




The need for the GREAT+ score to predict relapse in Graves' disease: a questionnaire among patients and internal medicine specialists

H. I. Jansen^{1,2,3} · C. Heuveling van Beek² · P. H. Bisschop^{3,4} · A. C. Heijboer^{1,2,3,5} · E. Bruinstroop^{3,4} · A. Boelen^{2,3,5} 

Received: 16 January 2024 / Accepted: 5 March 2024
© The Author(s) 2024

Abstract

Purpose Graves' disease (GD) is an auto-immune cause of hyperthyroidism. First-line treatment often consists of a 12–18 month course of antithyroid drugs (ATD). After discontinuation of ATD, GD relapses in approximately 50% of patients. The 'Graves recurrent event after therapy+' (GREAT+) score may predict individual relapse chances after ATD discontinuation more accurately based on clinical and laboratory parameters at diagnosis. We investigated the need for the GREAT+ score through an online questionnaire among GD patients and physicians treating GD.

Methods An anonymous online questionnaire was distributed to patients and physicians between June 2022 and August 2023.

Results The questionnaire was completed by 532 patients and 44 physicians. Results showed that 94% of patients were interested in knowing their GREAT+ score at the start of treatment. 55% would consider definite treatment (radioiodine/thyroidectomy) as first-line treatment in case of a high relapse chance. 98% of the physicians indicated the GREAT+ score would support patient counseling. 84% may change their advice for first-line treatment if a patient has a high relapse chance based on the score.

Conclusion Patients and physicians considered the GREAT+ score as a valuable addition to the current available information which could change treatment decisions. Therefore, external validation of the GREAT+ score is justified to implement this score in clinical practice.

Keywords Graves' disease · Relapse chance · Prediction model · Antithyroid drugs

Introduction

Graves' disease (GD) is an auto-immune disease and the most common cause of hyperthyroidism in iodine-sufficient areas with a prevalence of 0.5–2% [1]. GD is characterized by the presence of autoantibodies that are directed against the thyroid-stimulating hormone (TSH) receptor (TRAb) and thereby able to activate this receptor. This leads to increased free thyroxine (fT4) and free triiodothyronine (fT3) concentrations and suppressed TSH concentrations. TRAbs are often measured to differentiate between GD and other causes of hyperthyroidism, where positive TRAb results (above a pre-defined cut-off value) guide towards GD [2, 3]. Furthermore, TRAbs are measured for follow-up GD treatment [3]. First-line treatment of GD often involves a 12–18 months course of antithyroid drugs (ATD) in the form of ATD titration therapy or combined with levothyroxine (LT4) (block-and-replace therapy). After discontinuation of ATD, GD relapses in approximately 50% of the patients [4, 5]. After a relapse, patients can also choose a more definite

✉ A. Boelen
a.boelen@amsterdamumc.nl

¹ Department of Laboratory Medicine, Endocrine Laboratory, Amsterdam UMC Location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

² Department of Laboratory Medicine, Endocrine Laboratory, Amsterdam UMC Location University of Amsterdam, Academic Medical Center, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

³ Amsterdam Gastroenterology, Endocrinology and Metabolism, Amsterdam, The Netherlands

⁴ Department of Endocrinology and Metabolism, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

⁵ Amsterdam Reproduction and Development Research Institute, Amsterdam, The Netherlands

form of treatment including radioiodine (RAI) treatment or (total) thyroidectomy, after which life-long LT4 treatment is necessary.

In 2016, the ‘Graves recurrent event after therapy’ (GREAT) and GREAT+ score were developed in the Department of Endocrinology of Amsterdam UMC to improve the prediction of relapse after discontinuation of ATD already at the start of the first-line treatment [6]. These scores were based on four clinical markers: age, fT4 concentration, TRAb concentration, and goiter size. The GREAT+ score included the genetic markers tyrosine-protein phosphatase non-receptor type 22 (*PTPN22*) C/T polymorphism and HLA subtypes *DQB1*02*, *DQA1*05* and *DRB*03* (Table 1). The GREAT score allocates patients in three different categories corresponding with a relapse chance of 16%, 44% and 68%. The GREAT+ score uses four categories corresponding with a relapse chance of 4%, 21%, 49% and 84% (Table 1). Scores to predict relapse after ATD discontinuation are

currently limited to the GREAT(+) score and the clinical severity score (CSS). The CSS score was developed by Bartalena et al. [7] and incorporated fT4 concentration, thyroid volume and the presence of Graves orbitopathy [7].

