitre grade, and presence of Graves' orbitopathy. These variables were selected *a priori* because they plausibly influence both treatment duration decisions and relapse risk. FT4 was selected as the principal biochemical marker of disease severity because it was available in all patients and had less missingness than FT3; inclusion of both variables was considered unlikely to materially improve confounding adjustment and may have introduced collinearity. Stabilised inverse probability of treatment weights (IPTW) were constructed and truncated at the 1st and 99th percentiles. IPTW creates a weighted pseudo-population in which measured baseline covariates are balanced between treatment groups, thereby reducing confounding by indication in observational data. Valid inference relies on the assumptions of no unmeasured confounding, correct specification of the propensity model, and sufficient overlap of propensity scores between groups.¹⁷ Balance was assessed using standardised mean differences (SMD). SMD <0.1 was considered indicative of good balance and <0.15 acceptable in propensity-score matched analyses.¹⁸ All weighted analyses used the survey-weighted approach with robust variance estimation.

Primary Statistical Analysis: Weighted Kaplan–Meier for 12-Month Relapse

The primary analysis estimated 12-month relapse risk using IPTW-weighted Kaplan–Meier methods and included all 369 patients with complete 12-month relapse data. The relapse probability at 12 months and its 95% CI were extracted from the weighted survival curve for each group. The adjusted risk difference (RD) was calculated (short – standard), with NI evaluated using two pre-specified margins:

- 5% absolute risk difference (strict margin)
- 10% absolute risk difference (exploratory, acknowledging limited power)

This study was explicitly designated as exploratory and proof-of-concept; therefore, NI analyses are considered hypothesis-generating.

Secondary Analyses

TRAb at ATD cessation: Group comparisons were made using crude analyses (Wilcoxon rank-sum test), weighted linear regression, and adjusted ANCOVA including baseline TRAb, FT4, age, and sex.

Long-term relapse: Long-term relapse risk was evaluated using IPTW-weighted Kaplan–Meier curves and an IPTW-weighted Cox proportional hazards model.

Sensitivity Analyses

12-month landmark Cox model: Patients who were relapse-free and under follow-up at 12 months were included in a landmark dataset. Time zero was redefined as 12 months post-ATD cessation, and relapse thereafter was analysed using IPTW-weighted Cox regression.

Restricted Mean Survival Time: RMST up to 60 months was compared between groups using RMST2 methodology on the unweighted dataset, supported by descriptive visual assessment of weighted survival curves.

Sensitivity analyses were performed on complete-case data for variables required in each model (347 patients). Missing data were minimal, with only goitre size missing in 22 patients (6%), primarily in patients whose consultations occurred remotely during the COVID-19 pandemic. Relapse rates did not differ significantly between included and excluded patients. Given the low level of missingness, a complete-case approach was used. All analyses were performed in R (version 4.5.2).

Results

Study Population

Among 998 patients with Graves' disease, 369 had baseline TRAb levels between 1.8 and <10 U/L, and also had complete 12-month post-ATD follow-up data (**Figure 1**). Baseline median (IQR) age was 46 (36–56) years, 82% were women, 24% were current smokers,

mean (SD) FT4 was 38 ± 13 pmol/L, FT3 15 ± 6 pmol/L, and median (IQR) TRAb was 4.2 (2.5–7.9) U/L. Baseline characteristics differed modestly by treatment duration, with patients who received shorter therapy being slightly younger and having marginally lower TRAb values, although several covariates showed imbalance prior to propensity adjustment (**Table 1**).

Of these, 134 (36%) received short-duration ATD and 235 (64%) received standard-duration ATD. After IPTW weighting, all baseline covariates achieved good balance (all SMD <0.1) (**Online Supplementary Figure 1**). Balance after propensity-score matching was acceptable but less complete, reflecting reduced sample size and categorical covariate structure (**Table 1**). Median follow-up was 58 months (IQR 14–88).

