BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers’ comments and the authors’ responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open’s open peer review process please email info.bmjopen@bmj.com
# Longitudinal association of thyroid-stimulating immunoglobulin levels with clinical characteristics in thyroid eye disease

| Journal:          | BMJ Open                          |
|-------------------|-----------------------------------|
| Manuscript ID     | bmjopen-2021-050337               |
| Article Type:     | Original research                 |
| Date Submitted by | 18-Feb-2021                       |
| Author:           |                                   |
| Complete List of Authors: | Ko, JaeSang; Yonsei University College of Medicine, Department of Ophthalmology  |
|                   | Kook, Koung Hoon; Ajou University School of Medicine, Department of Ophthalmology |
|                   | Yoon, Jin-Sook; Yonsei University Hospital, Ophthalmology |
|                   | Woo, Kyung In; Samsung Medical Center, Ophthalmology |
| Keywords:         | Oculoplastics < OPHTHALMOLOGY, Thyroid disease < DIABETES & ENDOCRINOLOGY, Orbital and lacrimal disorders < OPHTHALMOLOGY |
I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd (“BMJ”) its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge (“APC”) for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author’s Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.
Longitudinal association of thyroid-stimulating immunoglobulin levels with clinical characteristics in thyroid eye disease

JaeSang Ko, MD, PhD,1 Koung Hoon Kook, MD, PhD,2 Jin Sook Yoon, MD, PhD,1
Kyung In Woo, MD, PhD,3* Jae Wook Yang, MD, PhD,4*

1Department of Ophthalmology, Severance Hospital, Institute of Vision Research, Yonsei University College of Medicine, Seoul, Republic of Korea
2Department of Ophthalmology, Ajou University School of Medicine, Suwon, Republic of Korea
3Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
4Department of Ophthalmology, Inje University College of Medicine, Inje University Busan Paik Hospital, Busan, Republic of Korea

* Correspondence author:
Kyung In Woo, MD, PhD
Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Ilwon-ro, Seoul 06351, Republic of Korea, E-mail: eyeminded@skku.edu

Jae Wook Yang, MD, PhD
Department of Ophthalmology, Inje University College of Medicine, Inje University Busan Paik Hospital, 75 Bokjiro, Busan 47392, Republic of Korea, E-mail: eyeyang@inje.ac.kr
Word count: 3423 words (excluding title page, abstract, references, figures, and tables.)
ABSTRACT

Objectives: The clinical course of thyroid eye disease (TED) is heterogeneous and predicting which patients might develop the severe sequelae of the disease is difficult. In this study, the longitudinal association between changes in serum thyroid stimulating hormone (TSH) receptor antibody (TRAb) levels and the course of disease activity and severity over time was evaluated.

Design: This was a multicentre, prospective, observational trial.

Setting: 15 tertiary care oculoplastic service centres in Korea.

Participants: 76 patients with newly diagnosed TED were included and followed-up with for 12 months. TSH-binding inhibitory immunoglobulin

Methods: The longitudinal association between changes in serum thyroid-stimulating hormone (TSH) receptor antibody (TRAb) levels and disease activity and severity over time was evaluated. Clinical characteristics and serum TRAb levels were evaluated at baseline and 6 and 12 months. The longitudinal association between serum TRAb levels and clinical activity score (CAS), NOSPECS (no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss) score, and proptosis was analyzed.

Results: Thyroid-stimulating immunoglobulin (TSI) and TSH-binding inhibitory immunoglobulin (TBII) levels decreased during the 1-year follow-up, whereas disease activity measured using CAS decreased mainly in the first 6 months. The disease severity measured using NOSPECS score and proptosis remained unchanged. Inter-person difference in TBII levels was associated with CAS, NOSPECS score, and proptosis over time; inter-person difference in TSI levels was associated with the NOSPECS score. Subgroup analysis of patients with baseline CAS ≥4 demonstrated that within-person changes in TSI levels affected the CAS and NOSPECS score.
Conclusions: Follow-up measurement of serum TSI and TBII levels may help evaluate the prognosis of TED; this can enable accurate clinical decision-making.

Keywords: thyroid eye disease; thyroid-stimulating immunoglobulin; TSH-binding inhibitory immunoglobulin; thyroid stimulating hormone receptor antibody

Strengths and limitations of this study

- This is the first study to investigate longitudinal correlation between thyroid-stimulating immunoglobulin levels and clinical characteristics in thyroid eye disease in a prospective setting.
- As such, this data provides a new perspective in managing and predicting prognosis of thyroid eye disease.
- This multicentre, prospective, observational trial was conducted among 15 tertiary care oculoplastic service centres in Korea.
- The findings of this study need to be interpreted with caution, because owing to the observational nature, baseline clinical characteristics, and treatment modalities during follow-up were not standardized, which may have affected the clinical characteristics over time.
INTRODUCTION

Thyroid eye disease (TED) is a component of autoimmune Graves' hyperthyroidism wherein the thyroid-stimulating hormone (TSH) receptor antibody (TRab) stimulates orbital and periorbital tissues. The clinical course of TED is heterogeneous. The natural course is benign in a considerable proportion of TED cases, but some TED cases may present with significant severe sequelae. Clues that can assist clinicians in determining TED prognosis are limited, making it difficult to ascertain which patients might develop the potentially blinding sequelae of TED.

The relationship between serum TRAb levels and clinical characteristics of TED has been elucidated. Serum TRAb levels correlate directly with the clinical features of TED; a high TRab level was shown to correlate with the prevalence of the disease and disease status. Furthermore, patients with higher baseline TRAb levels demonstrated a higher risk of severe disease course during a 1-year follow-up. However, TRAb levels differ over time and decreases to a large extent in some patients but not in others. Most previous studies measured serum TRAb level only at the baseline; however, using only baseline values leads to losing information provided by the serum TRAb level, which changes over time, leading to limited assessment of the relationship between thyroid autoantibodies and TED.

Currently, two established assays are available for measuring TRAb levels: the competitive TSH-binding inhibitory immunoglobulin (TBII) assay and functional thyroid-stimulating immunoglobulin (TSI) bioassay. The former uses the ability of TRab to inhibit the binding of radiolabelled TSH to the TSH-receptors, and the latter measures cyclic adenosine monophosphate production after TRab binds to the TSH receptor, enabling functional TRab identification. Several studies evaluating the correlation between TRab levels and the clinical course of TED revealed different predictive values between TBII and TSI levels; most results showed that disease activity and the
severity of TED are more closely correlated with the TSI bioassay than with TBII assay. Longitudinal studies assessing the association between TRAb levels and the clinical course of TED are limited. The inflammatory phase of TED reportedly correlates with changes in measured TSI levels, indicating that serial TSI level measurements may be an adjunct in assessing the clinical inflammatory activity of TED, and follow-up measurements of TBII levels also reportedly allow assessment of the prognosis of TED. However, the association and potential predictive value of TSI levels during the clinical course of TED have not yet been demonstrated in a prospective setting.

Here, we aimed to investigate whether changes in serum TRAb (TBII and TSI) levels over time are associated with the activity and severity of TED during the disease, revealing the correlation between the time-dependent changes in TRAb levels in an individual patient and the clinical course of the disease.

METHODS

Subjects

This multicentre, prospective, observational study was conducted by the Korean Society of Ophthalmic Plastic and Reconstructive Surgery from April 2017 to February 2019. During the study period, the 15 participating Korean tertiary care oculoplastic service centres enrolled 91 patients with Graves’ disease and TED. The complete data set and written informed consent of each patient were obtained before starting the trial. Institutional ethical review committee approval of each participating institute was obtained. This study followed the tenets of the Declaration of Helsinki.

Patients with clinically overt TED less than 6 months after symptom onset were recruited. Patients with previous history of TED treatment (except conservative treatments such as the use of artificial tears), uncertain cases, and patients aged less
than 19 years were excluded. Clinical data were recorded, and a blood sample was
drawn at inclusion and after 6 and 12 months from inclusion. After the enrolment,
participants received the necessary management for TED at the discretion of each
clinician, including conservative management, systemic steroid, and/or orbital
radiotherapy. No surgical procedure was performed on the enrolled patients during the
study period.

The activity and severity of the disease were assessed by an oculoplastic
specialist blinded to the laboratory data at each participating institute. The general
ophthalmic assessment included an examination of anterior and posterior eye
segments, measurement of proptosis using Hertel exophthalmometry, and evaluation
of ocular movement.\textsuperscript{20} The clinical activity score (CAS) was calculated based on the
classic signs of inflammation and comprised seven items.\textsuperscript{21} The CAS ranged from 0 to
7. The severity of the disease was assessed using the modified NOSPECS (no signs
or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle
involvement, corneal involvement, sight loss) score.\textsuperscript{12} The modified NOSPECS severity
score was calculated as the sum of each class present and ranged from 1 to 18. The
more severely affected eye (eye with more proptosis at enrolment or eye with higher
CAS in case of similar extent of proptosis in both eyes) of a patient was selected for the
study.

**TSH receptor Autoantibody assays**

The blood sample collected at each visit was centrifuged at each participating institute
and transferred to a central laboratory for analysis. TRAb levels were measured using
a third-generation TBII assay with the automated Cobas electrochemiluminescence
immunoassay system (Elecsys, Roche Diagnostics GmbH, Penzberg, Germany)
according to the manufacturer’s instructions. The cut-off value of positivity using this
system was 1.75 IU/L. Serum Mc4-TSI level was measured using the Thyretain™ TSI reporter bioassay (Diagnostic Hybrids, Inc., Athens, OH, USA) according to the manufacturer’s instructions. Results were considered positive with specimen-to-reference ratio >140% of the reference control.

**Data analysis and statistics**

The gathered data were statistically analyzed for a correlation between the changes in clinical indicators of TED, as reflected in the NOSPECS score, CAS, and proptosis, and the change in measured serum TRAb levels over time in individual patients.\(^22\) We assessed the effect of serum autoantibody level on clinical activity and severity of TED. We investigated the tendency of the serum autoantibody levels and clinical parameters over time using analysis of variance (ANOVA). Two approaches for modeling repeatedly measured outcomes were adopted. First, to assess how the baseline serum autoantibody level affected the clinical characteristics over time, we analyzed our data using a linear mixed model with time-dependent covariates. Age, sex, smoking status, and the treatment modality of patients (conservative, systemic steroid, and systemic steroid with orbital radiotherapy) were included as covariates in the model. Second, the serum autoantibody levels were used as a covariate, including the baseline level and all repeated measurements. One particularly salient strength offered by longitudinal data is the ability to disaggregate between-person and within-person effects in the regression of an outcome on a time-varying covariate;\(^23\) this allowed the separation of time-varying within-person change from time-invariant between-person (inter-person) difference in TRAb levels based on the clinical characteristics of the disease.\(^24\) This approach enabled the study of the correlation between autoantibody values at initial presentation and clinical characteristics at further follow-ups, and the study of correlations between the time-dependent changes in autoantibody values and changes
in clinical characteristics. Analyses were performed using the SAS software, version 9.4 (SAS Institute, Inc., Cary, NC, USA). Two-sided $P$-value < 0.05 was considered to indicate statistical significance.

