European Thyroid Journal

Research Article

Eur Thyroid J 2020;9(suppl 1):3–16
DOI: 10.1159/000510700

Received: June 25, 2020
Accepted: August 4, 2020
Published online: September 21, 2020

Proposal for Standardization of Primary and Secondary Outcomes in Patients with Active, Moderate-to-Severe Graves' Orbitopathy

Luigi Bartalena[a] Wilmar M. Wiersinga[b]

[a]Department of Medicine and Surgery, University of Insubria, Varese, Italy; [b]Department of Endocrinology and Metabolism, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Keywords
Graves' orbitopathy · Patient-reported outcome · Clinician-reported outcome · Exophthalmos · Clinical Activity Score · Quality of life

Abstract
Standardization of treatment outcomes in randomized clinical trials (RCTs) for active, moderate-to-severe Graves' orbitopathy (GO) is needed to make results of different RCTs comparable and to draw sound conclusions on the efficacy of a given treatment. Both subjective patient-reported outcome (PRO) and objective clinician-reported outcome (CRO) are important in this regard. In this paper, it is proposed that primary PRO should be the evaluation of treatment-related changes in the quality of life by the use of a validated and disease-specific questionnaire (GO-QoL). The proposed primary CRO is a revised composite index, which includes only objective items and provides an overall assessment of the effects of treatment. Secondary outcomes should also be provided in RCTs to show the effects of treatment on individual features of GO, as well on persistence of activity (by the 7-item Clinical Activity Score), safety, relapses of GO, need for subsequent medical and/or surgical treatments, and other indicators (orbital volume, cytokines, TSH receptor antibody levels). Assessment of the overall response to treatment by primary and secondary outcomes should be made 3 months after treatment completion.

Introduction

Ever since the report of Lord Brain in 1955, glucocorticoids have been the mainstay in the medical treatment of Graves' orbitopathy (GO) [1]. Intravenous glucocorticoid (IVGC) therapy is much more efficacious than placebo [2] and, because more effective than oral GC therapy, is recommended as first-line treatment for active, moderate-to-severe GO in current guidelines [3]. Novel treatment modalities that might be as efficacious as steroids or even better, have become available in the last few years. Rituximab is a promising agent, although conflicting outcomes have been reported [4, 5]. Teprotumumab and tocilizumab have potent efficacy as tested in placebo-controlled, randomized clinical trials (RCTs) [6, 7], but a

Luigi Bartalena
Endocrine Unit, Department of Medicine and Surgery
University of Insubria, ASST dei Sette Laghi
Viale Borri, 57, IT–21100 Varese (Italy)
luigi.bartalena@uninsubria.it
or
Wilmar M. Wiersinga
Department of Endocrinology and Metabolism, Academic Medical Center
University of Amsterdam
Meibergdreef 9, NL–1105AZ Amsterdam (The Netherlands)
w.m.wiersinga@amsterdamumc.nl

karger@karger.com
www.karger.com/etj

Karger

head-to-head comparison of these agents with IVGC has not been done yet. It is expected that a number of RCTs will be published in the near future evaluating which of these novel agents can replace steroids by demonstrating greater efficacy, better safety, and greater cost-effectiveness. One should hope the outcomes of these various RCTs are directly comparable to each other. This hope seems rather illusory as outcome measures differ substantially in the aforementioned RCTs [2, 4–7]. The aim of the present paper is to propose primary and secondary outcome measures for standard use in the treatment of active, moderate-to-severe GO.

The FDA provides "Drug Development Tools" qualification programs that help to optimize drug development and evaluation [8]. The FDA distinguishes between clinician-reported outcomes (CRO) based on clinical observation and interpretation by a trained clinician, and patient-reported outcomes (PRO) based on a report that comes directly from the patient about the status of the patient's symptoms or functioning without amendment or interpretation of the patient's response by a clinician or anyone else [9]. Discrepancies between PRO and CRO can be explained by the fact that the degree of daily functioning and general health as perceived by the patient is not only determined by the severity of signs and symptoms but also by the characteristics of the individual and the environment, such as expectations, coping ability, motivation, social support, and the physician-patient relationship [10]. Assessment of PRO can be of important additional value in the description of disease severity because they highlight a different outcome level of interest. PRO can be important indicators of treatment success and are especially useful in the evaluation of treatments that include side effects [11]. Regulatory authorities consequently ask quite frequently for both subjective outcome assessment (by PRO) and objective outcome assessment (by CRO) as qualification for the efficacy of a drug.

Methods

We followed a practical guideline how to select outcome measurement instruments for inclusion in a "Core Outcome Set" (Table 1) [12, 13]. The construct of interest is the outcome of treatment of active, moderate-to-severe GO. We searched the literature to find existing outcome measurement instruments published between 2000 and 2020. For PRO, we restricted ourselves to disease-specific quality-of-life (QoL) questionnaires and disregarded the less sensitive general QoL questionnaires. For CRO, we confined ourselves to RCTs in patients with active, moderate-to-severe GO. The quality of identified outcome measurements was evaluated as far as possible by criteria given by the COSMIN consortium and the FDA (Table 1) [9, 13–16]. Outcome measurement instruments were eligible for inclusion in the "Core Outcome Set" if there was at least high-quality evidence for good content validity and good internal consistency (or evidence for test-retest reliability) and if the instrument was feasible [13]. High quality rating of evidence in this context is defined as consistent findings in multiple studies of at least good quality OR 1 study of excellent quality AND a total sample size of ≥100 patients [13]. We selected 1 PRO as subjective primary outcome measure and 1 CRO as objective primary outcome measure for inclusion in the "Core Outcome Set."

