Sinha *et al. Thyroid Research* (2025) 18:27
https://doi.org/10.1186/s13044-025-00238-7

Thyroid Research

REVIEW Open Access

Comparing antithyroid drugs vs. radioactive iodine in paediatric Graves' disease: literature review

Akshat Sinha[1,2*], Reuben Oza[1,2], Brandon Karamveer Sangha[1,2], Arshia Akhavan-Mofrad[1,2] and Arvin Suddhi[1,2]

Abstract

Introduction Paediatric Graves' disease (PGD) is an autoimmune condition, which if left untreated, can result in cardiac complications. National Institute for Health and Care Excellence (NICE) Guidance (NG145) advocates the use of antithyroid drugs (ATD) as first-line therapy for PGD, with a consultation to consider a move to definitive therapy in the form of radioactive iodine (RAI) or thyroidectomy if the initial 2-year course failed to achieve normal thyroid function. We aim to evaluate the effectiveness, adverse events, and potential predictors of remission for ATD and RAI in treating PGD.

Methods A thorough guideline search of NICE Evidence and Royal College of Physicians (RCP) guidelines and policy was conducted to yield a guideline relevant to our review question. A literature search of the Cochrane Library, MEDLINE, EMBASE and PubMed, alongside a clear inclusion and exclusion criteria was utilised to generate systematic reviews and primary literature exploring the efficacy and adverse effects (AEs) of ATD and RAI. Our guideline, systematic reviews and primary literature were appraised using AGREE-II, AMSTAR 2 and CASP respectively.

Results The search strategy yielded one NICE guideline (NG145) published in November 2019, two systematic reviews published after November 2019 and four primary studies, published after the most recent systematic review (August 2020). All studies concluded that ATD and RAI are effective treatment options for PGD. With regards to AEs, RAI and ATD were safe treatment options, with the latter having the least severity of complications.

Conclusions In patients who have been identified to have predictors of remission, we agree with NG145 and ATD should be offered as first-line treatment. However, for those who do not have characteristics aligning with the predictors of remission, RAI should be offered as first-line therapy. Future studies should investigate the effect of biochemical parameters to identify predictors of remission, to aid the choice of treatment in paediatric Graves' disease treatment.

Keywords Paediatric Graves' disease, Graves' disease, Antithyroid drugs, Thionamides, Radioactive iodine, Ablation, Thyroid

*Correspondence:
Akshat Sinha
axs1745@student.bham.ac.uk
[1]College of Medicine and Health, University of Birmingham, Birmingham, UK
[2]Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Trust, Birmingham, UK

Introduction

Paediatric Graves' disease (PGD) is an autoimmune condition responsible for 10–15% of thyroid disorders in children and adolescents [1]. Graves' disease (GD) is caused by thyroid stimulating immunoglobulin (TSI) acting upon the thyroid-stimulating hormone receptor (TSHR) to stimulate the production and release of thyroid hormone, resulting in hyperthyroidism, and ultimately, if left untreated, thyrotoxicosis [2]. Consequently, patients may present with systemic symptoms and signs, including weight loss and fatigue, as well as muscle weakness, anxiety and ophthalmopathy [3]. In addition to these systemic symptoms, arguably the most dangerous sequelae of untreated thyrotoxicosis are the cardiac implications. Manifestations include atrial fibrillation and heart failure, highlighting the importance of early diagnosis and management [2, 3].

The wide-ranging effects of GD highlight the necessity for an effective treatment plan to successfully manage this disease as these consequences pose a burden for the patient, as well as National Health Service (NHS) resources. Subsequently, NICE clinical guidance [NG145] 1.6.18 details recommendations for first-line treatment of PGD with antithyroid drugs (ATDs) for a course of two years. If successful, ATD therapy is either stopped or decreased and subsequently the need for further treatment is reviewed. In the event of relapse [NG145] 1.6.19 states that surgery or radioactive iodine (RAI) should be discussed as alternative treatments [4].

ATDs work by either inhibiting the synthesis or action of thyroid hormones, with only thionamides prescribed in the UK [5]. Thionamides inhibit the Thyroid peroxidase (TPO) enzyme which plays a vital role in the synthesis of thyroid hormones [5]. Conversely, RAI emits beta-radiation causing DNA damage to thyrocytes leading to cell death [6]. Total thyroidectomy involves excision of the whole thyroid gland, thereby nullifying the synthesis and release of thyroid hormones [7].

Despite offering ATDs as the definitive first-line therapy being standard practice in England, there is evidence to support the claim that the chance of achieving remission after a two year course of ATDs is less than 30%, with further evidence stating certain features of the patient, i.e., serum level of TSI, predispose them to a better or worse chance of achieving remission with ATDs [1, 8]. This demonstrates that a broad advocation of ATDs as the first-line therapy for PGD may not be the most suitable option for all patients, and a more targeted approach may be required. Radioiodine and surgery both hold advantages over ATDs, such as the fact that radioiodine offers a definitive treatment of hyperthyroidism, while surgery can treat hyperthyroidism more immediately than either of the other two options [9]. However, they both have their disadvantages, such as the occurrence of hypothyroidism post-treatment, or the long-term effects of radioiodine which are lacking a clear evidence base [9]. Although evidence showing that surgery is a definitive treatment for PGD we are not considering this in our evidence review as an option for first line treatment because it is rarely used in paediatrics due to its associated complications [10, 11].

The primary aim of this review is to establish whether ATD or RAI is more effective in treating PGD, including the severity and frequency of adverse events (AEs). Through evaluation and critical appraisal of current evidence regarding the relative efficacy and AEs of ATDs and radioiodine in paediatrics, we aim to compare the optimal first-line treatment options for PGD.

