



Long-term follow-up of treatment outcomes in Graves' disease and toxic nodular disease

David Veríssimo¹ · Beatriz Pereira¹ · Joana Vinhais¹ · Catarina Ivo¹ · Ana C. Martins¹ · João N. e Silva¹ · Dolores Passos¹ · Luís Lopes¹ · João J. de Castro¹ · Mafalda Marcelino¹

Received: 14 June 2024 / Accepted: 7 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Purpose Hyperthyroidism guidelines have not been updated over the past five years, despite numerous data on the subject, and recent studies providing a wide variation in treatment success rates. We aim to compare the effectiveness and safety of treatment modalities in patients with Graves' disease or toxic nodular disease.

Methods Single center retrospective cohort study of Graves' disease and toxic nodular disease patients treated between 1983 and 2023.

Results A total of 411 patients were treated for hyperthyroidism, 245 due to Graves' disease and 166 due to or toxic nodular disease, followed for a median of 7 years. In Graves' disease, 90.2% were treated with antithyroid drugs over 250 cycles, achieving 41.7% cumulative remission. Half of all relapses (50.9%) occurred in the first year, 76.3% in the first three years, and 98.3% within nine years. Treatment periods of 12–24 months showed higher remission and lower relapse rates than longer periods. I-131 was used in 103 cycles with 82.5% remission and 7.1% relapse. A total of 29 thyroidectomies resulted in 100% remission, with no relapse. In toxic nodular disease, surgery was the most frequently used treatment (54.5%), followed by I-131 (37.1%).

Conclusion Our findings support antithyroid drugs as the preferential first-line treatment for Graves' disease, allowing for euthyroidism with minimal adverse effects. Given the propensity for relapse, we suggest a rigorous monitoring, particularly within the first three years. In toxic nodular disease, surgery should be the preferred option, with I-131 being reserved for single adenomas and small goiters.

Keywords Antithyroid drugs · Hyperthyroidism · Graves' disease · Thyrotoxicosis · Toxic nodular disease

Introduction

Hyperthyroidism is a subtype of thyrotoxicosis, caused by an excessive synthesis and secretion of thyroid hormones. The most frequent causes of hyperthyroidism are Graves' disease (GD), resulting in an excessive thyroid stimulation, and toxic nodular disease (TND) (including toxic multinodular goiter and solitary toxic adenoma), responsible for an autonomous thyroid hormone secretion [1].

GD represents 70% of cases of hyperthyroidism, with an annual incidence of 20–50 cases per 100,000 people per

year.² This disease is more frequent in women (6–10:1 female/male prevalence) and has an age peak between 30–50 years [2, 3].

Toxic multinodular goiter and solitary adenoma represent most of the remaining hyperthyroidism cases (10 and 6% respectively). The prevalence of TND increases with age and is more frequent in older patients, especially in iodine-deficient populations [1].

There are three distinct treatment options: antithyroid drugs (ATD), radioablation with I-131 and surgery. These treatments can be applied alone or as a combined therapy, depending on the cause, severity, initial response to treatment, geographical influences, and patients' own preference [3].

In GD, ATD have been the treatment of choice in European countries, despite achieving lower remission rates (50–55 vs. 62–93% with I-131 vs. 100% post-surgery) and higher relapse rates (52–53% vs. 8–15 after I-131 and

✉ David Veríssimo
d.verissimo87@gmail.com

¹ Department of Endocrinology, Portuguese Armed Forces Hospital, Lisboa, Portugal

0–10% after surgery) [2, 4, 5]. This preference for ATD may be explained by the low rate of associated complications, such as hypothyroidism. To improve the high relapse rate, predictive scores to help selecting the initial therapy have been proposed, with limited results [6].

On the contrary, TND is usually managed with I-131 and surgery, as treatment with ATD is associated with poor outcomes and high relapse rates [1]. I-131 is often preferred as it achieves a higher percentage of success (94%) than surgery, with a smaller risk of hypothyroidism (although still high in toxic multinodular goiter and when more than one dose of I-131 is applied) [1, 2].

Although several studies analyzing the treatment outcomes in both GD and TND have been published to date, neither American or European guidelines have been revised for the past years 5 years, and recent studies have presented a wide variation of success rates [1, 4].

The aim of this study is to characterize and compare the effectiveness and safety of the three available treatment choices (ATD, I-131 and surgery) in patients diagnosed with GD and TND, followed in the last 40 years in a military hospital in Portugal. We hope that data collected from this long-time span, will add to the evidence obtained to date.

Material e methods

Study design

Retrospective cohort study of the clinical records of patients with hyperthyroidism due to GD or TND, treated in the Department of Endocrinology of the Portuguese Armed Forces Hospital, a military tertiary hospital, in Lisbon, between 01/03/1983 and 19/07/2023.

The medical records underwent manual analysis, and only those containing a comprehensive initial observation establishing the diagnosis, a clear treatment decision, and legible medical notes with laboratory assessments throughout the entire follow-up period were included in our study.

