

RESEARCH

Machine learning identifies baseline clinical features that predict early hypothyroidism in patients with Graves' disease after radioiodine therapy

Lian Duan^{1,*}, Han-Yu Zhang^{2,*}, Min Lv^{1,*}, Han Zhang², Yao Chen², Ting Wang¹, Yan Li¹, Yan Wu³, Junfeng Li⁴ and Kefeng Li⁵ 

¹Department of Nuclear Medicine, Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi, China

²Changzhi Medical College, Changzhi, Shanxi, China

³Department of Clinical Laboratory, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai, Shandong, China

⁴Department of Radiology, Heping Hospital Affiliated to Changzhi Medical College, Changzhi, Shanxi, China

⁵School of Medicine, University of California, San Diego, California, USA

Correspondence should be addressed to Y Wu or J Li: yt6688082@163.com or lijunfeng@czmc.edu.cn

*(L Duan, H-Y Zhang and M Lv contributed equally to this work)

Abstract

Background and objective: Radioiodine therapy (RAI) is one of the most common treatment solutions for Graves' disease (GD). However, many patients will develop hypothyroidism as early as 6 months after RAI. This study aimed to implement machine learning (ML) algorithms for the early prediction of post-RAI hypothyroidism.

Methods: Four hundred and seventy-one GD patients who underwent RAI between January 2016 and June 2019 were retrospectively recruited and randomly split into the training set (310 patients) and the validation set (161 patients). These patients were followed for 6 months after RAI. A set of 138 clinical and lab test features from the electronic medical record (EMR) were extracted, and multiple ML algorithms were conducted to identify the features associated with the occurrence of hypothyroidism 6 months after RAI.

Results: An integrated multivariate model containing patients' age, thyroid mass, 24-h radioactive iodine uptake, serum concentrations of aspartate aminotransferase, thyrotropin-receptor antibodies, thyroid microsomal antibodies, and blood neutrophil count demonstrated an area under the receiver operating curve (AUROC) of 0.72 (95% CI: 0.61–0.85), an F1 score of 0.74, and an MCC score of 0.63 in the training set. The model also performed well in the validation set with an AUROC of 0.74 (95% CI: 0.65–0.83), an F1 score of 0.74, and a MCC of 0.63. A user-friendly nomogram was then established to facilitate the clinical utility.

Conclusion: The developed multivariate model based on EMR data could be a valuable tool for predicting post-RAI hypothyroidism, allowing them to be treated differently before the therapy. Further study is needed to validate the developed prognostic model at independent sites.

Key Words

- ▶ Graves' disease
- ▶ radioactive iodine therapy
- ▶ hypothyroidism
- ▶ machine learning
- ▶ electronic medical data
- ▶ predictive model

Endocrine Connections
(2022) 11, e220119

Introduction

Graves' disease (GD) is an autoimmune disease with the overproduction of thyroid-stimulating hormones (TSH) induced by the activation of thyrotropin-receptor antibodies (TRAb) (1). It is the most common cause of hyperthyroidism, which has significant adverse effects on many organ systems, such as cardiovascular, metabolism, growth, and gastrointestinal (2). Epidemiological studies showed that the incidence of Graves' hyperthyroidism in the United States is around 20–40 cases per 100,000 individuals per year (3, 4).

Current treatment for GD patients includes antithyroid drugs (ATDs), radioiodine therapy (^{131}I therapy) (RAI), and thyroidectomy (5). ATDs tend to have a higher failure rate. In addition, approximately half of patients treated with ATDs relapse within 2 years of drug withdrawal and require additional treatment such as RAI (6, 7). Although there was no significant difference in the recurrence of hyperthyroidism between RAI and surgery (8), post-thyroidectomy complications may occur at various incidence rates such as hypocalcemia. RAI is increasingly considered as a safe and inexpensive modality of treatment for GD (9). Although most of the side effects of RAI are mild, many patients will develop early hypothyroidism (HT) within 6 months after RAI. Given that patients with post-RAI HT may require lifelong thyroid hormone replacement therapy, it is essential to identify the GD patients with a high risk of early HT before RAI.

The pathophysiological mechanisms of the development of early HT after RAI are complicated, and the risk factors remain largely unknown. Several potential biomarkers, including thyroid mass and 24-h radioactive iodine uptake (RAIU), had been reported to have predictive values for early HT (10). Some studies also suggested that the female gender was a potential risk factor for HT (11, 12). However, these studies were limited by small sample size, low prognostic accuracy, and lack of external validation. In addition, the development of previous models has typically relied on conventional modeling approaches. Real-world electronic medical records (EMR), including demographic information, diagnoses, medications, laboratory tests, and radiologic findings, have the potential to accelerate data-driven clinical decision support (13). Strategies to enhance the prediction accuracy of the personalized responses to RAI using EMR are urgently needed.

Machine learning, a data analysis technique that develops algorithms to predict outcomes by 'learning' from data, is increasingly emphasized as a competitive alternative to regression analysis (14). Moreover, machine learning

has the potential to outperform conventional regression, possibly by its ability to capture nonlinearities and complex interactions among multiple predictive variables (15). However, to date, the application of machine learning approaches in thyroid-related diseases has primarily focused on the ultrasonic diagnosis or the optimization of levothyroxine dose after thyroidectomy, but prediction of early HT after RAI has not been reported (16, 17).

In this study, we developed a multi-feature predictive model based on the real-world EMR using the combination of multiple machine learning approaches for the prediction of GD patients with the occurrence of early HT within 6 months after RAI. The prognostic accuracy of the developed model was further validated in another cohort, and a clinician-friendly nomogram was then established to compute the probability of post-RAI HT.

Materials and methods

Patients

A total of 840 patients who received their first RAI in our hospital from January 2016 to June 2019 were retrospectively evaluated. The diagnosis of hyperthyroidism was based on the 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis (18). Inclusion criteria include: (i) signs and symptoms of hypermetabolic state such as palpitations and increased perspiration caused by thyrotoxicosis (See Supplementary Table 1, see section on [supplementary materials](#) given at the end of this article); (ii) diffuse enlargement of the thyroid gland (confirmed by physical examination and imaging examination); (iii) low serum TSH concentration and high serum thyroid hormone concentration; (iv) the presence of exophthalmos or pretibial myxedema; and (v) increased ^{131}I rate (RAIU) or enhanced thyroid uptake function revealed by thyroid radionuclide imaging. Exclusion criteria include: (i) patients with hyperthyroidism of other etiology, such as toxic polynodular goiter and solitary toxic adenoma; (ii) patients who did not complete the follow-up at 6 months after RAI; (iii) a history of thyroidectomy; and (iv) a history of treatment with RAI. The study protocol was approved by the Ethics Committee of Changzhi Medical College (IRB approval no.: CZMC-20191026) and conducted in accordance with the World Medical Association Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects. Written consent was obtained from all patients before RAI.