Table 1. Criteria for selection of outcome measurement instruments for inclusion in a "Core Outcome Set" (COS) [12, 13]

Criterion	Definition
Test-retest reliability	Difference of results in the repeated procedure
Internal consistency	The extent to which items comprising a scale measure the same concept
Content validity	The clinical outcome assessment measures the concept of interest in the context of use
Construct validity	Correlations with clinical characteristics
Ability to detect change	Based on the effect size (mean change divided by the mean SD of the pretreatment score)
Cross-cultural validity	The degree to which the performance of a translated or culturally adapted measurement instrument is an adequate reflector of the original instrument
Feasibility	Comprehensibility, interpretability, ease of administration, length, completion time

Primary Outcomes

Patient-Reported Outcomes

A literature search identified 7 disease-specific PRO instruments used in GO (Table 2) [17–23]. The **GO-QoL** contains 16 questions, 8 on the consequences of double vision and decreased visual acuity on visual functioning and 8 on the psychosocial consequences of a changed appearance. The questions are answered on a 3-point Likert scale (1 point for "yes, seriously limited," 2 points for "yes, a little limited," and 3 points for "no, not limited at all"). The points given to questions 1–8 and 9–16 are added up to obtain 2 raw scores ranging from 8 to 24 points. The 2 raw scores are then transformed to 2 total scores, 1 for visual functioning and 1 for appearance, by the formula: total score = (raw score − 8)/16 × 100. The range is from

Table 2. Characteristics of PRO used in clinical trials of patients with GO

	GO-QoL [17]	TAO-QoL [18]	GO-QLS [19]	NEI VFQ-25 [20]	ThyPRO [21][#]	TED-QoL [22]	OX-TED QoL [23]
Description of PRO							
Questions, *n*	16	90	9	25	84	3	7
Measurement properties							
Test-retest reliability	+	−			(+)	+	
Internal consistency	+	−	+		(+)		
Content validity	+	−	?	−	?	+	
Construct validity	+	−	?		(+)	+	
Ability to detect change	+	?			(+)	+	
Cross-cultural validity	+				(+)		
Feasibility	+	−	+	+	−	+	+
Conclusion							
Overall quality	High	Very low	Low	Low	(High)	High	Low
Used in open trials in GO	Yes $n > 40$	Yes $n = 1$	No	Yes $n = 1$	No	Yes $n = 2$	Yes $n = 1$
Used in RCTs in GO	Yes $n = 8$	No	No	No	No	No	No
Recommended in active moderate-to-severe GO	Yes	No	No	No	No	No	No

+, positive rating; ?, indeterminate rating; −, negative rating; blank box, no data; [#], properties between brackets because they have been validated for benign thyroid diseases but not specifically for GO. PRO, patient-reported outcomes; GO, Graves' orbitopathy; QoL, quality-of-life; RCT, randomized clinical trial.

0 to 100, higher scores indicating better health (online suppl. Table 1; see online Supplementary Materials) [17].

Test-retest reliability was evaluated in 89 GO patients [24]. Mean difference on visual functioning scale was 4.0 and on appearance scale 2.5. Substantial intraclass correlation coefficients (0.83 for visual functioning and 0.87 for appearance) reflect that errors of measurement are relatively small. *Internal consistency* is demonstrated by principal component analysis and high Cronbach's α for both summary scores [17]. Evidence of *content validity* has been described in detail [17]. Questions considered relevant for GO patients were discussed with patients and experienced physicians. The questionnaire was subsequently evaluated in 112 consecutive GO patients, encompassing the whole spectrum of mild-to-severe GO and from active to inactive GO. Correlations of the 2 GO-QoL scales with clinical characteristics support *construct validity*. This is illustrated by a significant increase in the visual functioning score from 27.1 to 47.4 ($p = 0.01$) but not in the appearance score (from 51 to 55, NS) after orbital decompression for dysthyroid optic neuropathy (DON); in contrast, after decompression for disfiguring exophthalmos, the change in visual functioning score was not significant (from 64.8 to 68.0, NS), whereas the appearance score increased significantly from 44.7 to 55.8 ($p < 0.001$) [25]. Longitudinal construct validity is derived from a 12-year follow-up study in GO patients [26]. *Ability to detect change* was ascertained for both nonsurgical and surgical treatments [25, 27]. Based on effect sizes, changes in GO-QoL scores of about 20 points could be interpreted as a large effect, of about 10 points as small to moderate effects, and of 3–4 points as very small effects. A minimal clinically important difference in either GO-QoL scale can be considered 6–10 points. The GO-QoL was originally phrased in Dutch language, requiring *cross-cultural validation* for application in other languages and societies. Translation into English and validation in English-speaking patients have been described in detail [28]. Except for the item "bicycling," which may be less