Patients with initial CAS of 4 or higher were included in the active subgroup and subjected to subgroup analysis, which included time-dependent changes in the clinical characteristics and laboratory results and time-dependent correlation analysis.

**Patient and public involvement**

No patient involved.

**RESULTS**

**Clinical characteristics and TSH receptor autoantibody levels**

Of the 91 patients enrolled in this study, 76 patients completed three clinical and laboratory evaluations. The clinical and biochemical characteristics of 76 patients with TED are shown in Table 1. During 12 months of follow-up, 24 patients (32%) were conservatively managed, 34 (45%) were prescribed systemic steroid (oral or intravenous), and 18 (24%) were managed with combined orbital radiation and systemic steroid. Using systemic steroids did not affect the serum TSI and TBII levels of patients ($P = 0.4225$ and 0.9634, respectively). TSI and TBII levels showed a statistically significant decrease during 12 months of follow-up. Regarding clinical characteristics, disease activity measured with CAS decreased during 12 months of follow-up, although the disease severity measured with NOSPECS and proptosis remained unchanged.

**Table 1. Clinical and Biochemical Characteristics of the Study Population**

| Variables | Baseline | 6-month | 12-month | $P$-value |
|-----------|----------|---------|----------|-----------|
|           |          |         |          |           |

9
Correlation between baseline TSH receptor autoantibody levels and clinical characteristics

The results of the linear mixed model with time-dependent covariate analysis of the correlation between baseline serum TRAb levels and clinical characteristics of TED are shown in Table 2. Lower baseline serum TBII level was associated with a future decrease in TED activity. Reduction in the severity of TED (NOSPECS score and proptosis) was associated with lower baseline serum TSI and TBII levels.
| CAS                      | NOSPECS                  | Proptosis                  |
|-------------------------|--------------------------|---------------------------|
| **Estimated change (SE)** | **P-value**              | **Estimated change (SE)** | **P-value**              | **Estimated change (SE)** | **P-value**              |
| TSI                     | 0.000 (0.001)            | 0.999                     | 0.003 (0.001)            | 0.016                     | 0.004 (0.002)            | 0.055                     |
| TBII                    | **0.029 (0.010)**        | **0.008**                 | **0.042 (0.019)**        | **0.029**                 | **0.062 (0.030)**        | **0.041**                |

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; SE, standard error; TSI, thyroid-stimulating immunoglobulin; TBII, TSHR-binding inhibitory immunoglobulin

The P-values were calculated using a linear mixed model with time-dependent covariates adjusted for age, sex, smoking status, and treatment modality.

The numbers in bold are statistically significant changes.

**Correlation between time-dependent changes in clinical characteristics and TSH receptor autoantibody levels**

Table 3 shows the effect of the time-dependent changes in TRAb levels on the clinical characteristics of TED. The effect of longitudinally measured TRAb level over time was disaggregated to inter-person differences and within-person change using a linear mixed model with time-dependent covariates. TSI level measured at multiple time points showed inter-person-relationship with serial changes in NOSPECS score, indicating that patients with low average serum TSI level during 1 year of follow-up showed more reduction in their NOSPECS score over time. TBII level also had inter-person relationship with CAS and NOSPECS score, indicating that patients with higher TBII levels during 1 year of follow-up showed prolonged disease activity and severity. Within-person changes in TSI and TBII levels did not show a statistically significant correlation with disease activity and severity.
Table 3. Effect of Inter-person Difference and Within-person Changes in Thyroid-Stimulating Hormone Receptor Autoantibody Levels on Clinical Characteristics: Longitudinal Association during a Year

| Variables | Effect of TSI on clinical characteristics | Effect of TBII on clinical characteristics |
|-----------|------------------------------------------|------------------------------------------|
|           | Estimated change (SE)                     | Estimated change (SE)                     |
|           |                                          |                                          |
| CAS       |                                          |                                          |
| Inter-person | 0.0013 (0.0007)       | 0.0013 (0.0109)       |
| Within-person | 0.0011 (0.0008)       | 0.0261 (0.0173)       |
| NOSPECS   |                                          |                                          |
| Inter-person | 0.0043 (0.0013)       | 0.0520 (0.0242)       |
| Within-person | 0.0002 (0.0007)       | 0.0249 (0.0152)       |
| Proptosis  |                                          |                                          |
| Inter-person | 0.0032 (0.0022)       | 0.0775 (0.0379)       |
| Within-person | -0.0010 (0.0007)      | 0.0017 (0.0167)       |

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; TSI, thyroid-stimulating immunoglobulin; TBII, TSHR-binding inhibitory immunoglobulin
† The $P$-values were calculated using a linear mixed model with time-dependent
covariates splitting within-person and between-person effects, adjusted for age, sex,
smoking status, and treatment modality.
The numbers in bold are statistically significant changes.

Correlation between time-dependent changes in clinical characteristics and TSH
receptor autoantibody levels: subgroup analysis

Twenty-five patients with initial CAS of ≥4 were included in the subgroup analysis and
this group was defined as the active subgroup, whereas others were included in the
inactive subgroup. Subgroup analysis was performed to evaluate the time-dependent
correlation between clinical characteristics of TED and TRAb levels in the active
subgroup. Clinical and biochemical characteristics of the active subgroup are shown in
Table 4. This subgroup had more females and smokers than the inactive subgroup
(both $P < 0.0001$). During 12 months of follow-up, 10 patients (40%) were prescribed
systemic steroid (oral or intravenous) and 15 (60%) were managed with combined
orbital radiation and systemic steroid. Using systemic steroids did not affect the serum
TSI and TBII levels of patients ($P = 0.9008$ and $0.4957$, respectively). Baseline TBII
level was higher in the active than in the inactive subgroups ($9.5 ± 9.5$ IU/L vs. 17.0
$±14.2$ IU/L, $P = 0.0125$), whereas baseline TSI level of the active subgroup was like
that of the inactive subgroup ($443.9 ± 210.5%$ vs. $453.2 ± 123.6%$, $P = 0.8096$). The
active subgroup had higher CAS ($4.4 ± 0.6$ vs. $1.9 ± 1.0$, $P < 0.0001$), higher
NOSPECS score ($5.6 ± 2.3$ vs. $3.1 ± 1.6$, $P < 0.0001$), and more proptosis ($19.2 ± 3.5$
mm vs. $16.9 ± 2.6$ mm, $P = 0.0020$) than the inactive group. Correlation analysis
between baseline TSH receptor autoantibody levels and clinical characteristics in the
active subgroup showed that lower baseline TBII level correlated with a decrease in
CAS and NOSPECS during the 1-year follow-up (estimated change $0.0523$, $P = 0.0078$
for CAS, and $0.0566$, $P = 0.0394$ for NOSPECS). Baseline TSI level did not show a
statistically significant correlation with clinical characteristics in the active subgroup.

Table 4. Clinical and Biochemical Characteristics: Subgroup Analysis of Patients with Active Thyroid Eye Disease at Baseline (n = 25)

| Variables       | Baseline     | 6-month      | 12-month     | P-value  |
|-----------------|--------------|--------------|--------------|----------|
| Age, y          | 43.6±11.84   |              |              |          |
| Sex             |              |              |              |          |
| Female, n (%)   | 10 (40.0)    |              |              | 0.3173†  |
| Male, n (%)     | 15 (60.0)    |              |              |          |
| Smoking status  |              |              |              |          |
| Non-smoker, n   | 17 (68.0)    |              |              | <.0719†  |
| (%)             |              |              |              |          |
| Smoker, n (%)   | 8 (32.0)     |              |              |          |
| TSI, SRR in %   | 453.2±123.6  | 342.5±139.7  | 297.9±137.0  | <.0001‡  |
| TBII, IU/L      | 17.0±14.2    | 9.4±11.6     | 7.6±10.4     | <.0001‡  |
| CAS             | 4.4±0.6      | 1.7±1.3      | 1.7±1.7      | <.0001‡  |
| NOSPECS score   | 5.6±2.3      | 5.0±2.3      | 4.9±2.4      | 0.5860‡  |
| Proptosis, mm   | 19.2±3.5     | 19.7±3.4     | 19.6±3.9     | 0.8942‡  |

TSI, thyroid-stimulating immunoglobulin; SRR, specimen-to-reference ratio; TBII, TSH-binding inhibitory immunoglobulin; CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss.

† The P-values were calculated between each group using Fischer’s exact test.
‡ The P-values were calculated between three visits using ANOVA.

Table 5 shows the effect of the time-dependent changes in TRAb levels on the clinical characteristics of TED in the active-disease subgroup. The effect of longitudinally measured TRAb level over time was disaggregated to inter-person differences and within-person change using a linear mixed model with time-dependent covariates. Within-person changes in TSI level measured at multiple time points
affected CAS and NOSPECS, whereas inter-person difference in TSI level was
unrelated to clinical characteristics. Thus, patients who showed more reduction in TSI
level during follow-up also showed more reduction in their CAS and NOSPECS scores
during the 1-year follow-up. In contrast, inter-person difference in TBII level correlated
with CAS and NOSPECS scores, and within-person changes in TBII level were
independent of disease severity and activity.
Table 5. Effects of Inter-person Differences and Within-person Changes in Thyroid-stimulating Hormone Receptor Autoantibody Levels on Clinical Characteristics: Subgroup Analysis of Patients with Active Thyroid Eye Disease at Baseline (n = 25)

| Variables | Estimated change (SE) | P-value† |
|-----------|-----------------------|----------|
| CAS       |                       |          |
| Inter-person | -0.0002 (0.0020)    | 0.9347   |
| Within-person | 0.0046 (0.0016)     | 0.0064   |
| NOSPECS   |                       |          |
| Inter-person | 0.0014 (0.0031)    | 0.6559   |
| Within-person | 0.0034 (0.0016)     | 0.0385   |
| Proptosis |                       |          |
| Inter-person | 0.0016 (0.0070)    | 0.8178   |
| Within-person | 0.0014 (0.0018)     | 0.4464   |

| Variables | Estimated change (SE) | P-value† |
|-----------|-----------------------|----------|
| CAS       |                       |          |
| Inter-person | 0.0470 (0.0146)     | 0.0023   |
| Within-person | 0.0243 (0.0284)    | 0.3959   |
| NOSPECS   |                       |          |
| Inter-person | 0.0517 (0.0245)     | 0.0402   |
| Within-person | 0.0397 (0.0261)    | 0.1354   |
| Proptosis |                       |          |
| Inter-person | 0.1067 (0.0566)   | 0.0657   |
| Within-person | -0.0046 (0.0290)  | 0.8742   |

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; TSI, thyroid-stimulating immunoglobulin; TBII, TSH-binding inhibitory immunoglobulin
The $P$-values were calculated using a linear mixed model with time-dependent
covariates splitting within-person and between-person effects, adjusted for age, sex,
smoking status, and treatment modality.
The numbers in bold are statistically significant changes.