After the assessment, 369 patients who did not meet the inclusion criteria were excluded, and a total of 471 patients were included in our analysis. The GD patients were randomly split into the training set and the validation set by a 2:1 ratio (310 patients in the training set and 161 in the validation set). A total of 138 clinical features and lab test results from the EMR were obtained for the enrolled patients, such as demographics, vital signs, symptoms, lab tests, RAIU, thyroid mass, and radiologic findings.

RAIU determination

Prior to RAIU calculation, all patients were requested to avoid taking any iodized food, drugs, and ATDs for at least 1 week. The patients were given Na¹³¹I solution orally at the dose of 74–370 kBq (2–10 μCi) on the empty stomach and fasted for 2 h. The laboratory background count rate and the standard source count rate were then measured first and then the thyroid radiation count rate was measured at the 3rd, 6th, and 24th h after the patient took the Na¹³¹I solution orally, 60 seconds for each time. The RAIU calculation formula is as follows:

$$\text{RAIU (\%)} = \frac{\text{Thyroid radiation count rate} - \text{background count rate}}{\text{Standard source count rate} - \text{background count rate}} \times 100\% \quad (1)$$

Determination of thyroid mass

Thyroid mass was estimated by thyroid ultrasonography combined with palpation. Briefly, patients lie flat on their back, and their necks were fully exposed. The thyroid size was measured using an ultrasound system by an experienced ultrasound technician. The thyroid mass was then estimated by two experienced nuclear medicine physicians based on ultrasonic data using the following formula, and the adjustment was verified by palpation.

$$W = \left(\frac{A}{6}\right) \times (L \times B \times H) \quad (2)$$

W (g) is the thyroid mass. L (cm), B (cm), and H (cm) are the length, width, and height of the thyroid gland, respectively.

Determination of ¹³¹I dose for RAI therapy

The ¹³¹I dose of RAI was calculated based on the thyroid mass, the radioactive activity per gram thyroid gland was set at 2.59–4.44 MBq (70–120 uCi) (5). The calculation formula is as follows:

$$A = \frac{W \times A_g}{\text{RAIU at 24 hours}} \quad (3)$$

A is the dose of ¹³¹I dose (mCi), W is the thyroid mass (g), A_g is the radioactivity per gram of the thyroid gland (μCi), and RAIU_{24h} is the RAIU at 24 h.

Follow-up and efficacy evaluation

The patients were followed up at 6 months after RAI treatment and divided into three groups according to the laboratory tests and clinical symptoms as follows: (1) The thyroid function was restored to be normal (euthyroidism, the normal group). The symptoms and signs of hyperthyroidism disappeared completely at 6 months after RAI treatment, and serum total triiodothyronine (TT₃), total thyroxine (TT₄), free triiodothyronine (FT₃), and free thyroxine (FT₄) concentration returned to normal. (2) The thyroid function was reduced (HT group). The signs and symptoms of HT occurred at 6 months after RAI treatment. The serum thyroid hormone concentration was lower than the normal, while TSH was higher than the normal. (3) Hyperthyroidism was not relieved at 6 months after RAI treatment, including (i) no improvement or aggravation in symptoms and signs of the patients, and no significant decrease in serum thyroid hormone concentrations; (ii) the symptoms of hyperthyroidism were relieved and signs partially disappeared. Serum TT₃, TT₄, FT₃, and FT₄ were significantly reduced but not restored to the normal concentration; (iii) the symptoms, and signs of hyperthyroidism after post-RAI remission within 6 months, and serum thyroid hormone concentrations increased again after remission.

Data analysis

All general statistical analyses were performed in GraphPad Prism 9.0 unless specified in the methods and legends. Continuous variables with normal distribution are expressed as mean ± s.d., while non-normally distributed continuous variables are expressed as the median and interquartile range (IQR). Categorical variables are expressed as value (%). Machine learning analysis was conducted in R 4.0.3 with the corresponding packages. A decision tree-based ensemble method (random forest, RF) was performed with 1000 trees. The mean decrease accuracy (MDA) scores were calculated. Partial least square discriminant analysis (PLS-DA) was conducted, followed by variable importance in projection (VIP) analysis. VIP ≥ 1.2 was considered statistically significant.

A forward selection and backward elimination approach was then used to determine the final optimal set of features. The performance of the predictive models was assessed using the receiver-operating characteristic curve (ROC) analysis, and 95% CIs for the area under ROC curves (AUROC) were calculated by bootstrap resampling (100 times). Classifier robustness was estimated by permutation tests (1000 times). The sensitivity, specificity, accuracy, positive predictive value (PPV, %), and negative predictive value (NPV, %), were calculated by 2×2 contingency table analyses. The F1 scores and Matthews correlation coefficient (MCC) were calculated in R 4.0.3, which provided complementary information to AUROC, especially with class-imbalanced datasets.

Results

Patient characteristics

As shown in [Table 1](#), 471 GD patients were enrolled in this study, and a total of 322 (68.4%) patients were female. The age of the enrolled patients was 44.9 ± 0.6 years old. The median course of the disease was 12 months (IQR: 3–72). Before RAI treatment, 357 (75.8%) patients had taken ATDs, among which 334 (93.5%) took methimazole orally, 21 (5.9%) took prothiouracil orally, and 2 (0.6%) took herbal medicine orally, with an average duration of 5 months. Among the GD complications before RAI, 105 cases (22.3%) had exophthalmos, 64 cases (13.6%) had hyperthyroid cardiopathy, 141 cases had an abnormal liver function, 157 cases (33.3%) with leukopenia, 52 cases (11.0%) with diabetes, 29 cases (6.2%) with drug allergy, and 23 cases (4.9%) with hypokalemia periodic paralysis. At 6 months after RAI, 122 cases (20.9%) were restored to normal thyroid function, 22 cases (4.7%) with no remission of hyperthyroidism, and 327 cases (69.4%) developed HT.