Patients

Inclusion criteria: patients diagnosed with hyperthyroidism due to GD or TND, minimum age 18 years, treated with ATD, I-131 or surgery. Minimum follow-up of 2 years after treatment. The diagnosis of GD required laboratory tests indicating hyperthyroidism, alongside the presence of at least one of the following: elevated TSH receptor auto-antibodies (TRABs), diffuse uptake on I-123 scintigraphy, or the manifestation of Graves' ophthalmopathy. TND diagnosis was confirmed by the presence of

hyperthyroidism in laboratory assessments, nodular disease detected via ultrasound, and focal uptake observed on I-123 scintigraphy.

Exclusion criteria: children (<18 years), pregnant or lactating patients. Patients treated with "block and replace" (combination of ATD and levothyroxine to allow for maximum doses of ATD while maintaining euthyroidism) were excluded to minimize bias.

Treatment protocol

Patients diagnosed with Graves' Disease (GD) were primarily managed with ATD or I-131, while surgery (total thyroidectomy) was reserved for concomitant nodular disease, patients with severe ophthalmopathy, ATD failure, or according to the patient's preference. The duration of ATD treatment ranged from a minimum of 12 months to 24 months and was preferably discontinued only after the normalization of TRABs. If disease remission was not achieved within 24 months, patients were preferentially treated with I-131 or surgery. Only those patients who either refused or were ineligible for I-131 or surgery continued ATD treatment beyond 24 months.

In TND, the preferred options were I-131 if the nodules were smaller than 3 cm or surgery in the remaining cases (total thyroidectomy, subtotal thyroidectomy or lobectomy depending on the nodules' location and extension). Treatment with ATD was mostly used to achieve disease control while waiting for a definitive treatment, in frail and older patients with mild (subclinical) disease, in those that refused or were ineligible for ¹³¹I or surgery.

The ATDs used were methimazole (MMI) and propylthiouracil (PTU), in a titration regimen to achieve euthyroidism.

I-131 activity was calculated based on the thyroid mass (for GD) or total toxic nodular mass (for TND) assessed by ultrasound, and on the 24-hour thyroid radioiodine uptake.

Effectiveness

Effectiveness was analyzed by comparing the disease control, remission, failure, and relapse rates of each treatment for GD and TND separately.

Disease control was defined as the normalization of TSH and FT4 values during each treatment protocol and the time to control as the time from treatment initiation to disease control.

Disease remission was defined as the maintenance of euthyroidism or hypothyroidism following the suspension of ATD for at least 3 months, following ¹³¹I for at least 6 months, or immediately after surgery. Time to remission was defined as the time from treatment initiation to remission.

In the 2016 American Thyroid Association (ATA) guidelines a patient is only considered to be in remission if they have had a normal serum TSH, free T4, and total T3 for 1 year after discontinuation of ATD therapy [1]. Although this seems reasonable and is used in the clinical practice, we chose to use the previously described definitions of effectiveness to properly evaluate the remission and relapse rates in the first year after discontinuing ATD therapy.

Treatment failure was considered for patients that couldn't achieve disease control, couldn't maintain euthyroidism or hypothyroidism after ATD suspension or following I-131 or surgery, and when ATD treatment had to be suspended due to adverse effects. Time to failure was defined as the time from treatment initiation to failure.

Disease relapse was defined as the diagnosis of hyperthyroidism following disease remission and time to relapse the time between disease remission and relapse of hyperthyroidism.

Safety

All the adverse effects described in the clinical records and related to each treatment were collected and analyzed.

Statistical analysis

Continuous data are presented as mean and standard deviations or median and interquartile ranges (IQR), according to their adaptation to a normal distribution, analyzed using the Shapiro-Wilk test. Differences among groups were assessed with Student's T-test or Mann-Whitney U-test.

Categorical variables are presented as absolute and relative frequencies and their associations studied using the Pearson's chi-square test and Fisher's exact test as appropriate.

Associations were adjusted for potential covariates, including age, sex, free T4 at diagnosis, TRABs at diagnosis and at end of treatment, thyroid volume, and World Health Organization (WHO) goiter grade, using logistic regression for categorical outcomes and linear regression for continuous outcomes.

All analyses were performed using SPSS software version 23®, with a significance level set to 5%.