DISCUSSION

This study, involving a large cohort of patients with TED, demonstrated the clinical
relevance of longitudinal follow-up of both TSI and TBII levels. Both TSI and TBII levels
decreased sustainably during the 1-year follow-up, whereas disease activity measured
using CAS decreased mainly between the first 6 months and stabilized thereafter. The
disease severity measured using NOSPECS and proptosis remained unchanged.

Multivariate analysis using a linear mixed model with time-dependent covariates
revealed that lower TBII level at presentation correlated with future decrease in disease
activity and severity, whereas lower TSI level at presentation correlated with a
decrease in disease severity but not activity. Although time-dependent changes in TSI
and TBII levels did not show significant within-person correlation with clinical
characteristics, subgroup analysis of patients with CAS ≥4 at baseline showed that
within-person changes in TSI level affected CAS and NOSPECS. This indicated that
patients who showed more decrease in their serum TSI level over time showed more
reduction in their disease severity and activity during the 1-year follow-up.

Multiple cross-sectional studies have demonstrated the relationship between
serum TRAb level and TED. Regarding the relationship between TED development
and TRAb level, the antibody level showed a positive correlation with TED prevalence
in studies with newly diagnosed untreated Graves' disease. A study dividing the
cohort of patients according to TSI quartiles showed that TED prevalence of TED
prevalence. Graves' disease patients with TED had higher TSI level than patients
without TED, and TED patients with extraocular muscle enlargement had higher TSI
level than patients without myopathy. Furthermore, TRAb level showed a positive relationship with the clinical characteristics of TED. Both TSI and TBII levels are related to disease activity and proptosis, whereas only TSI level correlated significantly with the degree of eyelid swelling, proptosis, and extraocular muscle enlargement. In a study involving patients with newly diagnosed untreated TED, TSI levels were significantly higher in patients with active and severe TED than in patients with mild TED and inactive disease; in addition, TSI bioassay showed significantly higher overall positivity among patients with active TED than in those with TBII assay. Longitudinal studies on TRAb levels and the clinical characteristics of TED are limited. A retrospective study showed that patients with higher baseline TRAb (TSI and TBII) levels demonstrated a higher risk of severe disease courses after 1 year. However, this study did not investigate the relationship between changes in antibody level and clinical characteristics over time, as laboratory analysis was performed only once at the initial presentation. Another retrospective chart analysis of 23 patients revealed that changes in the inflammatory phase of TED statistically correlated with changes in measured TSI level, indicating that serial TSI measurements may be adjunct in assessing the clinical inflammatory activity of TED. Similar findings were reported in a prospective study by Eckstein et al., who demonstrated the association between TBII levels and clinical characteristics of TED, where TBII levels were higher in patients with a severe course of TED (CAS ≥4, NOSPECS ≥5) than in those with a mild course (CAS <4, NOSPECS <5). In this study, TBII level measurement at multiple time points during follow-up was performed. Results indicated that certain cut-off TBII values may be defined as good and bad prognostic markers at every time point of the disease after the first 4 months of the disease, demonstrating a prognostic value for measuring TBII levels at various time points. Similar to that in the present study, Eckstein et al. showed a decrease in TBII level over time but demonstrated no
The correlation between the extent of TBII level reduction over time and clinical characteristics.

In terms of the natural course of Graves’ disease, the disease may fluctuate in activity, and occasionally patients may spontaneously become euthyroid over time. Furthermore, treatment of the disease affected disease activity. Patients treated with surgery (thyroidectomy) or medication showed a gradual decrease in serum TRAb level, and after one year, 50–60% of patients had entered remission of TSH receptor autoimmunity with the disappearance of TRAb from the serum. Similarly, untreated TED improves spontaneously with time in most patients, which is described as ‘Rundles’s curve’. Furthermore, TRAb levels decrease with time in the natural history of TED. However, most of these studies were performed before the introduction of the bioassay; hence, little is known about the alteration in TSI levels with time. Furthermore, studies analyzing the correlation between changes in serum TRAb levels and clinical course over time are limited.

In this study, we assessed the correlation between the disease state of TED and decrease in the TRAb level over time. Correlations between these outcomes were disaggregated to within-person effect and inter-person effects using a linear mixed model with time-dependent covariates, as the measured outcomes over time within each patient are not independent of one another.

Inter-person effect implies that the lower the average TRAb level during the disease course, the more the reduction in dependent variables (CAS, NOSPECS, and proptosis) over time. Within-person effect indicates that dependent variables (CAS, NOSPECS, and proptosis) tend to decrease over time as TRAb levels decrease in individual patients during the disease course. Our analysis revealed that an inter-person difference in TRAb level correlated significantly with disease severity and activity, indicating that patients with low TSI and TBII levels show a reduction in...
disease severity and activity during 1 year of follow-up. These results are meaningful because the mean disease severity (measured with NOSPECS and proptosis) did not show statistically significant reductions. Disease severity is expected to decrease in the future in patients with low baseline TRAb (both TSI and TBII) levels.

Within-person analysis indicated that within-person changes in TRAb level in each individual correlated with the disease course over time. No statistically significant results were obtained in the within-person association analysis, which indicated that although patients with low TRAb showed better prognosis, the decrease in serum TRAb level in one individual patient over time did not affect the clinical course. However, subgroup analysis of patients with CAS ≥4 at presentation demonstrated that within-person changes in serum TSI level correlated significantly with CAS and NOSPECS, whereas baseline TSI level showed no statistically significant correlation with clinical characteristics (CAS and NOSPECS score) in the active subgroup. Thus, we inferred that if TSI levels decreased appropriately with time in patients with active disease at presentation, disease activity and severity will also decrease, leading to a better prognosis. The reduction in TSI levels over time has more value in evaluating the better prognosis of the disease than the lower initial TSI level.

The study has certain limitations. Owing to the observational nature, baseline clinical characteristics and treatment modalities during follow-up were not standardized, and hence, differences in management may have affected the clinical characteristics over time. Therefore, while using the linear mixed model, variables that may affect the disease course of TED, such as age, sex, treatment modality, and smoking, were included as covariates and the adjusted results were presented. Moreover, it was demonstrated that the treatment strategy of patients did not affect their TRAb levels. Another limitation of this study is that the subtyping of TED (lipogenic or myogenic subtype orbitopathy) could not be differentiated owing to the lack of imaging study.
Further, factors that can affect TRAb levels, such as treatment modalities of Graves’ disease and thyroid hormone levels of patients, were not considered in this analysis. Therefore, further interventional prospective studies with standardized management strategies and comprehensive evaluation of the Graves’ disease status are warranted.

Indicators for evaluating the activity and severity of thyroid ophthalmopathy have always been controversial. CAS and modified NOSPECS scores have been challenged as reliable measures of disease status as they sum the clinical features of the disease with equal weight. For example, some features may worsen, whereas others may improve as the disease progresses. Consequently, a summative score such as the modified NOSPECS score may mask the effect of the severity of any one measure of the disease and its change over time. However, they have been regarded as reliable measures of disease activity and severity in the clinical setting, and widely used as traditional indicators in evaluating activity and severity of TED. To avoid the limitations associated with the summative scoring system, proptosis was separately evaluated to measure disease severity.

In conclusion, the trend of TSI and TBII level reduction and clinical characteristics of TED were demonstrated in this study. TSI and TBII levels showed correlation with the change in clinical characteristics of TED over time. Furthermore, in specific patients (CAS ≥4 at presentation), time-dependent within-person changes in serum TSI level correlated with disease severity and activity. This indicated that active TED patients with less TSI level reduction over time would probably show prolonged disease activity and more severe disease. Hence, clinicians may consider more aggressive or prolonged immunosuppressive therapy.
Acknowledgments

The following investigators have contributed their data to this multicentre study: Sang In Khwarg (Seoul National University Hospital, Seoul, Korea); Koung Hoon Kook (Ajou University School of Medicine, Suwon, Korea); Sang Duck Kim (Wonkwang University School of Medicine, Iksan, Korea); Sehyun Baek (Korea University College of Medicine, Seoul, Korea); Jun Hyuk Son (Yeungnam University College of Medicine, Daegu, Korea); Suk-Woo Yang (Seoul St. Mary’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea); Jae Wook Yang (Busan Paik Hospital, Inje University College of Medicine, Busan, Korea); Kyung In Woo (Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea); Helen Lew (CHA Bundang Medical Center, CHA University, Seongnam, Korea); Jin Sook Yoon (Yonsei University College of Medicine, Seoul, Korea); Sung Bok Lee (Chungnam National University College of Medicine, Daejeon, Korea); Jeong Kyu Lee (Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, Korea); Jae Woo Jang (Kim’s Eye Hospital, Konyang University College of Medicine, Seoul, Korea); Hee Young Choi (Pusan National University College of Medicine, Busan, Korea); Mi Jung Chi (Gachon University Gil Medical Center, Incheon, Korea).

We gratefully acknowledge the statistical support of Eun Hwa Kim from the Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine. The biochemical analysis was supported by Dow Biomedica, Inc. (Seoul, Republic of Korea). The sponsor had no role in the design or conduct of this research, and no conflicting relationship exists for any author.

Author contributions

K I Woo and J W Yang conceived of the presented idea. K H Kook developed the theory and performed the computations. J Ko and JS Yoon verified the analytical methods and
wrote the main manuscript text. All authors provided critical feedback and helped shape
the research, analysis, and reviewed the manuscript.

**Conflicts of interest**

The authors declare no conflicts of interest.