Predictor importance evaluation using multiple machine learning algorithms

We used multiple machine learning algorithms to rank the importance of baseline clinical features prior to RAI in predicting early HT at 6 months after RAI ([Fig. 1A](#)). Random forest (RF) is a widely used machine learning algorithm for the construction of predictive models due to its resilience to high dimensionality, insensitivity to noise, and robustness to overfitting ([19, 20](#)). We applied RF to identify which variables have more determinant impact on the prediction outcomes. The importance of the variables contributed to the classification between the HT, and the

normal control (NC) group was ranked by MDA scores ([Fig. 1B](#) for the training set and 1C for the validation set). PLS-DA was then used for feature reduction, and partial separation of baseline clinical features was achieved between patients with and without HT at 6 months after RAI in both the training and the validation set ([Fig. 1D and E](#)). The top 15 discriminant features in the PLS-DA model were ranked by their VIP scores ([Fig. 1F and G](#)). Mann–Whitney *U* test was further applied to verify the machine learning results, and the features were selected based on the criteria of MDA >0 , PLS-DA VIP ≥ 1.2 , and Mann–Whitney *U* test, $P < 0.05$ ([Fig. 1H and I](#)).

Table 1 Patient characteristics.

Variables	Values
<i>n</i>	471
Male (%)	149 (31.6%)
Age, mean (s.d.) (Years)	44.9 ± 0.6
Duration of hyperthyroidism (months), median (IQR)	12 (3–72)
Thyroid texture, <i>n</i> (%)	
Hard	172 (36.5%)
Soft	299 (63.5%)
Thyroid nodule, <i>n</i> (%)	
Yes	126 (26.8%)
No	345 (73.2%)
Anti-thyroid drugs (ATDs), <i>n</i> (%)	
Yes	357 (75.8%)
No	114 (24.2%)
Classification of ATDs, <i>n</i> (%)	
Methimazole	334 (93.5%)
Propylthiouracil	21 (5.9%)
Herbal medicine	2 (0.6%)
ATDs duration (months), median (IQR)	5.0 (0.20–35.0)
Presence of eye disease, <i>n</i> (%)	
Yes	105 (22.3%)
No	366 (77.7%)
Hyperthyroid cardiopathy, <i>n</i> (%)	
Yes	64 (13.6%)
No	407 (86.4%)
Abnormal liver function, <i>n</i> (%)	
Yes	141 (29.9%)
No	330 (70.1%)
Leucopenia, <i>n</i> (%)	
Yes	157 (33.3%)
No	314 (66.7%)
Diabetes, <i>n</i> (%)	
Yes	52 (11.0%)
No	419 (89.0%)
Drug allergy, <i>n</i> (%)	
Yes	29 (6.2%)
No	442 (93.8%)
Hypokalemia periodic paralysis, <i>n</i> (%)	
Yes	23 (4.9%)
No	448 (95.1%)

Data are mean \pm s.d., *n* (%), or median and interquartile range (IQR).

