Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effect of Different Therapeutic Approaches on Quality of Life of Patients with Graves' disease: A Systematic Review and Meta-Analysis Running title: Treatment and Quality of Life in Graves' Disease

Shervin Mossavarali
Tehran University of Medical Sciences
Ali Vaezi
Tehran University of Medical Sciences
Dorsa Ghorban Sarvi
Tehran University of Medical Sciences
Niloofar Mirdamadi
Tehran University of Medical Sciences
Shahrzad Mohseni
Tehran University of Medical Sciences
Mahnaz Pejman Sani
mpsani@sina.tums.ac.ir

Tehran University of Medical Sciences

Article

Keywords: Graves' Disease, Quality of Life, Treatment; Surgery, Radioiodine, Antithyroid Drug

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9661309/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Graves' disease (GD) impairs quality of life (QoL), and the effects of different treatments including antithyroid drugs (ATD), radioiodine (RAI), and surgery on QoL remain unclear. This study aimed to assess and compare QoL outcomes following these therapies. We conducted a PRISMA-based systematic review and meta-analysis, searching PubMed, Embase, Scopus, and Web of Science to July 30, 2025, for studies assessing QoL after GD treatment using generic SF-36 or thyroid-specific ThyPRO questionnaires. Mean scores were pooled using random-effects models, and pairwise comparisons were made for each domain. 16 studies were included. All trials showed moderate risk of bias, while observational studies demonstrated moderate to serious risk of bias. Among SF-36 domains, physical functioning had the highest scores for ATD (66.24 [43.97–88.50]) and RAI (84.02 [78.53–89.52]), while social functioning was highest after surgery (72.32 [55.62–89.01]). Vitality was the lowest-scoring domain for both ATD (55.90 [44.63–67.16]) and RAI (59.35 [50.35–68.34]), whereas bodily pain was lowest for surgery (51.11 [22.70–79.51]). In ThyPRO, tiredness was the worst domain for all treatments (ATD 40.04 [31.31–48.78]; RAI 32.50 [21.05–43.95]; surgery 41.90 [39.49–44.31]). The best outcomes were observed for goiter symptoms after surgery (7.08 [0.42–13.74]) and RAI (12.03 [5.94–18.12]), and for sex life after ATD (11.59 [6.49–16.69]). Surgery showed significantly better results than RAI for depression and emotional susceptibility, while ATD was superior to RAI for sex life. All treatments improved QoL, each excelling in specific domains. These differences may assist clinicians and patients in understanding treatment-related QoL outcomes.

Introduction

Graves' disease (GD) is an autoimmune disorder characterized by hyperthyroidism, with a considerable global prevalence of approximately 1.0–1.5% [1,2]. The elevation of thyroid hormones causes many signs and symptoms, such as fatigue, anxiety, depression, muscle weakness, erectile dysfunction, menstrual irregularities, and resting tremor, all of which clearly affect physical capacity, daily functioning, work productivity, and quality of life (QoL)[3]. Although GD is generally manageable, its complications, such as orbitopathy, can further reduce QOL, and shorten life expectancy [4,5]. Due to these complications, patients with GD often experience an impairment in QOL even after treatment, without returning to the level of the normal population [6]. Therefore, it is important to determine how different therapeutic approaches affect various aspects of QoL and how these effects differ among patients with GD.

The three main therapeutic approaches for GD include antithyroid drugs (ATD), radioiodine (RAI), and surgery. In recent decades, ATD has been the first-line treatment in Asia and Europe, while RAI has been the preferred choice in the United States [3]. Therapeutic decision-making in GD is primarily guided by an evaluation of disease characteristics, treatment-related risks and benefits, and established clinical indications, which may favor a specific modality in certain settings. When more than one therapeutic option is considered clinically appropriate, clinical judgment and patient preferences play an important role in selecting the optimal approach. A recent study reported that ATD has become increasingly the

first choice for both patients and clinicians, with the remission rate being the most important factor influencing treatment decisions[7,8]. However, there remains limited consensus on which treatment modality provides the greatest improvement in QoL[9].

Given this uncertainty, the present study aims to systematically review the available literature on QoL outcomes following different therapeutic approaches for GD and to perform a meta-analysis to compare these outcomes across treatment modalities. By evaluating QoL within an evidence-based clinical framework, this study aims to offer additional evidence that can help inform clinicians and patients when QoL considerations are weighed alongside established clinical factors in the management of Graves' disease.

Methods

Search Strategy and Selection Criteria

This study was designed as a systematic review and meta-analysis in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (**Supplementary Table 1**) [10]. The study protocol was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO) under registration number CRD420251111708. Moreover, the ethics committee of Endocrinology and Metabolism Research Institute reviewed and accepted the study protocol (ethical code: IR.TUMS.EMRI.REC.1404.078).

We systematically searched four electronic databases—PubMed, Embase, Scopus, and Web of Science—from inception to July 30, 2025. Keywords and Medical Subject Headings (MeSH) relevant to GD (or diffuse toxic goiter), treatment modalities, and QoL were applied (**Supplementary Table 2**). Only studies published in English were included. The reference lists of eligible studies and relevant reviews were screened, and a manual search was conducted to identify additional records.

Observational and trial studies were eligible if they included participants with a confirmed GD diagnosis, reporting QoL after ATD, RAI, or surgery. We excluded abstracts, conference proceedings, commentaries, literature reviews, non–full-text publications, limited case series, and case reports. Studies that used the 36-Item Short Form Health Survey (SF-36) or the Thyroid-Specific Patient Reported Outcome (ThyPRO) questionnaires, including validated translations, were included. Studies focusing on specific subgroups of GD (e.g., Graves orbitopathy (GO) or thyrotoxic heart disease), those using study-specific (custom) questionnaires, or those using non-validated questionnaires were excluded.

SF-36 is a general questionnaire with eight subscales and two component scales, whereas ThyPRO is a disease-specific instrument comprising 84 items that measure QoL across 13 subscales, one composite scale, and one overall QoL item. Both are typically self-administered and assess physical, mental, and social domains, with scores ranging from 0 to 100, higher scores indicate better QoL in SF-36, whereas lower scores indicate better QoL in ThyPRO[11,12].