Results

Between March 1983 and April 2023, 411 patients were treated at our department for hyperthyroidism, 245 due to GD and 166 due to TND, followed for a median of 7 [3; 15] and 7 [3; 13] years, respectively. In GD, 146 patients

Table 1 Patients baseline laboratory results

Laboratory results	Graves' disease	Toxic nodular disease	Reference
TSH, μ UI/mL	0.005 [0.003–0.01]	0.016 [0.005–0.051]	0.350–4.940
FT4, ng/dL	3.00 [2.13–4.41]	1.49 [1.12–2.20]	0.7–1.48
FT3 pg/mL	7.49 [5.08–10.66]	3.62 [2.28–5.39]	1.71–3.1
T3, ng/mL	4.31 [3.38–5.85]	2.54 [1.67–4.21]	0.35–1.93
TRABs, UI/L	11 [4.6–20.9]	–	<1.6

Square brackets are medians [quartiles]

Table 2 Patients baseline radiological and nuclear characteristics

Graves' disease	
Ultrasound, <i>n</i> = 245	
Normal	6 (2.4)
Diffuse goiter	157 (64.1)
Nodular goiter	82 (33.5)
Volume, mL	18.31 [12.50–28.65]
I-123 Scintigraphy, <i>n</i> = 146	
Diffuse uptake	139 (95.2)
Diffuse and focal uptake	7 (4.8)
Toxic nodular disease	
Ultrasound, <i>n</i> = 166	
Single Adenoma	34 (20.5)
Multinodular goiter	132 (79.5)
Volume, mL	24.2 [15.4–42.9]
I-123 Scintigraphy, <i>n</i> = 164	
Unifocal uptake	118 (72)
Multifocal uptake	46 (28)

n (%) are absolute and relative frequencies; Square brackets are medians [quartiles]

(59.6%) were monitored for over five years, with 93 patients (37.9%) followed for more than a decade. In the TND cohort, 103 patients (62%) were tracked for over five years, and 55 patients (33.1%) for more than ten years.

Both groups showed female predominance (62% in GD and 69.9% in TND). Median age at diagnosis was 44 [31–54] years in the GD group and 63 [51–71] years in TND.

Baseline laboratory, radiological and nuclear characteristics are summarized in Tables 1 and 2.

Effectiveness

Results regarding effectiveness are summarized in Table 3.

A total of 564 treatments were performed, 382 for GD and 182 for TND.

Table 3 Effectiveness analysis of Graves' disease and toxic nodular disease treatments

	Graves' Disease	Toxic nodular disease
Total treatments	<i>n</i> = 382	<i>n</i> = 182
First-line treatment		
Antithyroid drugs	221 (90.2)	14 (8.4)
I-131	22 (9.0)	62 (37.1)
Surgery	2 (0.8)	91 (54.5)
Antithyroid drugs	<i>n</i> = 250	<i>n</i> = 14
Methimazole	176 (70.4)	12 (85.7)
Propylthiouracil	74 (29.6)	2 (14.3)
Duration, months	19 [14–33.3]	10 [4.5 – 25]
Positive TRAbs at end of treatment ^a	131 (65.8)	–
Disease control	222 (88.8)	11 (78.6)
Time to control, months	4 [2–6]	4.5 [2–10.75]
Treatment failure	116 (46.4)	8 (57.1)
Adverse effects	14 (12.1)	0
No control	28 (24.2)	3 (37.5)
No remission	74 (63.7)	5 (62.5)
Disease remission	134 (53.6)	6 (42.9)
Time to remission, months	22.5 [17.8–33]	36 [1–40]
Disease relapse	59 (23.6)	2 (14.3)
Time to relapse, month	12 [7–33]	–
I-131	<i>n</i> = 103	<i>n</i> = 72
Second cycle	9 (8.7)	2 (2.8)
Activity, mCi	8 [6–10]	8 [6–10]
Treatment failure	17 (16.5)	6 (8.3)
Disease remission	85 (82.5)	66 (91.7)
Time to remission, months	2 [1–4]	4 [1.75–5]
Remission duration, months	74 [25–156.5]	38 [8–81.5]
Disease relapse	6 (7.1)	1 (1.5)
Time to relapse, months	9 [3.5–48]	6
Surgery	<i>n</i> = 29	<i>n</i> = 96
Total	29 (100)	62 (64.6)
Subtotal/Lobectomy	0	34 (35.4)

^aFrom a total of 199 patients with TRAbs measured at the end of treatment. *n* (%) are absolute and relative frequencies); Square brackets are medians [quartiles]

In GD, the most frequent first-line option was ATD (90.2%), with only a small fraction initially treated with I-131 (9%) and surgery (0.8%).

Overall, a total of 250 treatment cycles with ATD were used in GD patients, mostly MMI (70.4%), for a median duration of 19 [14–33.3] months. Disease control was obtained in 88.8%, with a median time to control of 4 [2–6]

months, treatment failure occurred in 46.4%, and remission was achieved in 53.6%, with a median time to remission of 22.5 [17.8–33] months.

Among patients who didn't achieve disease control (24.2% of all treatment failures), the main reason (63.7%) was difficulty in maintaining adequately high doses of ATD for a sufficient duration without inducing hypothyroidism. In 28.5% of cases, patients were not capable of complying with the treatment protocol, and only a minority (7.8%) failed to attain control despite receiving maximum doses of ATD.

In the 134 patients that achieved remission, 44% relapsed after a median time of 12 [7–33] months. All relapses occurred within a 14-year follow-up period (Fig. 1A, B), with a cumulative remission of 41.7% at the end of follow-up (33 years after treatment, Fig. 1A). More than half (50.9%) were observed in the first year after treatment, 76.3% in the first three years, 86.4% in the first five years and 98.3% within nine years (Fig. 1B).