Primary Outcome: 12-Month Relapse

Relapse by 12 months occurred in 23/134 (17.2%) in the short-duration ATD group and 48/235 (20.4%) in the standard-duration ATD group, with a crude risk difference of –3.2% (95% CI –11.1% to +4.7%). Among patients who relapsed within 12 months, the median (IQR) time to relapse was 7.0 (4.1 – 9.9) months in the short-duration group and 6.0 (4.7 – 7.3) months in the standard-duration group.

The corresponding IPTW-weighted relapse risk at 12 months (primary analysis) was 16.7% and 18.2%, respectively, with an adjusted RD of –1.5% (95% CI –11.0% to +8.0%), which met the 10% NI margin, but not 5% margin (**Online Supplementary Figure 2**).

In the 1:1 PS-matched pairs ($n=104$ per group), relapse occurred in 19.2% and 16.3%, respectively, with an RD of +2.9% (95% CI –7.3% to +13.1%) which did not meet the 10% NI margin.

Therefore, across analyses, point estimates suggested little difference in relapse risk between the short- and standard-duration treatment groups. Two of the three analyses (crude and IPTW) supported NI under the 10% margin (**Figure 2**). Given sample size limitations, all findings are interpreted as exploratory.

Secondary Outcomes:

296 *TRAb at ATD Cessation*

297 The median (IQR) TRAb values in the short- and standard-duration groups were nearly
298 identical: 0.92 (0.72) U/L and 0.90 (0.87) U/L, respectively (**Figure 3A**), when ATD was
299 discontinued. The corresponding proportion of patients who had positive TRAb values (≥ 1.8
300 U/L) were 15.7% and 15.1%, respectively (**Figure 3B**).

301 In adjusted analyses, the IPTW-adjusted difference (log scale) in TRAb levels at ATD
302 cessation was +0.24, with a geometric mean ratio of 1.27 (95% CI 1.01–1.59). In ANCOVA
303 (adjusted for baseline TRAb, FT4, age, sex and smoking status), the geometric mean ratio
304 was 1.29 (95% CI 1.06–1.57). Thus, short-duration therapy was associated with slightly
305 higher adjusted mean log-TRAb, but absolute TRAb values remained well within the normal
306 range, and relapse outcomes were comparable.

307 Hypothyroidism following ATD cessation was uncommon, with overt hypothyroidism
308 occurring in one patient in the short-duration group and two patients in the standard-
309 duration group during the first 12 months of follow-up.

310 *Long-Term Relapse*

311 Over total follow-up, 136 patients relapsed; the IPTW-weighted hazard ratio for long-term
312 relapse was 0.92 (95% CI 0.65–1.31) (**Figure 4**). There was no evidence of a difference in
313 long-term relapse risk between treatment groups. The corresponding weighted 3-year
314 cumulative relapse was 29.4% and 37.3%, respectively, with curves diverging early but
315 converging after approximately 18 months.

316 **Sensitivity Analyses**

317 *12-Month Landmark Analysis*

318 A total of 283 patients were relapse-free at 12 months and included in the landmark cohort.
319 During subsequent follow-up, 95 late relapses occurred and late relapse risk (post-12
320 months) was comparable in both groups with a HR of 0.96 (95% CI 0.61–1.50) (**Online**
321 **Supplementary Figure 3**). This indicates that relapse risk beyond the first year was similar
322 between treatment groups.

Restricted Mean Survival Time (RMST)

At 60 months, RMST was 45.2 months in the short-duration and 42.2 months in the standard-duration groups, with a difference of +3.1 months (95% CI –2.1 to +8.4; $p=0.13$) (**Online Supplementary Figure 4**). Thus, patients on shorter therapy spent slightly more relapse-free time over 5 years, though it was not statistically significant.

Overall, across all analytic approaches, shorter ATD duration (<12 months) performed comparably to standard duration (12–18 months) in patients with Graves' disease and baseline TRAb levels between 1.8 and 10.0 U/L. The study is exploratory but consistently shows no evidence that shorter treatment increases relapse risk, particularly beyond the first year.