Table 1 Population, intervention, comparator and outcome (PICO) framework used to generate review question

	Description
Population	Children with Graves' disease undergoing first line treatment
Intervention	Radioactive iodine
Comparator	Antithyroid drugs
Outcome	Remission (number whose thyroid returns to normal function)

Review question

An initial scope of evidence in this field of research was conducted via the NICE evidence database to determine whether there was adequate literature in the field to review. The success in this initial scoping allowed pursuit of this area of research, in which use of the PICO (Population, Intervention, Comparator and Outcome) framework shown in Table 1 created a sound basis for the development of our review question [12]. The current recommended first line treatment according to NICE guideline NG145 advises the use of antithyroid drugs, hence this treatment has been used as the comparator in the PICO framework [4].

Methods: literature searches

Literature searches for systematic reviews and primary research were conducted through the following databases: Cochrane library, MEDLINE, EMBASE and PubMed. The search criteria for these searches are shown in supplementary Tables 1 and 2, with inclusion and exclusion criteria also displayed in supplementary Tables 3 and 4. Consequently, systematic reviews and primary research was selected for analysis.

Guideline

The search of the NICE database and RCP yielded thirty-two documents of which NG145 - Thyroid Disease: Assessment and Management was selected [4]. It is a detailed generalised guideline with a vast amount of

information regarding the diagnosis and treatment of thyroid disorders and disease.

Guideline appraisal

Four independent assessors used the AGREE-II tool to critically appraise the NICE guideline as this is the optimal number of assessors [13]. AGREE-II was used because it is routinely employed for appraising guidelines [14]. The scores attributed to each of the six domains of AGREE-II are detailed in the appendices (Supplementary Material 1).

Scope and purpose

The scope is clearly defined [15]. There is a clear population to which the guidance can be applied, as well as having a setting of NHS-funded healthcare providers. Outcomes were concisely expressed and well-written. There is also discussion of available treatments and associated comparators. A range of conditions are explored in different age groups of the population, while also excluding certain populations, e.g. neonates. Additionally, there is advice as to who the guideline is intended to be used for.

Stakeholder involvement

There were varying levels of stakeholder involvement from the multidisciplinary team to wider engagement [16]. Each of them had their name, discipline, location, and institution denoted [17]. The process is robust where the team produce the guidance, which is then opened for stakeholder comment.

Rigour of development

This guideline was rigorously developed. A comprehensive search strategy was employed, and the relevant databases were searched. It was decided to use specific MeSH (Medical Subheading) terms and timeframes. Only papers written in English were reviewed which were cross-checked via reference reading of highly relevant papers. Predefined inclusion and exclusion criteria were stipulated, and studies assessed methodologically using the appropriate tool: GRADE (+GRADE CERQual), QUIPS and CASP [18]. Recommendations were developed in an orderly fashion by the committee from the available evidence. It is said they considered the benefits, harms, and cost of each course of action. They state the guideline is subject to being updated if needed.

Clarity of presentation

Presentation was succinct and not ambiguous. Relevant interventions are explored and targeted towards the populations in question. They make use of caveats to describe contraindications in certain situations that may differ from normal management. Separately, recommendations are summarised in a section and grouped together with key further questions and reviews.

Applicability

There was a brief mention of barriers and concerns raised by stakeholders; however, it was concluded that it was not significantly more difficult to access services depending on which group you were in. NICE itself provides a generic auditing tool and implementation support. Summaries and the NICE pathway showed other simplified versions of the guideline [19]. Health economists played a part in the development by conducting a cost-consequence analysis [20]. Systematic checking of the model calculations by two specialists occurred.

Editorial independence

While assessing editorial independence it was apparent that NICE funded the National Guideline Centre (NGC) to produce this guideline [18]. Moreover, the RCP hosted it. It is unlikely that their views will have influenced the guideline as they are separate independent bodies. Positively, the interests of all committee members were publicly declared. To finalise the four independent authors saw this guideline as a distinguished piece of work with only minor flaws and therefore would recommend it.

Systematic review search

The systematic review search was conducted independently by three separate authors (AA, BKSS and AS) using the electronic databases EMBASE, MEDLINE, PubMed, and the Cochrane Library (Supplementary Table 1).

'Paediatric Graves' Disease' was used as a MeSH term to yield a total of twenty-two systematic reviews, as shown in Supplementary Fig. 1. Filter included 'systematic review' and a publication date after November 2019, as the relevant guideline (NG145) was published at this date. The eligibility of the two selected systematic reviews was determined by collaborative analysis and comparison of abstracts between AA, BKSS and AS to assess whether the studies fulfilled the pre-defined inclusion and exclusion criteria (Supplementary Table 3). The final two systematic reviews chosen each look at one of the possible treatments used for PGD (ATD and RAI). All reviews included explored efficacy and adverse events.

GD is less common in children and adolescents than in adults, thus reviews exclusively containing randomised trials solely within the paediatric population are limited in number. This, along with the obvious ethical issues involved in paediatric treatment meant we included a combination of randomised and non-randomised studies, with the majority being observational. This enabled us to look at the duration of treatment, dosage differences and side effects in our outcome measure. The systematic

reviews were appraised using the AMSTAR-2 tool, with the review search authors (AA, BKSS, AS) collaborating with authors (RO, ASS) [21]. Each pairs' findings were presented to the remaining authors to discuss any disparities, drawing conclusions regarding any biases which may have affected the credibility of results.

Inclusion criteria & exclusion criteria

We chose an age range of 1–18 years to yield the greatest number of studies. The review excluded neonatal Graves' disease patients (age 0–1 year) due to differences in aetiology and treatment approach. Reviews that had varying methods of administering treatments were also included to increase the number of studies available for analysis.