relevant outside Holland, there were no major problems. Answers "never learned to ride a bike" or "no driver's license" are scored as missing values. When there are missing values for some items, GO-QoL scores are calculated by the formula: Total score = (raw score − #)/2 × # times 100 (in which # is the number of completed items) [28]. The recent CIRTED trial further validated the English version of the GO-QoL [29]. The GO-QoL is now available in 19 languages: Chinese [30–32], Danish, Dutch, English, French, Georgian, German [33], Greek, Italian, Korean [34], Persian [35], Polish [36], Portuguese [37], Romanian, Russian, Serbian, Spanish [38], Swedish, and Turkish (some can be downloaded for free from www.eugogo.eu). Formal validation of the translated versions has been published for the English, Persian, Polish, and Portuguese versions of the GO-QoL [29, 35–37]. The GO-QoL finally satisfies *feasibility* criteria. The questionnaire is handed over to the patient prior to the consultation, while sitting in the waiting room. It takes about 3 min to complete the GO-QoL [22].

A German **TAO-QoL** (Thyroid-Associated Ophthalmopathy – QoL) questionnaire had 90 items [18]. Its development was based on contributions from clinicians, without input of patients; content validity is thus low. Construct validation has not been performed. Internal consistency is low in view of a low Cronbach's α of 0.63 in a study among 104 patients undergoing orbital decompression [39]. Scores did not correlate with clinical variables.

The **GO-QLS** (GO Quality-of-Life Scale) was developed in the USA [19]. The original questionnaire contained 105 items, derived from the SF-12, DSQL (Dermatology-Specific Quality of Life), NEI-VFG (National Eye Institute Visual Function Questionnaire), and visual functions specific to GO. The questionnaire was completed by 253 GO patients and "validated by administering it to 33 healthy subjects." Patients had no conceptual input, raising doubt about the content validity. To select the most appropriate items for the GO-QLS, a pool of candidate items was based on appropriate response scales, no excessive skewness, few missing values, and face validity for their applicability. Items were selected that correlated best and discriminated best with GO severity scales and possessed good face validity. Items were assigned points based on correlation rank (best 10 received 10–1 points), discriminating rank (best 10 received 10–1 points), and good face validity (3 points). Items that exceeded 10 points were included in the GO-QLS. The proposed GO-QLS (scaled from 0 to 100) consists of 9 questions. The internal consistency of the scale is good (Cronbach's α of 0.89). The correlation between GO-QLS and GO severity scores is reasonable: Pearson's correlation coefficients with clinical severity outcomes regarding neuropathy, cosmesis, myopathy, and keratopathy are −0.49, −0.32, −0.28, and −0.21, respectively [19].

The 25-item **NEI-VFQ** (National Eye Institute Visual Function Questionnaire) was evaluated for its potential suitability to measure HRQL in 30 GO patients, although it was not specifically designed for this purpose [20]. More than two-thirds of GO patients thought the instrument lacked items relevant to their disease, such as altered appearance and ocular discomfort. Patients with diplopia had lower scores than patients without diplopia [20], and scores were lower in the presence of DON than in the absence of DON [40]. There is strong doubt about the validity, reliability, and responsiveness of the NEI VFQ-25 in GO [41].

ThyPRO (Thyroid PRO) measures a range of QoL aspects relevant to patients with benign thyroid diseases (hyperthyroidism, hypothyroidism, goiter, and GO), as identified during patient and expert interviews [21]. It consists of 84 items summarized in 13 scales. Each item is rated on a 0–4 Likert scale, from no problems/symptoms = 0 to severe problems/symptoms = 4. The average score of items in a scale is divided by 4 and multiplied by 100 to yield thirteen 0–100 scales, with higher scores indicating worse health status. One of the 13 ThyPRO scales is on eye symptoms. Among the sample of 66 hyperthyroid patients treated to become euthyroid, 9 (14%) had GO; their ThyPRO score on subscale eye symptoms changed from baseline 19 ± 22 to 13 ± 15 at 6 months [42]. Further data on the item GO are not available. ThyPRO demonstrated good responsiveness in patients with hyperthyroidism, autoimmune hypothyroidism, or nontoxic goiter [42], but its responsiveness to specific treatment of GO cannot be judged properly. Its overall good measurement properties thus do not apply to GO [21, 42–45], and the ratings of ThyPRO in Table 2 are therefore given within brackets. Feasibility of the ThyPRO might become better with the introduction of a short version [46] and the use of electronic questionnaires [47].

The **TED-QoL** (Thyroid Eye Disease QoL) is a 3-item questionnaire designed in Canada [22]. The 3 questions are as follows: How is your eye disease currently interfering with your overall quality of life? How is your eye disease currently affecting your ability to carry out daily activities? How is your eye disease currently affecting your satisfaction with your appearance? Patients are asked to answer on a scale from 0 (does not interfere) to 10 (completely interferes). Content validity was assessed through a pilot study and interviews with specialists and patients.