**Funding**

This research received no specific grant from any funding agency in the public,
commercial or not-for-profit sectors

**Data Availability Statement**

The datasets generated during and/or analyzed during the current study are not publicly
available due to the IRB recommendation not to send patient data into the public domain
but are available from the corresponding author upon reasonable request.
REFERENCES

1. Bahn RS. Graves’ ophthalmopathy. *N Engl J Med* 2010;362(8):726-38. doi: 10.1056/NEJMr0905750

2. Naik VM, Naik MN, Goldberg RA, et al. Immunopathogenesis of thyroid eye disease: emerging paradigms. *Surv Ophthalmol* 2010;55(3):215-26. doi: 10.1016/j.survophthal.2009.06.009 [published Online First: 2010/04/14]

3. Jang SY, Shin DY, Lee EJ, et al. Relevance of TSH-receptor antibody levels in predicting disease course in Graves’ orbitopathy: comparison of the third-generation TBII assay and Mc4-TSI bioassay. *Eye (Lond)* 2013;27(8):964-71. doi: 10.1038/eye.2013.120

4. Neigel JM, Rootman J, Belkin RI, et al. Dysthyroid optic neuropathy. The crowded orbital apex syndrome. *Ophthalmology* 1988;95(11):1515-21. doi: 10.1016/s0161-6420(88)32978-7 [published Online First: 1988/11/01]

5. Khoo DH, Ho SC, Seah LL, et al. The combination of absent thyroid peroxidase antibodies and high thyroid-stimulating immunoglobulin levels in Graves’ disease identifies a group at markedly increased risk of ophthalmopathy. *Thyroid* 1999;9(12):1175-80. doi: 10.1089/thy.1999.9.1175 [published Online First: 2000/01/26]

6. Noh JY, Hamada N, Inoue Y, et al. Thyroid-stimulating antibody is related to Graves’ ophthalmopathy, but thyrotropin-binding inhibitor immunoglobulin is related to hyperthyroidism in patients with Graves’ disease. *Thyroid* 2000;10(9):809-13. doi: 10.1089/thy.2000.10.809 [published Online First: 2000/10/21]

7. Ponto KA, Kanitz M, Olivo PD, et al. Clinical relevance of thyroid-stimulating immunoglobulins in graves’ ophthalmopathy. *Ophthalmology* 2011;118(11):2279-85. doi: 10.1016/j.joptha.2011.03.030 [published Online First: 2011/06/21]

8. Gerding MN, van der Meer JW, Broenink M, et al. Association of thyrotrophin receptor antibodies with the clinical features of Graves’ ophthalmopathy. *Clin Endocrinol (Oxf)* 2000;52(3):267-71. doi: 10.1046/j.1365-2265.2000.00959.x [published Online First: 2000/03/16]

9. Woo YJ, Jang SY, Lim TH, et al. Clinical Association of Thyroid Stimulating Hormone Receptor Antibody Levels with Disease Severity in the Chronic Inactive Stage of Graves’ Orbitopathy. *Korean J Ophthalmol* 2015;29(4):213-9. doi: 10.3341/kjo.2015.29.4.213

10. Eckstein AK, Plicht M, Lax H, et al. Clinical results of anti-inflammatory therapy in Graves’ ophthalmopathy and association with thyroidal autoantibodies. *Clin Endocrinol (Oxf)* 2004;61(5):612-8. doi: 10.1111/j.1365-2265.2004.02143.x [published Online First: 2004/11/04]

11. Kampmann E, Diana T, Kanitz M, et al. Thyroid Stimulating but Not Blocking Autoantibodies Are Highly Prevalent in Severe and Active Thyroid-Associated

24
Orbitopathy: A Prospective Study. *Int J Endocrinol* 2015;2015:678194. doi: 10.1155/2015/678194 [published Online First: 2015/07/30]

12. Eckstein AK, Plicht M, Lax H, et al. Thyrotropin receptor autoantibodies are independent risk factors for Graves' ophthalmopathy and help to predict severity and outcome of the disease. *J Clin Endocrinol Metab* 2006;91(9):3464-70. doi: 10.1210/jc.2005-2813 [published Online First: 2006/07/13]

13. Cappelli C, Gandossi E, Castellano M, et al. Prognostic value of thyrotropin receptor antibodies (TRAb) in Graves' disease: a 120 months prospective study. *Endocr J* 2007;54(5):713-20. doi: 10.1507/endocrj.k06-069 [published Online First: 2007/08/07]

14. Sanders J, Evans M, Premawardhana LD, et al. Human monoclonal thyroid stimulating autoantibody. *Lancet* 2003;362(9378):126-8. doi: 10.1016/s0140-6736(03)13866-4 [published Online First: 2003/07/18]

15. Filetti S, Foti D, Costante G, et al. Recombinant human thyrotropin (TSH) receptor in a radioreceptor assay for the measurement of TSH receptor autoantibodies. *J Clin Endocrinol Metab* 1991;72(5):1096-101. doi: 10.1210/jcem-72-5-1096 [published Online First: 1991/05/01]

16. Lytton SD, Li Y, Olivo PD, et al. Novel chimeric thyroid-stimulating hormone-receptor bioassay for thyroid-stimulating immunoglobulins. *Clin Exp Immunol* 2010;162(3):438-46. doi: 10.1111/j.1365-2249.2010.04266.x [published Online First: 2010/11/13]

17. Lytton SD, Kahaly GJ. Bioassays for TSH-receptor autoantibodies: an update. *Autoimmun Rev* 2010;10(2):116-22. doi: 10.1016/j.autrev.2010.08.018 [published Online First: 2010/09/03]

18. Jang SY, Shin DY, Lee EJ, et al. Clinical characteristics of Graves' orbitopathy in patients showing discrepancy between levels from TBII assays and TSI bioassay. *Clin Endocrinol (Oxf)* 2014;80(4):591-7. doi: 10.1111/cen.12318 [published Online First: 2013/09/17]

19. Dragan LR, Seiff SR, Lee DC. Longitudinal correlation of thyroid-stimulating immunoglobulin with clinical activity of disease in thyroid-associated orbitopathy. *Ophthalmic Plast Reconstr Surg* 2006;22(1):13-9. doi: 10.1097/01.iop.0000192649.23508.f7 [published Online First: 2006/01/19]

20. European Group on Graves O, Wiersinga WM, Perros P, et al. Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. *Eur J Endocrinol* 2006;155(3):387-9. doi: 10.1530/eje.1.02230 [published Online First: 2006/08/18]

21. Bartalena L, Baldeschi L, Dickinson AJ, et al. Consensus statement of the European group on Graves' orbitopathy (EUGOGO) on management of Graves' orbitopathy. *Thyroid*
22. Guerra MW, Shults J, Amsterdam J, et al. The analysis of binary longitudinal data with time-dependent covariates. *Stat Med* 2012;31(10):931-48. doi: 10.1002/sim.4465 [published Online First: 2012/01/17]

23. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu Rev Psychol* 2011;62:583-619. doi: 10.1146/annurev.psych.093008.100356 [published Online First: 2009/07/07]

24. Hoffman L, Stawski RS. Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development* 2009;6(2-3):97-120.

25. Laurberg P, Wallin G, Tallstedt L, et al. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radiiodine: a 5-year prospective randomized study. *Eur J Endocrinol* 2008;158(1):69-75. doi: 10.1530/EJE-07-0450 [published Online First: 2008/01/02]

26. Codaccioni JL, Orgiazzi J, Blanc P, et al. Lasting remissions in patients treated for Graves' hyperthyroidism with propranolol alone: a pattern of spontaneous evolution of the disease. *J Clin Endocrinol Metab* 1988;67(4):656-62. doi: 10.1210/jcem-67-4-656 [published Online First: 1988/10/01]

27. Pinchera A, Liberti P, Martino E, et al. Effects of antithyroid therapy on the long-acting thyroid stimulator and the antithyroglobulin antibodies. *J Clin Endocrinol Metab* 1969;29(2):231-8. doi: 10.1210/jcem-29-2-231 [published Online First: 1969/02/01]

28. Takamura Y, Nakano K, Uruno T, et al. Changes in serum TSH receptor antibody (TRAb) values in patients with Graves' disease after total or subtotal thyroidectomy. *Endocr J* 2003;50(5):595-601. doi: 10.1507/endocrj.50.595 [published Online First: 2003/11/14]

29. Menconi F, Profilo MA, Leo M, et al. Spontaneous improvement of untreated mild Graves' ophthalmopathy: Rundle's curve revisited. *Thyroid* 2014;24(1):60-6. doi: 10.1089/thy.2013.0240 [published Online First: 2013/08/29]

30. Bartley GB. Rundle and his curve. *Arch Ophthalmol* 2011;129(3):356-8. doi: 10.1001/archophthalmol.2011.29 [published Online First: 2011/03/16]

31. Kim JW, Woo YJ, Yoon JS. Is modified clinical activity score an accurate indicator of diplopia progression in Graves' orbitopathy patients? *Endocr J* 2016;63(12):1133-40. doi: 10.1507/endocrj.EJ16-0165 [published Online First: 2016/09/16]

32. Douglas RS, Tsirbas A, Gordon M, et al. Development of criteria for evaluating clinical response in thyroid eye disease using a modified Delphi technique. *Arch Ophthalmol* 2009;127(9):1155-60. doi: 10.1001/archophthalmol.2009.232 [published Online First: 2009/09/16]

33. Frueh BR. Why the NOSPECS classification of Graves' eye disease should be abandoned.
with suggestions for the characterization of this disease. *Thyroid* 1992;2(1):85-8. doi: 10.1089/thy.1992.2.85 [published Online First: 1992/01/01]

34. Barrio-Barrio J, Sabater AL, Bonet-Farriol E, et al. Graves’ Ophthalmopathy: VISA versus EUGOGO Classification, Assessment, and Management. *J Ophthalmol* 2015;2015:249125. doi: 10.1155/2015/249125 [published Online First: 2015/09/10]

35. Jang SY, Shin DY, Lee EJ, et al. Correlation between TSH receptor antibody assays and clinical manifestations of Graves’ orbitopathy. *Yonsei Med J* 2013;54(4):1033-9. doi: 10.3349/ymj.2013.54.4.1033 [published Online First: 2013/05/28]

36. Eckstein AK, Lax H, Losch C, et al. Patients with severe Graves’ ophthalmopathy have a higher risk of relapsing hyperthyroidism and are unlikely to remain in remission. *Clin Endocrinol (Oxf)* 2007;67(4):607-12. doi: 10.1111/j.1365-2265.2007.02933.x [published Online First: 2007/09/21]
Longitudinal association of thyroid-stimulating immunoglobulin levels with clinical characteristics in thyroid eye disease

| Journal: | BMJ Open |
|----------|----------|
| Manuscript ID | bmjopen-2021-050337.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 25-Mar-2022 |
| Complete List of Authors: | Ko, JaeSang; Yonsei University College of Medicine, Department of Ophthalmology  
Kook, Koung Hoon; Ajou University School of Medicine, Department of Ophthalmology; Nune Eye Hospital  
Yoon, Jin-Sook; Yonsei University Hospital, Ophthalmology  
Woo, Kyung In; Samsung Medical Center, Ophthalmology  
Yang, Jae Wook; Inje University College of Medicine, Dept. of Ophthalmology |
| Primary Subject Heading: | Ophthalmology |
| Secondary Subject Heading: | Diabetes and endocrinology |
| Keywords: | Oculoplastics < OPHTHALMOLOGY, Thyroid disease < DIABETES & ENDOCRINOLOGY, Orbital and lacrimal disorders < OPHTHALMOLOGY |
I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd (“BMJ”) its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge (“APC”) for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author’s Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.
Longitudinal association of thyroid-stimulating immunoglobulin levels with clinical characteristics in thyroid eye disease

JaeSang Ko,¹ Koung Hoon Kook,²,³ Jin Sook Yoon,¹ Kyung In Woo,⁴* Jae Wook Yang⁵*, Korean Society of Ophthalmic Plastic & Reconstructive Surgery

¹Department of Ophthalmology, Severance Hospital, Institute of Vision Research, Yonsei University College of Medicine, Seoul, Republic of Korea
²Department of Ophthalmology, Ajou University School of Medicine, Suwon, Republic of Korea
³Nune Eye Hospital, Seoul, Republic of Korea
⁴Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
⁵Department of Ophthalmology, Inje University College of Medicine, Inje University Busan Paik Hospital, Busan, Republic of Korea

* Corresponding authors:

Kyung In Woo, MD, PhD
Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Ilwon-ro, Seoul 06351, Republic of Korea, E-mail: eyeminded@skku.edu

Jae Wook Yang, MD, PhD
Department of Ophthalmology, Inje University College of Medicine, Inje University Busan Paik Hospital, 75 Bokjiro, Busan 47392, Republic of Korea, E-mail: eyeyang@inje.ac.kr

Word count: 3461 words (excluding title page, abstract, references, figures, and tables.)
ABSTRACT

Objectives: The clinical course of thyroid eye disease (TED) is heterogeneous and predicting patients who may develop the severe sequelae of the disease is difficult. In this study, we evaluated the longitudinal association between changes in serum thyroid-stimulating hormone (TSH) receptor antibody (TRAb) levels and course of disease activity and severity over time.