The studies must have taken place in large tertiary centres in economically developed countries. To define a less developed country we used the list of nations that are part of the G20 summit [22]. An exception was made for studies where Graves' disease prevalence was higher or conducted in tertiary centres, resulting in publication in European or American journals. The above points also apply to the inclusion and exclusion criteria for the selection of primary studies.

Primary study search

A search was undertaken to identify pertinent primary literature published after August 2020, as data published after this was not included in the systematic reviews. Authors AA, BKSS and AS conducted searches via the databases MEDLINE, EMBASE, Cochrane Library and PubMed.

A combination of MeSH terms was used to procure a large number of studies. This yielded thirty-four studies after duplicates from the combinations of search terms were removed and the titles were screened. Further studies were then removed based on our inclusion and exclusion criteria, resulting in the selection of four studies (Supplementary Table 4). Supplementary Table 2 details the search strategy. In addition, we prioritised studies that evaluated both efficacy and adverse events.

As shown four retrospective cohorts were chosen (Supplementary Fig. 2). This study design allowed for dosage variations to be investigated, rather than just unexposed versus exposed participants. Each study was independently appraised using the CASP framework by the five authors (AA, BKSS, AS, ASS, RO) [23]. Views of the five authors regarding the primary literature were explored collectively, allowing for the assembly of the results into one fluent appraisal. Authors agreed regarding the inclusion of selected studies, and which provided the strongest evidence.

Review findings

Our searches yielded one guideline, two systematic reviews and four primary studies (four cohort) which met our strict predefined inclusion criteria shown in supplementary Figs. 1 & 2.

Systematic review results

A summary of the results of both systematic reviews are shown below in Table 2.

Systematic review appraisal

No systematic reviews that directly met our inclusion and exclusion criteria compared ATDs and RAI in terms of efficacy and AEs. Therefore, we have included two systematic reviews, with each solely assessing one treatment (Table 3). AMSTAR2 reported a critically low and low rating for the systematic reviews included, which is largely due to the absence of RCTs, highlighting the need for higher quality evidence in this field. Further reasoning for the scores of each review can be attributed to the fact that they both evaluated retrospective cohort studies, which are of lower quality of evidence compared to prospective studies. A positive of the reviews is that they both state they are the most extensive and up to date reviews in this field, thus making their findings more reliable. Van Lieshout et al. was rated as the highest quality of reviews as it accounted for the impact of heterogeneity, coupled with the fact that it included the largest number of participants [25].

Table 2 Results of systematic reviews

Study Reference	Number of Study Participants	Cohort Studies included	RCTs included	Overall Findings
Lutterman et al. [24] 2021	1,283	23	0	Treating patients with 11-15MBq iodine-131 per gram of thyroid tissue is an effective treatment option when aiming to achieve hypothyroidism. Efficacy seems to increase with dosage and activity of RAI. Short-term & long-term side effects are a rare occurrence in radioactive iodine treatment.
JM van Lieshout et al. [25] 2021	3,057	24	5	Intention to treat analysis (ITTA) showed an overall remission rate of 28.8% in methimazole treated patients. Going up to 75% as treatment duration rises to 9 years • Occurrence of adverse effects: 17.6% • Occurrence of major side effects: 1.1%•

Table 3 Appraisal of systematic reviews

Study	Advantages	Disadvantages	Percentage (%) of criteria met on AMSTAR 2
Lutterman et al. [24] 2021	• Most in depth systematic review of the effectiveness and occurrence of adverse effects in PGD to date • Study selection was performed independently and blinded to reduce reporting bias • Data extraction was performed in duplicate • Justification of exclusion criteria provided • Use of the Critical Appraisal Skills Program (CASP) checklist in critical appraisal of cohort studies	• No randomised control trials included • Unable to determine effect of confounders due to lack of bias assessment tool • No meta-analysis of results included and therefore unable to determine heterogeneity through statistical analysis • Unaware of the impact of heterogeneity on the results • End points for efficacy were not consistently defined, thus it is difficult to compare included studies	Critically Low
J M van Lieshout et al. [25] 2021	• Largest study design with highest number of participant population ($n = 3,057$) • Data was standardised via recalculating remission rates using ITTA to overcome heterogeneity • Only studies assessed to be at a low risk of bias (determined via use of the CASP checklist) were included	• 82.6% of included study participants were female meaning results may not be generalisable to a typical hospital setting, where proportions of males: females may differ) • Confounding factors, such as study participant characteristics were not included for all cohort studies affecting the generalisability of the systematic review results	Low

Both systematic reviews detailed a comprehensive set of inclusion/exclusion criteria. Study selection and data extraction were performed in duplicate, minimising the possibility of bias. Lutterman et al. included studies published in languages other than English, allowing for a more expansive literature search [24]. Whereas Van Lieshout et al., excluded studies not published in English [25].

Lutterman et al. and Van Lieshout et al. both excluded studies at a high/moderate risk of bias via use of the CASP checklist [24, 25]. Both reviews included summary tables detailing study characteristics. However, neither accounted for the impact of confounding factors, such as age and ethnicity. The presence of confounding factors could influence remission rates, affecting the generalisability of review findings, and therefore reducing validity of the results.

In assessing the efficacy of the treatment options, only Lutterman et al. incorporated studies which assessed efficacy based on different treatment outcomes, namely euthyroidism & hypothyroidism [24]. This inhibited merging of the results, so a pooled estimate of effect was not able to be determined. In contrast, the Van Lieshout et al. review standardised data by recalculating remission rates using ITTA to overcome heterogeneity and hence, provided a pooled estimate of effect [25].

Primary study results

A summary of the results of the four primary study results are shown below in Table 4.