Discussion

In this exploratory analysis of patients with Graves' disease and lower baseline TRAb levels, ATD treatment for less than 12 months appeared to provide relapse outcomes comparable to the standard 12–18 months of therapy recommended by international guidelines. Importantly, relapse remained relatively common even in this lower-TRAb cohort, highlighting that lower immunological activity should not be equated with negligible long-term risk. Current recommendations are largely based on evidence derived from unselected Graves' disease populations, with limited attention to baseline TRAb levels. The rationale for applying a universal 12–18-month course to all patients rests on the assumption that extending treatment improves the probability of achieving durable immunological remission. However, this assumption has not been rigorously tested in patients with lower baseline immunological activity, particularly those with lower TRAb levels at diagnosis. Our study directly addresses this evidence gap and provides initial real-world insights into how treatment duration might be personalised using early immunological markers.

A central motivation for this work is the recognition that Graves' disease represents a spectrum of immunological activity rather than a uniform entity. Patients with lower TRAb titres tend to have less aggressive disease, greater likelihood of spontaneous immune downregulation, and more rapid normalisation of thyroid function tests following treatment initiation.^{8,14} In such patients, prolonged ATD exposure may offer diminishing marginal

benefit once immunological remission has been reached. Although adjusted analyses showed slightly higher TRAb levels at ATD cessation in the short-duration group, values remained within the reference range in both groups and TRAb positivity rates were similar, suggesting limited additional immunological benefit from extending treatment beyond 12 months. Taken together, these observations suggest that a “biomarker-guided” approach to ATD discontinuation may be feasible, whereby therapy is stopped once TRAb has normalised, rather than after a fixed calendar duration.

The findings also highlight an important consideration regarding non-inferiority assessment in observational settings. Because short-duration therapy is not established as equivalent, a strict non-inferiority framework requires strong justification for the chosen margin. We selected a primary non-inferiority margin of +5%, reflecting the approximate 20% baseline relapse risk in mild disease and representing a clinically conservative threshold. A more permissive secondary margin of +10% was prespecified to allow exploratory interpretation in the context of modest sample size and expected imprecision. Although our adjusted analysis did not satisfy the stricter margin, most estimates remained well within the more permissive boundary and did not suggest any clinically meaningful excess early relapse risk with shorter therapy. Importantly, failure to satisfy a very narrow margin does not imply inferiority; it simply reflects limited statistical precision. Strict non-inferiority testing requires high precision and narrow confidence intervals; consequently, observational studies, particularly those using propensity-based adjustment, are often underpowered for definitive non-inferiority conclusions and are best interpreted as proof-of-concept.¹⁹

Long-term relapse patterns also provide context for clinical decision-making. In Graves’ disease, most relapses occur within the first 12–24 months after ATD cessation.⁸ If extended therapy were to substantially modify the natural history of the autoimmune process, one would expect persistent divergence between groups after the treatment window. Instead, subsequent relapse trajectories were similar regardless of whether therapy extended beyond 12 months, suggesting that any potential influence of treatment duration on relapse risk is most likely to occur during the early post-withdrawal period. The landmark and RMST analyses further supported this interpretation and illustrated how longer follow-up can complement early outcome measures in understanding relapse dynamics.

This study also underscores the value of real-world clinical data for examining treatment decisions that are difficult to evaluate in randomised trials. Treatment duration in practice is influenced by a combination of clinician preference, patient characteristics, symptom profile, and biochemical response. These contextual factors are rarely captured in a trial environment but are central to understanding why some patients receive shorter or longer treatment. By using inverse-probability weighting and propensity-score matching, we attempted to address confounding by indication and create more comparable groups. Weighting achieved excellent covariate balance across multiple clinically relevant predictors, strengthening internal validity. While observational methods cannot fully eliminate unmeasured confounding, they provide an important complement to trial evidence and allow exploration of clinically relevant hypotheses in populations that are not typically studied.