The GREAT score has been validated in three independent cohorts [8–10], but the GREAT+ score has not been validated yet. Nevertheless, the addition of a fourth category seems extremely valuable in distinguishing relapse chances, especially in the lowest and the highest categories. Before starting an external validation of the GREAT+ score, we aimed to investigate the need for implementation of the GREAT+ score among patients treated for GD and physicians treating GD.

Methods

The need for the GREAT+ score was assessed by a questionnaire for current and past GD patients and physicians treating

Table 1 Determinants GREAT+ score and GREAT+ categories with relapse chances

Determinants	Score	
Age (years)		
≥40	0	
<40	1	
fT4 (pmol/L)		
<40	0	
≥40	1	
TRAb (IU/L)		
<6.0	0	
6.0–19.9	1	
≥20.0	2	
Goiter size		
Grade 0-I	0	
Grade II-III	2	
<i>PTPN22</i>		
Wild type	0	
C/T	1	
HLA polymorphisms		
0	0	
1–2	2	
3	3	
Score GREAT+	Category	Relapse chance (%)
0–2	I	4
3–4	II	21
5–6	III	49
7–10	IV	84

This table is based on the description of the GREAT+ score in the paper of Vos et al. [6].

fT4 free thyroxine, TRAb thyroid-stimulating hormone receptor antibodies, *PTPN22* tyrosine-protein phosphatase non-receptor type 22, HLA human leukocyte antigen

GD patients in the Netherlands. The questionnaire was distributed and completed between June 2022 and August 2023 and supported by the Dutch Thyroid Organization (Schildklier Organisatie Nederland; SON), Dutch Thyroid Research Foundation and the Dutch Society for Endocrinology (NVE) by posting the link to the questionnaire on either their website, social media account, or newsletter. Respondents were not approached personally and could anonymously and voluntarily complete the questionnaire. Therefore, no informed consent from each subject was necessary. Furthermore, the study was not subject to the Medical Research Involving Human Subjects Act since it does not impose any act or mode of behavior on the subjects.

Patient questionnaire

The patient questionnaire consisted of three sections (Supplementary Table 1). The first section concerned questions on age, sex, and medication (except thyroid medication). The second section concerned 14 questions and dealt with patients' experience with GD. Before the third section, information regarding the GREAT+ score was provided. After taking note of this information, nine specific questions regarding the GREAT+ score were asked. Closed questions were often followed by open questions.

Physician questionnaire

The physician questionnaire consisted of two sections (Supplementary Table 2). The first section included four questions regarding experience as a GD treating physician. Before the second section, information regarding the GREAT+ score was provided. After taking note of this information, 14 questions regarding the GREAT+ score were asked. Closed questions were followed by open questions.

Analysis

The results of both questionnaires were fully anonymous. Questions that subsequently proved redundant were excluded from the analysis. Since most questions were closed, results were merely depicted as a percentage of the total respondents. IBM SPSS Statistics 28.0 (Chicago, IL, USA) and GraphPad Prism 9.3.1 Software were used to analyze the results. Answers from the open questions were used to get insight into the rationale of answers from closed questions, but were not statistically assessed.

Results

Patient questionnaire

Table 2 shows the baseline characteristics of the 532 patients who completed the questionnaire. Mean age was 51 years and almost all patients were female (93%). Many patients have had a relapse (48.5%). Interestingly, patients considered GD symptoms as extremely burdensome (mean 4.24 on a 1–5 scale), while side-effects of treatment modalities patients received were considered less burdensome (mean 2.89, 2.73 and 3.56 on a 1–5 scale for, respectively, ATD, RAI and thyroidectomy). A score of 1 indicated no symptoms/side effects; a score of 2 indicated that symptoms/side effects influenced the patients' life a bit but did not disrupt daily life; 3 indicated that the symptoms/side effects caused some adjustments in daily life; 4 indicated that the symptoms/side effects continuously disrupted daily life; 5 indicated that the symptoms/side effects prevented the individual from carrying out any of their normal daily activities.

Almost 94% of the respondents wanted to know their GREAT+ score, while a minority (6%) did not (Fig. 1). 55.3% would consider definite treatment based on their GREAT+ result. This even increased up to 63% if their score would fall in category IV (84% chance of relapse).