Data Extraction

All records were imported into EndNote for reference management and duplicate removal. Title and abstract screening were performed independently by two authors. Full texts of potentially relevant articles were then assessed for eligibility. Data extraction was independently conducted by the same authors using a predesigned form, collecting information on author, year, country, study design, patient accrual interval, sample size, age, treatment type, follow-up period, questionnaire type, relapse, and pre- and post-treatment scores. Any disagreements were resolved by discussion or, if necessary, by a third author.

Risk of Bias Assessment

Risk of bias was independently assessed by three authors. For randomized trials, the Cochrane Risk of Bias 2 (RoB 2) tool was used, evaluating potential bias across domains such as randomization, deviations from intended interventions, missing data, outcome measurement, and selective reporting [13]. For non-randomized studies, the ROBINS-I V2 tool was applied, assessing bias due to confounding, intervention classification, participant selection, deviations from intended interventions, missing data, outcome measurement, and reporting[14]. Visualization of the risk-of-bias assessments was performed using the robvis tool[15]. Any discrepancies were resolved by consensus or, if necessary, adjudicated by another authors.

Statistical Analysis

QoL scores were extracted from the eligible studies. Only studies that reported outcomes as mean and standard deviation (SD), or in a format that could be validly converted to mean and SD, were included. The data were transformed into mean and SD using established methods and subsequently standardized to mean and standard error (SE) for analysis, applying the formula $SE = SD / \sqrt{n}$[16]. For studies that reported sample size as a range, the median value of the range was used as the effective sample size to facilitate inclusion in the meta-analysis[6]. When studies reported multiple post-treatment time points, the longest follow-up value was used. We prioritized data points where euthyroidism was achieved; however, as thyroid status was not consistently reported across all time points in many studies, the final follow-up was utilized as a proxy to ensure that patients had the maximum opportunity to reach a stable euthyroid state [17]. If a study had multiple arms for the same intervention (e.g., subtotal and total thyroidectomy), they were combined into a single “overall surgery” group for analysis[18].

Subsequent meta-analyses for each questionnaire domain were conducted using a random-effects model, with results reported as mean values and 95% confidence intervals (CIs). Heterogeneity among studies was assessed using the Higgins I^2 and chi-square-based Q statistic. Pairwise comparisons between treatment groups were performed for each domain. Statistical significance was assessed using z-tests for mean differences. A p-value < 0.05 was considered statistically significant. According to the

Cochrane Handbook recommendations, publication bias was not assessed, and funnel plots were not generated because fewer than 10 studies were available for each domain[19]. All analyses were performed using STATA version 17 (StataCorp, College Station, TX, USA). Studies with both serious risk of bias and outlier data were removed for the sensitivity analysis, and the meta-analysis was subsequently rerun, with p-values recalculated accordingly.

Results

Study Characteristics

As Fig. 1 demonstrates, a total of 1,459 records were identified through database searches (PubMed, Scopus, Web of Science, and Embase). After removal of duplicates, 935 records remained for screening. Following title and abstract screening, 31 full-text articles were assessed for eligibility. The main reasons for exclusion in this stage were studies reporting patients with hyperthyroidism not specific to GD, those that did not stratify QoL by treatment, and those using study-specific (custom) or non-validated questionnaires (Supplementary Table 3). Moreover, two eligible studies were identified through citation searching. In total, 16 studies met the inclusion criteria and were included in the systematic review.

Among the 16 included studies, nine studies utilized the SF-36 questionnaire [18,20–27], five employed the ThyPRO [28–32], and two applied both instruments [6,17]. Also, three evaluated QoL following ATD therapy [20,21,28], two after RAI [17,30], and two after surgery [18,29]. Three studies compared QoL between RAI and ATD [22,23,31], two compared surgery and ATD [24,25], and four compared QoL across all three treatment modalities [6,26,27,32]. Multiple studies originated from the same country; notably, Sweden, Denmark, and Uzbekistan each contributed two studies. In total, there were three randomized trials [17,21,22] and 13 observational studies [6,18,20,23–32]. Patient age ranged from 12 to 80 years [18,28], with sample sizes spanning from 14 to 975 participants [6,21]. Follow-up duration varied substantially, ranging from 4 weeks to 25 years [21,32].

Risk of Bias in the Included Studies

The risk of bias assessment showed that among the non-randomized studies, the most frequent source of serious bias was participant selection, affecting nearly half of the included studies, with overall ratings ranging from moderate (6 studies) to severe (7 studies) (Supplementary Table 4 and Fig. 2A–B). For the randomized trials, all three had low risk of bias in outcome measurement and missing data, but moderate concerns regarding the randomization process and selection of the reported results, leading to an overall moderate risk of bias (Supplementary Table 5 and Fig. 2C–D).

Key Findings from the Included Studies

As shown in Table 1, all treatment modalities for GD were associated with improvements in QOL, but with some differences. In one study, ATD improved nearly all SF-36 subscales except bodily pain [20] and led to improvements in ThyPRO domains, with the exception of cosmetic complaints in another study [28]. RAI was associated with improvements in all SF-36 subscales after 12 months and in almost all ThyPRO domains, except for sex life [17]. Surgery improved all SF-36 subscales [29], while a one study found no significant difference in these scores between preserve-function and ablative surgery [18].

Between ATD and RAI treatments, regarding SF-36, some studies found similar outcomes between the two treatments [22,23], while one study reported poorer scores for RAI in nearly all domains [6]. In ThyPRO, one study showed that RAI led to better outcomes than ATD in goiter symptoms, emotional susceptibility, and impaired daily life [31], whereas two studies found that RAI was associated with worse scores in some domains such as, sex life, and appearance [6,32].

Comparing ATD and surgery, in SF-36, one study found no difference [27], and two similar Uzbek studies showed better scores for surgery in physical functioning, vitality, emotional role limitation (role-emotional), and mental health [24,25]. In ThyPRO, surgery patients reported worse outcomes than those treated with ATD, reporting more hyperthyroid, hypothyroid, eye, and appearance-related symptoms [6,32].

When evaluating surgery versus RAI, in SF-36, a study found no difference [27]. In ThyPRO, one study found that RAI scored worse than surgery for goiter symptoms, hyperthyroid symptoms, tiredness, anxiety, depression, emotional functioning (role-emotional), impaired social life, impaired sex life, and impaired daily life [6], while another ThyPRO study reported no difference [32]. More details on study characteristics and main findings of the studies are presented in Table 1.