When comparing MMI (70.4% ATD treatments) with PTU (29.6%), there was no difference in disease control (91.5% vs. 82.4%, *p* = 0.078), treatment failure (46.6 vs. 45.9, *p* = 0.844), disease remission (53.4 vs. 54.1, *p* = 0.844) or relapse (43.6 vs. 45, *p* = 0.998).

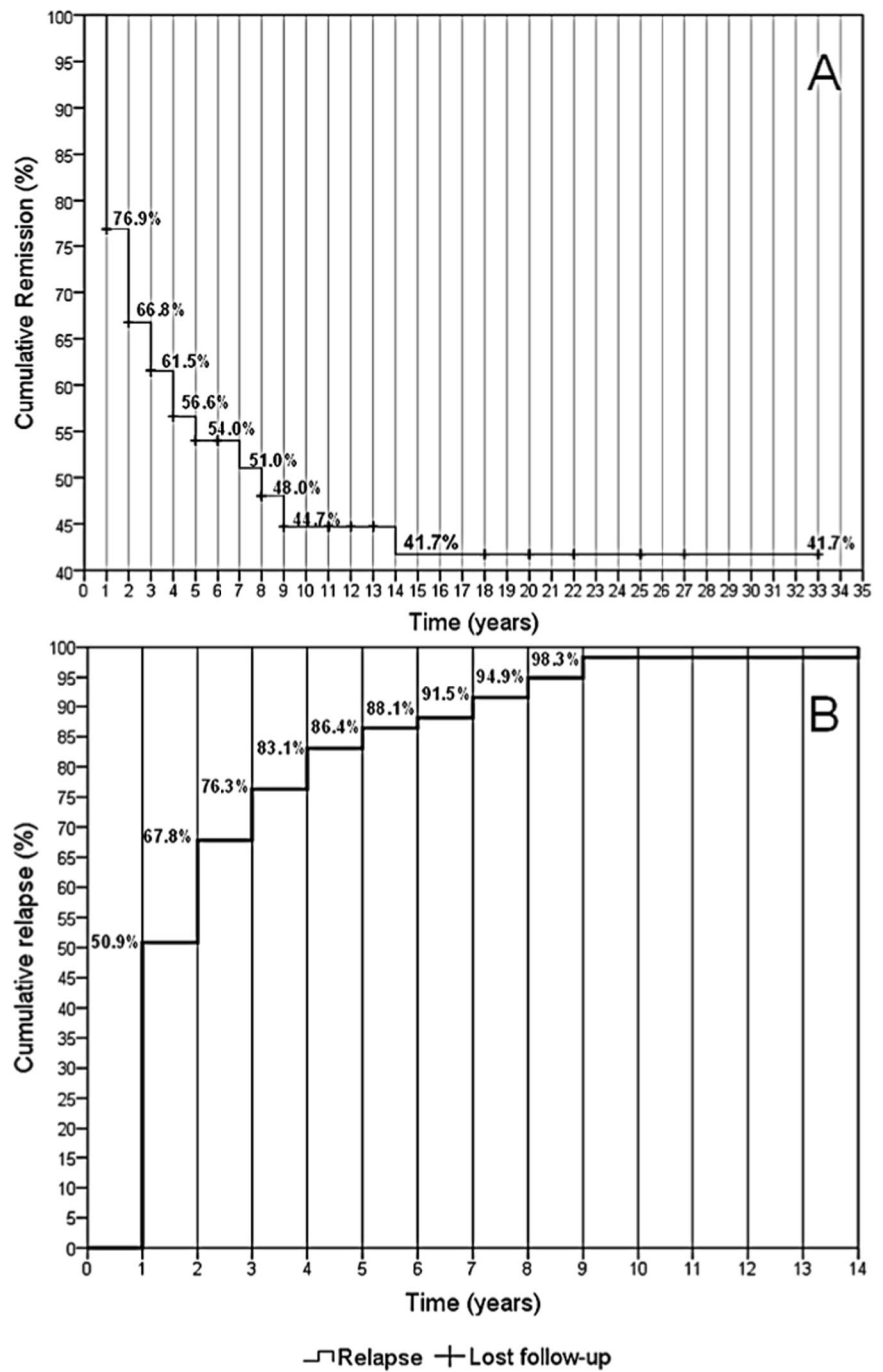
ATD treatment periods of 12–24 months were associated with higher remission rates (69.9 vs. 45.9%, *p* = 0.002) and lower relapse rates (38.8 vs. 56.4%, *p* = 0.046) than periods longer than 24 months. There was no difference between patients treated for 12–18 and those treated for 19–24 months (remission 68 vs. 73%, *p* = 0.523, relapse 42 vs. 30%, *p* = 0.995) (Fig. 2).