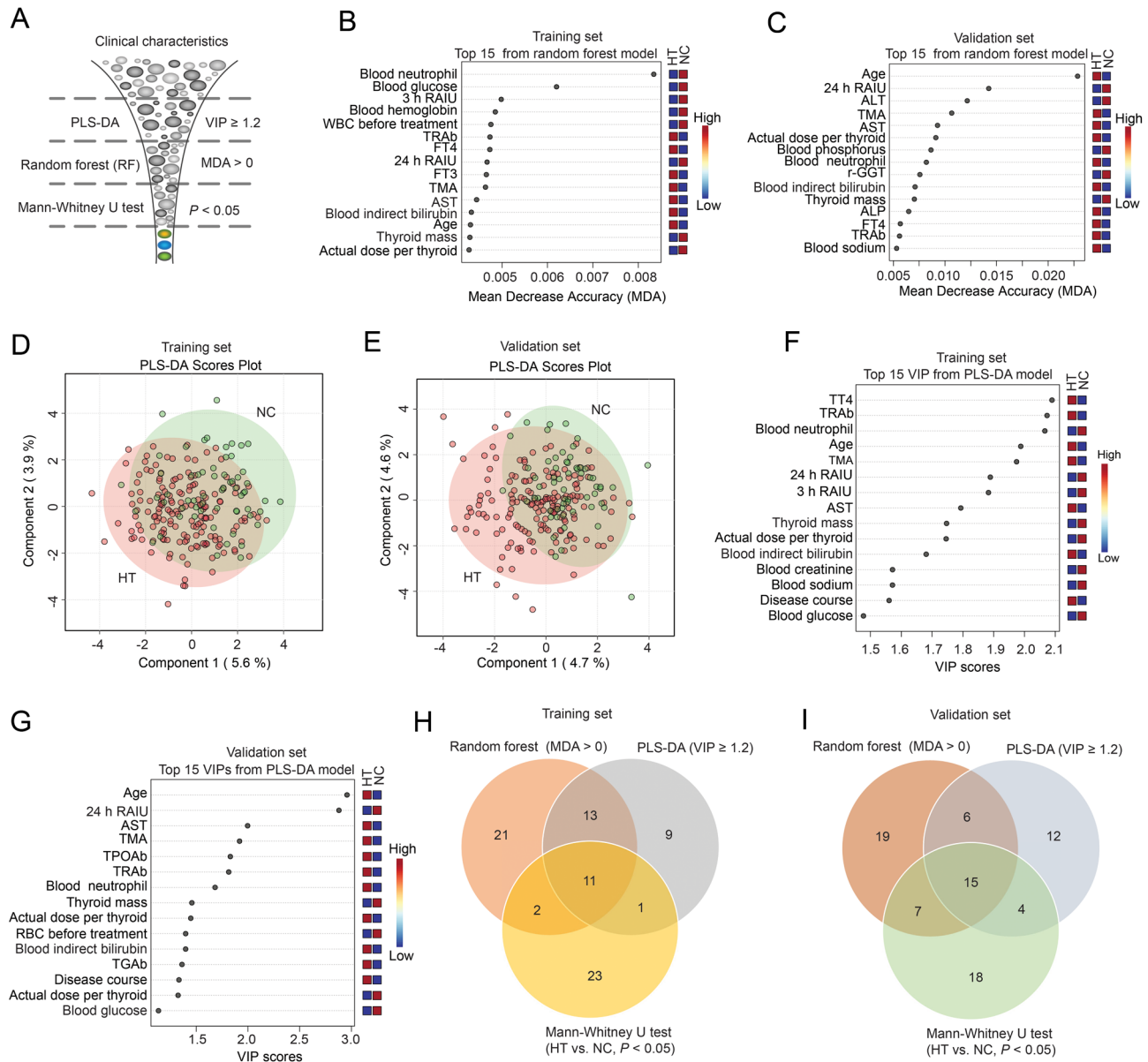


Figure 1

The selection of baseline clinical features from EMR data using multiple machine learning methods for predicting future early hypothyroidism at 6 months after RAI. (A) The flow chart of feature selection using the combinations of multiple machine learning methods. (B) Top 15 features ranked by MDA scores in the random forest model for classification between the HT group ($n = 222$) and the NC group ($n = 98$) in the training set. (C) Top 15 features ranked by MDA scores in the random forest model for classification between the HT group ($n = 105$) and the NC group ($n = 56$) in the validation set. (D) PLS-DA showed the partial separation of baseline EMR data between the HT group ($n = 222$) and the NC group ($n = 98$) in the training set. (E) PLS-DA 2-D plot for the separation of baseline EMR data in the HT group ($n = 105$) from that in the NC group ($n = 56$) in the validation set. (F) The top 15 differential features between the HT group ($n = 222$) and the NC group ($n = 98$) ranked by VIP scores in the training set. (G) The top 15 differential features between the HT group ($n = 105$) and the NC group ($n = 56$) ranked by VIP scores in the validation set. (H) The selection of optimal clinical features in the training set. (I) The selection of optimal clinical features in the validation set. EMR, electronic medical record; RAI, radioiodine therapy; MDA, mean decrease accuracy; HT, hypothyroidism; NC, the normal control group; PLS-DA, partial least squares discriminant analysis; VIP score, variable Importance in projection score.

Assessment of the performance of the developed multi-feature model for predicting future early hypothyroidism at 6 months after RAI

Figure 2A, B, C, D, E, F and G show the group differences (HT vs NC) of the selected seven features in the predictive model for evaluating the risk of HT in the training set ($n = 310$ patients). We found that age, TRAb, thyroid microsomal antibodies (TMA), and aspartate aminotransferase (AST) were significantly higher for GD patients with post-RAI HT than those in the normal group without HT after I-131 therapy (Fig. 2A, B, C and D). In contrast, the pre-RAI concentrations of 24-h RAIU, thyroid mass, and blood neutrophil count in GD patients with post-RAI HT were dramatically lower compared with those in the normal group (Fig. 2E, F and G). The performance of the single and multi-feature models in the training set was assessed using the ROC analysis and confusion matrices. Here, 24-h RAIU alone had low accuracy, sensitivity, NPV, PPV, F1 score, and MCC score (Table 2). Surprisingly, an integrated model with the combination of seven features before I-131 therapy performed well for predicting future early HT at 6 months after RAI with an AUROC of 0.72 (95% CI:

0.61–0.85) (Fig. 2H and I). The predictive accuracy, sensitivity, and specificity of the developed multi-feature model in the training set were 71.2% (95% CI: 61.9–81.3%), 74.5% (95% CI: 68.3–84.8%), and 76.4% (95% CI: 67.8–85.1%), respectively (Table 2). In addition, the developed multi-feature model yielded a PPV of 72.9% (95% CI: 65.1–79.5%), a NPV of 69.6% (95% CI: 51.4–83.2%), an F1 score of 0.77, and an MCC score of 0.65. The classifier robustness was further estimated by permutation tests (1000 times) with a P value of 0.01 (Fig. 2I).

Validation of the developed multi-feature model

The developed multi-feature predictive model was then validated in the validation set of another 161 GD patients (105 in the HT group and 56 in the NC group). Consistently, high pre-RAI age, TRAb, TMA, and AST in GD patients tend to develop HT within 6 months (Fig. 3A, B, C and D). Similar to the training set, we also observed significantly lower concentrations of 24-h RAIU, thyroid mass, and blood neutrophil in GD patients with HT after RAI compared with those in the NCs (Fig. 3E, F and G).

