Journal of the Endocrine Society, 2022, **6**, 1–11
https://doi.org/10.1210/jendso/bvac023
Advance access publication 16 February 2022
Clinical Research Article

The Prediction Model Using Thyroid-stimulating Immunoglobulin Bioassay For Relapse of Graves' Disease

Han-Sang Baek,[1] Jaejun Lee,[5] Chai-Ho Jeong,[2] Jeongmin Lee,[4] Jeonghoon Ha,[1] Kwanhoon Jo,[3] Min-Hee Kim,[4] Jae Hyoung Cho,[1] Moo Il Kang,[1] and Dong-Jun Lim[1]

[1]Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[2]Division of Endocrinology and Metabolism, Department of Internal Medicine, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[3]Division of Endocrinology and Metabolism, Department of Internal Medicine, Incheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[4]Division of Endocrinology and Metabolism, Department of Internal Medicine, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[5]Department of Internal Medicines, Armed Forces Goyang Hospital, Goyang, Republic of Korea

Correspondence: Dong-Jun Lim, MD, PhD, Division of Endocrinology and Metabolism, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea. E-mail: ldj6026@catholic.ac.kr.

Abstract

Objective: Thyroid-stimulating immunoglobulin (TSI) bioassay has a better ability to predict the relapse rate of Graves' disease (GD) than the thyroid-stimulating hormone (TSH)-binding inhibitory immunoglobulin method in terms of measuring the TSH receptor antibody. However, the optimal TSI bioassay cutoff for predicting relapse after antithyroid drug (ATD) withdrawal is not well evaluated.

Methods: This retrospective study enrolled GD patients who had been treated with ATD and obtained their TSI bioassay <140% from January 2010 to December 2019 in a referral hospital.

Results: Among 219 study subjects, 86 patients (39.3%) experienced relapse. The TSI bioassay value of 66.5% significantly predicted the relapse of GD ($P = 0.049$). The group with a TSI bioassay value > 66.5% were expected to show a 23.8% relapse rate at 2 from ATD withdrawal, and the group with a TSI < 66.5% had a 12.7% relapse rate based on Kaplan-Meier curves analysis. The TSI bioassay showed a good ability to predict relapse GD in the female group ($P = 0.041$) but did not in the male group ($P = 0.573$). The risk scoring based on the nomogram with risk factors for GD relapse, which was constructed to overcome the limitation, increased the predictive ability of GD relapse by 11.5% compared to the use of the TSI bioassay alone.

Conclusions: The cutoff value of the TSI bioassay to predict GD relapse should be lower than that for diagnosing GD. However, as the single use of the TSI bioassay has limitations, a nomogram with multiple risk factors including TSI bioassay could be helpful to predict GD relapse.

Key Words: Graves' disease, immunoglobulins, thyroid-stimulating, recurrence, nomograms

Graves' disease (GD) is one of the most well-known autoimmune thyroid diseases [1]. It is well known that the thyrotropin receptor antibody (TSH-R-Ab) plays an important role in the pathogenesis of GD by causing thyroid stimulation and inducing hyperthyroidism [2]. However, this TSH-R-Ab has a different action from the thyroid-stimulating hormone (TSH) receptor: stimulation or blocking [1-3]. Stimulating TSH-R-Ab activates the 3′,5′-cyclic adenosine 5′-monophosphate pathway to stimulate the TSH receptor, thus inducing thyroid growth and increasing thyroid hormone production [2, 4]. On the other hand, blocking TSH-R-Ab acts as an antagonist to the TSH receptor [2, 3].

There are 2 assays for TSH-R-Ab detection: the competitive thyrotropin-binding inhibitory immunoglobulin (TBII) assay and the thyroid stimulatory immunoglobulin (TSI) bioassay [1]. Immunoglobulins that inhibit the binding of radiolabeled TSH to the TSH receptor could be detected by the TBII assay [5, 6]. The problem is that this assay measures thyroid-blocking immunoglobulins as well as TSIs [6]. On the contrary, the TSI bioassay could differentiate between stimulating TSH-R-Ab and blocking TSH-R-Ab [7, 8]. The TSI bioassay can measure the 3′,5′-cyclic adenosine 5′-monophosphate produced when TSI stimulates the TSH receptor [5]. Although the TBII assay has limitations, TBII offers an accurate diagnosis of GD, and the TSI bioassay is predictive of extrathyroidal manifestations. [6, 9, 10].

For the treatment of GD, there are 3 options: surgery, radioactive iodine treatment (RAI), or antithyroid drug (ATD) [4, 10]. While surgery or RAI treats GD by destroying thyroid tissue, ATD inhibits the synthesis of thyroid hormone to treat GD without destroying the thyroid structure. This is an advantage of ATD and a limitation simultaneously; the relapse from remaining thyroid tissue is always a concern [11]. According to previous studies, the relapse rate after ATD withdrawal almost approached 50% [12, 13]. In addition, many clinical factors such as male sex, younger age, smoking,

Received: 3 November 2021. **Editorial Decision:** 10 February 2022. **Corrected and Typeset:** 15 April 2022

severe hyperthyroidism, large goiter, and orbitopathy are associated with a high relapse rate [14]. In addition, there is debate about ATD use during pregnancy because it could be harmful for embryonic development [15].

Furthermore, TSH-R-Ab levels showed a good ability to predict relapse and disease course in previous studies [16, 17]. In these studies, the TBII assay was used to measure TSH-R-Abs. Because it measures both stimulating and blocking antibodies, the TSI bioassay method appeared to be more accurate in predicting the course of disease [18, 19]. Kwon et al showed that the TSI bioassay could better predict relapse after withdrawal from ATD [20]. However, they did not measure 2 assays (TBII and TSI bioassay) simultaneously in 1 person and used a predetermined cutoff point of the TSI bioassay derived from the diagnosis of GD, not based on the prognosis of GD. Because they only used the positivity of the assay without quantitative measurement, the exact cutoff value to predict relapse was difficult to find.

Although the TSI bioassay has a better ability to predict relapse of GD, it is not known whether the TSI bioassay cutoff value for diagnosing GD and predicting relapse is the same. Therefore, in this study, we tried to achieve the optimal TSI bioassay cutoff value to predict relapse after withdrawal from ATD in patients with the results of 2 assays. Furthermore, we tried to make a prediction model with confounding factors for the relapse of GD.

Methods

This study adhered to the tenets of the Declaration of Helsinki and was approved by the institutional review board of the Seoul St. Mary's Hospital (Seoul, Korea) (KCR21RASI0731). Permission to use hospital data was granted by the institutional review board of Seoul St. Mary's Hospital. Due to the retrospective nature of the study, the requirement to obtain informed consent was waived by the institutional review board of St. Mary's Hospital.

We tried to find patients whose physician stopped ATD after checking that the TSI bioassay was <140%, which is the TSI bioassay cutoff value to diagnose GD in our hospital [21]. For this reason, among 1831 patients who had TSI bioassay results and underwent ATD therapy between 2010 and 2019, 1160 were excluded because the value of the TSI bioassay was >140%. The remaining 671 medical records were reviewed. Patients who did not have follow-up data (n = 102) or whose start date for ATD is unknown (n = 98) and whose final diagnosis is not GD (n = 66) were excluded. Patients who never stopped ATD or whose period from TSI bioassay measurement to ATD stop was >3 months were excluded (n = 158). So, only patients with a TSI bioassay result <3 months before ATD was stopped were included for the analysis. One patient was excluded due to the lack of data from the TBII assay. Patients who received thyroid surgery or RAI were also excluded (n = 3). All patients had both the TSI bioassay result and the TBII result. The patient selection flow chart is summarized in Figure 1.