Open questions revealed several motives for knowing the GREAT+ score. Examples are providing additional information, better expectations about GD relapse, reason for an alternative treatment decision or indication for laboratory testing based on information regarding relapse chance. Patients who did not want to know the score responded it would not change their treatment decision or knowing the score would make them feel less confident during treatment.

Physician questionnaire

44 physicians completed the questionnaire (Table 2). Ten (22.7%) physicians indicated they already used a tool to predict relapse chance, with four of them specifically using the GREAT score. Figure 1 shows that 70.5% of the physicians reported that the GREAT+ score would make them more likely to recommend definite treatment as first-line therapy, especially if the GREAT+ score would be category IV (84% of the physicians). 88.6% responded affirmatively that the GREAT+ score could help them in counseling patients with a pregnancy wish for more suitable treatment. Almost all physicians (97.7%) confirmed that the GREAT+ score would be beneficial in communication with their patients.

The open questions revealed that the GREAT+ score could make physicians better positioned to provide a

Table 2 Baseline characteristics of participating patients and physicians

Characteristics	Outcome measure	Outcome
Patients		
Respondents	<i>N</i>	532
Age	Mean years (SD)	51 (12.7)
Sex	Female, <i>n</i> (%)	496 (93.2)
Relapse after treatment (ATD, RAI, incomplete thyroidectomy)	No, <i>n</i> (%)	141 (26.5)
	Yes, <i>n</i> (%)	258 (48.5)
	Just started treatment, so not known, <i>n</i> (%)	133 (25)
How did you experience the symptoms of GD?	Scale 1–5	Mean 4.24, median 4.0 (<i>n</i> = 532)
What GD treatment(s) did you receive?	ATD, <i>n</i> (%)	339 (63.7)
	RAI, <i>n</i> (%)	24 (4.5)
	Thyroidectomy, <i>n</i> (%)	6 (1.1)
	ATD and RAI, <i>n</i> (%)	132 (24.8)
	ATD and thyroidectomy, <i>n</i> (%)	24 (4.5)
	RAI and thyroidectomy, <i>n</i> (%)	1 (0.2)
	ATD, RAI and thyroidectomy, <i>n</i> (%)	6 (1.1)
Did you experience side-effects of ATD?	Scale 1–5	Mean 2.89, median 3 (<i>n</i> = 380)
Did you experience side-effects of RAI?	Scale 1–5	Mean 2.73, median 2 (<i>n</i> = 97)
Did you experience side-effects of thyroidectomy	Scale 1–5	Mean 3.56, median 4 (<i>n</i> = 18)
Physicians		
Respondents	<i>N</i>	44
Specialized in endocrinology	Yes, <i>n</i> (%)	40 (90.9)
Years of experience	Mean	10.1
Number of patients treated for GD per year	Mean	41
Block-and-replace therapy as first line of treatment	<i>N</i> (%)	37 (84.1)
Using a tool for indication relapse chance	Yes, <i>n</i> (%)	10 (22.7)

SD standard deviation, *ATD* antithyroid drug, *RAI* radioactive iodine, *GD* Graves' disease

well-informed recommendation for definite treatment, especially for women in their reproductive age. Other physicians stated they would never start with definite treatment, and the GREAT+ score would not change this. Furthermore, physicians indicated the GREAT+ score could especially be of added value in expectation management and providing personalized information.

Discussion

The results of our questionnaire showed that both patients and physicians considered the GREAT+ score a valuable addition to the current available information for GD treatment which may change treatment decisions. Therefore, an external validation of the GREAT+ score is useful and needed to eventually implement the GREAT+ score in clinical practice.