Applying the remission definition defined by the 2016 American Thyroid Association guidelines (normal serum levels of TSH, free T4, and total T3 for 1 year post discontinuation of ATD therapy), yields similar outcomes between 12–18-month and 19–24-month ATD duration (*p* = 0.257 for remission and *p* = 0.803 for relapse). However, when using this stricter criterion, the previously observed difference between 12–24-month-group and >24-month-group dissipates (remission 42.7 vs. 30.6, *p* = 0.076, and relapse 34 vs. 50%, *p* = 0.173).

Positive TRAbs at the end of ATD treatment and WHO goiter grade were associated with a higher relapse rate (OR 7.202; *p* < 0.001 and OR 3.079; *p* = 0.002, respectively). Neither gender, age, baseline FT4, baseline TRAbs concentration, or thyroid volume impacted remission or relapse rates (Table 4).

I-131 was used in GD for a total of 103 cycles, with a median activity of 8 [6–10] mCi. Treatment failure was observed in 17.5% and remission in 82.5% following a median of 2 [1–4] months. Relapse occurred in 7.1%, after a median of 9 [3.5–48] months. Nine patients (8.7%) required a second cycle to achieve remission. Compared to ATD

Fig. 1 Cumulative remission¹ (A) and relapse¹ (B) following antithyroid drugs treatment. ¹Only included patients that relapsed



treatment, I-131 was associated with higher disease remission (82.5 vs. 53.6%, $p < 0.0001$) lower treatment failure (17.5 vs. 46.4%, $p < 0.0001$) and lower relapse rate (7.1 vs. 44%, $p < 0.001$).

A total of 29 total thyroidectomies were performed in GD patients, resulting in 100% remission and 100% hypothyroidism, with no relapse.

Due to treatment failure or disease relapse, 137 patients underwent more than one treatment: 78.8% were submitted to a definitive treatment, either I-131 (59.1%) or

surgery (19.7%), while 21.2% completed a second ATD treatment cycle.

In TND, surgery was the most frequently used first-line treatment (54.5%), followed by I-131 (37.1%), and only a small percentage was treated with ATD (8.4%). Overall, 96 surgical procedures were performed in these patients, 64.6% total thyroidectomies and 35.4% lobectomies, achieving 100% remission rate. Most were due to multinodular goiter (72.1%) but were also performed in large volume toxic adenomas (27.9%). I-131 was the treatment

Fig. 2 Association between antithyroid drugs treatment duration and failure, remission and relapse rates

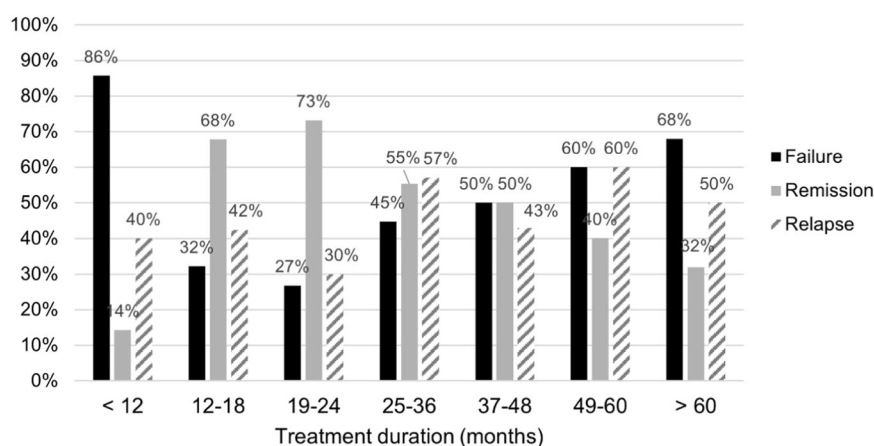


Table 4 Antithyroid drugs treatment efficacy predictor analysis

Variable	Odds ratio	<i>p</i> -value
Remission rate		
Male sex	0.858	0.580
Age < 40	0.852	0.561
FT4 ≥ 40 pmol/L	1.214	0.624
TRAbs > 3X upper limit at diagnosis	0.593	0.120
Positive TRAbs at end of treatment	0	0.996
Thyroid volume	1.014	0.321
WHO Goiter grade ≥ 2	1.226	0.452
Relapse rate		
Gender	0.802	0.569
Age < 40	1.265	0.538
FT4 ≥ 40 pmol/L	2.658	0.074
TRAbs > 3X upper limit at diagnosis	1.023	0.957
Positive TRAbs at end of treatment	6.970	<0.001
Thyroid volume	0.971	0.110
WHO Goiter grade ≥ 2	4.071	<0.001

choice in 72 cases, using a median activity of 8 [6–10] mCi, rendering 91.7% remission (77.8% in toxic multinodular goiter and 95.5% in toxic adenoma), with only 1.5% relapse. Two patients required a second I-131 cycle to achieve remission.

Fourteen patients with TND were treated with ATD, of which eleven (78.6%) achieved disease control, six (42.9%) managed to stop their medication and achieve remission. Only two of these relapsed.

Safety

The adverse effects of each treatment are summarized in Table 5.