ATD withdrawal was defined as the patient not taking ATD for at least 3 months. Patients who stopped ATD were divided into relapse (R) and nonrelapse (NR) groups. Relapse was defined as biochemical or clinical hyperthyroidism leading the clinician to restart ATD, free thyroxine (fT4) higher than the upper normal limit or TSH lower than the lower normal limit, and newly developed GD symptoms such as palpitation and sweating. If the patient did not show a relapse of the disease at least 1 year after stopping ATD, it was considered nonrelapse even if there were no data after the last follow-up data [22-24]. Patients who stopped ATD for <3 months in the relapse group [5], and patients who had a shorter follow-up duration than 1 year in the NR group [19] were excluded. Finally, 219 remaining patients (158 women and 61 males) were analyzed.

Graves' orbitopathy was defined when the study subjects had a medical history of the diagnosis of it by the ophthalmologist. If there was a prescription for levothyroxine during ATD treatment, it was considered concomitant T4 replacement.

Laboratory Tests

fT4 and TSH were performed in 2 ways: (1) using the Beckman Coulter immunoradiometric assay (IRMA) kit (Immunotech, Prague, Czech Republic) and (2) using the ADIVA Centaur electrochemiluminescence immunoassay kit (ECLIA; Siemens Healthcare Diagnostic Inc. Munich, Germany). The normal ranges were as follows: TSH of 0.55 to 4.78 uIU/mL in ECLIA and 0.17 to 4.05 IU/mL in IMRA and fT4 of 0.89 to 1.76 ng/dL in ECLIA and 0.89 to 1.79 ng/dL in IRMA. (RRID: AB_2895179, AB_2895183 in ECLIA and AB_2895185, AB_2895187 in IRMA)

TBII was measured in 2 ways. First, the Elecsys/Cobas electrochemiluminescence immunoassay kit (Roche Diagnostics, Mannheim, Germany) is a third-generation TBII assay. This assay measures the inhibition of binding of the labeled monoclonal antibody clone M22 to the TSH receptor with a positive value >1.75 IU/L (RRID:AB_2801453). Second, TRAK human radioimmunoassay (RAI) kit (BRAHMS Thermo Scientific, Henningsdorf, Germany) is a second-generation TBII assay. Detection is based on the ability of TBII to prevent the binding of labeled TSH to the TSH receptor with a positive value >1.5 IU/L.

The TSI titer was measured by the Thyretain™ TSI reporter bioassay (Diagnostic Hybrids, Inc., Athens, OH, USA). The Thyretain kit is based on Chinese hamster ovary cells transfected with chimeric TSH receptors, which has amino acids 262 to 335 substituted with 73 amino acids from the rat luteinizing hormone receptor (Mc4) [25, 26]. Mc4 was designed to limit the effect of TBI that exists coincidentally with TSI in up to 25% of patients with GD, which can interfere with TSI measurements. The results of the TSI bioassay are reported as specimen-to-reference ratio percentages (SRR%), calculated as follows: SRR% = (mean TSI specimen/ mean TSI reference) × 100. A specimen was considered positive if SRR was ≥140% [25].

Statistical Analysis

The 2 groups of patients (those who experienced relapse and those who did not experience relapse) were compared using the *t*-test or the chi-square test. To obtain the optimal cutoff value for the TSI bioassay to predict relapse of GD, receiver operating characteristic (ROC) curve analysis was performed. To obtain the odds ratio, Pearson's chi-square test was used. Kaplan-Meier curves were used to obtain the overall relapse rate using the log-rank *P*-value. The logistic regression model was used to adjust for confounders. Statistical significance was set at $P \leq 0.05$. SPSS v.24 (IBM Corp., New York, NY; formerly SPSS Inc., Chicago, IL, USA) was used for statistical analyses. The R software version 4.1.1 (R Project for

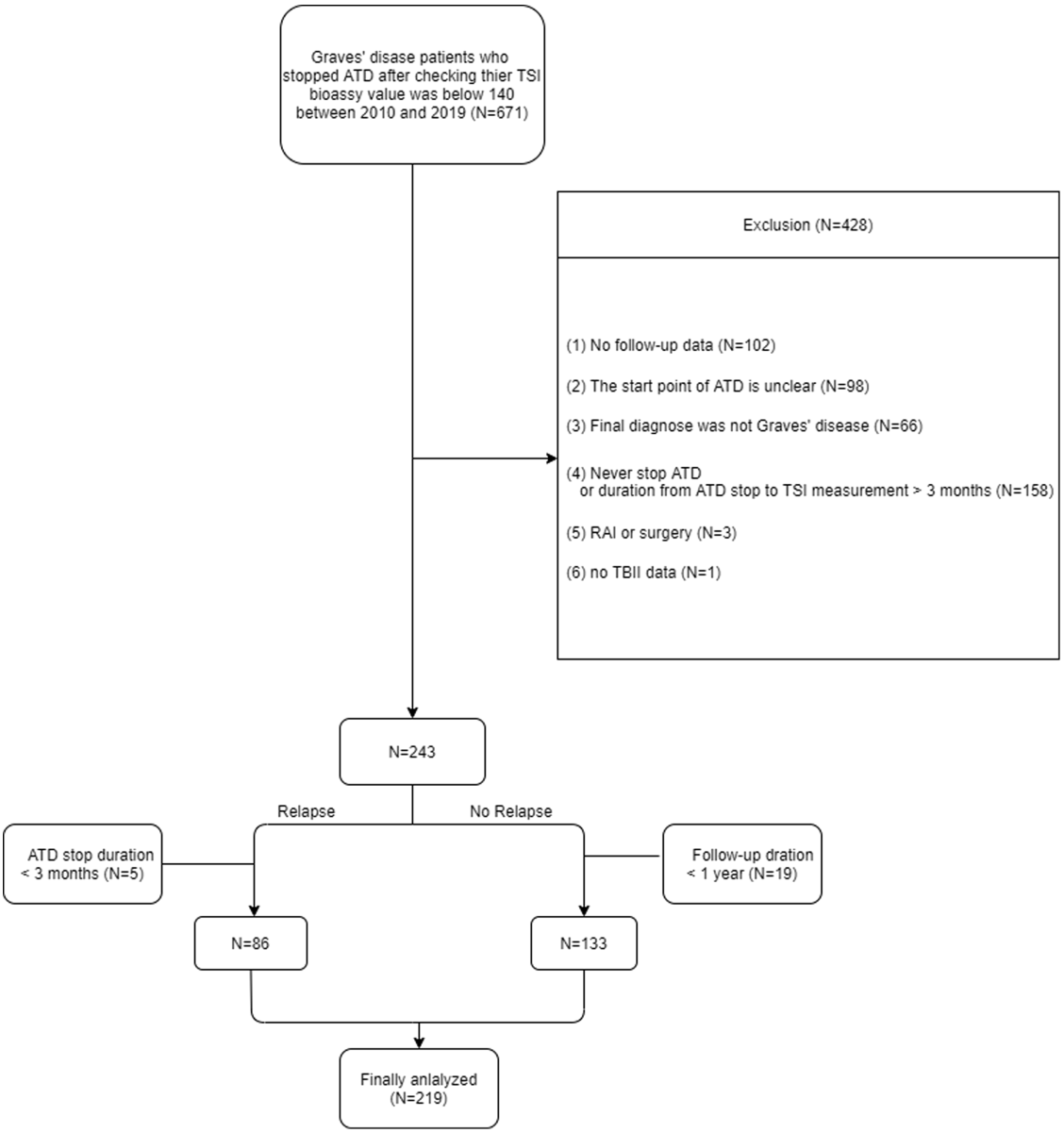

Figure 1. Flow chart of study subjects. Initially, 671 patients who had a history of prescription of antithyroid drug (ATD) and thyroid-stimulating immunoglobulin bioassay < 140%. The final 219 subjects were analyzed after exclusion 452 subjects with each exclusion criteria.