Construct validity was ascertained in 100 new consecutive GO patients through correlation between items on the TED-QoL and on GO-QoL and GO-QLS. TED-QoL correlated strongly with the other questionnaires for corresponding items: Pearson's correlations were 0.71 and 0.62 for appearance, 0.69 and 0.66 for visual functioning, and 0.53 for overall QoL. Intraclass correlation coefficients demonstrated good test-retest reliability of the 3 questionnaires – TED-QoL: 0.81, 0.74, and 0.87; GO-QoL: 0.81 and 0.82; and GO-QLS: 0.74, 0.86, and 0.67. There is only moderate correlation between items on all 3 questionnaires and GO severity assessed by VISA scores [22]. Responsiveness of the TED-QoL to rehabilitative surgery (orbital decompression, strabismus surgery, and palpebral aperture narrowing) is good as assessed from pre- and postoperative scores for the same subjects [48]. TED-QoL is significantly faster to complete (1.6 min vs. GO-QoL 3.1 min vs. GO-QLS 2.7 min, $p < 0.0001$) [22].

The **OX-TED** (Oxford-Thyroid Eye Disease) QoL questionnaire consists of 7 questions, each with a score out of 10, with 1 indicating no concerns and 10 indicating major concerns. The higher the total score (maximum 70), the greater the impact of GO on the patient's QoL [23]. There are no data how OX-TED was developed or validated, nor on measurement properties. In 12 patients with active GO who were treated early with low-dose rituximab, the average total score was 45 prior to treatment, decreasing nonsignificantly to 37 post-treatment at an average follow-up of 6 months.

Epicrises

There have been 2 systematic reviews on PRO in the field of GO [49, 50]. The first was published in 2004 and evaluated 31 questionnaires for people with vision impairments. Only a few questionnaires demonstrated sufficient psychometric quality. The results section states "The GO-QoL for people with Graves' ophthalmopathy is the questionnaire that rates best on our criteria overall, with a '+' for most aspects" [49]. The second was published in 2016 and evaluated 23 questionnaires on thyroid-specific HRQL. In the conclusion section, the authors state "After reviewing the present literature and critically examining published HRQL instruments, the ThyPRO, GO-QoL and ThyTSQ were the three with the greatest number of positive ratings according to nine quality assessment criteria of measurement properties. The ThyPRO was recommended to assess HRQL in patients with benign thyroid disease while the overall measurement properties of GO-QoL and ThyTSQ were satisfactory in measuring HRQL in GO and hypothyroidism, respectively" [50].The authors of the TED-QoL also suggest to use the GO-QoL in research as it enquires about a larger range of issues and is less prone to measurement error than a 3-item questionnaire might be [22]. EUGOGO made recommendations for assessing response to intervention in clinical trials in 2006, and the GO-QoL was recommended for assessment of the PRO as a primary outcome [51]. The recommendation was included in the formal ETA/EUGOGO guidelines for the management of GO in 2016 [3]. These recommendations are all in line with our analysis of the properties of the 7 PRO instruments in GO (Table 2). The GO-QoL clearly emerges as the instrument with the best overall quality. The GO-QoL has been used successfully as PRO to assess subjective primary outcome in 8 RCTs: in mild GO (comparing selenium with placebo) [52], in active, moderate-to-severe GO (comparing intravenous with oral methylprednisolone) [53], comparing different doses of IVGC [54], comparing rituximab with IVGC [4], comparing IVGC with or without mycophenolate [55], comparing teprotumumab with placebo [6], comparing tocilizumab with placebo [7], and in the CIRTED trial comparing oral prednisone with orbital irradiation, azathioprine, and placebo [29]. Using the notation of the GRADE system [56], we recommend the use of the GO-QoL as PRO for all interventions in active, moderate-to-severe GO (1, ØØØØ). This is not to say the GO-QoL should be used only in immunosuppressive treatment modalities for GO. On the contrary, the GO-QoL has been applied successfully as PRO for nonmedical treatment options as well, like orbital bony decompression [25, 57–59], orbital fat decompression [60], orbital bony + fat decompression [61], strabismus surgery [25, 27, 62], eyelid surgery [25], and orbital irradiation [25, 63].

Clinician-Reported Outcome

The following are the objective features that should be considered for CRO at baseline and in evaluating treatment outcomes in patients with active, moderate-to-severe GO.

Eyelid Aperture

This is the vertical distance between lid margins in millimeter in the mid-pupil position. The patient should look in primary position, sitting relaxed and with distant fixation. In addition, upper and lower eyelid position relative to the limbus might be indicated. The least significant change (increase/decrease) in eyelid aperture should be 2 mm [51].

Exophthalmos

This defines protrusion of the eyes, usually measured by a Hertel exophthalmometer, using the same intercanthal distance at baseline and follow-up visits. Exophthalmos is present when exophthalmometer readings are ≥3 mm above the upper normal values [51, 64]. Normal values are different in different ethnic groups (Asians < Caucasians < black people). The least significant change in exophthalmos (increase/decrease) should be 2 mm [51].