Nevertheless, several limitations warrant careful consideration. This was a retrospective analysis based on routinely collected clinical data, and residual confounding remains possible, particularly from variables not captured in the dataset such as medication adherence or subtler clinical cues (e.g. asymptomatic thyroid eye disease) that influence physicians' decisions. Missing data were minimal; only goitre size was missing in 6% of patients, resulting in exclusion of these individuals from weighted analyses because inverse-probability weighting requires complete covariate information. Although the proportion excluded was small and relapse rates did not differ between included and excluded patients, complete-case analysis may still introduce some risk of selection bias. Additionally, our findings apply specifically to patients with baseline TRAb levels between 1.8 and 10.0 U/L and should not be extrapolated to individuals with more active, severe, or clinically complicated Graves' disease, or to the distinct subgroup of patients with TRAb-negative Graves' disease. Finally, because this was not a randomised controlled trial, causal conclusions must remain tentative, and the analyses should be interpreted as exploratory rather than confirmatory.

Despite these limitations, the present findings offer clinically meaningful insights. For patients with Graves' disease and lower baseline TRAb levels whose TRAb levels normalise with treatment, continuing ATDs beyond 12 months may offer limited additional value.

Earlier cessation could reduce pill burden, improve quality of life, and minimise exposure to rare but potentially serious adverse effects such as agranulocytosis or hepatotoxicity. However, decisions regarding treatment duration should remain individualised, particularly for patients with moderate or severe disease, persistent TRAb elevation, ophthalmopathy, or large goitre. Future research should focus on prospective validation of TRAb-guided treatment algorithms and randomised comparisons of early cessation versus standard duration in patients stratified by baseline immunological activity, including evaluation of alternative TRAb thresholds that may better identify candidates for shorter treatment.

In summary, this exploratory analysis suggests that, in patients with Graves' disease and lower baseline TRAb levels, shorter ATD courses may yield relapse outcomes broadly comparable to those achieved with standard 12–18-month therapy. While strict non-inferiority was not demonstrated under a conservative margin, the overall findings support further evaluation of personalised, biomarker-guided approaches to ATD duration.

Acknowledgments:

None

Author contributions

Salman Razvi: Conceptualization (lead); formal analysis (lead); writing – original draft (lead); writing – review and editing (equal).

Mia Holley: Formal analysis (equal); writing – review and editing (equal).

Hannah Wheeler: Writing – review and editing (equal).

Jonathan Vernazza: Writing – review and editing (equal).

Kathryn Stewart: Writing – review and editing (equal).

Simon H.S. Pearce: Writing – review and editing (equal).

Kilimangalam Narayanan: Writing – review and editing (equal).

Vasileios Tsalidis: Writing – review and editing (equal).

Author Disclosure Statement

439 Salman Razvi: No competing financial interests exist.
440 Mia Holley: No competing financial interests exist.
441 Hannah Wheeler: No competing financial interests exist.
442 Jonathan Vernazza: No competing financial interests exist.
443 Kathryn Stewart: No competing financial interests exist.
444 Simon H.S. Pearce: No competing financial interests exist.
445 Kilimangalam Narayanan: No competing financial interests exist.
446 Vasileios Tsalidis: No competing financial interests exist.

447

448 **Funding statement**

449 Salman Razvi: Funded by a National Institute for Health and Care Research Advanced
450 Fellowship.

451 Mia Holley: No funding to declare.

452 Hannah Wheeler: No funding to declare.

453 Jonathan Vernazza: No funding to declare.

454 Kathryn Stewart: No funding to declare.

455 Simon Pearce: Part-funded by a Medical Research Council grant.

456 Kilimangalam Narayanan: no funding information to declare.

457 Vasileios Tsalidis: No funding information to declare.

458

459 **Data availability statement:** De-identified individual participant data that underlie the
460 results reported in this article will be made available to bona fide researchers on reasonable
461 request, subject to institutional approvals, data-sharing agreements, and compliance with

relevant data protection and governance requirements. Requests should be directed to the corresponding author. No additional data are available.