Statistical Computing, Vienna, Austria) was used to build the nomogram. Graphs were produced using Prism version 8.02 (GraphPad Software Inc., La Jolla, CA, USA). For graphical improvement of nomogram, we used an open website made by Dr. Jeehyoung Kim (Department of Orthopedic Surgery, Seoul Sacred Heart General Hospital, Seoul, Korea), available at https://tinyurl.com/Cox-Logistic-Nomogram.

Nomogram

Nomograms graphically represent a complex mathematical formula [27]. In medical research, the nomogram identifies risk factors for a specific event and generates a prognostic model. Each factor is assigned a score in a nomogram based on the estimated regression coefficients calculated from a complex statistical model such as logistic regression or the Cox proportional hazard model. The more important the factor is, the higher the score is, and the most important factor is assigned 100 points. In addition, the length of the lines is proportional to the impact in the model [28].

Results

Clinical Characteristics of Study Subjects With Relapse

Among 219 study subjects, 86 patients (39.3%) experienced relapse (Table 1). Male sex and methimazole (MZ) use showed more frequent relapse than female sex and carbimazole (CM) use. Younger age and lower TSH levels by IRMA at the end of ATD were also predictive factors for GD relapse despite not being statistically significant. On the other hand, serum levels of the TBII and TSI bioassays at the time of withdrawal from ATD between the R and NR groups were not significantly different. The presence of Graves' orbitopathy also showed no difference between the 2 groups. In the NR group (n = 133), the median period from ATD stop to last follow-up

Table 1. Baseline characteristics of 219 patients who stopped antithyroid drugs after their thyroid-stimulating immunoglobulin bioassay was <140%

	Total (n = 219)	No Relapse (n = 133)	Relapse (n = 86)	*P*-value
Age	45.3 ± 13.3	46.7 ± 13.2	43.2 ± 13.3	0.055
Female sex	158 (72.1)	106 (79.7)	52 (60.5)	0.011
Graves' orbitopathy	22 (10.0)	12 (9.0)	10 (11.6)	0.692
Median period from ATD stop to relapse (R) or last follow-up (NR) (month)[a]	19 (11-33)	23 (15.5-44)	11.5 (5.75-23)	<0.001
History of GD treatment	36 (16.4)	21 (15.8)	15 (17.4)	0.892
ATD regimen when stopping				0.002
Methimazole	166 (75.8)	91 (68.4)	75 (87.2)	
Carbimazole	38 (17.4)	32 (24.1)	6 (7.0)	
PTU	15 (6.8)	10 (7.5)	5 (5.8)	
Switch during treatment	18 (8.2)	12 (9.0)	6 (7.0)	0.775
Duration of ATD usage	31.0 ± 28.5	31.0 ± 29.5	31.0 ± 27.0	0.997
Concomitant replacement of T4 during ATD	67 (30.6)	44 (33.1)	23 (26.7)	0.399
Thyroid function tests				
fT4 at ATD stop				
ECLIA (0.89-1.76 ng/dL)	1.2 ± 0.2	1.2 ± 0.2	1.2 ± 0.2	0.139
IRMA (0.89-1.79 ng/dL)	1.3 ± 0.2	1.4 ± 0.2	1.3 ± 0.2	0.106
TSH at the ATD stop				
ECLIA (0.55-4.78 uIU/mL)	2.3 ± 1.7	2.2 ± 1.5	2.5 ± 2.1	0.482
IRMA (0.17-4.05 uIU/L)	2.5 ± 1.8	2.8 ± 1.7	2.2 ± 1.8	0.054
fT4 in relapse				
ECLIA (0.89-1.76 ng/dL)			2.6 ± 1.3	
IRMA (0.89-1.79 ng/dL)			2.5 ± 0.7	
TSH at relapse				
ECLIA (0.55-4.78 uIU/mL)			0.1 ± 0.4	
IRMA (0.17-4.05 uIU/L)			0.1 ± 0.1	
Thyroid autoantibodies				
TBII				
ECLIA (>1.75 IU/L)	1.0 ± 1.1	1.0 ± 1.1	1.2 ± 1.0	0.372
RAI (>1.5 IU/L)	3.5 ± 6.5	4.1 ± 7.0	2.8 ± 5.7	0.270
TBII positivity	44 (20.1)	24 (18.0)	20 (23.3)	0.443
TSI, %	68.9 ± 28.0	67.3 ± 29.0	71.4 ± 26.5	0.290

Data are given as mean ± SD or n (%) unless otherwise noted.
Abbreviations: ATD, antithyroid drug; ECLIA, electrochemiluminescence immunoassay; fT4, free T4; GD, Graves' disease; IRMA, immunoradiometric assay; NR, nonrelapse group; PTU, propylthiouracil; R, relapse group; TBII, thyrotropin-binding inhibitory immunoglobulin binding to thyrotropin; TSH, thyrotropin; TSI, thyroid-stimulating immunoglobulin.
[a]Data are given as median (interquartile range).

was 23 months (interquartile range 15.5-44): 24 patients were followed for 1 year (they stopped the visit), 48 patients were followed for 2 years, and 61 patients were followed >2 years. In the R group (n = 86), the median period from ATD stop to relapse was 11.5 months (interquartile range 5.75-23): 43 patients experienced relapse within 1 year after ATD withdrawal, 23 patients relapsed within 2 years, and the remaining 20 patients experienced relapse after 2 years following ATD withdrawal (Fig. 2). Of the NR and R groups, 15.8% and 17.4%, respectively, had a history of 1 cycle of ATD treatment,

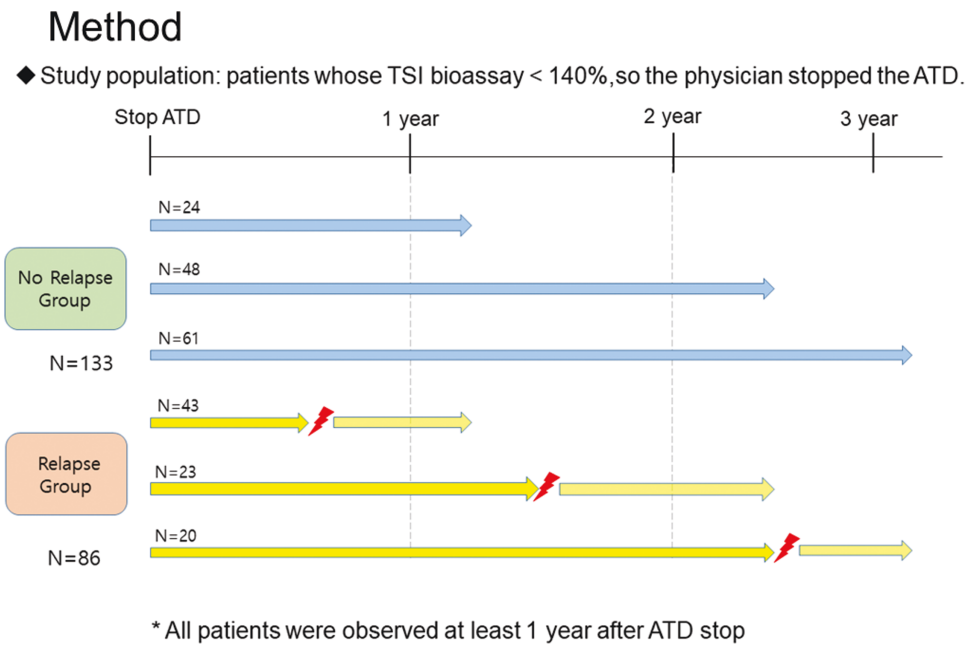

Figure 2. Study population. In the nonrelapse group (n = 133), 24 patients were followed for 1 year (they stopped the visit), 48 patients were followed 2 years, and 61 patients were followed >2 years. In the relapse group (n = 86), 43 patients experienced relapse within 1 year after the antithyroid drug (ATD) withdrawal, 23 patients relapsed within 2 years, and the remaining 20 patients experienced relapse 2 years after the ATD withdrawal.