Extraocular Muscle Function

Subjective diplopia scores have been proposed, such as the Bahn-Gorman score [65]. They are useful in daily clinical practice, but, being patient dependent, do not fulfill criteria for objective CRO. Objective measures of extraocular muscle function include the prism cover test, the field of binocular vision, and the uniocular fields of fixation (ductions). While the former are useful when both eyes are involved [64], ductions independently assess ocular excursions in 4–6 directions of gaze in each eye using bowl or arc perimeter and expressing them in degrees [64]. Accordingly, ductions should be reported, the least significant change (increase/decrease) being 8° in at least 1 direction of gaze [51].

Visual Function

DON is absent in the definition of moderate-to severe GO: its presence categorizes GO as sight-threatening (or very severe) [3]. However, optic nerve involvement should be assessed at baseline by determination of best-corrected visual acuity, color vision, visual field, optic disc appearance, and relative afferent pupillary defect. In the evaluation of treatment outcome, appearance of significant abnormalities in 1 or more of the above features should be regarded as treatment failure and progression of GO from moderate-to-severe to sight-threatening (or very severe).

Corneal changes (punctate keratopathy or corneal ulcers) may occur but should be regarded as secondary to globe exposure. The presence of corneal breakdown moves GO from moderate-to-severe to sight-threatening (or very severe) and represents an emergency situation [3].

Soft Tissue Changes

Soft tissue changes, without the subjective pain (either spontaneous or with gaze), and therefore including eyelid edema, eyelid erythema, edema of caruncle, conjunctival hyperemia, chemosis, could be considered, if based on comparison with a color picture atlas, as objective manifestations and, therefore, be included in the CRO. In other words, a new 5-item Clinical Activity Score (CAS) might replace the non-fully objective 7-item CAS (see Secondary Outcomes section). GO should be considered active if CAS is ≥3/5. An 1-point change in the 5-item CAS might be regarded as significant.

Table 3 reports RCTs of the management of active, moderate-to-severe GO published in the period 2000–2020 (2, 4–7, 53–55, 66–81). Primary outcomes to define response to a given treatment were very heterogeneous. In some instances [4, 5, 7, 68–71, 74, 75], the 7-item CAS changes defined response as the only primary outcome or one of them. In many instances, the primary outcome was defined by a composite index, which included both objective measures (lid aperture, exophthalmos, and eye motility) and CAS changes; in the first study on teprotumumab [6], the primary outcome to define response was a composite index consisting of a decrease in exophthalmos of ≥2 mm plus a decrease in the CAS of ≥2 points. In the second study on teprotumumab from same multicenter consortium, given the striking effect on exophthalmos shown in the first study, the primary outcome was represented only by a decrease in exophthalmos ≥2 mm [81]. Only in 1 study [54] did changes in the QoL, assessed by the disease-validated GO-QoL questionnaire, represent a primary outcome contributing to definition of response to treatment. Secondary outcomes were also heterogeneous, including both subjective and objective features and were not always the same (Table 3). In some cases, CAS changes, included in the composite index (primary outcome), were also independently evaluated among secondary outcomes; in other words, they were counted twice. This is to some extent conceivable, because inactivation of GO (CAS change) is part of an overall response, which includes modifications in objective features. However, objective and subjective (or semi-subjective) features should be kept distinct.

A revised composite index as a representation of the overall objective response to treatment for moderate-to-severe and active GO might be composed as follows:

- Measurement of palpebral aperture: a significant change would be a decrease/increase of ≥2 mm;
- Measurement of exophthalmos: a significant change would be a decrease/increase of ≥2 mm;
- Evaluation of ocular motility: a significant change would be an increase/decrease of at least 8°; the Bahn-Gorman score should not be considered, at least in RCTs, because it is a patient-dependent subjective outcome;
- Five-item CAS (excluding both spontaneous and gaze-evoked pain) using the EUGOGO color atlas: a significant change would be a decrease/increase of ≥1 point.

Table 3. Primary outcomes used to define positive response to immunosuppressive therapies for active, moderate-to-severe GO in randomized clinical trials