Pooled QoL Scores from Meta-Analysis (SF-36 and ThyPRO Subscales)

As presented in Table 2, in SF-36 subscales, RAI and ATD patients scored highest in physical functioning and social functioning compared to other domains (RAI: physical functioning 84.02, 95% CI 78.53–89.52; social functioning 80.95, 95% CI 73.10–88.79; ATD: physical functioning 66.24, 95% CI 43.97–88.50; social functioning 65.78, 95% CI 48.81–82.76). The surgery group showed relatively higher scores in social functioning (72.32, 95% CI 55.62–89.01) and emotional role limitation (70.79, 95% CI 58.79–82.79) compared to other domains. ATD and surgery patients reported lowest scores in vitality (ATD 55.90, 95% CI 44.63–67.16; surgery 55.41, 95% CI 47.10–63.71) and bodily pain (ATD 56.76, 95% CI 37.25–76.27; surgery 51.11, 95% CI 22.70–79.51), whereas RAI patients' lowest domains were vitality (59.35, 95% CI 50.35–68.34) and general health perception (65.61, 95% CI 59.74–71.48) (Figure. 3A). None of the pairwise comparisons reached statistical significance. Moreover, ATD scores showed the highest heterogeneity, with I^2 ranging from 98.62% to 99.64%, whereas RAI had the lowest heterogeneity, with I^2 ranging from 0% to 92.25%. Surgery exhibited intermediate heterogeneity across subscales. Detailed I^2 statistics for each treatment group are reported in Supplementary Tables 6.

Table 2
Pooled Quality of Life Scores from Meta-Analysis (SF-36 and ThyPRO Subscales)

SF-36			
	Antithyroid Drug	**Surgery**	**Radioiodine**
Physical functioning	66.24 (43.97, 88.50)	67.22 (40.32, 94.12)	84.02 (78.53, 89.52)
Physical role limitation	60.66 (44.80, 76.52)	70.75 (48.91, 92.58)	76.30 (68.49, 84.11)
Bodily pain	56.76 (37.25, 76.27)	51.11 (22.70, 79.51)	68.57 (65.94, 71.20)
General health perception	63.30 (52.71, 73.88)	63.05 (53.88, 72.21)	65.61 (59.74, 71.48)
Vitality	55.90 (44.63, 67.16)	55.41 (47.1, 63.71)	59.35 (50.35, 68.34)
Social functioning	65.78 (48.81, 82.76)	72.32 (55.62, 89.01)	80.95 (73.10, 88.79)
Emotional role limitation	62.92 (48.19, 77.65)	70.79 (58.79, 82.79)	78.23 (69.61, 86.85)
Mental health	62.19 (53.63, 70.75)	66.55 (62.77, 70.32)	67.07 (60.09, 74.05)
ThyPRO			
Goiter symptoms	13.71 (2.08, 25.34)	7.08 (0.42, 13.74)	12.03 (5.94, 18.12)
Hyperthyroid symptoms	22.26 (11.76, 32.77)	18.96 (13.27, 24.64)	23.05 (18.43, 27.67)
Hypothyroid symptoms	25.65 (13.51, 37.8)	24.57 (16.96, 32.18)	26.24 (20.13, 32.34)
Eye symptoms	18.99 (7.93, 30.06)	17.92 (10.37, 25.48)	21.75 (14.28, 29.21)
Tiredness	40.04 (31.31, 48.78)	32.5 (21.05, 43.95)	41.9 (39.49, 44.31)
Cognitive impairment	22.33 (11.26, 33.4)	16.79 (10.05, 23.53)	23.52 (15.12, 31.92)
Anxiety	24.79 (12.88, 36.7)	20.45 (16.66, 24.23)	25.39 (16.62, 34.16)
Depression	25.21 (16.78, 33.64)	20.11 (17.36, 22.85)	24.65 (21.14, 28.16)
Emotional susceptibility	32.61 (16.31, 48.91)	20.56 (15.8, 25.31)	32.68 (23.15, 42.21)
Social life	14.44 (-0.34, 29.21)	8.16 (7.07, 9.25)	14.6 (5.65, 23.54)

SF-36			
	Antithyroid Drug	**Surgery**	**Radioiodine**
Daily life	16.05 (2.25, 29.85)	9.3 (5.94, 12.67)	15.29 (8.33, 22.25)
Sex life	11.59 (6.49, 16.69)	17.02 (9.75, 24.29)	21.85 (18.38, 25.31)
Cosmetic complaints	20.25 (-0.51, 41.01)	11.63 (5.79, 17.47)	24.24 (12.63, 35.85)
Overall QoL	25.69 (7.51, 43.86)	18.68 (16.77, 20.58)	25.22 (15.16, 35.28)

All data are presented as mean (95% confidence interval).

Mean follow-up for SF-36 groups: medical group = 6.8 years; surgical group = 9.06 years; radioiodine group = 7.06 years.

Mean follow-up for ThyPRO groups: medical group = 9.2 years; surgical group = 11.8 years; radioiodine group = 9.5 years.

Note 1: Physical Component Scale and Mental Component Scale scores for SF-36 are not presented in SF-36 section of this table, as they were derived from a different set of studies with fewer available reports; these results are provided separately in the Supplementary File.

Note 2: Composite scale of ThyPRO scores is not presented this table, as they were derived from a different set of studies with fewer available reports; these results are provided separately in the Supplementary File.

Note 3: Li et al., 2024 and Mangelen et al., 2017 did not report the scores for sex life domain.

Note 4: Täeb et al., 2016 did not report the scores for overall QoL domain.