Table 2. Odds ratio to relapse of Graves' disease according to each cutoff value of thyroid-stimulating immunoglobulin and inhibitory immunoglobulin binding to thyrotropin

	Odds ratio	95% CI	*P*-value
TBII positivity	1.376	0.706-2.683	0.347
Cutoff based on the mean of TSI (66.87%)	2.063	1.189-3.579	0.010
Cutoff based on the median of TSI (66%)	1.690	0.977-2.923	0.060
Cutoff based on ROC curve (66.5%)	2.063	1.189-3.79	0.010

TBII, thyrotropin-binding inhibitory immunoglobulin; TSI, thyroid-stimulating immunoglobulin; ROC, receiver operating characteristic.

Table 3. Logistic regression analysis of risk factors for relapse of Graves' disease

	Univariate			Multivariate		
	HR	95% CI	*P*-value	HR	95% CI	*P*-value
Age	0.980	0.960-1.001	0.056	0.980	0.957-1.003	0.052
Male to female	2.567	1.403-4.698	0.002	2.476	1.277-4.803	0.007
Graves' orbitopathy	0.327	0.547-3.221	0.532	1.046	0.388-1.046	0.0.929
ATD regimen[a]						
Carbimazole	0.228	0.090-0.573	0.002	0.234	0.089-0.615	0.003
PTU	0.607	0.199-1.852	0.380	0.647	0.197-2.124	0.473
Duration of ATD usage	1.000	0.991-1.010	0.997	1.003	0.992-1.013	0.630
Concomitant T4 replacement during ATD	0.738	0.406-1.344	0.321	0.714	0.370-1.379	0.316
TBII positivity	1.376	0.706-2.683	0.348	1.186	0.568-2.615	0.611
TSI positivity based on ROC curve	2.063	1.189-3.579	0.010	1.992	1.095-3.623	0.022

Abbreviations: HR, hazard ratio; ATD, antithyroid drug; PTU, propylthiouracil; ROC, receiver operating characteristic; TBII, thyrotropin-binding inhibitory immunoglobulin; TSI, thyroid-stimulating immunoglobulin.
[a]All compared to methimazole.

which is not statistically significant. Most of the study subjects were treated with MZ (68.4% in the NR group and 87.2% in the R group). Twelve in the NR group and 6 in the R group changed to ATD during treatment. The main reason for the change in ATD was CM from MZ for further reduction in ATD dose [9], propylthiouracil (PTU) for pregnancy [6], and minor adverse events such as skin reaction after MZ. In both groups, ATD was discontinued when the thyroid function test showed a euthyroid state. In this cohort, the TSI bioassay was the main indicator of withdrawal of ATD. So, when even TBII was positive, ATD was discontinued. In the NR group, 24 cases (18.0%) showed positive TBII, and in the R group, 20 cases (23.3%) showed positive TBII ($P = 0.443$)

Clinical Values of the TBII Assay and the TSI Bioassay to Predict Relapse of Graves' Disease

We wanted to predict GD relapse according to TBII positivity and each TSI bioassay cutoff value (Table 2). Positive TBII levels showed an odds ratio of 1.376 without statistical significance ($P = 0.347$). The cutoff value of the positive TSI bioassay was calculated in 3 ways: the mean value of the TSI bioassay in the study population (66.87%), the median value of the TSI bioassay in the study population (66%), and the value obtained from the ROC curve (66.5%). In ROC curve analysis, above the cutoff value (66.5%) to predict the relapse of GD showed 57.0% sensitivity and 60.9% specificity, and area under the curve (AUC) was 0.557. The cutoff value from the mean of the TSI bioassay values and the ROC curve showed statistical significance with the odds ratio with the relapse of the disease. ($P = 0.010$ in both groups).

We built a logistic regression model to find the risk factors for relapse of GD (Table 3). Male sex and TSI bioassay value greater than 66.5% significantly increase the risk of relapse. [hazard ratio (HR) 2.476, $P = 0.007$ in male sex; HR 1.992, $P = 0.022$ in the TSI >66.5% group). Younger age increased the risk of GD relapse with borderline significance (HR 0.980, P-value = 0.052). The TBII positivity or the duration of ATD usage did not show a relationship with relapse. Graves' orbitopathy increases the risk of relapse without statistical significance. Concomitant replacement of thyroid hormone reduced the risk without statistical significance. In Kaplan-Meier curve analysis, the TSI bioassay value of 66.5% significantly predicted GD relapse ($P = 0.049$) (Fig. 3). The TSI bioassay of the ≥66.5% and <66.5% groups at the time of ATD stop were expected to show a relapse rate of 23.8% and 12.7%, respectively, at 2 years after stopping ATD ($P = 0.033$). At 5 years after stopping ATD, it was expected that the TSI bioassay ≥66.5% group would show a relapse rate of 40.3%, and the TSI bioassay <66.5% group would show a relapse rate of 24.3% ($P = 0.017$).

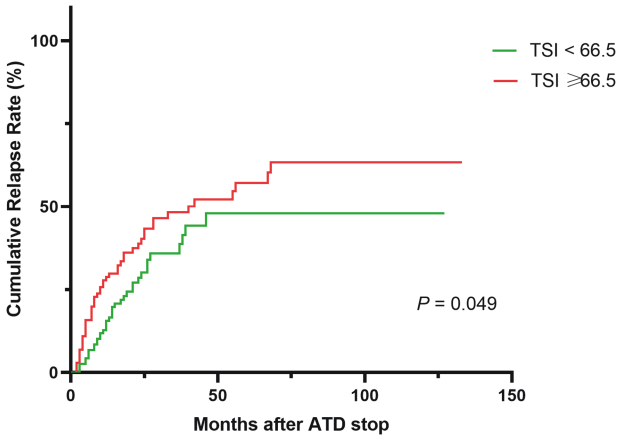

Figure 3. Cumulative relapse rate curve according to the thyroid-stimulating immunoglobulin (TSI) cutoff with 66.5%. The TSI value of 66.5% significantly predicted the relapse of Graves' disease ($P = 0.049$).

TSI Bioassay Cutoff Value to Predict Relapse of GD According to Sex

Because male sex significantly increased the risk of GD relapse, we separately analyzed data by sex (Table 3, Fig. 4). The TBII assay positivity and the TSI bioassay values did not show differences between the 2 sexes. In the male group, the optimal cutoff value of the TSI bioassay to predict relapse was the same at 66.5%, obtained from the whole study cohort. In ROC curve analysis, with the optimal value, the TSI bioassay in the male group showed 58.8% sensitivity, 66.7% specificity, and AUC 0.557.