ATD treatment in GD patients resulted in a small percentage of adverse effects (12%), mostly minor, and

Table 5 Treatment-related adverse effects

	Graves' Disease	Toxic nodular disease
Antithyroid drugs	30 (12)	0
Gastrointestinal	3 (1.2)	–
Skin	6 (2.4)	–
Hematological	2 (0.8)	–
Liver	7 (2.8)	–
Rheumatological	2 (0.8)	–
Hypothyroidism	10 (4)	–
I-131	74 (71.8)	18 (25)
Hypothyroidism	74 (71.8)	18 (25)
Surgery	29 (100)	78 (81.3)
Permanent hypoparathyroidism	7 (24.1)	8 (10.3)
Hypothyroidism	29 (100)	66 (84.6)
Recurrent laryngeal nerve injury	–	4 (5.1)

n (%) are absolute and relative frequencies

primarily hypothyroidism (4%). No adverse effects were observed in TND patients.

I-131 treatment was associated with a lower incidence of hypothyroidism compared to surgical treatment in both GD and TND patients (71.8 vs. 100%, $p = 0.001$, and 25 vs. 84.6%, $p < 0.001$, respectively). There were no records of I-131-associated malignancies during the follow-up period.

Surgical treatment resulted in permanent hypoparathyroidism in 24.1% of GD patients and 10.3% of TND patients.

Discussion

Our main goal was to characterize and compare the effectiveness and safety of ATD, I-131 and surgery on patients diagnosed with GD and TND.

Table 6 Comparison of the treatment outcomes in Graves' disease with the literature

Treatment	Follow-up Years	Remission %	Relapse %
Graves' Disease			
Anti-thyroid drugs			
Veríssimo et al. ^a , <i>n</i> = 250	7 [3; 15]	53.6	44
Sjolin et al., <i>n</i> = 596 [9]	8 ± 0.9	58.9	31.7
Zarif et al., <i>n</i> = 60 [10]	–	53	–
Sundaresh et al., <i>n</i> = 118 [11]	1.5 ± 2.2	54.2	27.4
Sundaresh et al., <i>n</i> = 667 [12]	4.8	–	52.7
Anagnostis et al., <i>n</i> = 135 [13]	4.7 ± 0.5	–	39
Mohlin et al., <i>n</i> = 219 [14]	10	–	43.5
I-131			
Veríssimo et al. ^a , <i>n</i> = 103	7 [3; 15]	82.5	7.1
Sjolin et al., <i>n</i> = 324 [9]	8 ± 0.9	81.5	–
Zarif et al., <i>n</i> = 40 [10]	–	95	–
Sundaresh et al., <i>n</i> = 543 [11]	1.5 ± 2.2	92	–
Sundaresh et al., <i>n</i> = 314 [12]	64	–	15
Surgery			
Veríssimo et al. ^a , <i>n</i> = 29	7 [3; 15]	100	0
Sjolin et al., <i>n</i> = 54 [9]	8 ± 0.9	96.3	–
Sundaresh et al., <i>n</i> = 19 [11]	1.5 ± 2.2	100	–
Sundaresh et al., <i>n</i> = 419 [12]	59	–	10

^aResults from our study; *n* (%) are absolute and relative frequencies); Plus-minus signs are means ± standard deviations; Square brackets are medians [quartiles].

In GD, ATD was the most used first-line treatment (90.2%), despite being associated with lower remission and higher relapse rates, compared to I-131 (53.6 vs. 82.5%, $p < 0.0001$ and 44 vs. 7.1%, $p < 0.001$, respectively). The preference for ATD treatment was due to the low percentage of associated adverse effects (12%), especially the small percentage of hypothyroidism (4%), a frequently observed consequence of both I-131 (71.8%) and surgery (75.9%). When first-line treatment failed or disease relapsed, I-131 was the preferred definitive treatment, due to the lower hypothyroidism rate compared to surgery (71.8 vs. 100%, $p = 0.001$), and no surgical complications.

The observed treatment outcomes generally overlap with the experience of other series (Table 6).

ATD treatment duration of 12–24 months was associated with the highest remission and lowest relapse. This supports what is recommended in the 2016 American Thyroid Association and in the 2018 European Thyroid Association guidelines [1, 4]. Interestingly, a recent prospective study by Azizi et al. showed that a group of patients treated for a median duration of 96 months experienced a significantly lower relapse rate compared to those receiving conventional treatment durations (15 vs. 53%, $p < 0.001$), despite there being no difference in remission rates between the two groups. [7] Similarly, Lertwattanak et al., demonstrated in a prospective study,

that a low dose treatment for 36 months beyond the conventional treatment duration were associated with lower relapse rates (11.0 vs. 41.2%, $p < 0.01$) [8]. In our study, patients treated for more than 24 months were mainly those more difficult to manage, or those that either refused or were ineligible for I-131 or surgery, posing a potential selection bias which may explain the lower remission and higher relapse rates. Additionally, the higher remission rate associated with a 12–24-month treatment duration was only observed when using our specific remission criteria. When applying the ATA 1-year post-discontinuation of ATD therapy remission criteria, there was no significant difference in remission rates between treatment durations of 12–24 months and those exceeding 24 months.