Primary study appraisal

Baseline characteristics

Nawongprom et al., Lee et al. and Song et al. included similar baseline characteristics such as age, gender, and family history of thyroid disease, with each of the studies including further characteristics [26, 27, 28]. Mizokami et al., on the other hand, does not report baseline characteristics [29]. This introduces the possibility of confounding to a greater extent than the other studies as multiple factors are not accounted for in the statistical analysis.

Namwongprom et al., Lee et al. and Song et al. studies present continuous data as mean ± SD, and categorical data as a % [26, 27, 28]. At baseline, only Lee et al. separated subjects into different arms (this was done based on their severity of Graves' disease and the resultant ATD dose received) [27]. Therefore, *p*-values were given to inform on statistical differences in baseline characteristics between each group, however, this was not possible in the other primary studies. Namwongprom et al. stated no statistically significant differences in baseline characteristics between participants, contributing to the external validity of this study [26].

The absence of information on ethnicity is a key point that prevents true generalisability to the general population, as Ehrhart et al. suggested in a 2018 article [30]. Ethnicity is not reported in any study, and as they are all conducted in East Asian countries, it can be problematic to assume these results will hold true in the UK, further hindering the strength of its evidence towards its application in the NICE guideline.

Study design

None of the studies were of a particularly large scale, however relatively, Namwongprom et al. included the lowest number of subjects [32], with Song et al. using the largest number of subjects (195). To lower the sampling variability, a larger sample size would be required, which would contribute to more precise estimates of treatment effect.

All studies included follow-up tests and examinations, to assess the efficacy and safety of each treatment.

Table 4 Result of primary studies

Study Reference	Study Design	Study participants	Type of Therapy	Results for effectiveness	Results for adverse events
Namwongprom et al. [26] 2021	Retrospective Cohort	32	Radioiodine: 24 h I-131	Hypothyroidism achieved 3–6 months after treatment in 65.6% of participants after single dose I-131	NA
H Lee et al. [27] 2021	Retrospective cohort	99	Methimazole (MMI)	Free thyroxine levels returned to normal after a mean time of: • 5.64 weeks for an initial dose of < 0.4 mg/kg/day (Group A) • 8.61 weeks for an initial dose of < 0.4-0.7 mg/kg/day (Group B) • 7.98 weeks for an initial for an initial dose of > 0.7 mg/kg/day (Group C)	Liver dysfunction (ALT/AST > 60IU/L) *Group A vs. B vs. C*: ***P = 0.034*** No events of ***serious liver failure*** Neutropenia (neutrophil count < 1000/mm3 *Group A vs. B vs. C*: ***P = 0.015*** No ***agranulocytosis***
Song et al. [28] 2021	Retrospective cohort	195	Methimazole (MMI) or Propylthiouracil (PTU)	More than six months of euthyroid status after terminating ATD treatment was defined as achieving remission. Cumulative remission rates: • Within 1 year of starting ATD = 3.3% • Within 3 years of starting ATD = 19.6% • Within 5 years of starting ATD = 34.1% • Within 7 years of starting ATD = 43.5% • Within 10 years of starting ATD = 50.6%	Total adverse events = 13.3% Most common: • Rash = 5.6% • Abnormal CBC = 2.6% (neutropenia), 0.5% (agranulocytosis) • Abnormal LFTs = 2.1% (↑ liver enzymes, no fulminant failure)
Mizokami et al. [29] 2020	Retrospective cohort	111	Radioiodine: I-131	Outcomes of thyroid levels: • Overt hypothyroidism = 91% • Subclinical hypothyroidism = 2% • Euthyroidism = 5% • Subclinical hyperthyroidism = 2%	Adverse events reported included: • Thyroid cysts = 4.27% • Iso-or hypo-echoic solid nodule(s) = 7.69% • 17.5% of patients followed up for 10 years or more developed newly detected solid thyroid nodules

Namwongprom et al. assessed thyroid levels in patients at 6 months whereas Song et al. exhibited a mean duration of follow-up of 5.9 ± 3.8 years, testing various biochemical markers such as thyroid function tests and complete blood count. The follow-up periods in Lee et al. and Mizokami et al. were 2 years and a median of 95 months respectively. Such an extended follow-up period increases the reliability and validity of the results. However, some cases in Mizokami et al. were only followed-up for a minimum period of 4 months, ranging to a maximum of 226 months. The major advantages of using longer follow-up periods are that they are more likely to detect any declines in the efficacy of a treatment, showing any relapses, and detect any adverse events which may not present until a later date.

Outcomes

Namwongprom et al., Lee et al. and Song et al. measured the time taken for free T4 levels to normalise, with Song et al. also looking for the presence of goitre and ophthalmopathy. Mizokami et al. used ultrasound imaging to determine thyroid status using the volumetric ellipsoid method.

A major limitation seen in many retrospective cohort studies is a high loss to follow-up. The loss to follow up can significantly reduce the validity of the results if the loss to follow up ratio is > 30%. High loss-to-follow-up ratios (> 30%) are known to compromise study validity by introducing attrition bias and reducing the reliability of findings [45]. Mizokami et al. and Song et al. reported a lost to follow up of less than 30%, suggesting the validity of their results will not be affected by lost to follow up. A drawback of Namwongprom et al. study is the omission of a lost to follow up ratio being reported, meaning this study could be prone to bias. Interestingly, Lee et al. decided to remove loss to follow up from their statistical analysis, using a so-called 'per-protocol' approach. This approach can be problematic as it leads to the possibility of attrition bias, reducing the external validity of the study (Table 5).

Discussion

Comparison of the effectiveness of ATD and RAI proved to be difficult as each review utilised different methods and assessed different outcome measures. Definitions of remission differed between included primary studies meaning that the overall risk ratios could not be calculated, but pooled estimates of effect were provided. However, due to the different treatment lengths and dosages, the effectiveness of radioiodine could not be directly compared with the effectiveness of ATD. Despite this, we were able to assess the effectiveness of each treatment option on an individual basis. In order to enable the direct comparison of ATD and RAI, future trials should incorporate core outcome sets, to standardise the chosen measures of treatment efficacy [31]. This will help policy makers to make more informed decisions with regards to the first-line treatment for PGD.