Design: This was a multicenter, prospective, observational study.

Setting: Fifteen tertiary care oculoplastic service centers in Korea.

Participants: Seventy-six patients with newly diagnosed TED were included and followed up for 12 months.

Methods: We evaluated clinical characteristics and serum TRAb levels at baseline, 6, and 12 months of TED diagnosis. Additionally, we analyzed longitudinal associations between the serum TRAb levels and clinical activity score (CAS), NOSPECS (no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss) score, and proptosis.

Results: Thyroid-stimulating immunoglobulin (TSI) and TSH-binding inhibitory immunoglobulin (TBII) levels decreased during the 1-year follow-up, whereas disease activity measured using CAS decreased mainly in the first 6 months. Disease severity measured using NOSPECS score and proptosis remained unchanged. Moreover, inter-person differences in TBII levels were associated with CAS, NOSPECS score, and proptosis over time, whereas inter-person differences in TSI levels were associated with NOSPECS score. Subgroup analysis of patients with a baseline CAS ≥4 demonstrated that within-person changes in TSI levels affected the CAS and NOSPECS score.

Conclusions: Follow-up measurement of serum TSI and TBII levels may help evaluate TED prognosis and enable accurate clinical decision-making.
Keywords: thyroid eye disease; thyroid-stimulating immunoglobulin; TSH-binding inhibitory immunoglobulin; thyroid stimulating hormone receptor antibody

Strengths and limitations of this study

- This is the first study to investigate longitudinal correlation between thyroid-stimulating immunoglobulin levels and clinical characteristics in thyroid eye disease (TED) in a prospective setting.
- Correlations between TED status and decreases in thyroid-stimulating hormone receptor antibody (TRAb) level were disaggregated to within-person and inter-person effects using a linear mixed model with time-dependent covariates.
- This approach enabled assessment of correlations between autoantibody values at initial presentation and clinical characteristics at follow-ups, as well as correlations between time-dependent changes in autoantibody values and changes in clinical features.
- The findings of this study should be interpreted with caution given the observational nature of the study and lack of standardization of the baseline clinical characteristics and treatment modalities during follow-up, which may have affected the clinical characteristics over time.
INTRODUCTION

Thyroid eye disease (TED) is a component of autoimmune Graves’ hyperthyroidism, wherein the thyroid-stimulating hormone (TSH) receptor antibody (TRAb) stimulates orbital and periorbital tissues. The clinical course of TED is heterogeneous, and the natural course is benign in a considerable proportion of TED cases, although some cases may present with substantially severe sequelae. Indications that can assist clinicians in determining TED prognosis are limited, such that ascertaining the patients who can develop the potentially blinding sequelae of TED is difficult.

Relationships between TED-related serum TRAb levels and clinical characteristics have been elucidated. Serum TRAb levels correlate directly with the clinical features of TED, with high TRAb level correlating with the disease prevalence and status. Additionally, patients with higher baseline TRAb levels demonstrated a higher risk of severe disease course during a 1-year follow-up. However, TRAb levels differ over time and decrease to a large extent in some patients but not others. Most studies have measured serum TRAb level only at the baseline; however, using only baseline values results in the loss of information provided by the serum TRAb level, which changes over time thereby limiting the assessment of relationships between thyroid autoantibodies and TED.

Currently, two established assays are available for measuring the TRAb levels: the competitive TSH-binding inhibitory immunoglobulin (TBII) assay and functional thyroid-stimulating immunoglobulin (TSI) bioassay. The former uses the ability of TRAb to inhibit the binding of radiolabeled TSH to TSH receptors, and the latter measures cyclic adenosine monophosphate production after TRAb binds to the TSH receptor, enabling functional TRAb identification. Several studies evaluating the correlation between TRAb levels and the clinical course of TED revealed different predictive values between TBII and TSI levels, with most showing that disease activity and TED severity are more closely correlated with the TSI bioassay than the TBII assay.
Longitudinal studies assessing the association between TRAb levels and the clinical course of TED are limited. The inflammatory phase of TED reportedly correlates with changes in measured TSI levels, indicating that serial measurements of TSI level may be an adjunct in assessing the clinical inflammatory activity of TED; additionally, follow-up measurements of TBII levels also reportedly allow assessment of TED prognosis. However, the association and potential predictive value of TSI levels during the clinical course of TED have not yet been demonstrated in a prospective setting.

Herein, we investigated whether changes in serum TRAb (TBII and TSI) levels over time are associated with TED activity and severity during the disease course. The findings revealed correlations between time-dependent changes in TRAb levels in an individual patient and the clinical course of the disease.

METHODS

Patients

This multicenter, prospective, observational study was conducted by the Korean Society of Ophthalmic Plastic and Reconstructive Surgery from April 2017 to February 2019. During the study period, 91 patients with Graves’ disease and TED were enrolled at the 15 participating Korean tertiary care oculoplastic service centers. The complete datasets and written informed consent of each patient were obtained before starting the study. Approval by the institutional ethics review committee of each participating institution was obtained, and the study was performed according to the tenets of the Declaration of Helsinki.

Patients with clinically overt TED at <6 months after symptom onset were recruited. Patients with a previous history of TED treatment (except conservative treatments, such as the use of artificial tears), uncertain cases, and patients aged <19 years were excluded.

Clinical data were recorded, and a blood sample was taken upon study inclusion and at 6 and 12 months after inclusion. After enrollment, the study participants received the
necessary management for TED at the discretion of each clinician, including conservative
management, systemic steroid administration, and/or orbital radiotherapy. No surgical
procedure was performed on the enrolled patients during the study period.

Disease activity and severity were assessed by an oculoplastic specialist blinded to
the laboratory data at each participating institute. The general ophthalmic assessment
included an examination of anterior and posterior eye segments, measurement of proptosis
using Hertel exophthalmometry, and evaluation of ocular movement. The clinical activity
score (CAS; range: 0–7) was calculated based on the classic signs of inflammation and
comprised seven items. Disease severity was assessed using the modified NOSPECS (no
signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle
involvement, corneal involvement, sight loss) score, which was calculated as the sum of
each class present and ranged from 1 to 18. For each patient, the more severely affected
eye (one with more proptosis at enrolment or with a higher CAS in cases of similar extents of
proptosis in both eyes) was selected for the study.

TRAb assays

The blood sample collected at each visit was centrifuged at each participating institution and
transferred to a central laboratory for analysis. TRAb levels were measured using a third-
generation TBII assay with the automated Cobas electrochemiluminescence immunoassay
system (Elecsys; Roche Diagnostics GmbH, Penzberg, Germany) according to manufacturer
instructions. The cutoff value of positivity using this system was 1.75 IU/L. Serum Mc4-TSI
level was measured using the Thyretain TSI reporter bioassay (Diagnostic Hybrids, Inc.,
Athens, OH, USA) according to manufacturer instructions. Results were considered positive
at a specimen-to-reference ratio >140% relative to the reference control.

Data analysis and statistical analysis
Data were statistically analyzed to determine correlations between changes in clinical indicators of TED, as reflected in the NOSPECS score, CAS, and proptosis, and change in measured serum TRAb levels over time in individual patients. We assessed the effect of serum autoantibody levels on clinical TED activity and severity and investigated alterations in serum autoantibody levels and clinical parameters over time using a linear mixed model. We used a linear mixed-effects model with an unstructured covariance structure to analyze the data. Additionally, we adopted two approaches for modeling repeatedly measured outcomes. First, we analyzed the data using a linear mixed model with time-dependent covariates to show the effect of baseline serum autoantibody levels on clinical characteristics over time. Time-dependent covariates are independent variables that include both within-subject and between-subject variations and can be used to make comparisons across populations and describe time trends and dynamic relationships between covariate and response. Patient age, sex, smoking status, and treatment modality (conservative, systemic steroid administration, and systemic steroid administration with orbital radiotherapy) were included as covariates in the model. Second, serum autoantibody levels were used as a covariate, including the baseline level and all repeated measurements. One particularly salient strength offered by longitudinal data is the ability to disaggregate between-person and within-person effects in the regression of an outcome on a time-varying covariate. This allowed the separation of time-varying within-person changes from time-invariant between-person (inter-person) differences in TRAb levels based on the clinical characteristics of the disease. Moreover, this approach enabled evaluation of correlations between autoantibody levels at initial presentation and clinical characteristics at further follow-ups, as well as correlations between time-dependent changes in autoantibody levels and changes in clinical characteristics. Analyses were performed using SAS software (v9.4; SAS Institute, Inc., Cary, NC, USA), and a two-sided $P < 0.05$ was considered statistically significant.
Patients with an initial CAS $\geq 4$ were included in the active subgroup and subjected to subgroup analysis, which included time-dependent changes in the clinical characteristics and laboratory results and time-dependent correlation analysis.

**Patient and public involvement**

No patients were involved.

**RESULTS**

**Clinical characteristics and TRAb levels**

Of the 91 patients enrolled in this study, 76 completed three clinical and laboratory evaluations. The clinical and biochemical characteristics of the 76 patients with TED are shown in Table 1. During the 12-month follow-up period, 24 patients (32%) were conservatively managed, 34 (45%) were prescribed systemic steroid administration (oral or intravenous), and 18 (24%) were managed with combined orbital radiation and systemic steroid administration. Using systemic steroids did not affect the serum TSI or TBII levels of the patients ($P = 0.4225$ and $0.9634$, respectively), and TSI and TBII levels decreased significantly during the 12-month follow-up. Regarding clinical characteristics, disease activity measured with CAS decreased during the 12-month follow-up period, although disease severity measured by NOSPECS score and proptosis remained unchanged.