Regarding post-ATD treatment follow-up, our results showed that over 75% of relapses occurred within the first three years, underscoring the importance of vigilant monitoring during this timeframe. Conversely, a mere 1.7% relapsed beyond nine years of follow-up, implying that this milestone could serve as a potential endpoint for patient surveillance. The only predictors of relapse identified in the present study were WHO goiter grade, and TRAbs at the end of treatment, both previously described in other series [1, 4, 6].

As for adverse effects, ATD treatment was associated with a low percentage (12%), especially an insignificant

percentage of hypothyroidism (4%), a frequently observed consequence of both I-131 (71.8%) and surgery (75.9%).

Patients submitted with I-131, using a median activity of 8 [6, 9–12] mCi achieved 82.5% remission with 7.1% relapse and resulting in 74% hypothyroidism. Other series, such as Zarif et al., and Sundaresh et al., achieved a higher remission rate (95 and 92% respectively), using a higher activity (16.5 mCi), resulting in a higher percentage of hypothyroidism (90%) [10, 11]. Comparing I-131 to thyroidectomy, although the former was associated with a lower remission rate and a higher relapse rate, more patients remained euthyroid at the end of follow-up (28.2 vs 0%, $p < 0.001$), while avoiding surgical complications.

TND patients were preferably treated with definitive first-line options (91.6%), particularly surgery, mainly due to larger thyroid volumes (24.2 [15.4–42.9] ml) and the high prevalence of multinodular goiter (79.5%). I-131 was used in toxic adenomas (61.1% of all I-131 treatments) and small volume toxic multinodular goiter (19.3 [13.1–29.3] ml). Remission rate was 91.7% (77.8% in toxic multinodular goiter and 95.5% in toxic adenoma), similar to the described in the literature (81.1 in toxic multinodular goiter and 93.7% in toxic adenoma).¹ As what was discussed for GD, despite a lower remission, I-131 was associated with lower hypothyroidism and absence of surgical complications, being the preferred option compared to surgery, in line with the recommendations of the American Thyroid Association [1].

ATD treatment was reserved for 14 patients that were ineligible or refused a definitive treatment, as theoretically ATD do not induce remission in patients with nodular disease. Although the 42.9% remission rate and 33.3% relapse rate are noteworthy, the limited sample size does not permit to draw definitive conclusions.

The strengths of this study are the large number of patients (411 patients, 245 GD and 166 TND), submitted to 564 treatment cycles, followed over a large period (median of 7 years), one of the largest European single center analyses. Despite these strengths, it is a retrospective study with a long-time span, where different treatments were available, and different guidelines were followed.

In conclusion, the present study supports ATD as the preferential first-line treatment for GD since they allow for normal thyroid function at the end of treatment in most patients, with few associated adverse effects, despite achieving remission in only half the population.

ATD treatment duration between 12–24 months was associated with the higher remission and lower relapse than longer periods, supporting current guidelines. However, recent prospective studies have provided evidence that long-term treatments are associated with lower relapse rates, without any significant difference in remission rates. This

disparity may be due to a selection bias in this study's population, and more prospective studies are necessary to comprehensively address this question.

While further studies are required to thoroughly assess optimal follow-up cut-off points, we propose a more intensive follow-up during the initial three years post-ATD treatment. We suggest reassessment every three months during the first year (based on a relapse rate of 50.9%), followed by semiannual evaluations for the subsequent two years (guided by the 76.3% relapse rate within the first three years). Afterward, annual check-ups are advised until the ninth year (cumulative 98.3% relapse rate), at which point ongoing monitoring becomes unnecessary.

In treatment failure or disease relapse, I-131 should be the preferred definitive treatment, due to a lower hypothyroidism rate compared to surgery, with no surgical complications.

In TND, definitive treatment should be the preferred option, namely surgery in high median thyroid volume and multinodular goiter, and I-131 in toxic adenomas and smaller goiters, in line with current guidelines.

Significance statement

Hyperthyroidism presents significant management and treatment challenges, and neither American nor European guidelines have been revised in the past five years. We conducted a four-decade retrospective cohort analysis to evaluate the effectiveness and safety of antithyroid drugs (ATD), radioablation with I-131, and surgery in patients with Graves' Disease (GD) and toxic nodular disease (TND). We found that 12–24 months of ATD treatment is associated with higher remission and lower relapse rates. In cases of treatment failure or relapse, I-131 is preferred for its lower hypothyroidism rates and lack of surgical complications. For TND, surgery is recommended for large goiters and I-131 for toxic adenomas. Our study, encompassing 411 patients and 564 treatment cycles, offers valuable insights into the real-world effectiveness and safety of these treatments.