As supported by the AMSTAR II scoring criteria [21], Van Lieshout et al. demonstrated the greatest strength of evidence in support of the effectiveness of the reviewed drug, in this case ATD. Van Lieshout et al. demonstrated a 75% effectiveness after a 9-year course of ATD. However, the NICE guideline previously recommended the use of ATD for 2 years with alternative definitive treatment considered at this point if remission has not yet been achieved [4]. After a 2-year course of ATD, Van Lieshout et al. reported a relapse of over 70%. Therefore, in practice, most patients would be required to change their treatment to RAI or surgery, and recent NICE evidence indicates a switch to a more definitive treatment at the last update in 2023 [4]. Consequently, it could be argued RAI would be better suited as a first line treatment in some. To establish whether this would be the case, one factor that is important to consider is the predictors of remission with respect to ATD. Multiple studies have shown how certain characteristics can predict whether remission will be achieved [32, 33, 34]. A correlation between initial thyrotropin receptor antibody (TRAb), thyroid peroxidase antibody (TPOAb), and total triiodothyronine (T3) at diagnosis and remission after 2 years of ATD therapy has been reported in both paediatric and adult populations [32, 33, 34]. While these adult studies provide valuable insights, their applicability to paediatric cases may be limited due to differences in disease presentation, treatment response, and long-term outcomes. We recommend measuring total T3, TPOAb, and TRAb levels in PGD patients at diagnosis to evaluate their potential as predictors of remission with ATD therapy. If these biochemical markers predict remission with the use of ATD we suggest their use as first line treatment. However, if the biomarkers do not predict remission from the use of ATDs, we advocate first line treatment with RAI. This will help to ensure that patients are offered the most effective treatment for them, thereby mitigating the consequences of unnecessary additional treatment, such as wasted resources and patient concerns over failed treatment.

Another differentiating factor for a definitive treatment option is the severity and frequency of AEs. AEs of medical management have been explored in both the systematic reviews and primary studies. Frequency of AEs in ATD therapy was reported to be low in Song et al., with a low severity of AEs reported (i.e. rash). This is further supported by Van Lieshout et al., with only a 1.1% occurrence of major adverse effects across studies. Lutterman et al. reported the short and long-term SEs of radioiodine treatment to be rare, with Mizokami et al.29 supporting

Table 5 Appraisal of primary studies

Study Reference	Advantages	Disadvantages
Namwong-prom et al. [26] 2021	• There are no significant differences in the baseline characteristics of patients recruited (excluding 24 h I-131 uptake and RI status), meaning the effects of confounding factors are minimised • Appropriate comparison of data using statistical analysis tests (Sample T Test and Kruskal-Wallis H test) • Clearly defined outcome	• Small sample size of 32 patients • All patients were recruited from a single institute and are therefore likely to be susceptible to selection bias. • Study design is retrospective and is thus susceptible. • No statistical significance or confidence intervals are reported for any outcomes. • All cases had previously taken ATD which may affect the outcome of radioactive iodine therapy
H Lee et al. [27] 2021	• The efficacy and adverse effects of a wide variety of ATD dosages were explored. • Relationship between dose of ATD and frequency of adverse events are reported along with statistical significance. • Clearly defined outcome • Continuous variable differences were compared with the post hoc Tukey tests	• Small sample size of 99 patients • Non-randomised allocation of patients, introducing the effect of confounding factors. • Cases were obtained from only one hospital which may not be representative of the entire population, introducing selection bias
Song et al. [28] 2021	• The largest sample size of 195 of all primary studies included. • Clearly defined exclusion criteria • A lengthy mean treatment duration of 4.7 +/- 3.4 years • Clearly defined definition of remission • Use of Cox regression model to adjust for confounding at analysis stage	• Over the study period, 28.3% of patients were lost to follow-up. • Predictors of remission could not be recognised due to inadequate sample size.
Mizokami et al. [29] 2020	• Median follow-up period of 95 months • Statistical analyses performed using the Mann- Whitney U test or Pearson's chi square test. • Accurate assessment of thyroid volumes with the use of ultrasonography	• Small sample size of 117 patients, with concurrent loss to follow-up of 25.6% of patients • Only a maximal limit of 13.5 mCi of I-131 can be administered to outpatients in Japan, which means activity of I-131 uptake in patients with large goitres is difficult to interpret. • Variations of TSH levels are likely to have affected thyroid volumes. • Limited knowledge of the effect of confounding variables due to an absence of a baseline characteristics table

a low occurrence of SE's. However, the AE's Mizokami et al. reported included thyroid cysts and thyroid nodules, proving to be more prevalent than those reported in ATD treatment [46]. This indicates the frequency of SE is similar in both ATD and RAI. It is important to emphasize that the adverse effects of surgery, such as recurrent laryngeal nerve injury and hypocalcaemia, can be more life-altering than those associated with ATD and RAI treatment [47].

As alluded to in the introduction, surgery has not been included in our evidence search for the following reasons. A recent systematic review in Western Europe reported on frequency of postoperative complications, with hypocalcaemia and right laryngeal nerve injury proving to be common complications [35]. It is likely that these complications will occur in an even greater frequency within the UK setting, compared to a Western European setting, because Western European countries, i.e. France & Germany, performed 45,000 and 60,000 total thyroidectomies respectively, compared to the UK, which performed 4,663 thyroidectomies for thyrotoxicosis [36, 37]. Therefore, the volume of thyroid surgeries performed in the UK is still relatively low. This, combined with the varying level of skill and experience amongst the surgeon performing the procedure, suggests that complications because of a total thyroidectomy are more likely to occur in a UK setting compared to in Western Europe [38]. Another consideration with regards to surgery is the patient preference, as it has been demonstrated there is a patient apprehension for surgical management with a preference for medical management [39]. Bearing in mind such limitations to surgery, we cannot recommend thyroidectomy as a first-line definitive treatment for PGD.