All treatments showed their worst ThyPRO scores in tiredness compared with other domains (ATD 40.04, 95% CI 31.31–48.78; surgery 32.50, 95% CI 21.05–43.95; RAI 41.90, 95% CI 39.49–44.31). Emotional susceptibility was also among the worst domains for ATD and RAI (ATD 32.61, 95% CI 16.31–48.91; RAI 32.68, 95% CI 23.15–42.21), and hypothyroid symptoms were relatively poor across treatments (ATD 25.65, 95% CI 13.51–37.80; surgery 24.57, 95% CI 16.96–32.18; RAI 26.24, 95% CI 20.13–32.34) (Table 2 and Fig. 3B). ATD had its most favorable scores in sex life (11.59, 95% CI 6.49–16.69) and goiter symptoms (13.71, 95% CI 2.08–25.34), whereas surgery and RAI scored best in goiter symptoms (surgery 7.08, 95% CI 0.42–13.74; RAI 12.03, 95% CI 5.94–18.12) and social life (surgery 8.16, 95% CI 7.07–9.25; RAI 14.60, 95% CI 5.65–23.54) compared with other domains. In pairwise comparisons, surgery was significantly better than RAI for depression ($p = 0.046$) and emotional susceptibility ($p = 0.026$), and ATD was significantly better than RAI for sex life ($p = 0.0011$); all other pairwise differences were not significant. ATD scores ($I^2 = 83.87–99.58\%$) and RAI scores ($I^2 = 58.64–99.05\%$) exhibited higher heterogeneity across domains, whereas surgery scores demonstrated lower heterogeneity ($I^2 = 0–98.48\%$) (Supplementary Tables 7**).**

Sensitivity analysis

After removing the scores from Urunbayeva et al. (2025) and Ismailov et al. (2020), the sensitivity analysis of SF-36 subscales showed that the patterns across treatments remained consistent. Except for the component summary scales, physical functioning remained the highest-scoring domain, while vitality continued to be among the lowest for most treatments. All pairwise comparisons of non-component domains remained statistically non-significant, execpt bodily pain. Surgery had significantly higher score in bodily pain domain compared to RAI (Supplementary Tables 8).

After removing the scores from Mangelen et al. (2017), the sensitivity analysis of ThyPRO subscales showed that the overall pattern remained the same: goiter symptoms and social life remained the most favorable domains, while tiredness remained the worst across treatments. In pairwise comparisons, surgery scored significantly better than RAI for depression ($p = 0.039$), and RAI scored significantly worse than ATD for hyperthyroid symptoms ($p < 0.001$), hypothyroid symptoms ($p < 0.001$), tiredness($p = 0.033$), depression ($p = 0.001$), and sex life ($p = 0.001$); all other comparisons were not statistically significant (Supplementary Tables 8).

Disscusion

To the best of our knowledge, this is the first systematic review and meta-analysis to comprehensively evaluate and compare the QoL outcomes following different treatment modalities for GD. Across the 16 included studies, we observed that all three modalities improved QoL. Our review of the literature showed that there have been discrepancies in comparing the QOL scores between different therapeutic approaches. In our analysis of SF-36, no significant differences were observed between treatment groups, while RAI and ATD were associated with the highest scores in physical functioning, whereas surgery yielded the highest scores in social functioning. Consistently, social life also emerged as one of the most favorable domains for surgery in ThyPRO assessments. In the ThyPRO, pairwise comparisons demonstrated that surgery was significantly better than RAI for depression and emotional susceptibility, and ATD was significantly better than RAI for sex life. In the sensitivity analysis, the overall pattern remained the same.

ATD has been the first-line and mainstay treatment for GD over recent decades in many regions, including Europe, South America, and Japan [3,33–35]. Although RAI remains the preferred treatment in the United States, the use of ATDs has been reported to be increasing [36]. ATD is associated with some minor (arthralgia, skin rash, and gastric intolerance) and major (agranulocytosis) adverse effects [37–39]. In our study, the sex life of patients treated with ATD was significantly better than that of patients treated with RAI. This difference may reflect that ATD is associated with a lower likelihood of permanent hypothyroidism than RAI or surgery[23], as hypothyroidism is known to cause sexual dysfunction[40]. Some evidence suggests transient impairment of male gonadal function after RAI [41]. Therefore, in sexually active patients, ATD may be a better choice than RAI.

RAI has been the preferred treatment in North America since the 1940s. It is associated with a lower relapse rate than ATD but has notable adverse effects, including bone loss, as well as increased risks of

cardiovascular and cerebrovascular events per multiple reports [6,42]. In our study, RAI had significantly worse scores than surgery in the depression and emotional susceptibility domains. It is reported that new or worsened GO is one of the most common adverse effects of RAI[42], and GO can negatively affect the psychological and emotional aspects of life[43]; however, these complications are primarily seen in mild cases and can be effectively mitigated with steroid prophylaxis[44]. Although the association between RAI and worse emotional scores is notable, it does not establish a direct causal link. Many symptoms and risks traditionally attributed to RAI may instead reflect the burden of the underlying hyperthyroidism rather than the treatment itself [44]; therefore, these results should be interpreted with caution regarding causality.

Surgery is the least commonly used treatment modality but provides the highest cure rate. It is typically reserved for cases in which other modalities are contraindicated. Surgery also carries risks, such as recurrent laryngeal nerve palsy and hypothyroid symptoms[3,45]. One of the best-performing domains for patients treated with surgery in our study was social functioning, while the worst domains were tiredness and lack of vitality, similar to the findings for RAI and ATD. However, surgery demonstrated higher scores in the cosmetic domain. Therefore, it may be an appropriate option for patients who prefer a definitive, single treatment and value aesthetic outcomes, particularly those at increased risk of developing GO or depression.

Many patients initially prefer ATD because of concerns regarding radioactive exposure and potential adverse effects of RAI, as well as apprehension about general anesthesia, scarring, and voice alterations following surgery. Nonetheless, relapse often leads patients to opt for surgery or RAI [39]. Financial considerations also play a key role, as a study showed that total thyroidectomy is the most most cost-effective modality in patients with relapse after 18 months of ATD [46]. Another study reported that thyroidectomy may be associated with better outcomes in terms of mortality and major adverse cardiovascular events compared to ATD[47]. These data, together with the risk profiles of each modality, reinforce the need for careful pre-treatment discussion. Accordingly, clinicians should combine the primary clinical assessment of treatment outcomes and safety with QoL priorities where applicable, creating a longitudinal pathway that accounts for relapse probabilities and long-term health implications to support informed, preference-concordant decisions[42].