Author	Subjects, *n*	Primary outcome(s)	Definition of positive response	Secondary outcome(s)
Kahaly et al. [66] (2000)	62	Composite index	Improvement of at least 3 of the following: (i) decrease in lid aperture >2 mm; (ii) decrease in exophthalmos >2 mm; (iii) decrease in intraocular pressure (IOP) (upgaze) >3 mm Hg; (iv) decrease in eye muscle area <5 mm^2; (v) absence of diplopia in primary gaze	Individual ocular features
Marcocci et al. [67] (2001)	82	Composite index	Improvement in at least 2 major criteria (decrease in lid aperture or exophthalmos ≥2 mm; change in grade of subjective diplopia; decrease in the CAS ≥ 2 points; improvement in visual acuity ≥1/10) and 1 minor criterion (soft tissue changes, self-assessment evaluation [improved, unchanged, worsened])	Individual ocular features, CAS
Macchia et al. [68] (2001)	51	OI,[1] exophthalmos, CAS	Not specified (descriptive analysis of changes in individual items, CAS, IOP)	See primary outcomes
Kauppinen-Mäkelin et al. [69] (2002)	33	CAS, individual items (exophthalmos, subjective diplopia, visual acuity, IOP)	Not specified (descriptive analysis of changes in individual items, CAS, IOP)	Need for orbital radiotherapy and surgery
Dickinson et al. [70] (2004)	50	OI, CAS, GO-QOL	Not specified (descriptive analysis of changes)	GO-QOL, individual ocular items
Wémeau et al. [71] (2005)	51	NOSPECS + CAS	Decrease in NOSPECS class + decrease or no change in CAS	CAS, GO-QOL, individual ocular items
Ng et al. [72] (2005)	15	NOSPECS grade, TES[2]	Decrease in NOSPECS class, decrease in TES	Non-validated questionnaire on QoL
Kahaly et al. [73] (2005)	70	Composite index	Improvement of at least 3 of the following: (i) decrease in lid aperture ≥2 mm; (ii) decrease in exophthalmos ≥2 mm; (iii) decrease in IOP (upgaze) ≥3 mm Hg; (iv) improvement in eye motility ≥10°; (v) disappearance of diplopia in primary gaze; (vi) improvement in visual acuity ≥2/10	Individual ocular items
Chang and Liao [74] (2006)	60	CAS	Mean (?) change in CAS	Individual ocular items (including retrobulbar tissue volume)
Stan et al. [75] (2006)	25	CAS	Not specified (descriptive analysis of changes)	Individual ocular items (including orbital volume)
Aktaran et al. [53] (2007)	52	Composite index	Improvement in at least 2 major criteria (decrease in lid aperture or exophthalmos ≥2 mm; change in grade of subjective diplopia; decrease in the CAS ≥ 2 points; improvement in visual acuity ≥1/10) and 1 minor criterion (soft tissue changes, self-assessment evaluation [improved, unchanged, worsened])	Individual ocular items
Menconi et al. [76] (2007)	60	Composite index	Improvement at 9 months of at least 2 of the following: (i) decrease in lid aperture ≥2 mm; (ii) decrease in exophthalmos ≥2 mm; (iii) decrease in CAS of ≥2 points; (iv) disappearance or improvement of subjective diplopia	Overall response at 3 months; individual ocular items
van Geest et al. [2] (2008)	15	Composite index	Improvement in 1 major criterion (improvement in diplopia grade according to NOSPECS class 4; improvement of eye motility of at least 8°; a decrease in CAS of ≥3 points) and/or 2 minor criteria (decrease in lid aperture ≥2 mm; decrease in exophthalmos ≥2 mm; improvement in NOSPECS class 2; a decrease in CAS of ≥2 points)	Individual ocular items
Bartalena et al. [54] (2012)	159	Composite index, GO-QoL	Composite Index: improvement of at least 2 of the following: decrease in lid aperture ≥3 mm; reduction in any of the class 2 NOSPECS of at least 2 grades; decrease in exophthalmos of ≥2 mm; improvement of subjective diplopia; decrease in CAS of ≥2 points GO-QOL: increase of ≥6 points in either (or both) subscales (functioning and appearance) Safety (adverse events)	CAS

Table 3 (continued)

Author	Subjects, *n*	Primary outcome(s)	Definition of positive response	Secondary outcome(s)
Zhu et al. [77] (2014)	80	Composite index	Improvement in at least 3 of the following: (i) decrease in lid aperture ≥3 mm; (ii) reduction in any of the class 2 NOSPECS of at least 2 grades; (iii) decrease in exophthalmos ≥2 mm; (iv) decrease in IOP ≥2 mm Hg; (v) improvement of subjective diplopia; (vi) decrease in CAS of ≥2 points; (vii) improvement in visual acuity by 1 Snellen line	CAS, adverse events, retreatments
Stan et al. [5] (2015)	21	CAS	Not specified (descriptive analysis of changes)	Individual ocular features (including retrobulbar tissue volume)
Salvi et al. [4] (2015)	31	CAS	Decrease in CAS ≥2 points or final CAS <3	Individual ocular features, GO-QOL, GO reactivation
He et al. [78] (2017)	40	Composite index	Improvement in 1 major criterion (improvement in diplopia grade according to NOSPECS class 4; improvement of eye motility of at least 8°; a decrease in CAS ≥3 points) and/or 2 minor criteria (decrease in lid aperture ≥2 mm; decrease in exophthalmos ≥2 mm; improvement in class NOSPECS class 4; a decrease in CAS ≥2 points)	Individual ocular features, CAS, IOP
Ye et al. [79] (2017)	158	Composite index	Improvement of at least 3 of the following: (i) decrease in CAS ≥2 points or disease inactivation (CAS <3); (ii) reduction in any of the class 2 NOSPECS of at least 1 grade; (iii) decrease in exophthalmos ≥2 mm; (iv) improvement in eye movements (not quantified); (v) improvement in subjective diplopia; (vi) increase in visual acuity ≥2/10	CAS, individual ocular features, adverse events, retreatments
Smith et al. [6] (2017)	87	Composite index	Decrease in exophthalmos ≥2 mm + decrease in CAS ≥2 points	Exophthalmos, CAS, GO-QoL, adverse events
Kahaly et al. [55] (2018)	164	Composite index	Improvement of at least 2 of the following: (i) eyelid swelling (using the EUGOGO atlas); (ii) decrease in CAS ≥ 2 points; (iii) decrease in eyelid aperture ≥2 mm; (iv) decrease in exophthalmos ≥2 mm; (v) improvement in subjective diplopia; (vi) improvement in eye motility ≥8°	Individual ocular features, CAS, adverse events
Rajendram et al. [80] (2018)	103	Composite index, OI[3]	Composite index: Improvement in ≥1 major criteria (improvement of ≥1 grade in diplopia score; improvement of >8° of eye motility; decrease in exophthalmos ≥2 mm) or ≥2 minor criteria (decrease in eyelid aperture ≥2 mm; improvement of ≥1 grade in soft tissue involvement; improvement of visual acuity of ≥1 line on Snellen chart; patient-judged subjective improvement OI	GO-QoL
Perez-Moreiras et al. [7] (2018)	32	CAS	Decrease in CAS ≥ 2 points (at week 16)	Decrease in CAS ≥2 points (at week 40), CAS <3, quality of life (SF-36, GO-QoL), composite index (post hoc)
Douglas et al. [81] (2020)	83	Exophthalmos	Decrease in exophthalmos ≥2 mm	Decrease in exophthalmos ≥2 mm + decrease in CAS ≥2 points; CAS 0–1, mean change in exophthalmos, change in subjective diplopia, GO-QoL