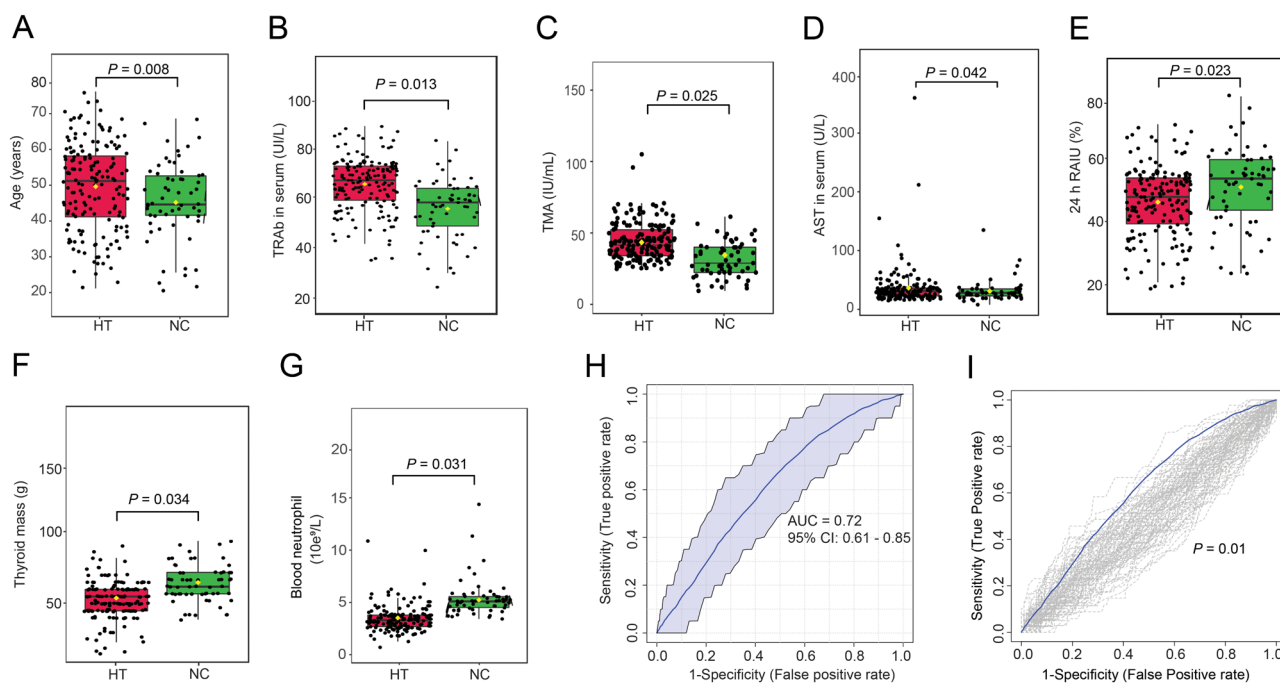


Figure 2

The selected clinical features in the developed multivariate model and the predictive performance in the training set evaluated by the ROC curve analysis. (A, B, C, D, E, F, and G) The significant differences of the selected pre-RAI clinical features between GD patients with post-RAI hypothyroidism (the HT group, $n = 222$) and the normal controls after RAI (the NC group, $n = 98$) in the training set. Mann-Whitney U test was performed to verify the results of machine learning. (H) Evaluation of the selected pre-RAI features for predicting post-RAI hypothyroidism in the training set. (I) A permutation test (1000 times) for the cross-validation of the ROC curve in the training set. RAI, radioiodine therapy; ROC, receiver operator characteristic curve; HT, hypothyroidism; NC, the normal control group; AST, aspartate aminotransferase; TRAb, thyrotropin-receptor antibodies; 24-h RAIU, radioactive iodine uptake at 24 h; TMA, thyroid microsomal antibodies; HT, hypothyroidism; NC, the normal control group.

Table 2 The predictive accuracy of the developed model for the development of early hypothyroidism at 6 months after radioiodine therapy.

Prediction models	Prediction model performance (mean and 95% CI)						
	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1 score	MCC
24-h RAIU	53.5 (48.7–58.2)	25.4 (19.9–31.5)	50.1 (31.9–68.1)	62.1 (53.1–70.4)	51.1 (48.7–53.2)	0.36	0.11
Multi-feature model (training set)	71.2 (61.9–81.3)	74.5 (68.3–84.8)	76.4 (67.8–85.1)	72.9 (65.1–79.5)	69.6 (51.4–83.2)	0.77	0.65
Multi-feature model (validation set)	74.4 (65.3–83.0)	70.8 (65.4–79.8)	73.5 (68.7–82.6)	75.7 (68.5–84.8)	77.3 (65.5–86.9)	0.74	0.63

Values calculated by standard 2 × 2 contingency table analysis. Multi-feature model includes patients' age, thyroid mass, 24-h RAIU, serum AST, TRAb, TMA, and blood neutrophil count.

24-h RAIU, radioactive iodine uptake at 24 h; AST, aspartate aminotransferase; MCC, Matthews correlation coefficient; NPV (%), negative predictive value; PPV (%), positive predictive value; TMA, thyroid microsomal antibodies; TRAb, thyrotropin-receptor antibodies.

An AUROC of 0.74 (95% CI: 0.65–0.83, permutation $P=0.03$) was obtained using the developed multi-feature model in the validation set. Additionally, the developed model had an accuracy of 71.4% (95% CI: 65.3–83.0), a sensitivity of 70.8% (95% CI: 65.4–79.8), a specificity of 73.5% (95% CI: 68.7–82.6), a PPV of 75.7% (95% CI: 68.5–84.8), a NPV of 77.3% (95% CI: 65.5–86.9), an F1 score of 0.74, and a MCC of 0.63 for predicting the risk of future development of HT after ¹³¹I therapy in the validation set (Table 2).

Construction of the nomogram

There is still no consensus regarding the most appropriate approach for treating Graves' hyperthyroidism with radioiodine. To facilitate the clinical utility, the features in the developed predictive model were then translated into a nomogram (Fig. 4). The nomogram maps the predicted risk probabilities into element points on a scale from 0 to 100 in a user-friendly graphical interface. The total points