The majority of patients reported they wanted to know their GREAT+ score at diagnosis. Patients often mentioned uncertainty regarding relapse chance was a large burden, especially since symptoms of active GD were

considered more burdensome than side effects of all treatment modalities. Therefore, additional information would contribute to set expectations about the likelihood of relapse. A study from 2019 investigating long-term effects of GD treatment showed that over 25% of GD patients did not feel recovered after a mean follow-up of 8 years [11]. This did not differ per treatment modality and side effects played an important role in this. In our study, we showed a significant burden of side effects for the three treatment modalities, although lower than the burden of GD symptoms. The side effects were rated similarly burdensome for ATD and RAI (median 2.89 and 2.73, respectively), while the burden of side effects for thyroidectomy was higher (mean 3.56). However, this may be caused by the lower number of responding participants with a thyroidectomy (*n* = 18). Previous research evaluated the quality of life of GD patients after receiving either ATD, RAI or thyroidectomy and mostly did not show a difference between the treatment modalities [12, 13], although the quality of life was significantly lower compared to the reference population [12, 14]. Only the study of Törring et al. [14] showed a lower quality of life in RAI-treated patients compared

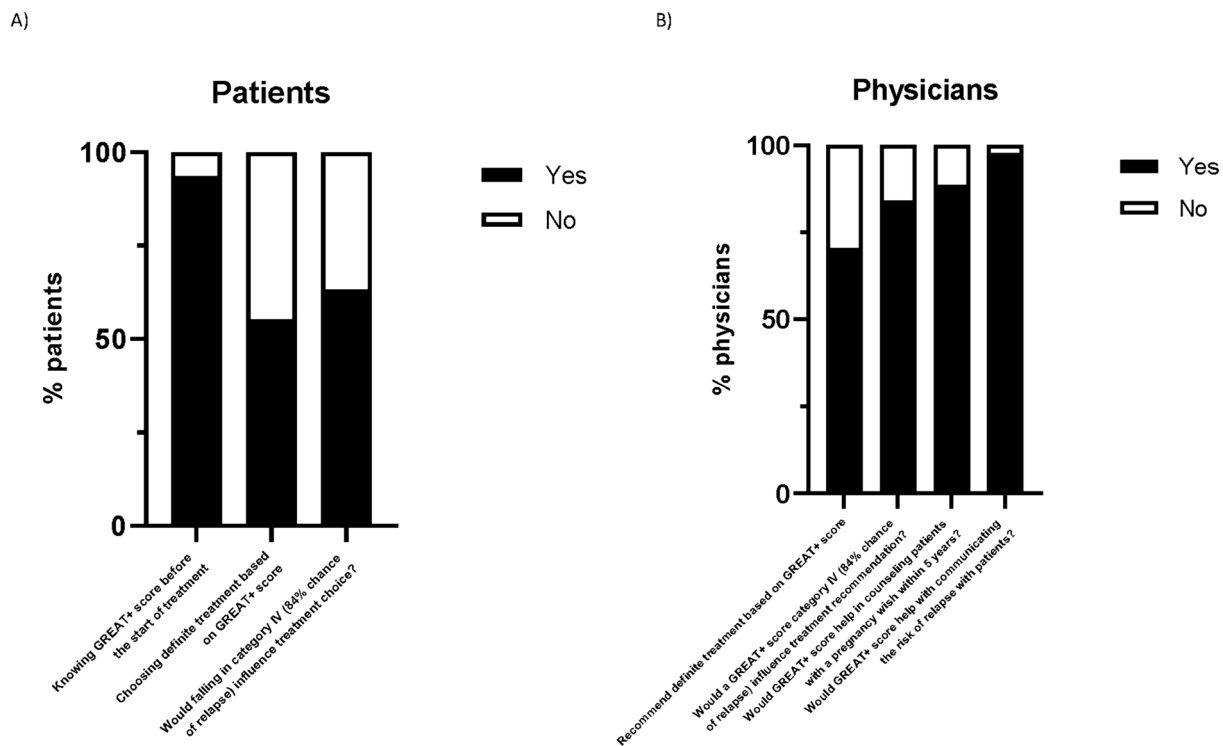


Fig. 1 Responses of **A** patients and **B** physicians to GREAT+ score questions

to ATD and surgery [14]. These findings show the large impact of GD before and after treatment.

The GREAT+ score would not change treatment decisions for all patients and open questions indicated that additional information about the disease is a large part of the added value of the GREAT+ score as well. Nevertheless, more than half of the responding patients reported they would more likely opt for definite treatment if they knew their GREAT+ score, especially if the score indicated a relapse chance of 84% (category IV). This indicated the importance of adding this fourth category to the GREAT+ score compared to the GREAT score. A previous study supported this finding by reporting that remission rate was the most important component in treatment decision for GD [15]. Also physicians were interested in the GREAT+ score, and they would advise more likely definite treatment if patients would fall in category IV. Furthermore, physicians reported that results of the GREAT+ score would help in counseling and treating patients with a pregnancy wish within the upcoming five years, which could lead to a different treatment regimen. A European questionnaire from 2010 showed that only 22% of physicians would advise definite therapy for newly diagnosed GD in women with a pregnancy wish, while this increased after a relapse to 80% [16]. Lastly, communicating the risk of GD relapse with patients was found to be a large advantage of the GREAT+ score. A recent study from the US found that in only 3% of their observed consultations

with patients with GD an individualized remission estimate was mentioned [17]. The GREAT+ score could, therefore, contribute to firstly discuss this remission or relapse estimate and secondly help to personalize treatment.