CAS, Clinical Activity Score; OI, Ophthalmopathy Index; IOP, intraocular pressure; TES, Total Eye Score; GO, Graves' orbitopathy; QoL, quality-of-life. [1] Calculated by giving 1–3 points (depending on severity) to each NOSPECS class (2–6) and then summing the points (highest score: 15). [2] Obtained by summing scores in each class (number of class × degree of severity [replacing letters a, b, c with numbers 1, 2, and 3, e.g., class 2c = 6]). Highest score: 60. [3] Based on a series of scores covering soft tissue changes, exophthalmos, eyelid aperture, IOP, subjective diplopia, corneal abnormalities, optic neuropathy. See [80, online suppl. Table 2] for the complex calculation of the score.

Table 4. Proposal for revision of the composite index as primary objective treatment outcome for active, moderate-to-severe Graves' orbitopathy

Item	Original composite index [54]	Revised composite index
Reduction of lid aperture	≥3 mm	≥2 mm
Soft tissue changes	Reduction in NOSPECS class 2 changes by at least 2 grades	Omit (overlap with "objective" CAS)
Reduction of exophthalmos	≥2 mm	≥2 mm
Eye motility	Increase of ≥8° or improvement of Bahn-Gorman diplopia score	Increase of ≥8° (omit subjective Bahn-Gorman diplopia score)
CAS	Reduction in 7-item CAS of ≥2 points	Reduction in 5-item CAS (no spontaneous or gaze-evoked pain) of ≥1 point

A response to treatment might be considered positive in the case of an improvement of at least 2 of the above features in 1 eye, without concomitant deterioration in the other eye. CAS, Clinical Activity Score.

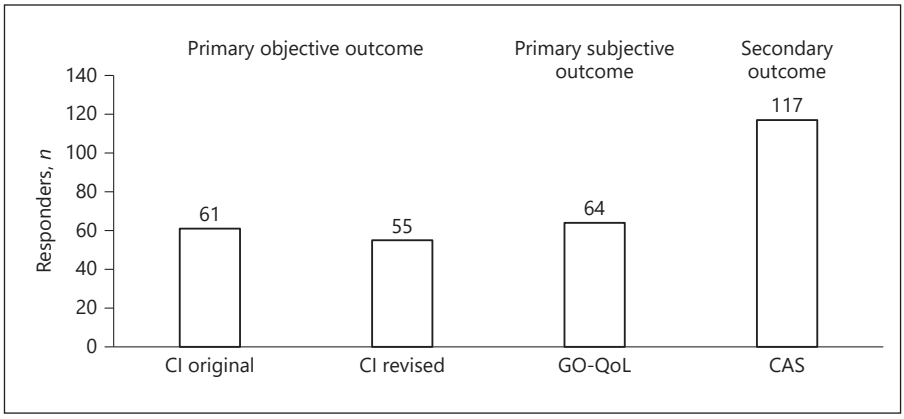

Fig. 1. Evaluation of the revised composite index. Data from Bartalena et al. [54] were reanalyzed using the original composite index and the revised composite index (see Table 4). For this purpose, all the enrolled patients ($n = 159$), included in 3 groups given low-dose, medium-dose, high-dose intravenous glucocorticoids were pooled. Number of responders using the original composite index and the revised composite index were not significantly different.