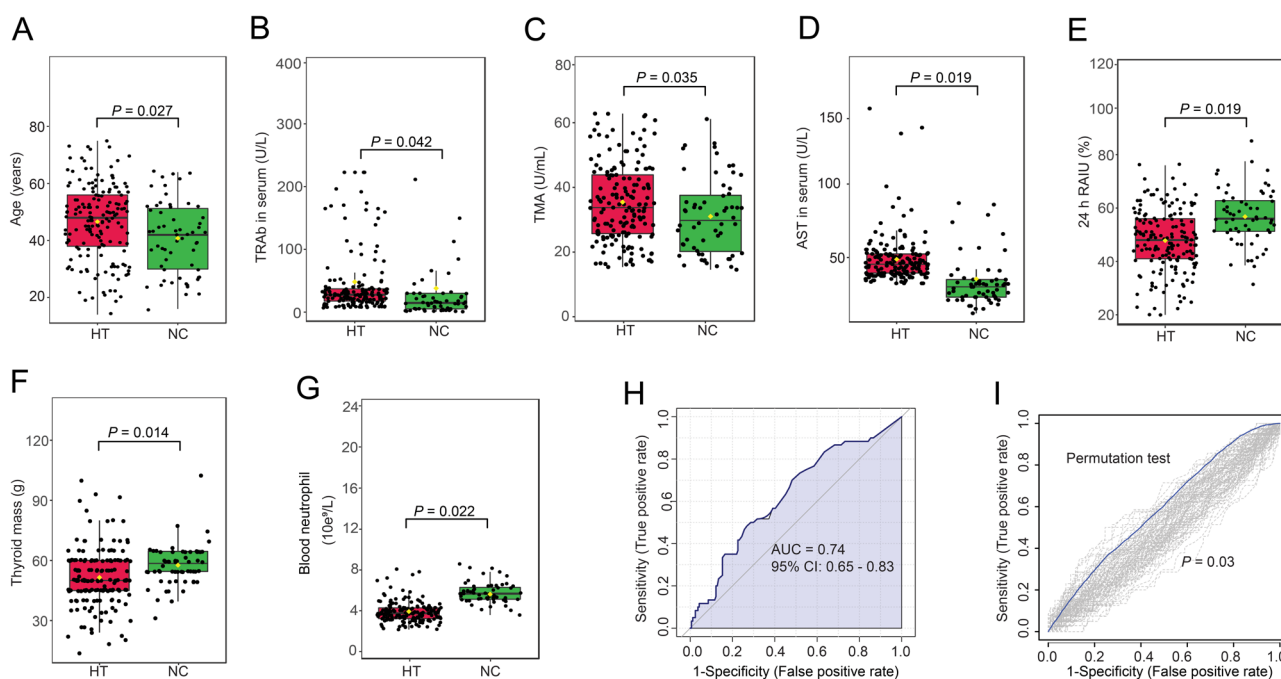


Figure 3

The selected clinical features for predicting post-RAI hypothyroidism in the developed multivariate model and the predictive performance in the validation set evaluated by the ROC curve analysis. (A, B, C, D, E, F, and G) The significant differences of the selected pre-RAI clinical features between GD patients with post-RAI hypothyroidism (HT group, $n = 105$) and the normal controls after RAI (NC group, $n = 56$) in the validation set. Mann-Whitney U test was performed to verify the results of machine learning. (H) The ROC curve for predicting post-RAI hypothyroidism in the validation set. (I) A permutation test (1000 times) for the cross-validation of the ROC curve in the validation set. RAI, radioiodine therapy; ROC, receiver operator characteristic curve; HT, hypothyroidism; NC, the normal control group; AST, aspartate aminotransferase; TRAb, thyrotropin-receptor antibodies; 24-h RAIU, radioactive iodine uptake at 24 h; TMA, thyroid microsomal antibodies.

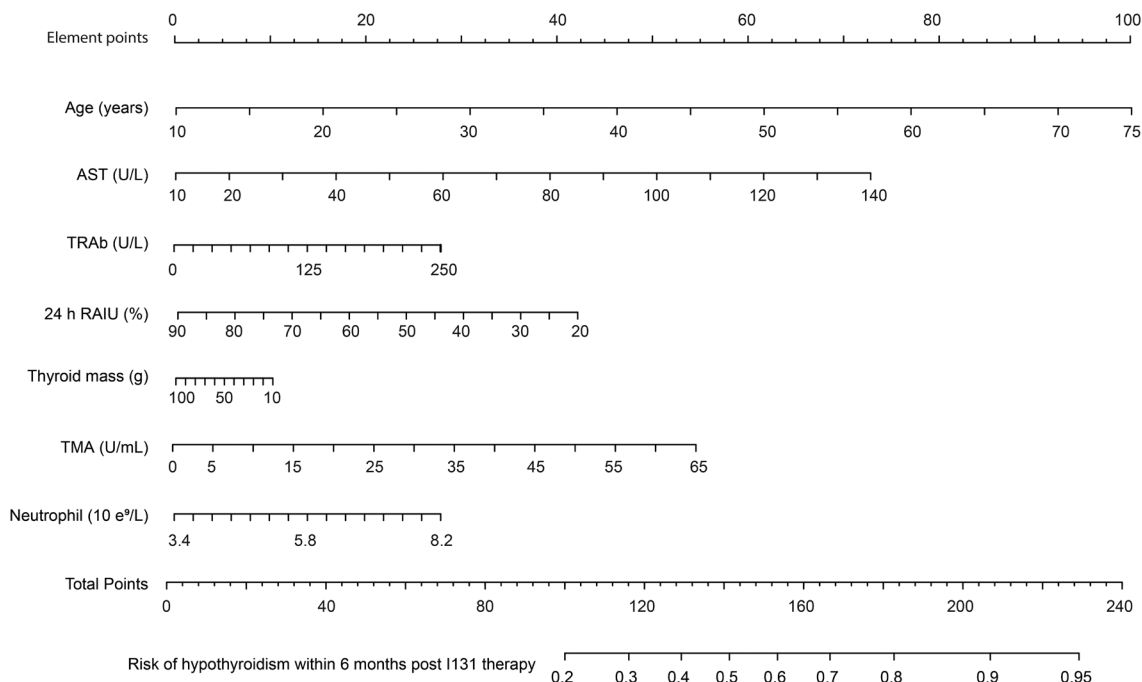


Figure 4

The nomogram for predicting post-RAI hypothyroidism using the clinical features and lab tests results before the therapy. RAI, radioiodine therapy; AST, aspartate aminotransferase; TRAb, thyrotropin-receptor antibodies; 24-h RAIU, radioactive iodine uptake at 24 h; TMA, thyroid microsomal antibodies.

accumulated by all variables in the nomogram correspond to the predicted probability of HT risk within 6 months after RAI.

Discussion

RAI has become one of the most economical and convenient treatment approaches for the treatment of GD (18). However, the potential occurrence of HT after RAI, which requires lifelong drug replacement, limits the clinical use of RAI (16). In this study, we successfully developed a multivariate model to predict HT after RAI using multiple machine learning methods and the baseline clinical characteristics before RAI. The clinical features in the multivariate model, include age, AST, TRAb, 24-h RAIU, thyroid mass, TMA, and blood neutrophil count. If replicated, this finding could have a significant impact on clinical practice by permitting the patients to be stratified according to their future risk of HT prior to RAI and new preventive treatments to be tested in clinical trials.

HT following RAI may induce deleterious symptoms and affect the quality of life such as memory impairment, cognitive decline, fatigue, depressed mood, or heart failure (21). Therefore, it is crucial to prevent HT in advance, while minimizing the risk of hyperthyroidism relapse.

In the context of precision medicine, our study will help developing a personalized treatment approach for RAI therapy. The overall principle is to maximize the benefits of RAI for patients with hyperthyroidism. Eventually, patients have no symptoms of thyroid dysfunction. The thyroid hormone concentrations were normalized and patients could maintain their own hypothalamic-pituitary-thyroid axis hormone concentration regulation without taking any ATDs.

GD is not only a common thyroid disorder but also an autoimmune disease. TRAb is an autoimmune antibody that stimulates the TSH receptor on thyroid cells, leading to increased secretion and synthesis of thyroid hormones. TMA is a complement binding antibody induced by microsomal antigen in the cytoplasm of thyroid epithelial cells. TRAb and TMA can be detected in the serum of more than 85% of GD patients, and elevated concentration of these antibodies are closely associated with the pathogenesis of GD (22). A previous study found that TRAb was an independent risk factor affecting the prognosis of the RAI treatment, and GD patients with positive TRAb before treatment had a lower remission rate after RAI treatment (23). Here, we demonstrated for the first time that TRAb and TMA may also play a predictive role in the occurrence of early HT after ¹³¹I therapy for Graves' hyperthyroidism.