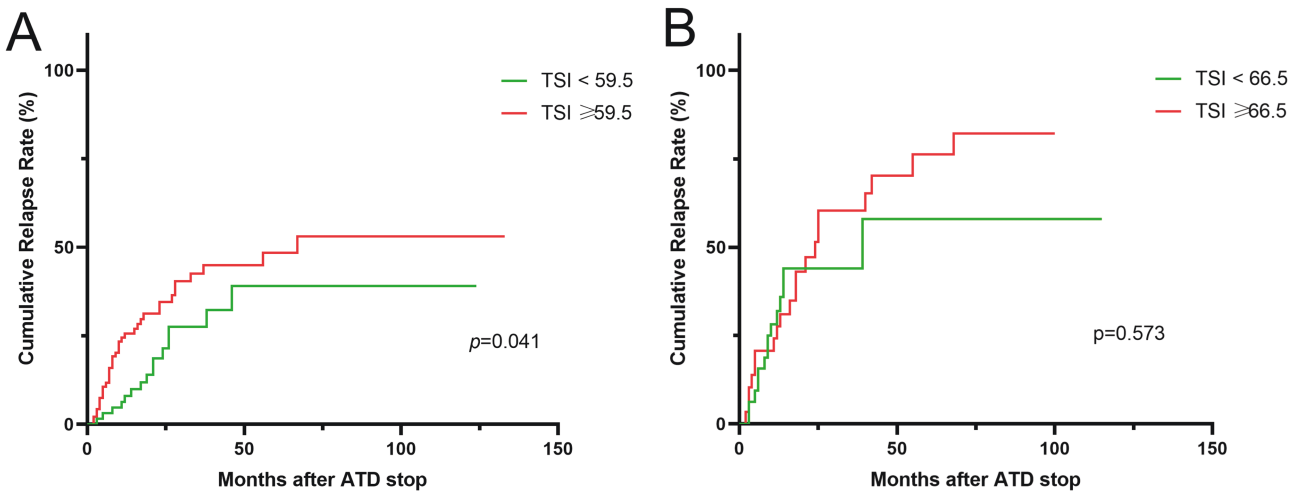

Figure 4. Cumulative relapse rate curve according to sex. (A) The thyroid-stimulating immunoglobulin (TSI) bioassay showed better predictive ability in the female group ($P = 0.041$). (B) TSI bioassay was unable to predict Graves' disease relapse significantly in the male group ($P = 0.573$).

However, in the female group, the optimal cutoff value of the TSI bioassay for the relapse was 59.5%, less than that of the entire study cohort. In ROC curve analysis, with optimal value, the TSI bioassay in the female group showed 71.2% sensitivity, 46.2% specificity, and AUC 0.560. The TSI bioassay showed better predictive ability in the female group ($P = 0.041$) (Fig. 4A) than in the male group, in which the TSI bioassay was unable to predict the relapse of GD significantly ($P = 0.573$) (Fig. 4B).

The ATD regimen at withdrawal and the duration of ATD use were associated with the risk of relapse.

In the logistic regression model, the ATD regimen at the withdrawal of the drug appears to be a possible factor for relapse of the disease. CM use in ATD reduces the risk of relapse (HR 0.248, $P = 0.004$). Because PTU was used only for female patients and the number of patients who used each medication was different, we did a propensity score-matching of MZ and CM usage. After adjusting the sex and age variation for matching, TBII with the RAI assay (3.0 ± 5.2 vs 0.9 ± 0.7, $P = 0.040$) and duration of ATD (22.6 ± 12.2 vs 36.1 ± 26.3, $P = 0.006$) showed a significant difference between the 2 groups of medications (Table 4). The difference was not shown before matching. Moreover, TBII positivity also showed the difference between MZ and CM use groups ($P = 0.050$). Simultaneous replacement of T4 was more common in the MZ group before (34.9% vs 15.8%, $P = 0.036$) and after (42.1% vs 15.8%, $P = 0.023$) propensity score-matching, but TSI bioassay did not show differences between the 2 groups ($P = 0.79$). Before and after the matching, the dose of MZ (3.0 ± 1.3 mg) was significantly higher than that of CM (2.3 ± 1.6 mg) when ATD stop.

Nomogram Construction for Risk Scoring for Relapse of GD and Prediction of Relapse

The single-use of the TSI bioassay to predict relapse had limitations. Therefore, we constructed a nomogram to show the probability of GD relapse. The nomogram graphically represents the numerical relationships between the risk of relapse and 4 main risk factors (age, sex, TBII positivity, and TSI bioassay positivity with a 66.5% cutoff value). For each risk factor, the assigned points were 100 for men, 62.6 for a younger age, 22.8 for TBII assay positivity, and 67.2 for TSI bioassay positivity (Fig. 5A). The points for each risk factor were assigned from the logistic regression model. For analysis, the age factor was divided by age 46, which was the median age of the study cohort. The higher the point in each factor is, the more important the factor is in relapse. For example, assume a 31-year-old (62.6) female (0) with negative TBII assay (0) and positive TSI bioassay (67.2). Her total score is 129.8 (Fig. 5B). Her risk of relapse is 47%. To verify the discrimination of the nomogram, we performed an ROC curve analysis. We obtained the risk scoring from the constructed nomogram for each of our patients in the study cohort. An optimal cutoff value (111.4) showed a sensitivity of 57%, specificity of 70.7%, and AUC 0.672. It was better able to predict GD relapse compared to the TSI bioassay single usage. The cumulative hazard curve between the 2 groups (risk scores >111.4 and <111.4) showed a significant difference ($P < 0.01$) (Fig. 6). At 2 years from stopping ATD, the risk score for the <111.4 group was expected to show an 11.5% relapse rate, and the risk score for the >111.4 group was expected to show a 27.3% relapse rate ($P = 0.004$).

Table 4. Clinical characteristics according to medication at withdrawal and propensity score correlated with age and sex results

	Before propensity score matching				After propensity score matching			
	MZ (n = 166)	CM (n = 38)	P-value	Standardized difference	MZ (n = 38)	CM (n = 38)	*P*-value	Standardized difference
Male sex	53 (31.9)	8 (21.1)	0.261	−0.267	8 (21.1)	8 (21.1)	1.000	0.000
Age	45.5 ± 13.5	46.5 ± 13.4	0.703	0.069	46.2 ± 13.5	46.5 ± 13.4	0.919	0.024
Duration of ATD usage (month)	29.8 ± 29.5	36.1 ± 26.3	0.229		22.6 ± 12.2	36.1 ± 26.3	0.006	
Concomitant T4 replacement during ATD	58 (34.9)	6 (15.8)	0.036		16 (42.1)	6 (15.8)	0.023	
Dose when ATD stop (mg)	3.1 ± 1.4	2.3 ± 1.6	0.003		3.0 ± 1.3	2.3 ± 1.6	0.031	
Thyroid autoantibodies								
TBII								
ECLIA (>1.75 IU/L)	1.2 ± 1.3	0.7 ± 0.4	0.009		1.1 ± 1.0	0.7 ± 0.4	0.327	
RAI (>1.5 IU/L)	3.2 ± 6.2	0.9 ± 0.7	<0.001		3.0 ± 5.2	0.9 ± 0.7	0.040	
TBII positivity	35 (21.1)	2 (5.3)	0.040		9 (23.7)	2 (5.3)	0.050	
TSI	68.1 ± 26.7	70.7 ± 31.5			70.6 ± 28.9	70.7 ± 31.5	0.979	
TSI positivity according to cutoff of 66.5%	79 (47.6)	15 (39.5)	0.468		18 (47.4)	15 (39.5)	0.643	
Relapse of disease	75 (45.2)	6 (15.8)	0.002		17 (44.7)	6 (15.8)	0.013	

Data are given as mean ± SD or n (%).
Abbreviations: ATD, antithyroid drug; CM, carbimazole; ECLIA, electrochemiluminescence immunoassay; IRMA, immunoradiometric assay; MZ, methimazole; TBII, thyrotropin-binding inhibitory immunoglobulin; TSI, thyroid-stimulating immunoglobulin.