**Table 1. Clinical and biochemical characteristics of the study population**

| Variables      | Baseline | 6 months | 12 months | $P$  |
|----------------|----------|----------|-----------|------|
| Age, y         | 42.4±12.0|          |           |      |
| Sex            |          |          |           |      |
| Female, n (%)  |          | 47 (61.84)|          |      |
| Male, n (%)    |          | 29 (38.16)|          |      |
| Smoking status |          |          |           |      |
Correlation between baseline TRAb levels and clinical characteristics

Table 2 shows the results of correlation analysis using a linear mixed model with time-dependent covariates between baseline serum TRAb levels and the clinical characteristics of TED. Lower baseline serum TBII level was associated with a future decrease in TED activity, and reduced TED severity (NOSPECS score and proptosis) was associated with lower baseline serum TSI and TBII levels.

### Table 2. Correlation between baseline serum TRAb levels and future changes in clinical characteristics

|        | CAS          | NOSPECS      | Proptosis  |
|--------|--------------|--------------|------------|
| TSI    | 0.000 (0.001)| 0.003 (0.001)| 0.004 (0.002)| 0.055 |
| TBII   | 0.029 (0.010)| 0.042 (0.019)| 0.062 (0.030)| 0.041 |

P-values were calculated using a linear mixed model with time-dependent covariates adjusted for age, sex, smoking status, and treatment modality. Numbers in bold denote statistically significant changes.

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; SRR, specimen-to-reference ratio; TSI, thyroid-stimulating immunoglobulin; TBII, thyroid-stimulating hormone receptor-binding inhibitory immunoglobulin.
Correlation between time-dependent changes in clinical characteristics and TRAb levels

Table 3 shows the effect of the time-dependent changes in TRAb levels on the clinical characteristics of TED. The effect of longitudinally measured TRAb levels over time was disaggregated to inter-person differences and within-person changes using a linear mixed model with time-dependent covariates. TSI levels measured at multiple time points showed inter-person relationships with serial changes in NOSPECS score, indicating that patients with a low average serum TSI levels during the 1-year follow-up showed a greater reduction in their NOSPECS score over time. Additionally, TBII level showed an inter-person relationship with CAS and the NOSPECS score, indicating that patients with higher TBII levels during the 1-year follow-up demonstrated prolonged disease activity and severity. Within-person changes in TSI and TBII levels did not show a statistically significant correlation with disease activity and severity.
### Table 3. Effect of inter-person differences and within-person changes in TRAb levels on clinical characteristics: longitudinal associations during a 1-year follow-up

| Variables | Estimated change (SE) | \(P^\dagger\) |
|-----------|-----------------------|----------------|
| **Effect of TSI on clinical characteristics** | | |
| CAS       |                       |               |
| Inter-person | 0.0013 (0.0007) | 0.0731 |
| Within-person | 0.0011 (0.0008) | 0.1510 |
| NOSPECS   |                       |               |
| Inter-person | 0.0043 (0.0013) | 0.0013 |
| Within-person | 0.0002 (0.0007) | 0.7700 |
| Proptosis  |                       |               |
| Inter-person | 0.0032 (0.0022) | 0.1410 |
| Within-person | -0.0010 (0.0007) | 0.1796 |
| **Effect of TBII on clinical characteristics** | | |
| CAS       |                       |               |
| Inter-person | 0.0392 (0.0109) | 0.0004 |
| Within-person | 0.0261 (0.0173) | 0.1334 |
| NOSPECS   |                       |               |
| Inter-person | 0.0520 (0.0242) | 0.0335 |
| Within-person | 0.0249 (0.0152) | 0.1036 |
| Proptosis  |                       |               |
| Inter-person | 0.0775 (0.0379) | 0.0428 |
| Within-person | 0.0017 (0.0167) | 0.3095 |

\(P^\dagger\)-values were calculated using a linear mixed model with time-dependent covariates adjusted for age, sex, smoking status, and treatment modality. Numbers in bold denote statistically significant changes.

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; TBII, thyroid-stimulating hormone receptor-binding inhibitory immunoglobulin;
Correlation between time-dependent changes in clinical characteristics and TRAb levels: subgroup analysis

Twenty-five patients with an initial CAS ≥4 were included in the subgroup analysis, with this group defined as the active subgroup and other patients included in the inactive subgroup. Subgroup analysis was performed to evaluate the time-dependent correlations between clinical characteristics of TED and TRAb levels in the active subgroup. Table 4 shows that clinical and biochemical characteristics of the active subgroup. This subgroup had more females and smokers than the inactive subgroup (both \( P < 0.0001 \)). During 12 months of follow-up, 10 patients (40%) were prescribed systemic steroid administration (oral or intravenous) and 15 (60%) were managed with combined orbital radiation and systemic steroid administration. Using systemic steroids did not affect the serum TSI and TBII levels of patients (\( P = 0.9008 \) and 0.4957, respectively). Additionally, baseline the TBII level was higher in the active than in the inactive subgroups (9.5 ± 9.5 IU/L vs. 17.0 ±14.2 IU/L, \( P = 0.0125 \)), whereas the baseline TSI level of the active subgroup was similar to that of the inactive subgroup (443.9 ± 210.5% vs. 453.2 ± 123.6%, \( P = 0.8096 \)). Moreover, the active subgroup had a higher CAS (4.4 ± 0.6 vs. 1.9 ± 1.0, \( P < 0.0001 \)), higher NOSPECS score (5.6 ± 2.3 vs. 3.1 ± 1.6, \( P < 0.0001 \)), and more proptosis (19.2 ± 3.5 mm vs. 16.9 ± 2.6 mm, \( P = 0.0020 \)) than the inactive group. Correlation analysis between baseline TRAb levels and clinical characteristics in the active subgroup showed that lower baseline TBII levels correlated with a decrease in CAS and NOSPECS score during the 1-year follow-up [estimated change (CAS and NOSPECS): 0.0523 (\( P = 0.0078 \)) and 0.0566 (\( P = 0.0394 \)), respectively]. Furthermore, baseline TSI level did not show a statistically significant correlation with clinical characteristics in the active subgroup.
Table 4. Clinical and biochemical characteristics: subgroup analysis of patients with active TED at baseline 
(n = 25)

| Variables          | Baseline   | 6 months   | 12 months   | P         |
|--------------------|------------|------------|-------------|-----------|
| Age, y             | 43.6±11.84 |            |             |           |
| Sex                |            |            |             |           |
| Female, n (%)      | 10 (40.0)  |            |             |           |
| Male, n (%)        | 15 (60.0)  |            |             |           |
| Smoking status     |            |            |             |           |
| Non-smoker, n (%)  | 17 (68.0)  |            |             |           |
| Smoker, n (%)      | 8 (32.0)   |            |             |           |
| TSI, SRR (%)       | 453.2±123.6| 342.5±139.7| 297.9±137.0 | <.0001†   |
| TBII, IU/L         | 17.0±14.2  | 9.4±11.6   | 7.6±10.4    | <.0001†   |
| CAS                | 4.4±0.6    | 1.7±1.3    | 1.7±1.7     | <.0001†   |
| NOSPECS score      | 5.6±2.3    | 5.0±2.3    | 4.9±2.4     | 0.0877†   |
| Proptosis, mm      | 19.2±3.5   | 19.7±3.4   | 19.6±3.9    | 0.3414†   |

† P-values were calculated between three visits using a linear mixed model.

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; SRR, specimen-to-reference ratio; TBII, thyroid-stimulating hormone-binding inhibitory immunoglobulin; TED, thyroid eye disease; TSI, thyroid-stimulating immunoglobulin.

Table 5 shows the effect of the time-dependent changes in TRAb levels on the clinical characteristics of TED in the active subgroup. The effect of longitudinally measured TRAb level over time was disaggregated to inter-person differences and within-person changes using a linear mixed model with time-dependent covariates. Within-person changes in TSI level measured at multiple time points affected the CAS and NOSPECS score, whereas inter-person differences in TSI level was unrelated to clinical characteristics. Thus, patients who showed greater reductions in TSI level during follow-up also showed greater
reductions in their CAS and NOSPECS score during the 1-year follow-up. By contrast, inter-
person differences in TBII level correlated with the CAS and NOSPECS score, whereas
within-person changes in TBII level were independent of disease severity and activity.
Table 5. Effects of inter-person differences and within-person changes in TRAb levels on clinical characteristics: subgroup analysis of patients with active TED at baseline \((n = 25)\)

| Variables | Estimated change \((SE)\) | \(P^\dagger\) |
|-----------|-----------------|-----------|
| **Effect of TSI on clinical characteristics** | | |
| CAS |  | |
| Inter-person | -0.0002 (0.0020) | 0.9347 |
| Within-person | **0.0046 (0.0016)** | **0.0064** |
| NOSPECS |  | |
| Inter-person | 0.0014 (0.0031) | 0.6559 |
| Within-person | **0.0034 (0.0016)** | **0.0385** |
| Proptosis |  | |
| Inter-person | 0.0016 (0.0070) | 0.8178 |
| Within-person | 0.0014 (0.0018) | 0.4464 |
| **Effect of TBII on clinical characteristics** | | |
| CAS |  | |
| Inter-person | **0.0470 (0.0146)** | **0.0023** |
| Within-person | 0.0243 (0.0284) | 0.3959 |
| NOSPECS |  | |
| Inter-person | **0.0517 (0.0245)** | **0.0402** |
| Within-person | 0.0397 (0.0261) | 0.1354 |
| Proptosis |  | |
| Inter-person | 0.1067 (0.0566) | 0.0657 |
| Within-person | -0.0046 (0.0290) | 0.8742 |

\(\dagger\) \(P\)-values were calculated using a linear mixed model with time-dependent covariates adjusted for age, sex, smoking status, and treatment modality. Numbers in bold denote statistically significant changes.

CAS, clinical activity score; NOSPECS, no signs or symptoms, only signs, soft tissue involvement, proptosis, extraocular muscle involvement, corneal involvement, sight loss; TBII, thyroid-stimulating hormone receptor-binding inhibitory immunoglobulin;
**DISCUSSION**

This prospective, observational study involving a large cohort of patients with TED demonstrated the clinical relevance of longitudinal follow-up of both TSI and TBII levels. Both TSI and TBII levels decreased sustainably during the 1-year follow-up, whereas disease activity measured using CAS decreased mainly within the first 6 months and stabilized thereafter. Notably, disease severity measured using NOSPECS score and proptosis remained unchanged. Additionally, multivariate analysis using a linear mixed model with time-dependent covariates revealed that lower TBII level at presentation correlated with future decreases in disease activity and severity, whereas lower TSI level at presentation correlated with a decrease in disease severity but not activity. Although time-dependent changes in TSI and TBII levels did not show significant within-person correlations with clinical characteristics, subgroup analysis of patients with a CAS ≥4 at baseline showed that within-person changes in TSI level affected the CAS and NOSPECS score. These findings indicated that patients who demonstrated greater decreases in serum TSI levels over time showed greater reductions in disease severity and activity during the 1-year follow-up.