Twenty-four-hour RAIU is regarded as an important factor with the contradictory results for its association with the outcome of radioiodine therapy. Our study found that patients with lower 24-h RAIU before RAI are more vulnerable to develop early HT after RAI. Consistent with our study, Damle and the co-authors observed a similar negative association between 24-h RAIU and the occurrence of HT at 12 months post-therapy (24). One explanation might be related to the method used for the calculation of the ^{131}I dose. When RAIU at 24 h of GD patients was lower, the calculated ^{131}I dose was often higher. Another possibility is that the GD patients with lower RAIU may have a lower degree of hyperthyroidism, and the possibility of HT after ^{131}I treatment is relatively high.

In this study, we found that thyroid quality is an important risk factor for post-RAI HT. Patients with low thyroid mass had a higher incidence of developing HT, which was consistent with the previous observation (25). In addition, we found that the softer the thyroid gland, the more prone to develop HT. It is likely that the hard thyroid gland is less sensitive to radiation.

Age is another important influencing factor for the outcome of RAI, and older patients tend to develop early HT than younger ones. This observation was opposite to two previous reports (26, 27). The discrepancy might be due to several reasons, such as the small size in previous studies, different age ranges, and complications. Further studies are warranted to investigate the effects of aging on the responses to RAI therapy.

The blood neutrophil count is a simple and cost-effective indicator of the host immune system and a useful prognostic marker in many inflammatory diseases. A recent study reported that a high neutrophil-to-lymphocyte (NLR) ratio is associated with the relapse in Graves' disease after antithyroid drug therapy (28). Beyond the literature, here, we showed that a low neutrophil count predicts the occurrence of HT after RAI. Different mechanisms could explain the potential relationship between Graves' HT and neutropenia. It is likely that dysregulated thyroid hormones inhibit the metabolism and proliferation of hematopoietic progenitor cells (29). Another mechanism of the reduction in circulating neutrophils could be due to the immune-mediated reduced survival of peripheral neutrophils (30).

This study had some limitations. First, this is a single-center study with a relatively small sample size (471 patients in 3 years). Additional validation should be considered in a more diverse demographic group and larger sample size. Secondly, the outcomes of machine learning analysis were not verified by conventional regression analysis model. Thirdly, the data were analyzed retrospectively using EMR

data, which can be plagued by missingness, inaccuracies, and changes in practice patterns over time. However, our study confirmed the utility of EMR data in a real-world clinical setting, showing good performance in predicting RAI outcomes even in the presence of missing data.

Conclusion

This study demonstrates the feasibility of multiple machine learning algorithms for analysis of EMR and lab tests data to accurately predict a early occurrence of HT 6 months after RAI. We have also developed a prognostic nomogram to facilitate the clinical utility. The new model might be useful for medical professionals to develop further treatment options and long-term follow-up plans for GD patients undergoing RAI.

Supplementary materials

This is linked to the online version of the paper at <https://doi.org/10.1530/EC-22-0119>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This project was funded by the Natural Science Foundation of Shanxi Province (201901D221113), Science and Technology Innovation Project of Shanxi Universities (2020L0374), and Doctoral Startup Fund of Changzhi Medical College (BS201905). The funder had no role in the design, execution, interpretation, or writing of the study (2020L0374).

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Changzhi Medical College (IRB approval: CZMC-20191026).

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Data availability

Data supporting the findings of this study are available from the corresponding authors upon request.

Author contribution statement

Conceptualization: J L, Y W, and K L; Writing – original draft: H Y Z, Y W, and K L; Writing – review and editing: L D, H Y Z, M L, H Z, Y C, T W, Y L, J L, K L; Methodology: L D, H Y Z, L V, Y W, K L; Data curation: L D, H Y Z, M L, H Z, Y C, T W, Y L, J L.