The GREAT+ score is unique in adding genetic markers which makes it better able to categorize patients into specific relapse chances. These four genetic markers are not included in the standard diagnostic workup of GD and additional costs of measuring these markers in the Netherlands are approximately €300,- per patient. On the other hand, as described by Vos et al. [6] the GREAT+ score could lower healthcare costs as well (e.g. better treatment compliance due to additional information provision, and lower treatment costs in case of lower recurrence rate) [6]. Furthermore, it could be debated whether genetic testing should be conducted for all three GREAT score categories or solely for category II. The cost-effectiveness of adding genetics to the GREAT score should thus be evaluated. Moreover, future research should make an effort to further personalize GD relapse chances, for example the use of machine learning could be further explored.

Strengths of this study were the inclusion of a large cohort of patients and the inclusion of physicians as well. A limitation is that we predominantly included patients who were members of the patient association and, therefore, more involved in their disease. Furthermore, the questions were asked retrospectively, so it would be interesting to know how

newly diagnosed patients would value the GREAT + score. Our data showed no obvious difference in interest in the GREAT+ score between patients who were in their first-line treatment, patients who experienced a relapse and patients who did not. Patients with a relapse only reported more often that if they would fall in category IV of the GREAT+ score, this would most likely influence their treatment choice (70.2%).

In conclusion, our questionnaire showed that the GREAT + score is of added value for both patients and physicians which justifies the validation and implementation of the GREAT + score in the near future.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40618-024-02358-7>.

Acknowledgements We would like to thank the Dutch Thyroid Organization (Schildklier Organisatie Nederland; SON), the Dutch Thyroid Research Foundation and the Dutch Society for Endocrinology (NVE) for helping with the distribution of the questionnaires.

Author contributions Heleen I. Jansen: Conceptualization, Formal analysis, Methodology, Project administration, Visualization, Writing—original draft. Cecile Heuveling van Beek: Conceptualization, Formal analysis, Methodology, Writing—review & editing. Peter H. Bisschop: Writing—review & editing. Annemieke C. Heijboer: Writing—review & editing. Eveline Bruinstroop: Conceptualization, Supervision, Writing—review & editing. Anita Boelen: Conceptualization, Supervision, writing—review & editing.

Funding This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Data availability Raw data may be shared upon reasonable request.

Declarations

Conflict of interest There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Ethical approval This study was not subject to the Medical Research Involving Human Subjects Act (WMO) as it does not impose any act or mode of behavior on the subjects.