Several limitations should be acknowledged for this study. First, we did not assess changes in QoL before and after treatment, as there were not enough studies that evaluated both pre- and post-treatment QoL across all questionnaire domains. Second, considerable variation existed among studies in terms of follow-up duration, surgical techniques, RAI doses, medication regimens, and types of ATD, which limited our ability to analyze these factors separately. Third, a high level of heterogeneity was observed across studies, and many potentially relevant studies were excluded due to inconsistent data reporting. Fourth, for the study that reported sample size as a range, we utilized the median value to allow for their inclusion in the quantitative synthesis. While this approach was chosen to avoid the loss

of significant data, it introduces a degree of imprecision in the weighting of those studies. Therefore, our results should be interpreted with caution.

In future studies, researchers should conduct large-scale longitudinal trials with adequate case numbers and extended follow-up periods to evaluate changes in QoL before and after each treatment modality separately. Such studies should assess short-, medium-, and long-term QoL outcomes across all domains of validated questionnaires to provide a more comprehensive understanding of how each treatment affects patients' well-being over time.

Conclusion

While the included studies reported that all therapeutic approaches were associated with improvements in QoL, specific differences emerged: ATD showed better outcomes in sexual functioning, whereas RAI was associated with lower scores in emotional and psychological domains. These findings may help clinicians and patients better understand the effects of each treatment on QoL. Treatment selection should remain a collaborative process that integrates clinical indications and safety considerations with patient preferences, long-term outcomes, local expertise, and financial factors to support informed decision-making.

Abbreviations

ATD: Antithyroid Drug

CIs: Confidence Intervals

GD: Graves' disease

MeSH: Medical Subject Headings

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROSPERO: International Prospective Register of Systematic Reviews

QoL: Quality of Life

RIA: Radioiodine

SD: Standard Deviation

SE: Standard Error

SF-36: 36-Item Short Form Health Survey

ThyPRO: Thyroid-Specific Patient Reported Outcome

Declarations

Author contributions statement

SH. Mossavarali. and A.V.: formal analysis, methodology, data curation, and review and editing; SH. Mossavarali and D.G.S: writing-original draft; SH. Mossavarali and SH. Mohseni: investigation, N.M: review and editing; SH. Mohseni: supervision; M.P.S: conceptualization and project administration. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Ethics approval

The ethics committee of Endocrinology and Metabolism Research Institute reviewed and accepted the study protocol (ethical code: IR.TUMS.EMRI.REC.1404.078).

Funding

There was no funding.

Author Contribution

SH. Mossavarali. and A.V.: formal analysis, methodology, data curation, and review and editing; SH. Mossavarali and D.G.S: writing-original draft; SH. Mossavarali and SH. Mohseni: investigation, N.M: review and editing; SH. Mohseni: supervision; M.P.S: conceptualization and project administration. All authors reviewed the manuscript.

Acknowledgment

None.

Data Availability

The aggregated data generated and analysed during the current study are included in this published article. The aggregated data supporting the findings of this review are available within the article and its supplementary materials. Additional information may be obtained from the corresponding author upon reasonable request.

References

1. Takasu, N. et al. Thyroid-stimulating antibody and TSH-binding inhibitor immunoglobulin in 277 Graves' patients and in 686 normal subjects. *J. Endocrinol. Invest.* **20**, 452–461. 10.1007/bf03348001 (1997).
2. Kahaly, G. J. et al. 2018 European Thyroid Association guideline for the management of Graves' hyperthyroidism. *Eur. thyroid J.* **7**, 167–186 (2018).
3. Kus, L. H., Hopman, W. M., Witterick, I. J. & Freeman, J. L. Quality-of-life outcomes in Graves disease patients after total thyroidectomy. *Ear Nose Throat J.* **96**, E8–e15. 10.1177/0145561317096004-501 (2017).
4. Choi, Y. J., Han, K., Cho, W. K., Jung, M. H. & Suh, B. K. Cancer and Mortality Risks of Graves' Disease in South Korea Based on National Data from 2010 to 2019. *Clin. Epidemiol.* **15**, 535–546. 10.2147/clep.S406361 (2023).
5. Palella, M. et al. Risk and Prognosis of Thyroid Cancer in Patients with Graves' Disease: An Umbrella Review. *Cancers (Basel)*. **15** 10.3390/cancers15102724 (2023).
6. Törring, O. et al. Impaired Quality of Life after Radioiodine Therapy Compared to Antithyroid Drugs or Surgical Treatment for Graves' Hyperthyroidism: A Long-Term Follow-Up with the Thyroid-Related Patient-Reported Outcome Questionnaire and 36-Item Short Form Health Status Survey. *Thyroid* **29**, 322–331. 10.1089/thy.2018.0315 (2019).
7. Pokhrel, B. & Bhusal, K. *in StatPearls (StatPearls Publishing Copyright © 2025* (StatPearls Publishing LLC., 2025).
8. van Kinschot, C. M. J. et al. Preferences of patients and clinicians for treatment of Graves' disease: a discrete choice experiment. *Eur. J. Endocrinol.* **184**, 803–812. 10.1530/eje-20-1490 (2021).
9. Conaglen, H. M., Tamatea, J. A. U., Conaglen, J. V. & Elston, M. S. Treatment choice, satisfaction and quality of life in patients with Graves' disease. *Clin. Endocrinol. (Oxf)*. **88**, 977–984. 10.1111/cen.13611 (2018).
10. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71. 10.1136/bmj.n71 (2021).
11. Watt, T. et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *Eur. J. Endocrinol.* **162**, 161–167. 10.1530/eje-09-0521 (2010).
12. Ware, J. E. Jr. & Sherbourne, C. D. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care*. **30**, 473–483 (1992).
13. Sterne, J. A. C. et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj* **366**, l4898. 10.1136/bmj.l4898 (2019).
14. *ROBINS-I V2 tool*, (2025). https://www.riskofbias.info/welcome/robins-i-v2
15. *Create publication quality risk-of-bias assessment figures*, https://mcguinlu.shinyapps.io/robvis/ (.
16. Hozo, S. P., Djulbegovic, B. & Hozo, I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med. Res. Methodol.* **5**, 13. 10.1186/1471-2288-5-13 (2005).