Multiple cross-sectional studies have demonstrated a relationship between serum TRAb level and TED. Regarding the relationship between TED development and TRAb level, antibody levels showed a positive correlation with TED prevalence according to studies of patients with newly diagnosed, untreated Graves’ disease. Another study using a patient cohort divided according to TSI quartiles showed that TED prevalence increased with each quartile of TSI levels. Additionally, a previous report indicated that patients with both Graves’ disease and TED showed higher TSI levels relative to those without TED, and
that TED patients with extraocular muscle enlargement also showed higher TSI levels than
patients without myopathy, as well as a positive relationship between TRAb levels and the
clinical characteristics of TED. Moreover, both TSI and TBII levels are reportedly associated
with disease activity and proptosis, whereas only TSI level correlated significantly with the
degree of eyelid swelling, proptosis, and extraocular muscle enlargement. In a study
involving patients with newly diagnosed and untreated TED, TSI levels were significantly
higher in patients with active and severe TED than in those with mild TED and inactive
disease, with TSI bioassays showed significantly higher overall positivity among patients
with active TED relative to results generated using the TBII assay.

Longitudinal studies on TRAb levels and the clinical characteristics of TED are
limited. A retrospective study showed that patients with higher baseline TRAb (TSI and TBII)
levels demonstrated a higher risk of severe disease course after 1 year; however, that study
did not investigate the relationship between changes in antibody level and clinical
characteristics over time, as laboratory analysis was performed only once at the initial
presentation. Another retrospective chart analysis of 23 patients revealed that changes in
the inflammatory phase of TED statistically correlated with changes in measured TSI level,
indicating that serial TSI measurements may be adjunct in assessing the clinical
inflammatory activity of TED. Similar findings were reported in a prospective study by
Eckstein et al. who demonstrated an association between TBII level and the clinical
characteristics of TED, with TBII levels higher in patients with a severe course of TED (CAS
≥4, NOSPECS score ≥5) relative to those with a mild course (CAS <4, NOSPECS score
<5). In this study, patients underwent measurements of TBII levels at multiple time points
during follow-up. The results revealed that certain cut-off TBII values could be defined as
good or bad prognostic markers at each time point of development after the first 4 months of
disease onset, demonstrating the prognostic value of measuring TBII levels at various time
points. Similarly, Eckstein et al. also reported a decrease in TBII level over time but
demonstrated no correlation between the extent of this reduction and clinical characteristics.

In terms of the natural course of Graves’ disease, the disease may fluctuate in activity, and occasionally patients can spontaneously become euthyroid over time. Treatment of the disease affects disease activity. Patients treated with surgery (thyroidectomy) or medication show a gradual decrease in serum TRAb level, and after one year, 50% to 60% of patients enter remission of TSH receptor autoimmunity accompanied by the disappearance of TRAb from serum. Similarly, untreated TED improves spontaneously over time in most patients (described as “Rundle’s curve”). Moreover, studies report that TRAb levels associated with TED also decrease over time. However, most of these studies were performed before the introduction of the bioassay; therefore, little is known about alterations in TSI levels over time. Furthermore, studies analyzing the correlation between changes in serum TRAb levels and clinical course over time are limited.

In this study, we assessed correlations between the TED status and decreases in TRAb levels over time. Correlations between these outcomes were disaggregated to within-person and inter-person effects using a linear mixed model with time-dependent covariates, as the measured outcomes over time within each patient are not independent of one another.

Inter-person effects imply that lower average TRAb levels during the disease course correlate a greater reduction in dependent variables (CAS, NOSPECS score, and proptosis) over time. Within-person effects suggest that dependent variables (CAS, NOSPECS score, and proptosis) tend to decrease over time along with decreases in TRAb levels in individual patients during the disease course. The present analysis revealed that an inter-person difference in TRAb level correlated significantly with disease severity and activity, indicating that patients with low TSI and TBII levels showed a reduction in disease severity and activity during the 1-year of follow-up. These results are meaningful, because the mean disease severity (measured using NOSPECS score and proptosis) did not show statistically
significant reductions. The findings suggest that low baseline TRAb (both TSI and TBII) levels might be predictive of future decreases in disease severity in these patients.

Within-person analysis indicated that within-person changes in TRAb level in each individual correlated with changes in the disease course over time. No statistically significant results were obtained in this analysis, suggesting that although patients with low TRAb levels showed better prognosis, the decrease in serum TRAb level in one individual patient over time did not affect the clinical course. However, subgroup analysis of patients with a CAS ≥4 at presentation demonstrated that within-person changes in serum TSI level correlated significantly with the CAS and NOSPECS score, whereas baseline TSI level showed no statistically significant correlation with clinical characteristics (CAS and NOSPECS score) in the active subgroup. Thus, we inferred that if TSI levels decreased appropriately over time in patients with active disease at presentation, disease activity and severity will also likely decrease, leading to a better prognosis. These findings suggest that reductions in TSI levels over time have a better prognostic value for evaluating improvements in disease status than a lower baseline TSI level.

This study has limitations. Given the observational nature of the study and the lack of standardization of the baseline clinical characteristics and treatment modalities during follow-up, differences in management may have affected the clinical characteristics over time. Therefore, variables that could affect the disease course of TED, such as age, sex, treatment modality, and smoking, were included as covariates in the linear mixed model, and the adjusted results were presented. Another limitation is that the inclusion criteria were relatively broad, and the subtyping of TED (lipogenic or myogenic subtype orbitopathy) could not be differentiated owing to the lack of imaging data. Furthermore, factors that can affect TRAb levels, such as treatment modalities of Graves' disease and thyroid hormone levels of patients, were not considered in this analysis. Therefore, further interventional prospective studies with standardized management strategies and comprehensive evaluation of Graves'
disease status are warranted.

Indicators for evaluating the activity and severity of thyroid ophthalmopathy have always been controversial. CAS and the modified NOSPECS score have been challenged as reliable measures of disease status, as they sum the clinical features of the disease with equal weight. For example, some features may worsen, whereas others may improve as the disease progresses. Consequently, a summative score, such as the modified NOSPECS score, might mask the effect of the severity of any one measure of the disease and its change over time. However, these scores have been regarded as reliable measures of disease activity and severity in the clinical setting and widely used as traditional indicators in evaluating TED activity and severity. To avoid the limitations associated with the summative scoring system, we separately evaluated proptosis to measure disease severity.

In conclusion, we demonstrated correlated trends associated with reductions in TSI and TBII levels and changes in clinical characteristics of TED over time. Furthermore, in patients with active TED (a CAS \( \geq 4 \) at presentation), time-dependent within-person changes in serum TSI level correlated with disease severity and activity, indicating that active TED patients with smaller reductions in TSI levels over time would likely show prolonged disease activity and more severe disease. These findings offer clinicians insight to inform considerations of more aggressive or prolonged immunosuppressive therapy for TED.
Acknowledgments

The following investigators contributed their data to this multicenter study: Sang In Khwarg (Seoul National University Hospital, Seoul, Korea); Kyoung Hoon Kook (Ajou University School of Medicine, Suwon, Korea); Sang Duck Kim (Wonkwang University School of Medicine, Iksan, Korea); Sehyun Baek (Korea University College of Medicine, Seoul, Korea); Jun Hyuk Son (Yeungnam University College of Medicine, Daegu, Korea); Suk-Woo Yang (Seoul St. Mary’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea); Jae Wook Yang (Busan Paik Hospital, Inje University College of Medicine, Busan, Korea); Kyung In Woo (Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea); Helen Lew (CHA Bundang Medical Center, CHA University, Seongnam, Korea); Jin Sook Yoon (Yonsei University College of Medicine, Seoul, Korea); Sung Bok Lee (Chungnam National University College of Medicine, Daejeon, Korea); Jeong Kyu Lee (Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, Korea); Jae Woo Jang (Kim’s Eye Hospital, Konyang University College of Medicine, Seoul, Korea); Hee Young Choi (Pusan National University College of Medicine, Busan, Korea); Mi Jung Chi (Gachon University Gil Medical Center, Incheon, Korea).

We gratefully acknowledge the statistical support of Eun Hwa Kim from the Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine. Biochemical analyses were supported by Dow Biomedica, Inc. (Seoul, Republic of Korea).

Author contributions

K.I.W. and J.W.Y. conceived of the study; K.H.K. developed and performed the analyses; J.K. and J.S.Y verified the analytical methods and wrote the main manuscript; all authors provided critical feedback and participated in the research, analysis, and review of the manuscript.
Competing Interests

The authors declare no conflicts of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Data availability statement

The datasets generated and/or analyzed during this study are not publicly available due to the Institutional Review Boards’ recommendation not to send patient data into the public domain but are available from the corresponding author upon reasonable request.

Ethics statements

Patient consent for publication

Not applicable.

Ethics approval

This study involves human participants and was approved by the institutional review board of the Severance Hospital, Yonsei University College of Medicine (Seoul, Korea) approved the study (approval number 4-2016-1117). Participants gave written informed consent to participate in the study before taking part.
REFERENCES

1. Bahn RS. Graves' ophthalmopathy. *N Engl J Med* 2010;362(8):726-38. doi: 10.1056/NEJMr0905750

2. Naik VM, Naik MN, Goldberg RA, et al. Immunopathogenesis of thyroid eye disease: emerging paradigms. *Surv Ophthalmol* 2010;55(3):215-26. doi: 10.1016/j.survophthal.2009.06.009 [published Online First: 2010/04/14]

3. Jang SY, Shin DY, Lee EJ, et al. Relevance of TSH-receptor antibody levels in predicting disease course in Graves' orbitopathy: comparison of the third-generation TBII assay and Mc4-TSI bioassay. *Eye (Lond)* 2013;27(8):964-71. doi: 10.1038/eye.2013.120

4. Neigel JM, Rootman J, Belkin RI, et al. Dysthyroid optic neuropathy. The crowded orbital apex syndrome. *Ophthalmology* 1988;95(11):1515-21. doi: 10.1016/s0161-6420(88)32978-7 [published Online First: 1988/11/01]

5. Khoo DH, Ho SC, Seah LL, et al. The combination of absent thyroid peroxidase antibodies and high thyroid-stimulating immunoglobulin levels in Graves' disease identifies a group at markedly increased risk of ophthalmopathy. *Thyroid* 1999;9(12):1175-80. doi: 10.1089/thy.1999.9.1175 [published Online First: 2000/01/26]

6. Noh JY, Hamada N, Inoue Y, et al. Thyroid-stimulating antibody is related to Graves' ophthalmopathy, but thyrotropin-binding inhibitor immunoglobulin is related to hyperthyroidism in patients with Graves' disease. *Thyroid* 2000;10(9):809-13. doi: 10.1089/thy.2000.10.809 [published Online First: 2000/10/21]

7. Ponto KA, Kanitz M, Olivo PD, et al. Clinical relevance of thyroid-stimulating immunoglobulins in graves' ophthalmopathy. *Ophthalmology* 2011;118(11):2279-85. doi: 10.1016/j.ophtha.2011.03.030 [published Online First: 2011/06/21]