Corresponding author

Salman Razvi, MD

Translational and Clinical Research Institute, Newcastle University,

Centre for Life, Central Parkway,

Newcastle upon Tyne, NE1 3BZ, United Kingdom

Email: salman.razvi@newcastle.ac.uk

References

1. Chaker L, Cooper DS, Walsh JP, et al. Hyperthyroidism. *Lancet*. 2024;403:768–780.
2. Wiersinga WM, Poppe KG, Effraimidis G. Hyperthyroidism: aetiology, pathogenesis, diagnosis, management, complications, and prognosis. *Lancet Diabetes Endocrinol*. 2023;11:282–298.
3. Villagelin D, Cooper DS, Burch HB. A 2023 international survey of clinical practice patterns in the management of Graves disease: a decade of change. *J Clin Endocrinol Metab*. 2024;109:2956–2966.
4. Azizi F, Abdi H, Mehran L, et al. Appropriate duration of antithyroid drug treatment as a predictor for relapse of Graves' disease: a systematic scoping review. *J Endocrinol Invest*. 2022;45:1139–1150.
5. Kahaly GJ, Bartalena L, Hegedüs L, et al. 2018 European Thyroid Association guideline for the management of Graves' hyperthyroidism. *Eur Thyroid J*. 2018;7:167–186.
6. 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;26:1343–1421.
7. National Institute for Health and Care Excellence. Thyroid disease: assessment and management (NG145). London: NICE; 2019 (updated 2023). Available from: <https://www.nice.org.uk/guidance/ng145> (accessed March 3, 2026)
8. Tun NNZ, Zammitt NN, Strachan MWJ, et al. Ten year outcome of anti-thyroid drug treatment for first episode Graves' thyrotoxicosis: the predictive importance of TRAb. *Thyroid*. 2025;35:612–618.
9. Vos XG, Endert E, Zwinderman AH, et al. Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with Graves' hyperthyroidism. *J Clin Endocrinol Metab*. 2016;101:1381–1389.
10. Pearce SHS. Spontaneous reporting of adverse reactions to carbimazole and propylthiouracil in the UK. *Clin Endocrinol*. 2004;61:589–594.

11. Cappelli C, Gandossi E, Castellano M, et al. Prognostic value of thyrotropin receptor antibodies (TRAb) in Graves' disease: a 120 months prospective study. *Endocr J*. 2007;54:713–720.
12. Jin M, Jang A, Kim CA, et al. Long-term follow-up result of antithyroid drug treatment of Graves' hyperthyroidism in a large cohort. *Eur Thyroid J*. 2023;12:e220226.
13. Karmisholt J, Andersen SL, Bulow-Pedersen I, et al. Predictors of initial and sustained remission in patients treated with antithyroid drugs for Graves' hyperthyroidism: the RISG study. *J Thyroid Res*. 2019;2019:5945178.
14. Chen Z, Xu J, Kang W, et al. High TRAb titer at diagnosis predicts persistent positivity and relapse in Graves' disease after prolonged antithyroid therapy. *Endocrinol Metab*. 2025;40:950–960.
15. World Health Organization. Assessment of iodine deficiency disorders and monitoring their elimination: a guide for programme managers. Geneva: World Health Organization; 2007. Available from: <https://apps.who.int/iris/handle/10665/43781> (accessed March 21, 2026)
16. Bartalena L, Baldeschi L, Dickinson AJ, et al. Consensus statement of the European group on Graves' orbitopathy (EUGOGO) on management of Graves' orbitopathy. *Thyroid*. 2008;18:333–346.
17. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
18. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28:3083–3107.
19. Cuzick J, Sasieni P. Interpreting the results of noninferiority trials – a review. *Br J Cancer*. 2022;127:1755–1759.