17. Täeb, D. et al. Quality of life, clinical outcomes and safety of early prophylactic levothyroxine administration in patients with Graves' hyperthyroidism undergoing radioiodine therapy: A randomized controlled study. *Eur. J. Endocrinol.* **174**, 491–502. 10.1530/EJE-15-1099 (2016).
18. Al-Adhami, A., Craig, W. & Krukowski, Z. H. Quality of life after surgery for graves' disease: Comparison of those having surgery intended to preserve thyroid function with those having ablative surgery. *Thyroid* **22**, 494–500. 10.1089/thy.2011.0153 (2012).
19. Higgins, J. P. T. *et al. Cochrane Handbook for Systematic Reviews of Interventions, version 6.5 (updated August 2024)*. (Cochrane).
20. Elberling, T. V. et al. Impaired health-related quality of life in Graves' disease.: A prospective study. *Eur. J. Endocrinol.* **151**, 549–555. 10.1530/eje.0.1510549 (2004).
21. Tagami, T. et al. Short-term effects of β-adrenergic antagonists and methimazole in new-onset thyrotoxicosis caused by graves' disease. *Intern. Med.* **51**, 2285–2290. 10.2169/internalmedicine.51.7302 (2012).
22. Azizi, F., Ataie, L., Hedayati, M., Mehrabi, Y. & Sheikholeslami, F. Effect of long-term continuous methimazole treatment of hyperthyroidism: comparison with radioiodine. *Eur. J. Endocrinol.* **152**, 695–701. 10.1530/eje.1.01904 (2005).
23. Villagelin, D., Romaldini, J. H., Santos, R. B., Milkos, A. B. & Ward, L. S. Outcomes in Relapsed Graves' Disease Patients Following Radioiodine or Prolonged Low Dose of Methimazole Treatment. *Thyroid* **25**, 1282–1290. 10.1089/thy.2015.0195 (2015).
24. Ismailov, S. I., Khayitboeva, K. K., Ruzibaev, R. Y. & Sapaev, D. S. Health quality assessment of patients with toxic-diffuse goiter. *Indian J. Forensic Med. Toxicol.* **14**, 7389–7394. 10.37506/ijfmt.v14i4.12816 (2020).
25. Urunbayeva, D., Umarova, D. & Bashirova, A. Assessment of The Quality of Life of Patients with Diffuse Toxic Goiter Against The Background of Drug Therapy and Surgical Treatment. *Sci. J. Appl. Med. Sci.* **4**, 101–109 (2025).
26. Riguetto, C. M., Neto, A. M., Tambascia, M. A. & Zantut-Wittmann, D. E. The relationship between quality of life, cognition, and thyroid status in Graves' disease. *ENDOCRINE* 63, 87–93, (2019). 10.1007/s12020-018-1733-y
27. Abraham-Nordling, M. et al. Graves' disease: A long-term quality-of-life follow up of patients randomized to treatment with antithyroid drugs, radioiodine, or surgery. *Thyroid* **15**, 1279–1286. 10.1089/thy.2005.15.1279 (2005).
28. Liew, H. L. et al. Psychometric properties of the thyroid-specific quality of life questionnaire ThyPRO in Singaporean patients with Graves' disease. *J. Patient Rep. Outcomes*. **5** 10.1186/s41687-021-00309-x (2021).
29. Bukvic, B. et al. Improved quality of life in hyperthyroidism patients after surgery. *J. Surg. Res.* **193**, 724–730. 10.1016/j.jss.2014.07.061 (2015).
30. Li, Y. et al. Quality of life and decision regret in patients with late-hypothyroidism after radioiodine treatment for Graves' disease. *Clin. Endocrinol.* **100**, 87–95. 10.1111/cen.14986 (2024).

31. Mangelen, S. F. & Cunanan, E. Health-Related Quality of Life (HRQoL) of Adult Filipinos with Graves' Disease cured by Radioiodine Therapy compared to those controlled by Antithyroid Drugs at University of Santo Tomas Hospital: A Pilot Study. *J. ASEAN Fed. Endocr. Soc.* **32**, 100–107. 10.15605/jafes.032.02.02 (2017).
32. Stokland, A. E. M. et al. Outcomes of Patients With Graves Disease 25 Years After Initiating Antithyroid Drug Therapy. *J. Clin. Endocrinol. METABOLISM*. **109**, 827–836. 10.1210/clinem/dgad538 (2024).
33. Glinoer, D., Hesch, D., Lagasse, R. & Laurberg, P. The management of hyperthyroidism due to Graves' disease in Europe in 1986. Results of an international survey. *Acta Endocrinol. Suppl. (Copenh)*. **285**, 3–23 (1987).
34. Romaldini, J. H. Case selection and restrictions recommended to patients with hyperthyroidism in South America. *Thyroid* **7**, 225–228. 10.1089/thy.1997.7.225 (1997).
35. Nagayama, Y., Izumi, M. & Nagataki, S. The management of hyperthyroidism due to Graves' disease in Japan in 1988. The Japan Thyroid Association. *Endocrinol. Jpn*. **36**, 299–314. 10.1507/endocrj1954.36.299 (1989).
36. Burch, H. B., Burman, K. D. & Cooper, D. S. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J. Clin. Endocrinol. Metab.* **97**, 4549–4558. 10.1210/jc.2012-2802 (2012).
37. Cooper, D. S. Antithyroid drugs. *N Engl. J. Med.* **352**, 905–917. 10.1056/NEJMra042972 (2005).
38. Watanabe, N. et al. Antithyroid drug-induced hematopoietic damage: a retrospective cohort study of agranulocytosis and pancytopenia involving 50,385 patients with Graves' disease. *J. Clin. Endocrinol. Metab.* **97**, E49–53. 10.1210/jc.2011-2221 (2012).
39. Chen, Y., Ren, Z., Zhou, X., Gao, P. A. & Systematic Review of Patient Preferences, Expectations, and Values for the Management and Treatment of Graves Disease. *Patient Prefer Adherence*. **19**, 2949–2957. 10.2147/ppa.S540261 (2025).
40. Shen, M. et al. Is There an Association Between Hypothyroidism and Sexual Dysfunction: A Systematic Review and Cumulative Analysis. *Sex. Med.* **9**, 100345. 10.1016/j.esxm.2021.100345 (2021).
41. Cai, Y., Yang, Y., Pang, X. & Li, S. The effect of radioactive iodine treatment for differentiated thyroid cancer on male gonadal function: a meta-analysis. *Endocr. Connect.* **12** 10.1530/ec-23-0299 (2023).
42. Sundaresh, V. et al. Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network meta-analysis. *J. Clin. Endocrinol. Metab.* **98**, 3671–3677. 10.1210/jc.2013-1954 (2013).
43. Woo, T. et al. The Effect of Ophthalmic Surgery for Graves' Orbitopathy on Quality of Life: A Systematic Review and Meta-Analysis. *Thyroid* **32**, 177–187. 10.1089/thy.2021.0411 (2022).
44. Bartalena, L., Gallo, D., Kahaly, G. J., Piantanida, E. & Tanda, M. L. Process to radioactive iodine treatment for Graves' hyperthyroidism: condemned or absolved? *J. Endocrinol. Investig.* **48**, 1927–1950. 10.1007/s40618-025-02653-x (2025).