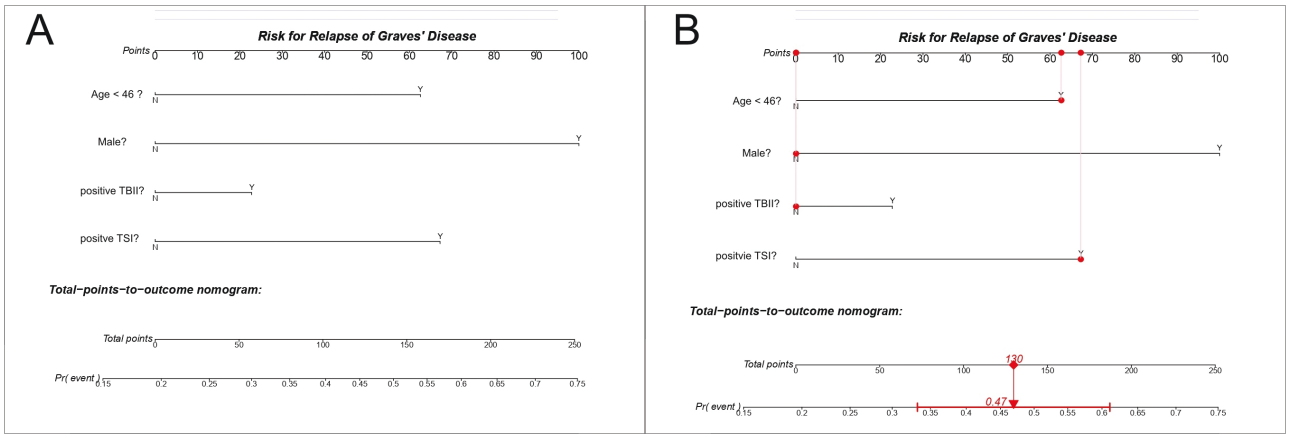

Figure 5. Nomogram construction for risk scoring to predict relapse of Graves disease. (A) Based on logistic regression, its score was assigned to each risk factor (100 for males, 62.6 for younger age, 22.8 for thyrotropin-binding inhibitory immunoglobulin (TBII) assay positivity, and 67.2 for thyroid-stimulating immunoglobulin (TSI) bioassay positivity with cutoff 66.5%). Pr(event) represents the probability of the event. (B) An example of applying the nomogram to risk scoring with 31 years of age (62.6) female (0.0) with negative TBII assay (0.0) and positive TSI assay (67.2). The total score for her is 129.8. (62.6 + 0.0 + 0.0 + 67.2) and her relapse risk is 47%.

Discussion

The finding of our study is summarized in Figure 7. In our study, a large proportion (39.3%) of the patients experienced a relapse of GD, although their TSI bioassay values showed below the cutoff (140%) originally defined to diagnose GD. In addition, serum levels of TSI bioassay at the time of withdrawal from ATD between the R and NR groups were not significantly different. In this context, one could assume that the cutoff point for TSI to diagnose GD and stop ATD should be different.

Kwon et al showed that TSI bioassay at withdrawal from ATD could predict relapse of GD, but the TBII assay could not [20]. However, the number of patients evaluated in their study was relatively small (TSI bioassay, n = 35; TBII assay, n = 39), and both the TSI and TBII bioassays were not measured simultaneously within 1 person. Liu et al showed that the TSI bioassay could be used to predict response to MZ treatment. However, they used the same TSI cutoff value to diagnose GD [18]. The prospective trial of Kahaly et al showed that the TSI bioassay was positive at week 24 after stopping ATD in the group of nonresponders to MZ or patients who experience relapse [19]. However, they also used the TSI bioassay cutoff value to diagnose GD.

Unlike previous studies, our study, through ROC curve analysis, unveiled the newly calculated TSI bioassay cutoff value to discriminate the relapse of GD was 66.5% with AUC 0.557. With this cutoff value, the TSI bioassay >66.5% group showed a relapse rate of 23.8% and the TSI bioassay <66.5% group were expected to show a relapse rate of 12.7% at 2 years from stopping ATD in the Kaplan-Meier curves model. The two groups were expected to show 40.3% and 24.3% relapse rates, respectively, at 5 years from the ATD stop. According to an observational study, the negative TBII assay was associated with a 58% risk of relapse 4 years after discontinuation of ATD [16]. Considering this result, the negative TSI bioassay with a cutoff value of 66.5% is expected to reduce the risk of relapse by about 20%.

However, the single use of the TSI bioassay to predict the relapse of GD had some limitations. In addition, despite of lower TSI cutoff value, the AUC (0.557) of TSI bioassay to predict the relapse of GD in the ROC curve analysis was similar to that (0.62) from 1 observational pilot study with a small population [29]. Especially, being of male sex significantly

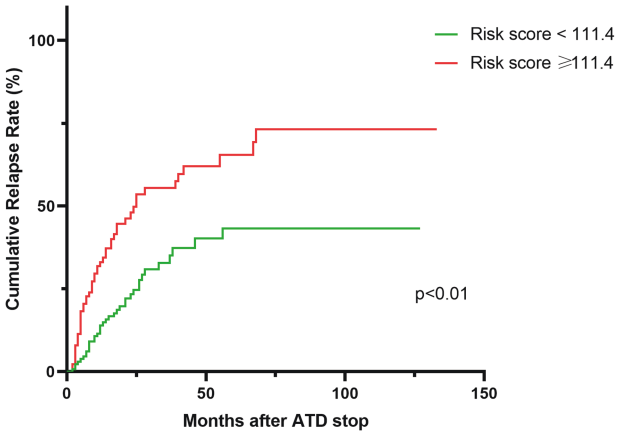

Figure 6. Cumulative relapse rate curve according to the risk score calculated from the constructed nomogram. The cumulative hazard curve between the 2 groups (risk score > 111.4 and risk score < 111.4) showed a significant difference ($P < 0.01$).

increased the risk of relapse (HR 2.569, $P = 0.005$). Meta-analysis or large population studies indicated male factor is a risk for relapse for GD [24, 30]. One previous study revealed that estrogen is associated with B-cell hyperactivity, causing severe autoimmune disease, and [31] Chailurkit et al showed that higher circulating estradiol is related to thyroid autoimmunity in men [32]. Furthermore, Ishido et al suggested skewed X chromosome inactivation was likely related to the prognosis of GD [33].

Other risk factors for relapse that should be discussed are age at diagnosis and the drug regimen at the end of ATD. Younger age increased the risk of GD relapse with borderline significance. Bano et al showed an association between increased TBII and increased risk of relapse in younger patients [34]. In the large population study in Japan, the TBII values have decreased with advancing age, which may explain the high probability of relapse at younger age in our study.