Concerns about long-term risks of RAI, such as potential effects on growth, fertility, and pubertal development, remain a significant consideration in paediatric patients. Safi et al. highlight that the potential radiation risks in growing children may differ from those in adults, necessitating careful evaluation [40]. Similarly, another study discussed the potential implications for fertility in adolescents, emphasizing the importance of age in determining suitability for RAI [41]. Ethical considerations regarding the use of a permanent treatment modality in a developing child must also be addressed. For instance, the risk of hypothyroidism post-RAI and the lifelong dependence on thyroid hormone replacement therapy may pose challenges for young patients.

The acceptability of RAI in children hinges on the perspectives of parents and caregivers. Parental apprehension also plays a pivotal role in the acceptability of RAI. Rivkees and Stephenson (2010) reported that families often prioritize safety and reversibility when choosing between ATDs and RAI [42]. In addition, the approach to definitive therapy further underscores the need for shared decision-making, particularly as families weigh the irreversible nature of RAI against the benefits of definitive therapy [43, 44]. However, there is limited literature directly exploring the views of paediatric patients themselves, highlighting the need for further research in this area.

Limitations

This review does not include direct head-to-head studies comparing ATD and RAI. Conclusions regarding relative effectiveness and safety are based on indirect comparisons across separate studies, which limits the strength of the findings. The variability in clinical guidelines across countries, particularly regarding ATD treatment duration, presents a limitation in generalizing these findings to the UK clinical setting, where NICE guidelines recommend ATD therapy for a maximum of two years. Existing studies on predictors of remission are small and outdated compared to recent systematic reviews. Further research is essential to validate predictors and optimize treatment recommendations.

The severity of adverse effects such as thyroid cysts and nodules versus ATD-associated effects like rash or neutropenia remains subjective without direct comparative studies. While both ATD and RAI are effective, the current evidence base, predominantly retrospective and limited in scale, necessitates cautious interpretation and further prospective studies.

Future research

The scope of literature included in this essay has been thoroughly reviewed, however certain questions need to be answered with further research. Novel studies should focus on conducting randomised controlled trials (RCTs) to ascertain the optimal dosage and type of ATD or RAI used. In addition, RCTs comparing the two treatments investigated in this review should also evaluate efficacy of thyroidectomy in PGD. Prospective cohort studies can predict levels of response to ATDs by using biochemical markers. Future research should also explore the perspectives of paediatric patients on each type of treatment through qualitative studies.

Conclusion

In conclusion, PGD is a disease that if left untreated, can lead to severe complications including arrhythmias and thyroid eye disease. This review was undertaken to explore the efficacy and associated AEs of ATDs against RAI in paediatric patients. The findings of this review demonstrate the effectiveness of all treatment options considered, with adverse effects being the most detrimental in cases of surgery, and of less medical concern in ATD and RAI. Based on the findings of this evidence

review, there is potential to stratify treatment options based on predictors of remission, where ATD will be offered as first line treatment to patients who satisfy these predictors, and RAI for those without predictors of remission. To ensure upmost reliability of predictors of remission, prospective cohort studies are required, to determine their validity and generalisability to the paediatric population.

As shown in the discussion section, further research questions which utilise a randomised methodology will strengthen the evidence-base regarding effectiveness and safety of treatment options for PGD. In particular, a systematic review which includes studies utilising a core outcome set to directly compare the effectiveness of all treatment options will allow a better comparison to be made regarding their relative efficacies, helping to better inform on the ideal first-line therapy for PGD.

Abbreviations

PGD	Paediatric Graves' disease
GD	Graves' Disease
TSI	Thyroid-stimulating Immunoglobulin
TSHR	Thyroid-stimulating hormone receptor
ATD	Antithyroid drug
AEs	Adverse Effects
TPO	Thyroid Peroxidase
RAI	Radioactive Iodine
MeSH	Medical Subject Headings
PICO	Population, Intervention, Comparator, Outcome
ITTA	Intention to Treat Analysis
TPOAb	Thyroid Peroxidase Antibody
TRAb	TSH Receptor Antibodies

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13044-025-00238-7.

Supplementary Material 1

Author contributions

A.S, B.K.S.S and A.A-M were prinicpally resposnible for carrying out searches and preparing supplementary Figs. 1 and 2, and supplementary Tables 1 and 2. R.O and A.S.S carried out appraisal of systematic reviews and primary studies.All authors equally participated in guideline appraisal and data extraction for outcome/ result analysis. Contributions were spread between the authors for the main manuscript text with A.S.S adopting a supervisory role.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 1 June 2024 / Accepted: 13 March 2025
Published online: 14 May 2025