45. Guo, Z., Yu, P., Liu, Z., Si, Y. & Jin, M. Total thyroidectomy vs bilateral subtotal thyroidectomy in patients with Graves' diseases: a meta-analysis of randomized clinical trials. *Clin. Endocrinol. (Oxf)*. **79**, 739–746. 10.1111/cen.12209 (2013).
46. In, H. et al. Treatment options for Graves disease: a cost-effectiveness analysis. *J Am Coll Surg* 209, 170–179.e171–172, (2009). 10.1016/j.jamcollsurg.2009.03.025
47. Quinn, L., Torlinska, B. & Boelaert, K. Systematic review of mortality and long-term major cardiovascular events (MACE) following different treatment approaches for hyperthyroidism. *Endocrine Abstracts* (2024).

Table

Table 1 is available in the Supplementary Files section.

Figures

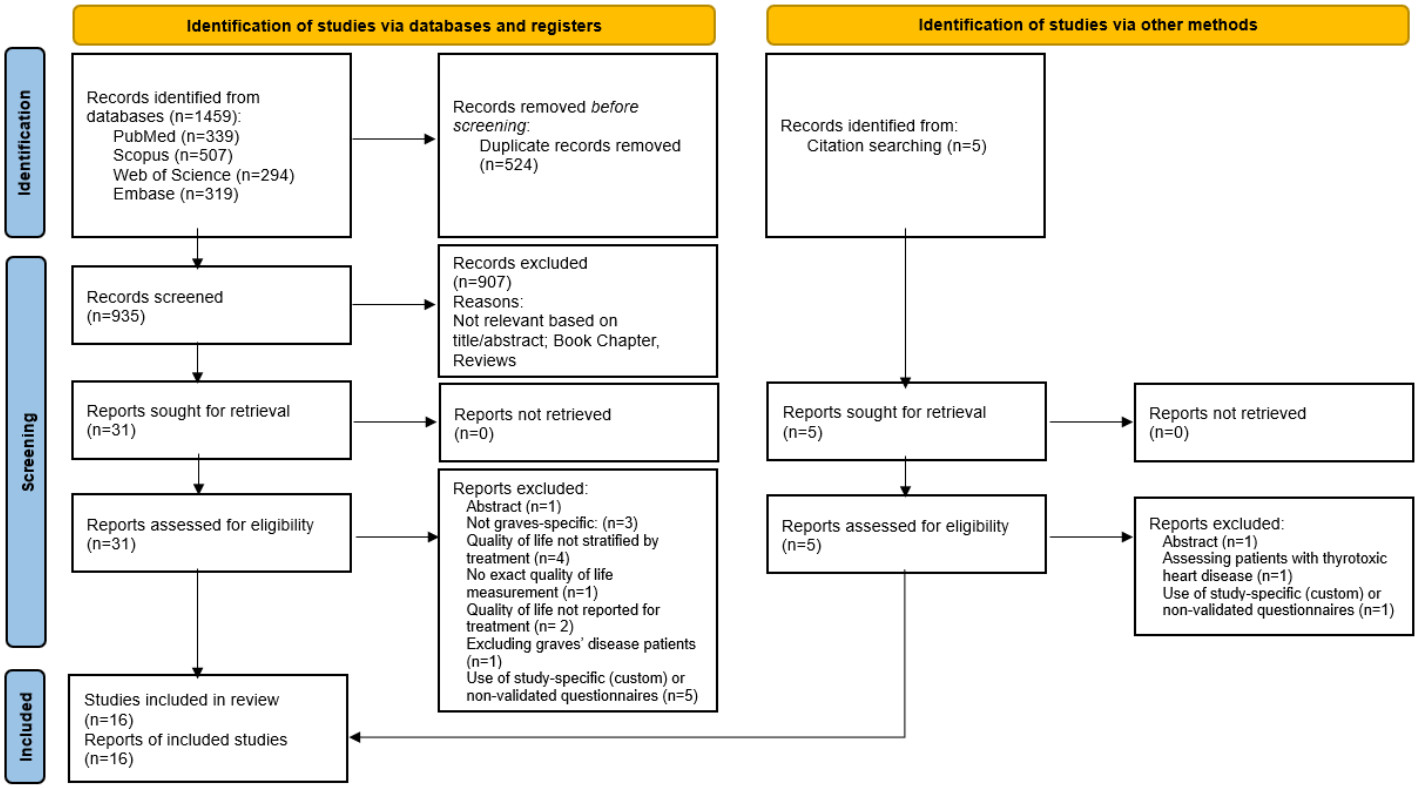


Figure 1