In our study, the use of CM at the time of ATD decreased the risk of relapse (HR 0.248, $P = 0.004$). Equivalent doses of ATD are 40 mg of CM, 30 mg of MZ, and 400 mg of PTU [4]. Although CM has a lower potency than MZ, it would be possible to use CM longer than other ATDs. After matching

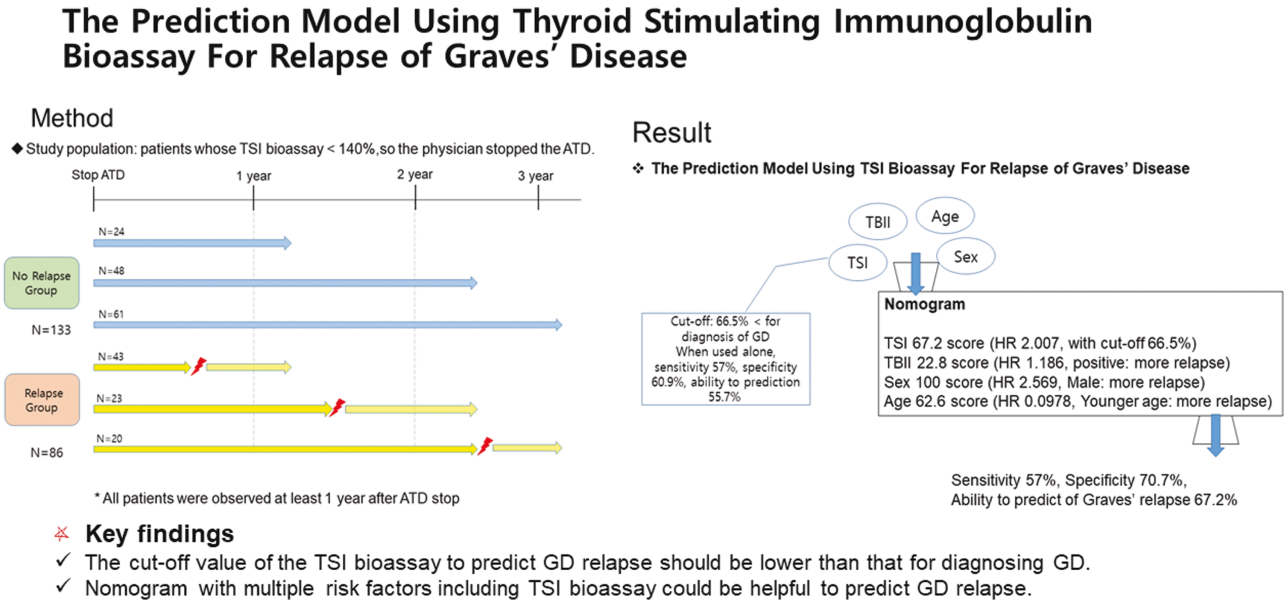

Figure 7. Graphic abstract of our study. Abbreviations: ATD, antithyroid drugs; GD, graves' disease; HR, hazard ratio; TBII; inhibitory immunoglobulin binding to thyrotropin; TSI, thyroid stimulating immunoglobulin.

the propensity score with sex and age, the use of ATD in the CM group showed a significantly longer duration. In randomized clinical trials, the long-term treatment with ATD increases remission rates [35]. The proportion of TBII positivity, concomitant use of T4, and actual dosage was higher in MZ group. It may be because physicians may tend to prescribe MZ more frequently than CM or PTU due to the difference in drug potency or might wait for the negative conversion of TBII with long-term use of CM. Concomitant use of T4 was probably because of the high efficacy of MZ and subsequent hypothyroidism.

From our results, the single usage of the TSI bioassay to predict GD relapse was difficult. In this context, we build the nomogram with multiple factors that could influence the risk of relapse. The logistic regression's score was assigned to each risk factor. When verified with the study cohort, the risk scoring gained by the constructed nomogram showed a good ability to predict the relapse of GD. With the optimal cutoff value (111.4), the nomogram-based risk scoring increased the ability to predict relapse by 11.5% compared to the single use of TSI.

Previously, Vos et al in 2016 developed a prediction model for the recurrence risk of GD [Graves' Recurrent Events After Therapy (GREAT score)], based on clinical data (age, goiter, serum fT4, and serum TBII level) as well as genetic predisposition of untreated 178 patient in a prospective, multicenter, observational study [36]. Although the TSI bioassay was not included in this model, the GREAT score was validated in a large retrospective observational study with 741 patients, which concluded it might help treatment selection in GD patients [37]. Second, a clinical severity score was developed in 2018 from an observational study of 387 consecutive, newly diagnosed GD patients and showed similar predictability to GREAT score, although it only included clinical parameters such as goiter, orbitopathy, and fT4 but did not include TBII [38, 39]. These 2 models at the time of diagnosis can predict the relapse after 1 course of ATD treatment so it may be helpful at the initial stage to decide whether to start ATD or to move on to definite treatments such as thyroidectomy or radioiodine therapy. On the other hand, as our prognostic model was generated based on study subjects on ATD just before stopping ATD, it may guide whether to stop ATD or to continue ATC in a long-term manner.

Our study has some limitations. First, it is a retrospective study and only included patients whose TSI bioassay value was <140%. Therefore, it is hard to say that our study cohort represents all patients with GD. Second, although we defined a minimum follow-up period of remission as 1 year, as the previous studies defined, it could be relatively short. Several patients did not visit the hospital after 1 year of follow-up from stopping ATD because they thought themselves to be in remission but possibly they may show relapse and visit another hospital for a check-up. Third, the smoking status or the presence of goiter, which could be a possible factor associated with relapse of GD, could not be included in our study. Finally, we were not able to compare the TSI values after stopping ATD to the baseline TSI values due to a lack of data. In Korea, the TSI bioassay is covered by insurance only in specific cases: (1) when the diagnosis of GD is unclear because the TBII is negative or not enough high; (2) to check the probability of GD relapse before ATD withdrawal; and (3) in patients with Graves' orbitopathy or in the third trimester in pregnancy. The diagnosis of GD was based on clinical history and TBII results. In our study population, only 6 patients had baseline TSI values: 2 patients in case 1, 3 patients in case 3, and 1 patient without insurance coverage. In the near future, a prognostic model with incorporation of baseline TSI, which was not measured in our study, may better predict a relapse of GD. We also could not evaluate free T3, total T4, total T3, ratio of T4/T3, thyroglobulin antibody, and thyroid peroxidase antibody because of insurance issue.

In conclusion, the TSI bioassay cutoff point to predict GD relapse should be lower than that to diagnose GD. The male factor and the medication regimen at the end of ATD could be attributed to the risk of relapse. Nomogram with multiple predictive factors such as age, sex, TBII, and TSI can predict GD relapse more effectively.

Acknowledgments

We appreciate the help for statistical support to refine the nomogram by Dr. Jeehyoung Kim (Department of Orthopedic Surgery, Seoul Sacred Heart General Hospital, Seoul, Korea).

Funding

This research was carried out without funding.

Author Contributions

H-S.B. and D-J.L. mainly designed the study. H-S.B. mainly wrote the manuscript. D-J.L. supervised the study and is the corresponding author. J.L. contributed to data analysis. All authors contributed to drafting or revising the article, gave final approval of the version to be published, agreed to the submitted journal, and agreed to be accountable for all aspects of the work.

Previous Presentation

A portion of this study was presented in abstract form at the SICEM 2021 (Seoul International Congress of Endocrinology and Metabolism 2021) in Busan, Korea.

Disclosures

The authors have nothing to disclose.