Author contributions D.V.: conceptualization, methodology, writing – original draft, review, and editing. B.R.P.: conceptualization, methodology, review, and editing. J.V.R.: conceptualization, methodology, review, and editing. C.I.: conceptualization, methodology, and review. A.C.M.: conceptualization, methodology, and review. D.P.: conceptualization, methodology, and review. L.L.: conceptualization, methodology, and review. J.J.C.: conceptualization, methodology, and review. M.M.: conceptualization, methodology, and review.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethics approval This study was approved by our center's ethics committee and complies with the Declaration of Helsinki.

Informed consent Confidentiality was kept throughout the study.

References

1. D.S. Ross, H.B. Burch, D.S. Cooper, M.C. Greenlee, P. Laurberg, A.L. Maia et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* **26**(10), 1343–1421 (2016). <https://doi.org/10.1089/thy.2016.0229>
2. W.M. Wiersinga, K.G. Poppe, G. Effraimidis, Hyperthyroidism: aetiology, pathogenesis, diagnosis, management, complications, and prognosis. *Lancet Diab. Endocrinol.* **11**(4), 282–298 (2023). [https://doi.org/10.1016/S2213-8587\(23\)00005-0](https://doi.org/10.1016/S2213-8587(23)00005-0)
3. D.M. Smith, S. Dutta, F. Ahmed, M.A. Thaha, Change in practice over four decades in the management of Graves' disease in Scotland. *J. Thyroid Res.* 9697849 (2016). <https://doi.org/10.1155/2016/9697849>
4. G.J. Kahaly, L. Bartalena, L. Hegedüs, L. Leenhardt, K. Poppe, S.H. Pearce, European Thyroid Association guideline for the management of graves' hyperthyroidism. *Eur. Thyroid J.* **7**(4), 167–186 (2018). <https://doi.org/10.1159/000490384>
5. E. Fanning, W.J. Inder, E. Mackenzie, Radioiodine treatment for graves' disease: a 10-year Australian cohort study. *BMC Endocr. Disord.* **18**, 94 (2018). <https://doi.org/10.1186/s12902-018-0322-7>
6. T. Struja, M. Kaeslin, F. Boesiger, R. Jutzi, N. Imahorn, A. Kutz et al. External validation of the GREAT score to predict relapse risk in Graves' disease: results from a multicenter, retrospective study with 741 patients. *Eur. J. Endocrinol.* **176**(4), 413–419 (2017). <https://doi.org/10.1530/EJE-16-0986>
7. F. Azizi, A. Amouzegar, M. Tohidi, M. Hedayati, D. Khalili, L. Cheraghi et al. Increased remission rates after long-term methimazole therapy in patients with Graves' disease: results of a randomized clinical trial. *Thyroid* **29**(9), 1192–200 (2019). <https://doi.org/10.1089/thy.2019.0180>
8. R. Lertwattanak, T. Kunavisarut, S. Sriussadaporn, benefits of long-term continuation of low-dose methimazole therapy in the prevention of recurrent hyperthyroidism in graves' hyperthyroid patients: a randomized prospective controlled study. *Int. J. Endocrinol.* **2022**, 1705740 (2022). <https://doi.org/10.1155/2022/1705740>
9. G. Sjölin, M. Holmberg, O. Törring, K. Byström, S. Khamisi, D. de Laval et al. The long-term outcome of treatment for Graves' Hyperthyroidism. *Thyroid* **29**(11), 1545–1557 (2019). <https://doi.org/10.1089/thy.2019.0085>
10. H.A. Zarif, S.S. Alam, A. Baashar, A. Alsharif, M. Alhilabi, Comparison between the treatment modalities for Graves' disease at King Abdulaziz Medical City, Jeddah. *Cureus* **12**(1), e6730 (2020). <https://doi.org/10.7759/cureus.6730>
11. V. Sundaresh, J.P. Brito, P. Thapa, R.S. Bahn, M.N. Stan, Comparative effectiveness of treatment choices for Graves' hyperthyroidism: a historical cohort study. *Thyroid* **27**(4), 497–505 (2017). <https://doi.org/10.1089/thy.2016.0343>
12. V. Sundaresh, J.P. Brito, Z. Wang, L.J. Prokop, M.N. Stan, M.H. Murad et al. Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network meta-analysis. *J. Clin. Endocrinol. Metab.* **98**(9), 3671–3677 (2013). <https://doi.org/10.1210/jc.2013-1954>
13. P. Anagnostis, F. Adamidou, S.A. Polyzos, S. Katargari, E. Karathanasi, C. Zouli et al. Predictors of long-term remission in patients with Graves' disease: a single center experience. *Endocrine* **44**(2), 448–453 (2013). <https://doi.org/10.1007/s12020-013-9895-0>
14. E. Mohlin, H.F. Nyström, M. Eliasson, Long-term prognosis after medical treatment of Graves' disease in a northern Swedish population 2000–2010. *Eur. J. Endocrinol.* **170**(3), 419–427 (2014). <https://doi.org/10.1530/EJE-13-0811>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.