Informed consent Since results of the questionnaire were fully anonymous and respondents were not personally approached for participation in the questionnaire, it was not necessary to obtain informed consent from each subject.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Davies TF, Andersen S, Latif R, Nagayama Y, Barbesino G, Brito M, Eckstein AK, Stagnaro-Green A, Kahaly GJ (2020) Graves' disease. *Nat Rev Dis Prim* 6(1):52. <https://doi.org/10.1038/s41572-020-0184-y>
- Vasileiou M, Gilbert J, Fishburn S, Boelaert K (2020) Thyroid disease assessment and management: summary of NICE guidance. *BMJ* 368:m41. <https://doi.org/10.1136/bmj.m41>
- Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, Rivkees SA, Samuels M, Sosa JA, Stan MN et al (2016) 2016 American Thyroid Association Guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* 26(10):1343–1421. <https://doi.org/10.1089/thy.2016.0229>
- Lee SY, Pearce EN (2023) Hyperthyroidism: a review. *JAMA* 330(15):1472–1483. <https://doi.org/10.1001/jama.2023.19052>
- Sundaresh V, Brito JP, Wang Z, Prokop LJ, Stan MN, Murad MH, Bahn RS (2013) Comparative effectiveness of therapies for graves' hyperthyroidism: a systematic review and network meta-analysis. *J Clin Endocrinol Metab* 98(9):3671–3677. <https://doi.org/10.1210/jc.2013-1954>
- Vos XG, Ender E, Zwinderman AH, Tijssen JGP, Wiersinga WM (2016) Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with graves' hyperthyroidism. *J Clin Endocrinol Metab* 101(4):1381–1389. <https://doi.org/10.1210/jc.2015-3644>
- Bartalena L, Masiello E, Magri F, Veronesi G, Bianconi E, Zerbini F, Gaiti M, Spreafico E, Gallo D, Premoli P et al (2016) The phenotype of newly diagnosed Graves' disease in Italy in recent years is milder than in the past: results of a large observational longitudinal study. *J Endocrinol Invest* 39(12):1445–1451. <https://doi.org/10.1007/s40618-016-0516-7>
- Masiello E, Veronesi G, Gallo D, Premoli P, Bianconi E, Rosetti S, Cusini C, Sabatino J, Ippolito S, Piantanida E et al (2018) Antithyroid drug treatment for graves' disease: baseline predictive models of relapse after treatment for a patient-tailored management. *J Endocrinol Invest* 41(12):1425–1432. <https://doi.org/10.1007/s40618-018-0918-9>
- Zuhur SS, Elbuken G, Yildiz I, Kadioglu P, Erol S, Sahin S, Kilinc F, Akbaba G, Topcu B, Altuntas Y (2019) External validation of the GREAT score in Turkish patients with graves' hyperthyroidism treated with the titration regimen method of antithyroid drugs: a multicenter study. *Horm Metab Res* 51(10):627–633
- Struja T, Kaeslin M, Boesiger F, Jutzi R, Imahorn N, Kutz A, Bernasconi L, Mundwiler E, Mueller B, Christ-Crain M et al (2017) External validation of the GREAT score to predict relapse risk in graves' disease: results from a multicenter, retrospective study with 741 patients. *Eur J Endocrinol* 176(4):413–419. <https://doi.org/10.1530/eje-16-0986>
- Sjölin G, Holmberg M, Törning O, Byström K, Khamisi S, de Laval D, Abraham-Nordling M, Calissendorff J, Lantz M, Hallengren B et al (2019) The long-term outcome of treatment for graves' hyperthyroidism. *Thyroid* 29(11):1545–1557. <https://doi.org/10.1089/thy.2019.0085>
- Abraham-Nordling M, Törning O, Hamberger B, Lundell G, Tallstedt L, Calissendorff J, Wallin G (2005) Graves' disease: a long-term quality-of-life follow up of patients randomized to treatment with antithyroid drugs, radioiodine, or surgery. *Thyroid* 15(11):1279–1286. <https://doi.org/10.1089/thy.2005.15.1279>
- Conaglen HM, Tamatea JAU, Conaglen JV, Elston MS (2018) Treatment choice, satisfaction and quality of life in patients with graves' disease. *Clin Endocrinol* 88(6):977–984. <https://doi.org/10.1111/cen.13611>

14. Törring O, Watt T, Sjölin G, Byström K, Abraham-Nordling M, Calissendorff J, Cramon PK, Filipsson Nyström H, Hallengren B, Holmberg M et al (2019) 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(3):322–331. <https://doi.org/10.1089/thy.2018.0315>
15. van Kinschot CMJ, Soekhai VR, de Bekker-Grob EW, Visser WE, Peeters RP, van Ginhoven TM, van Noord C (2021) Preferences of patients and clinicians for treatment of graves' disease: a discrete choice experiment. *Eur J Endocrinol* 184(6):803–812. <https://doi.org/10.1530/eje-20-1490>
16. Poppe K, Hubalewska-Dydejczyk A, Laurberg P, Negro R, Vermiglio F, Vaidya B (2012) Management of hyperthyroidism in pregnancy: results of a survey among members of the European Thyroid Association. *Eur Thyroid J* 1(1):34–40. <https://doi.org/10.1159/000336101>
17. Hidalgo J, Vallejo S, El Kawkgi OM, Ospina NMS, Maraka S, Brito JP (2022) Individualized graves' disease remission rates conversations: a videographic analysis of medical encounters. *Endocrine* 76(1):237–239. <https://doi.org/10.1007/s12020-022-02989-x>

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