References

- Yang F, Teng W, Shan Z, Guan H, Li Y, Jin Y, Hu F, Shi X, Tong Y, Chen W, *et al.* Epidemiological survey on the relationship between different iodine intakes and the prevalence of hyperthyroidism. *European Journal of Endocrinology* 2002 **146** 613–618. (<https://doi.org/10.1530/eje.0.1460613>)
- van der Vaart J, Bosmans L, Sijbesma SF, Knoops K, van de Wetering WJ, Otten HG, Begthel H, Borel Rinkes IHM, Korving J, Lentjes EGWM, *et al.* Adult mouse and human organoids derived from thyroid follicular cells and modeling of Graves' hyperthyroidism. *PNAS* 2021 **118** e2117017118. (<https://doi.org/10.1073/pnas.2117017118>)
- Smith TJ & Hegedus L. Graves' disease. *New England Journal of Medicine* 2016 **375** 1552–1565. (<https://doi.org/10.1056/NEJMra1510030>)
- Hussain YS, Hookham JC, Allahabadia A & Balasubramanian SP. Epidemiology, management and outcomes of Graves' disease-real life data. *Endocrine* 2017 **56** 568–578. (<https://doi.org/10.1007/s12020-017-1306-5>)
- Jiang NY, Lin YS, Guan H, Tan J, Li L, Gao Z, Lu H, Wu Y, Guan M, Yuan WH, *et al.* 131I guidelines for treating Graves' disease (2013 edition). *Labeled Immunoassays and Clinical Medicine* 2014 **21** 92–104. (<https://doi.org/10.3760/cma.j.issn.2095-2848>)
- Burch HB & Cooper DS. Management of Graves disease: a review. *JAMA* 2015 **314** 2544–2554. (doi: [10.1001/jama.2015.16535](https://doi.org/10.1001/jama.2015.16535))
- Sjolin G, Holmberg M, Torring O, Bystrom K, Khamisi S, de Laval D, Abraham-Nordling M, Calissendorff J, Lantz M, Hallengren B, *et al.* The long-term outcome of treatment for Graves' hyperthyroidism. *Thyroid* 2019 **29** 1545–1557. (<https://doi.org/10.1089/thy.2019.0085>)
- Sundaresh V, Brito JP, Wang Z, Prokop LJ, Stan MN, Murad MH & Bahn RS. Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 3671–3677. (<https://doi.org/10.1210/jc.2013-1954>)
- Sundaresh V, Brito JP, Thapa P, Bahn RS & Stan MN. Comparative effectiveness of treatment choices for Graves' hyperthyroidism: a historical cohort study. *Thyroid* 2017 **27** 497–505. (<https://doi.org/10.1089/thy.2016.0343>)
- Hu RT, Liu DS & Li B. Predictive factors for early hypothyroidism following the radioactive iodine therapy in Graves' disease patients. *BMC Endocrine Disorders* 2020 **20** 76. (<https://doi.org/10.1186/s12902-020-00557-w>)
- Boelaert K, Syed AA, Manji N, Sheppard MC, Holder RL, Gough SC & Franklyn JA. Prediction of cure and risk of hypothyroidism in patients receiving 131I for hyperthyroidism. *Clinical Endocrinology* 2009 **70** 129–138. (<https://doi.org/10.1111/j.1365-2265.2008.03291.x>)
- Xing YZ, Zhang K & Jin G. Predictive factors for the outcomes of Graves' disease patients with radioactive iodine. *Bioscience Reports* 2020 **40** BSR20191609. (<https://doi.org/10.1042/BSR20191609>)
- Ayala Solares JR, Diletta Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, Pinho Gomes AC, Payberah AH, Zottoli M, Nazarzadeh M, *et al.* Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics* 2020 **101** 103337. (<https://doi.org/10.1016/j.jbi.2019.103337>)
- Miller DD & Brown EW. Artificial intelligence in medical practice: the question to the answer? *American Journal of Medicine* 2018 **131** 129–133. (<https://doi.org/10.1016/j.amjmed.2017.10.035>)
- Dimopoulos AC, Nikolaidou M, Caballero FF, Engchuan W, Sanchez-Niubo A, Arndt H, Ayuso-Mateos JL, Haro JM, Chatterji S, Georgousopoulou EN, *et al.* Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Medical Research Methodology* 2018 **18** 179. (<https://doi.org/10.1186/s12874-018-0644-1>)
- Seib CD, Chen J & Iagaru A. Shifting trends and informed decision-making in the management of Graves' disease. *Thyroid* 2020 **30** 351–354. (<https://doi.org/10.1089/thy.2020.0114>)
- Chen SS, Zaborek NA, Doubleday AR, Schaefer SC, Long KL, Pitt SC, Sippel RS & Schneider DE. Optimizing levothyroxine dose adjustment after thyroidectomy with a decision tree. *Journal of Surgical Research* 2019 **244** 102–106. (<https://doi.org/10.1016/j.jss.2019.06.025>)
- Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, Rivkees SA, Samuels M, Sosa JA, Stan MN, *et al.* 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* 2016 **26** 1343–1421. (<https://doi.org/10.1089/thy.2016.0229>)
- Guo X, Li Y, Li H, Li X, Chang X, Bai X, Song Z, Li J & Li K. An improved multivariate model that distinguishes COVID-19 from seasonal flu and other respiratory diseases. *Aging* 2020 **12** 19938–19944. (<https://doi.org/10.18632/aging.104132>)
- Cui S, Li L, Zhang Y, Lu J, Wang X, Song X, Liu J & Li K. Machine learning identifies metabolic signatures that predict the risk of recurrent angina in remitted patients after percutaneous coronary intervention: a multicenter prospective cohort study. *Advanced Science* 2021 **8** 2003893. (<https://doi.org/10.1002/advs.202003893>)
- Taieb D, Bournaud C, Eberle MC, Catargi B, Schwartz C, Cavarec MB, Faugeron I, Toubert ME, Benisvy D, Archange C, *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. *European Journal of Endocrinology* 2016 **174** 491–502. (<https://doi.org/10.1530/EJE-15-1099>)
- Kamath C, Adlan MA & Premawardhana LD. The role of thyrotrophin receptor antibody assays in Graves' disease. *Journal of Thyroid Research* 2012 **2012** 525936. (<https://doi.org/10.1155/2012/525936>)
- Zheng W, Jian T, Guizhi Z, Zhao M & Ren W. Analysis of ¹³¹I therapy and correlation factors of Graves' disease patients: a 4-year retrospective study. *Nuclear Medicine Communications* 2012 **33** 97–101. (<https://doi.org/10.1097/MNM.0b013e32834d3bb9>)
- Damle N, Bal C, Kumar P, Reddy R & Virkar D. The predictive role of 24 h RAIU with respect to the outcome of low fixed dose radioiodine therapy in patients with diffuse toxic goiter. *Hormones* 2012 **11** 451–457. (<https://doi.org/10.14310/horm.2002.1377>)
- Szumowski P, Abdelrazek S, Sawicka AK, Mojsak M, Kostecki J, Sykała M & Myśliwiec J. Radioiodine therapy for Graves' disease – retrospective analysis of efficacy factors. *Endokrynologia Polska* 2015 **66** 126–131. (<https://doi.org/10.5603/EP.2015.0019>)
- Yang D, Xue J, Ma W, Liu F, Fan Y, Rong J, Yang A & Yu Y. Prognostic factor analysis in 325 patients with Graves' disease treated with radioiodine therapy. *Nuclear Medicine Communications* 2018 **39** 16–21. (<https://doi.org/10.1097/MNM.0000000000000770>)
- Fanning E, Inder WJ & Mackenzie E. Radioiodine treatment for Graves' disease: a 10-year Australian cohort study. *BMC Endocrine Disorders* 2018 **18** 94. (<https://doi.org/10.1186/s12902-018-0322-7>)
- Kim M, Kim BH, Jang MH, Kim JM, Kim EH, Jeon YK, Kim SS & Kim JJ. High neutrophil-to-lymphocyte ratio is associated with relapse in Graves' disease after antithyroid drug therapy. *Endocrine* 2020 **67** 406–411. (<https://doi.org/10.1007/s12020-019-02137-y>)
- Ahmed SS & Mohammed AA. Effects of thyroid dysfunction on hematological parameters: case controlled study. *Annals of Medicine and Surgery* 2020 **57** 52–55. (<https://doi.org/10.1016/j.amsu.2020.07.008>)
- Scappaticcio L, Maiorino MI, Maio A, Esposito K & Bellastella G. Neutropenia in patients with hyperthyroidism: systematic review and meta-analysis. *Clinical Endocrinology* 2021 **94** 473–483. (<https://doi.org/10.1111/cen.14313>)

Received in final form 10 April 2022

Accepted 22 April 2022

Accepted Manuscript published online 22 April 2022