8. Gerding MN, van der Meer JW, Broenink M, et al. Association of thyrotrophin receptor antibodies with the clinical features of Graves' ophthalmopathy. *Clin Endocrinol (Oxf)* 2000;52(3):267-71. doi: 10.1046/j.1365-2265.2000.00959.x [published Online First: 2000/03/16]

9. Woo YJ, Jang SY, Lim TH, et al. Clinical Association of Thyroid Stimulating Hormone Receptor Antibody Levels with Disease Severity in the Chronic Inactive Stage of Graves' Orbitopathy. *Korean J Ophthalmol* 2015;29(4):213-9. doi: 10.3341/kjo.2015.29.4.213

10. Eckstein AK, Plicht M, Lax H, et al. Clinical results of anti-inflammatory therapy in Graves' ophthalmopathy and association with thyroidal autoantibodies. *Clin Endocrinol (Oxf)* 2004;61(5):612-8. doi: 10.1111/j.1365-2265.2004.02143.x [published Online First: 2004/11/04]

11. Kampmann E, Diana T, Kanitz M, et al. Thyroid Stimulating but Not Blocking Autoantibodies Are Highly Prevalent in Severe and Active Thyroid-Associated Orbitopathy: A Prospective Study. *Int J Endocrinol* 2015;2015:678194. doi: 10.1155/2015/678194 [published Online
12. Eckstein AK, Plicht M, Lax H, et al. Thyrotropin receptor autoantibodies are independent risk factors for Graves' ophthalmopathy and help to predict severity and outcome of the disease. *J Clin Endocrinol Metab* 2006;91(9):3464-70. doi: 10.1210/jc.2005-2813 [published Online First: 2006/07/13]

13. Cappelli C, Gandossi E, Castellano M, et al. Prognostic value of thyrotropin receptor antibodies (TRAb) in Graves' disease: a 120 months prospective study. *Endocr J* 2007;54(5):713-20. doi: 10.1507/endocrj.k06-069 [published Online First: 2007/08/07]

14. Sanders J, Evans M, Premawardhana LD, et al. Human monoclonal thyroid stimulating autoantibody. *Lancet* 2003;362(9378):126-8. doi: 10.1016/s0140-6736(03)13866-4 [published Online First: 2003/07/18]

15. Filetti S, Foti D, Costante G, et al. Recombinant human thyrotropin (TSH) receptor in a radioreceptor assay for the measurement of TSH receptor autoantibodies. *J Clin Endocrinol Metab* 1991;72(5):1096-101. doi: 10.1210/jcem-72-5-1096 [published Online First: 1991/05/01]

16. Lytton SD, Li Y, Olivo PD, et al. Novel chimeric thyroid-stimulating hormone-receptor bioassay for thyroid-stimulating immunoglobulins. *Clin Exp Immunol* 2010;162(3):438-46. doi: 10.1111/j.1365-2249.2010.04266.x [published Online First: 2010/11/13]

17. Lytton SD, Kahaly GJ. Bioassays for TSH-receptor autoantibodies: an update. *Autoimmun Rev* 2010;10(2):116-22. doi: 10.1016/j.autrev.2010.08.018 [published Online First: 2010/09/03]

18. Jang SY, Shin DY, Lee EJ, et al. Clinical characteristics of Graves' orbitopathy in patients showing discrepancy between levels from TBII assays and TSI bioassay. *Clin Endocrinol (Oxf)* 2014;80(4):591-7. doi: 10.1111/cen.12318 [published Online First: 2013/09/17]

19. Dragan LR, Seiff SR, Lee DC. Longitudinal correlation of thyroid-stimulating immunoglobulin with clinical activity of disease in thyroid-associated orbitopathy. *Ophthalmic Plast Reconstr Surg* 2006;22(1):13-9. doi: 10.1097/01.iop.0000192649.23508.f7 [published Online First: 2006/01/19]

20. European Group on Graves O, Wiersinga WM, Perros P, et al. Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. *Eur J Endocrinol* 2006;155(3):387-9. doi: 10.1530/eje.1.02230 [published Online First: 2006/08/18]

21. Bartalena L, Baldeschi L, Dickinson AJ, et al. Consensus statement of the European group on Graves' orbitopathy (EUGOGO) on management of Graves' orbitopathy. *Thyroid* 2008;18(3):333-46. doi: 10.1089/thy.2007.0315 [published Online First: 2008/03/18]

22. Guerra MW, Shults J, Amsterdam J, et al. The analysis of binary longitudinal data with time-dependent covariates. *Stat Med* 2012;31(10):931-48. doi: 10.1002/sim.4465 [published Online First: 2012/01/17]

23. Lalonde TL. Modeling time-dependent covariates in longitudinal data analyses. *Innovative Biostatistics* 2004;1:1-20.
statistical methods for public health data: Springer 2015:57-79.

24. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu Rev Psychol* 2011;62:583-619. doi: 10.1146/annurev.psych.093008.100356 [published Online First: 2009/07/07]

25. Hoffman L, Stawski RS. Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development* 2009;6(2-3):97-120.

26. Laurberg P, Wallin G, Tallstedt L, et al. TSH-receptor autoimmunity in Graves’ disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol* 2008;158(1):69-75. doi: 10.1530/EJE-07-0450 [published Online First: 2008/01/02]

27. Codaccioni JL, Orgiazzi J, Blanc P, et al. Lasting remissions in patients treated for Graves’ hyperthyroidism with propranolol alone: a pattern of spontaneous evolution of the disease. *J Clin Endocrinol Metab* 1988;67(4):656-62. doi: 10.1210/jcem-67-4-656 [published Online First: 1988/10/01]

28. Pinchera A, Liberti P, Martino E, et al. Effects of antithyroid therapy on the long-acting thyroid stimulator and the antithyroglobulin antibodies. *J Clin Endocrinol Metab* 1969;29(2):231-8. doi: 10.1210/jcem-29-2-231 [published Online First: 1969/02/01]

29. Takamura Y, Nakano K, Uruno T, et al. Changes in serum TSH receptor antibody (TRAb) values in patients with Graves’ disease after total or subtotal thyroidectomy. *Endocr J* 2003;50(5):595-601. doi: 10.1507/endocrj.50.595 [published Online First: 2003/11/14]

30. Menconi F, Profilo MA, Leo M, et al. Spontaneous improvement of untreated mild Graves’ ophthalmopathy: Rundle’s curve revisited. *Thyroid* 2014;24(1):60-6. doi: 10.1089/thy.2013.0240 [published Online First: 2013/08/29]

31. Bartley GB. Rundle and his curve. *Arch Ophthalmol* 2011;129(3):356-8. doi: 10.1001/archophthalmol.2011.29 [published Online First: 2011/03/16]

32. Kim JW, Woo YJ, Yoon JS. Is modified clinical activity score an accurate indicator of diplopia progression in Graves’ orbitopathy patients? *Endocr J* 2016;63(12):1133-40. doi: 10.1507/endocrj.EJ16-0165 [published Online First: 2016/09/16]

33. Douglas RS, Tsirbas A, Gordon M, et al. Development of criteria for evaluating clinical response in thyroid eye disease using a modified Delphi technique. *Arch Ophthalmol* 2009;127(9):1155-60. doi: 10.1001/archophthalmol.2009.232 [published Online First: 2009/09/16]

34. Frueh BR. Why the NOSPECS classification of Graves’ eye disease should be abandoned, with suggestions for the characterization of this disease. *Thyroid* 1992;2(1):85-8. doi: 10.1089/thy.1992.2.85 [published Online First: 1992/01/01]

35. Barrio-Barrio J, Sabater AL, Bonet-Farriol E, et al. Graves’ Ophthalmopathy: VISA versus EUGOGO Classification, Assessment, and Management. *J Ophthalmol* 2015;2015:249125. doi: 10.1155/2015/249125 [published Online First: 2015/09/10]
36. Jang SY, Shin DY, Lee EJ, et al. Correlation between TSH receptor antibody assays and clinical manifestations of Graves' orbitopathy. *Yonsei Med J* 2013;54(4):1033-9. doi: 10.3349/ymj.2013.54.4.1033 [published Online First: 2013/05/28]

37. Eckstein AK, Lax H, Losch C, et al. Patients with severe Graves' ophthalmopathy have a higher risk of relapsing hyperthyroidism and are unlikely to remain in remission. *Clin Endocrinol (Oxf)* 2007;67(4):607-12. doi: 10.1111/j.1365-2265.2007.02933.x [published Online First: 2007/09/21]
## STROBE Statement—Checklist of items that should be included in reports of cohort studies

| Item No | Recommendation |
|---------|----------------|
| **Title and abstract** | (a) Indicate the study’s design with a commonly used term in the title or the abstract  
  Line #32  
  
  (b) Provide in the abstract an informative and balanced summary of what was done and what was found  
  Line #26-50 |
| **Introduction** | Explain the scientific background and rationale for the investigation being reported  
  Line #97-107 |
| **Objectives** | State specific objectives, including any prespecified hypotheses  
  Line #79-87 |
| **Methods** | Present key elements of study design early in the paper  
  Line #118-125 |
| **Setting** | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection  
  Line #111-117 |
| **Participants** | (a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  
  Line #118-120  
  
  (b) For matched studies, give matching criteria and number of exposed and unexposed  
  Not applicable |
| **Variables** | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable  
  Line #126-136  
  
  Data sources/measurement | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group  
  Line #111-117 |
| **Bias** | Describe any efforts to address potential sources of bias  
  Line #111-114, Line 126-127 |
| **Study size** | Explain how the study size was arrived at  
  Line #111-114, Line #126-136, Line #136-146 |
| **Quantitative variables** | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why  
  Line #126-136, Line #136-146 |
| **Statistical methods** | (a) Describe all statistical methods, including those used to control for confounding  
  Line #148-174  
  
  (b) Describe any methods used to examine subgroups and interactions  
  Line #175-177  
  
  (c) Explain how missing data were addressed  
  Line #184-186  
  
  (d) If applicable, explain how loss to follow-up was addressed  
  Not applicable  
  
  (e) Describe any sensitivity analyses  
  Not applicable |
Results

Participants 13* (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
Line #184-186
(b) Give reasons for non-participation at each stage
Line #184-186
(c) Consider use of a flow diagram
Not applicable
Descriptive data 14* (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
Line #195
(b) Indicate number of participants with missing data for each variable of interest
Line #184-186
(c) Summarise follow-up time (eg, average and total amount)
Line #186-189
Outcome data 15* Report numbers of outcome events or summary measures over time
Not applicable
Main results 16 (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
Not applicable
(b) Report category boundaries when continuous variables were categorized
Not applicable
(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Line #241-246
Other analyses 17 Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Line #224-246
Discussion
Key results 18 Summarise key results with reference to study objectives
Line #267-280
Limitations 19 Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Line #335-276
Interpretation 20 Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Line #377-383
Generalisability 21 Discuss the generalisability (external validity) of the study results
Line #364-366
Other information
Funding 22 Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based
Line #414-416
*Give information separately for exposed and unexposed groups.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.