Data Availability

Some or all data sets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. Kamath C, Adlan AM, Premawardhana LD. The role of thyrotrophin receptor antibody assays in graves' disease. *J Thyroid Res*. 2012;2012:525936.
2. Morshed SA, Davies TF. Graves' disease mechanisms: the role of stimulating, blocking, and cleavage region TSH receptor antibodies. *Horm Metab Res*. 2015;47(10):727-734.
3. Giuliani C, Saji M, Bucci I, Napolitano G. Bioassays for TSH receptor autoantibodies, from FRTL-5 cells to TSH receptor-LH/CG receptor chimeras: the contribution of Leonard D. Kohn. *Front Endocrinol (Lausanne)*. 2016;7:103.
4. Badiu C. Williams Textbook of Endocrinology [book review]. *Acta Endocrinol (Buchar)*. 2019;15(3):416.
5. Barbesino G, Tomer Y. Clinical review: clinical utility of TSH receptor antibodies. *J Clin Endocrinol Metab*. 2013;98(6):2247-2255.
6. Tozzoli R, Bagnasco M, Giavarina D, Bizzaro N. TSH receptor autoantibody immunoassay in patients with Graves' disease: improvement of diagnostic accuracy over different generations of methods. Systematic review and meta-analysis. *Autoimmun Rev*. 2012;12(2):107-113.
7. Diana T, Wüster C, Kanitz M, Kahaly GJ. Highly variable sensitivity of five binding and two bio-assays for TSH-receptor antibodies. *J Endocrinol Invest*. 2016;39(10):1159-1165.
8. Diana T, Wüster C, Olivo PD, *et al*. Performance and specificity of 6 immunoassays for TSH receptor antibodies: a multicenter study. *Eur Thyroid J*. 2017;6(5):243-249.
9. Scappaticcio L, Trimboli P, Keller F, Imperiali M, Piccardo A, Giovanella L. Diagnostic testing for Graves' or non-Graves' hyperthyroidism: a comparison of two thyrotropin receptor antibody immunoassays with thyroid scintigraphy and ultrasonography. *Clin Endocrinol (Oxf)*. 2020;92(2):169-178.
10. Kahaly GJ. Management of Graves thyroidal and extrathyroidal disease: an update. *J Clin Endocrinol Metab*. 2020;105(12):3704-3720.
11. Chung HK. Medical treatment of Graves' disease. *Int J Thyroidol*. 2019;12(2):79-84.
12. Cooper DS. Antithyroid drugs. *N Engl J Med*. 2005;352(9):905-917.
13. Feldt-Rasmussen U, Schleusener H, Carayon P. Meta-analysis evaluation of the impact of thyrotropin receptor antibodies on long term remission after medical therapy of Graves' disease. *J Clin Endocrinol Metab*. 1994;78(1):98-102.
14. Wiersinga WM. Graves' disease: can it be cured? *Endocrinol Metab (Seoul)*. 2019;34(1):29-38.
15. Van Herck SL, Geysens S, Bald E, *et al*. Maternal transfer of methimazole and effects on thyroid hormone availability in embryonic tissues. *J Endocrinol*. 2013;**218**(1):105-115.
16. Tun NNZ, Beckett G, Zammitt NN, Strachan MWJ, Seckl JR, Gibb FW. Thyrotropin receptor antibody levels at diagnosis and after thionamide course predict Graves' disease relapse. *Thyroid*. 2016;26(8):1004-1009.
17. Quadbeck B, Hoermann R, Roggenbuck U, *et al*. Sensitive thyrotropin and thyrotropin-receptor antibody determinations one month after discontinuation of antithyroid drug treatment as predictors of relapse in Graves' disease. *Thyroid*. 2005;15(9):1047-1054.
18. Liu K, Fu Y, Li T, *et al*. Clinical efficacy of thyroid-stimulating immunoglobulin detection for diagnosing Graves' disease and predictors of responsiveness to methimazole. *Clin Biochem*. 2021;97:34-40.
19. Kahaly GJ, Diana T, Kanitz M, Frommer L, Olivo PD. Prospective trial of functional thyrotropin receptor antibodies in Graves disease. *J Clin Endocrinol Metab*. 2020;105(4):e1006-e1014.
20. Kwon H, Kim WG, Jang EK, *et al*. Usefulness of measuring thyroid stimulating antibody at the time of antithyroid drug withdrawal for predicting relapse of Graves disease. *Endocrinol Metab (Seoul)*. 2016;31(2):300-310.
21. Lytton SD, Li Y, Olivo PD, Kohn LD, Kahaly GJ. Novel chimeric thyroid-stimulating hormone-receptor bioassay for thyroid-stimulating immunoglobulins. *Clin Exp Immunol*. 2010;162(3):438-446.
22. Wang SY, Wang CT, Tien KJ, Chang CC, Liu TH. Thyroid-stimulating hormone receptor antibodies during follow-up as remission markers in childhood-onset Graves' disease treated with antithyroid drugs. *Kaohsiung J Med Sci*. 2020;36(4):281-286.
23. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*. 2016;26(10):1343-1421.
24. Suzuki N, Noh JY, Yoshimura R, *et al*. Does age or sex relate to severity or treatment prognosis in Graves' disease? *Thyroid*. 2021;31(9):1409-1415.
25. Li Y, Kim J, Diana T, Klasen R, Olivo PD, Kahaly GJ. A novel bioassay for anti-thyrotrophin receptor autoantibodies detects both thyroid-blocking and stimulating activity. *Clin Exp Immunol*. 2013;173(3):390-397.
26. Watson PF, Ajjan RA, Phipps J, Metcalfe R, Weetman AP. A new chemiluminescent assay for the rapid detection of thyroid stimulating antibodies in Graves' disease. *Clin Endocrinol (Oxf)*. 1998;49(5):577-581.
27. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16:e173-e180.
28. Kang EJ, Lee JY. Nomogram plot for predicting chronic otitis media in Korean adults. *J Korean Data Inf Sci Soc*. 2017;28:899-910.
29. Struja T, Jutzi R, Imahorn N, *et al*. Comparison of five TSH-receptor antibody assays in Graves' disease: results from an observational pilot study. *BMC Endocr Disord*. 2019;19(1):38.

30. Zuhur SS, Yildiz I, Altuntas Y, *et al*. The effect of gender on response to antithyroid drugs and risk of relapse after discontinuation of the antithyroid drugs in patients with Graves' hyperthyroidism: a multicentre study. *Endokrynol Pol*. 2020;71(3):207-212.
31. Da Silva JA. Sex hormones, glucocorticoids and autoimmunity: facts and hypotheses. *Ann Rheum Dis*. 1995;54(1):6-16.
32. Chailurkit LO, Aekplakorn W, Ongphiphadhanakul B. The relationship between circulating estradiol and thyroid autoimmunity in males. *Eur J Endocrinol*. 2013;170(1):63-67.
33. Ishido N, Inoue N, Watanabe M, Hidaka Y, Iwatani Y. The relationship between skewed X chromosome inactivation and the prognosis of Graves' and Hashimoto's diseases. *Thyroid*. 2015;25(2):256-261.
34. Bano A, Gan E, Addison C, *et al*. Age may influence the impact of TRAbs on thyroid function and relapse-risk in patients with Graves disease. *J Clin Endocrinol Metab*. 2019;104(5):1378-1385.
35. Azizi F, Amouzegar A, Tohidi M, *et al*. Increased remission rates after long-term methimazole therapy in patients with Graves' disease: results of a randomized clinical trial. *Thyroid*. 2019;29(9):1192-1200.
36. Vos XG, Endert E, Zwinderman AH, Tijssen JGP, Wiersinga WM. Predicting the Risk of Recurrence Before the Start of Antithyroid Drug Therapy in Patients With Graves' Hyperthyroidism. *J Clin Endocrinol Metab*. 2016;101(4):1381-1389.
37. Struja T, Kaeslin M, Boesiger F, *et al*. External validation of the GREAT score to predict relapse risk in Graves' disease: results from a multicenter, retrospective study with 741 patients. *Eur J Endocrinol*. 2017;176(4):413-419.
38. Bartalena L, Masiello E, Magri F, *et al*. 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*. 2016;39(12):1445-1451.
39. Masiello E, Veronesi G, Gallo D, *et al*. Antithyroid drug treatment for Graves' disease: baseline predictive models of relapse after treatment for a patient-tailored management. *J Endocrinol Invest*. 2018;41(12):1425-1432.