The PRISMA flow diagram of the systematic review

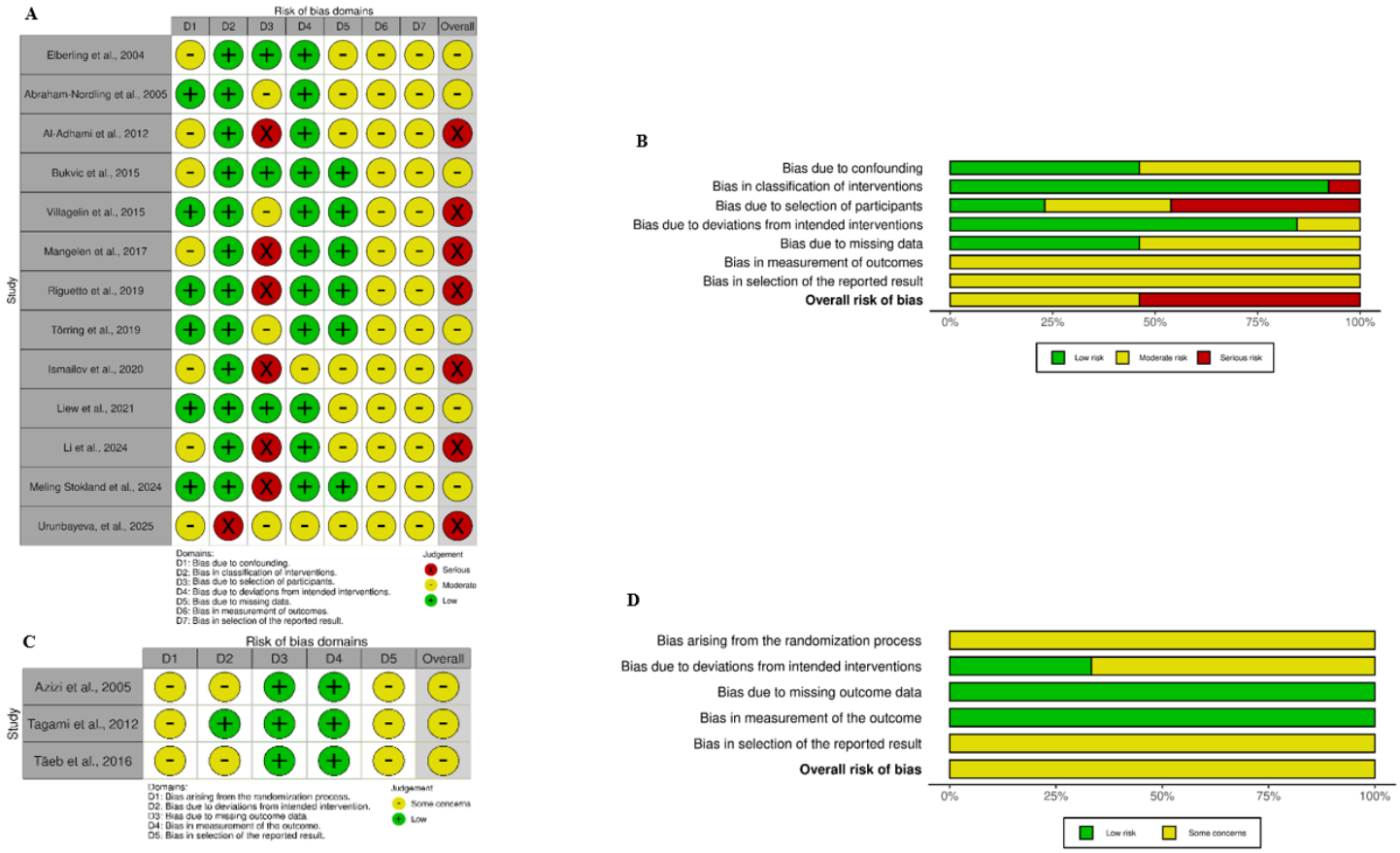


Figure 2

Risk of bias assessment: **A.** Risk of bias in non-randomized studies assessed using the ROBINS-I V2 tool, **B.** Summary plot of the overall risk of bias across all domains for the non-randomized studies, **C.**Risk of bias in randomized studies assessed using the RoB2 tool **D.**Summary plot of the overall risk of bias across all domains for the randomized studies

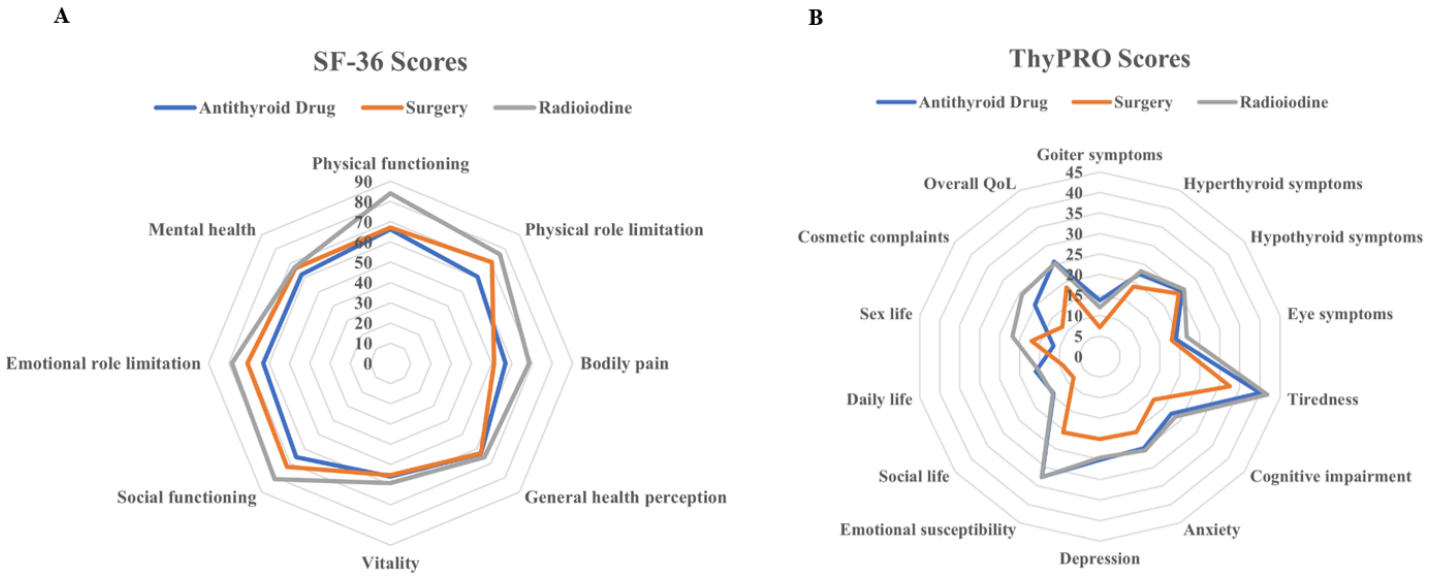


Figure 3

Pooled Quality of Life Scores from Meta-Analysis: **A.** SF-36 subscales and **B.** ThyPRO subscales.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GraphicalAbstract.png
- Supplementary1.docx
- Table1.docx
- Supplementary2.docx