References

1. Lee HS, Hwang JS. The treatment of Graves' disease in children and adolescents. Ann Pediatr Endocrinol Metab. 2014 Sep [cited 2022 Feb 2];19(3):126. Available from: /pmc/articles/PMC4208256/
2. Pokhrel B, Bhusal K. Graves Disease. StatPearls. 2021 Jul 21 [cited 2022 Feb 2]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK448195/
3. Pearce EN. Diagnosis and management of thyrotoxicosis. BMJ Br Med J. 2006 Jun 10 [cited 2022 Feb 2];332(7554):1373. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1476727/
4. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145
5. Abdi H, Amouzegar A, Azizi F, Antithyroid Drugs. Iran J, Pharm Res, IJPR. 2019 Sep 1 [cited 2022 Feb 2];18(Suppl1):12. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7393052/
6. Mumtaz M, Lin LS, Hui KC, Khir ASM. Radioiodine I-131 For The Therapy Of Graves' Disease. Malays J Med Sci. 2009 [cited 2022 Feb 2];16(1):33. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3336179/
7. Schüssler-Fiorenza CM, Bruns CM, Chen H. The surgical management of Graves' disease. J Surg Res. 2006 Jun 15 [cited 2022 Feb 2];133(2):207–14. Available from: https://pubmed.ncbi.nlm.nih.gov/16458922/#:~:text=Materials and methods%3A The three, radioactive iodine therapy%2 C and surgery.
8. Bauer AJ. Approach to the Pediatric Patient with Graves' Disease: When Is Definitive Therapy Warranted? J Clin Endocrinol Metab. 2011 Mar 1 [cited 2022 Feb 2];96(3):580–8 Available from: https://academic.oup.com/jcem/article/96/3/580/2596398
9. Cooper DS. Hyperthyroidism. Lancet. 2003 Aug 9 [cited 2022 Feb 2];362(9382):459–68. Available from: https://www.sciencedirect.com/science/article/pii/S0140673603140731
10. Lee JA, Grumbach MM, Clark OH. The Optimal Treatment for Pediatric Graves' Disease Is Surgery. J Clin Endocrinol Metab. 2007 Mar 1 [cited 2022 Feb 2];92(3):801–3. Available from: https://academic.oup.com/jcem/article/92/3/801/2596860
11. Rivkees SA. Pediatric Graves' Disease: Controversies in Management. Horm Res Paediatr. 2010 Nov [cited 2022 Feb 2];74(5):305–11. Available from: https://www.karger.com/Article/FullText/320028
12. Brown D. A Review of the PubMed PICO Tool: Using Evidence-Based Practice in Health Education. Health Promot Pract. 2020 Jul 21 [cited 2022 Feb 16];21(4):496–8. Available from: https://journals.sagepub.com/doi/https://doi.org/10.1177/1524839919893361
13. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. AGREE II: Advancing guideline development, reporting and evaluation in health care. CMAJ. 2010 Dec 14 [cited 2022 Feb 2];182(18). Available from: https://www.agreetrust.org/agree-ii/
14. Hoffmann-Eßer W, Siering U, Neugebauer EAM, Brockhaus AC, Lampert U, Eikermann M. Guideline appraisal with AGREE II: Systematic review of the current evidence on how users handle the 2 overall assessments. PLoS One. 2017 Mar 30 [cited 2022 Feb 2];12(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5373625/
15. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, History, Final Scope. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/documents/final-scope
16. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, History, Stakeholder List. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/documents/stakeholder-list-3
17. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, History, Committee Member List. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/documents/committee-member-list
18. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, Evidence, Methods. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/evidence/methods-pdf-6967432045
19. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, Tools and Resources, Visual

Summary. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/resources/visual-summary-pdf-6965793901
20. National Institute for Health and Care Excellence. Thyroid disease: assessment and management. NICE guidelines NG145, Evidence, Cost Consequence Analysis. London: National Institute for Health and Care Excellence. 2019 [cited 2022 Feb 2]. Available from: https://www.nice.org.uk/guidance/ng145/evidence/costconsequenceanalysis-pdf-6967432046
21. AMSTAR -. Assessing the Methodological Quality of Systematic Reviews. [cited 2022 Feb 4]. Available from: https://amstar.ca/Amstar_Checklist.php
22. Akashi H, Ishizuka A, Lee S, Irie M, Oketani H, Akashi R. The role of the G20 economies in global health. Glob Heal Med. 2019 Oct 31 [cited 2022 Feb 4];1(1):15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7731051/
23. CASP CHECKLISTS - CASP -. Critical Appraisal Skills Programme. [cited 2022 Feb 4]. Available from: https://casp-uk.net/casp-tools-checklists/
24. Lutterman SL, Zwaveling-Soonawala N, Verberne HJ, Verburg FA, Van Trotsenburg ASP, Mooij CF. The Efficacy and Short- and Long-Term Side Effects of Radioactive Iodine Treatment in Pediatric Graves' Disease: A Systematic Review. Eur Thyroid J. 2021 Jul [cited 2022 Feb 4];10(5):353–63. Available from: https://www.karger.com/Article/FullText/517174
25. Van Lieshout JM, Mooij CF, van Trotsenburg ASP, Zwaveling-Soonawala N. Methimazole-induced remission rates in pediatric Graves' disease: a systematic review. Eur J Endocrinol. 2021 Jul 7 [cited 2022 Feb 4];185(2):219–29. Available from: https://eje.bioscientifica.com/view/journals/eje/185/2/EJE-21-0077.xml
26. Namwongprom S, Dejkhamron P, Unachak K. Success rate of radioactive iodine treatment for children and adolescent with hyperthyroidism. J Endocrinol Invest. 2021 Jun 24 [cited 2022 Feb 4];44(3):541–5. Available from: https://link.springer.com/article/10.1007%2Fs40618-020-01339-w
27. Lee HG, Yang EM, Kim CJ. Efficacy and adverse events related to the initial dose of methimazole in children and adolescents with Graves' disease. Ann Pediatr Endocrinol Metab. 2021 Sep 30 [cited 2022 Feb 4];26(3):199–204. Available from: https://e-apem.org/journal/view.php?doi=10.6065/apem.2142046.023
28. Song A, Kim SJ, Kim MS, Kim J, Kim I, Bae GY et al. Long-Term Antithyroid Drug Treatment of Graves' Disease in Children and Adolescents: A 20-Year Single-Center Experience. Front Endocrinol (Lausanne). 2021 Jun 14 [cited 2022 Feb 4];12. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8236938/#B19
29. Mizokami T, Hamada K, Maruta T, Higashi K, Tajiri J. Long-Term Outcomes of Radioiodine Therapy for Juvenile Graves Disease with Emphasis on Subsequently Detected Thyroid Nodules: A Single Institution Experience from Japan. Endocr Pract. 2020 Jul 1 [cited 2022 Feb 4];26(7):729–37. Available from: https://www.endocrinepractice.org/article/S1530-891X(20)36011-0/fulltext
30. Ehrhart MD, Guthrie IR, Bucheli JR, Burge MR. What is the Role of Race and Ethnicity in the Development Of Thionamide-Induced Neutropenia? J Endocrinol Thyroid Res. 2018 Feb 28 [cited 2022 Feb 4];3(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6394871/
31. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev. 2016 Jan 20 [cited 2022 Feb 16];5(1):1–4. Available from: https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-016- 0188-6.
32. Glaser NS, Styne DM, Group for the O of PE of NCCGDS. Predicting the Likelihood of Remission in Children With Graves' Disease: A Prospective, Multicenter Study. Pediatrics. 2008 Mar [cited 2022 Feb 4];121(3):e481–8. Available from: https://publications.aap.org/pediatrics/article-abstract/121/3/e481/72820/Predictingthe- Likelihood-of-Remission-in-Children?redirectedFrom=fulltext.
33. Karmisholt J, Andersen SL, Bulow-Pedersen I, Carlé A, Krejbjerg A, Nygaard B. Predictors of Initial and Sustained Remission in Patients Treated with Antithyroid Drugs for Graves' Hyperthyroidism: The RISG Study. J Thyroid Res. 2019 Jan 3 [cited 2022 Feb 16];2019. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6335719/
34. Young ET, Steel NR, Taylor JJ, Stephenson AM, Stratton A, Holcombe M et al. Prediction of Remission after Antithyroid Drug Treatment in Graves' Disease. QJM An Int J Med. 1988 Feb 1 [cited 2022 Feb 16];66(2):175–89. Available from: https://academic.oup.com/qjmed/articleabstract/66/2/175/1593890?redirectedFrom=fulltext
35. Zaat AS, Derikx JPM, Zwaveling-Soonawala N, Van Trotsenburg ASP, Mooij CF. Thyroidectomy in Pediatric Patients with Graves' Disease: A Systematic Review of Postoperative Morbidity. Eur Thyroid J. 2021 Mar [cited 2022 Feb 4];10(1):39–51. Available from: https://etj.bioscientifica.com/view/journals/etj/10/1/ETJ511345.xml
36. Fortuny JV, Guigard S, Karenovics W, Triponez F. Surgery of the thyroid: recent developments and perspective. Swiss Med Wkly 2015 31. 2015 Jul 26 [cited 2022 Feb 4];145(31). Available from: https://smw.ch/article/doi/smw.2015.14144
37. Chadwick D, Bch BM, Kinsman R, Walton P, Frcp M. The British Association of Endocrine & Thyroid Surgeons Prepared by on behalf of The British Association of Endocrine & Thyroid Surgeons 2017 [cited 2022 Feb 4]; Available from: https://www.baets.org.uk/wp-content/uploads/BAETS-Audit-National-Report-2017.pdf
38. Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. Ann Surg. 1998 Sep [cited 2022 Feb 4];228(3):330. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1191485/
39. Dent J, Brun J, Fendrick AM, Fennerty MB, Janssens J, Kahrilas PJ et al. An evidence-based appraisal of reflux disease management — the Genval Workshop Report. Gut. 1998 Apr [cited 2022 Feb 4];44(suppl 2): S1–16. Available from: https://gut.bmj.com/content/44/suppl_2/S1
40. Persons EM, Hussein MH, Herrera M, Pinion D, Webster A, Pineda E, et al. The impact of radioactive iodine on outcomes among pediatric and adolescent thyroid cancer patients: A SEER database analysis. Cancers. 2025;17(1):107. https://doi.org/10.3390/cancers17010107.
41. Nies M, Cantineau AEP, Arts EGJM, van den Berg MH, van Leeuwen FE, Muller Kobold AC, et al. Long-Term effects of radioiodine treatment on female fertility in survivors of childhood differentiated thyroid carcinoma. Thyroid. 2020;30(8):1169–76. Epub 2020 Apr 9. PMID: 32079487.
42. Rivkees SA, Stephenson K. Radioiodine in the management of pediatric graves' disease: Evidence-based experience and perspectives. Horm Res Paediatr. 2010;73(2):81–9.
43. Bauer AJ. Approach to the pediatric patient with graves' disease: when is definitive therapy warranted? J Clin Endocrinol Metab. 2011;96(3):580–8.
44. Lee HS, Hwang JS. The treatment of graves' disease in children and adolescents. Ann Pediatr Endocrinol Metab. 2014;19(3):126–30.
45. Kristman VL, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751–60.
46. Ma C, Xie J, Wang H, Li J, Chen S. Radioiodine therapy versus antithyroid medications for graves' disease. Cochrane Database Syst Rev. 2016;2(2):CD010094. https://doi.org/10.1002/14651858.CD010094.pub2. PMID: 26891370; PMCID: PMC10517434.
47. Zaat AS, Derikx JPM, Zwaveling-Soonawala N, van Trotsenburg ASP, Mooij CF. Thyroidectomy in pediatric patients with graves' disease: A systematic review of postoperative morbidity. Eur Thyroid J. 2021;10(1):39–51. https://doi.org/10.1159/000511345. Epub 2020 Nov 17. PMID: 33777818; PMCID: PMC7983567.

Publisher's note