The revised composite index is illustrated in Table 4. A response to treatment might be considered positive in the case of an improvement of at least 2 of the above features in 1 eye, without concomitant deterioration in the other eye. Deterioration would be defined by the occurrence of DON or worsening of at least 2 of the 4 components of the revised composite index. Deterioration represents at any time an indication to move to other treatments, either medical or surgical. In an attempt to verify whether the above modification of the composite index is feasible, we reevaluated the results of an RCT on the use on different doses of IVGC [54] with the revised composite index and compared them with data obtained using the original composite index. For this purpose, we pooled the 3 groups of patients ($n = 159$) treated with a low-dose, medium-dose, or high-dose IVGC regimen. Figure 1 shows that the number of responders did not significantly differ, suggesting that changes in the revised composite index, entirely objective, very well describe the overall changes in the different components of GO.

It may be argued that a composite index might not faithfully or completely reflect what happened to the patient's eyes. For example, inflammatory manifestations might improve but extraocular muscle dysfunction, heralded by diplopia, might persist. Or extraocular muscle dysfunction and inflammation might subside, while exophthalmos might be only partially affected by treatment. This objection is correct but can be overcome by arguing that residual manifestations have an inevitable impact on the QoL of patients, which would be unveiled and underscored by the PRO using GO-QoL. This brings us back to the central and fundamental role of patient self-assessment.

Secondary Outcomes

Clinical Activity Score

The natural history of GO encompasses an initial inflammatory phase (active disease), followed by a static phase (plateau), and finally by a progressive regression of inflammation (burnt-out or inactive disease) [82]. In most cases, inactivation of GO by no means implies complete *restitutio ad integrum*, and therefore, rehabilitative surgery is often required after medical treatment in patients who have had full-blown, moderate-to-severe and active GO [3]. The active phase of the disease is generally responsive to immunosuppressive treatments, which are completely useless in the inactive phase [3]. The CAS is a tool introduced more than 30 years ago by Mourits et al. [83] to differentiate inflammatory (active) GO from non-inflammatory (inactive GO). The original CAS included 10 items, the last 3 of which are indicative of recent progression of the ocular disease, but less useful for evaluating the effects of treatment. In the 7-item CAS, 1 point is attributed to each item (eyelid edema, eyelid hyperemia, conjunctival hyperemia, chemosis, caruncle edema, spontaneous ocular pain, and gaze-evoked pain), if present; therefore, the CAS (sum of the points) can range from 0 (no activity) to 7 (maximal activity). Active GO is defined by a CAS ≥ 3/7 [3]. The CAS was originally intended as a tool to identify patients who had a high chance of responding to immunosuppressive treatment [84], and in a more recent study by Yang et al. [85], it proved to be also a good predictor for additional immunosuppressive treatments or additional rehabilitative surgeries in patients who had not responded satisfactorily to a first course of immunosuppressive treatment. As mentioned above, changes in the CAS have been largely used as a primary treatment outcome measure in many randomized and non-randomized studies of the management of active, moderate-to-severe GO. Undoubtedly, the CAS is a very simple tool that can be used by an endocrinologist or an ophthalmologist in the office, both in the initial assessment and in follow-up visits, also outside specialized centers. The CAS has, however, some limitations, including: (i) it is binary (present/absent), with no possibility to grade changes in severity of each item; (ii) it may be difficult to distinguish inflammatory from congestive changes and, therefore, it may not always accurately reflect inflammation/activity of the disease; and (iii) it does not predict the possible development of DON. On the other hand, some of these limitations can be overcome. With the exception of pain, either at rest or with eye movements, which is entirely subjective (patient dependent), the other 5 items can be made objective if patient's soft tissue changes are compared with a picture atlas (see the color picture atlas available on the EUGOGO website at www.eugogo.eu). This comparison may also help distinguish soft tissue changes due to inflammation from those related to congestion. Thus, if the 7-item CAS cannot be included, as a whole, either in PRO or in CRO, a revised 5-item CAS, without spontaneous and gaze-provoked ocular pain, can be included in the CRO, as suggested in a previous paragraph. Accordingly, we suggest that when evaluating treatment outcomes in both observational and RCTs, (i) changes in the 7-item CAS should not be considered as a primary treatment outcome and (ii) changes in the revised 5-item CAS can be included in the primary CRO as part of the revised composite index. The 7-item CAS maintains its validity as a secondary outcome to orientate the treatment strategy toward additional immunosuppressive treatments (CAS ≥ 3/7) or rehabilitative surgical procedures (CAS < 3/7).

Individual Items of GO and Other Indicators

In addition to an overall assessment of treatment outcome, report of details of individual ocular features should be included among secondary outcomes, thereby contributing to identify, in the context of a generally favorable (or unfavorable) response to a given treatment, which objective components were more (or less) responsive. See, for example, the apparently striking difference (in the absence of a head-to-head comparative study) in terms of exophthalmos decrease in patients treated with teprotumumab [7, 81] or IVGC [54]. These individual items may include eyelid aperture, exophthalmos, subjective and objective eye motility abnormalities, intraocular pressure, optic nerve involvement, orbital volume assessment, measurement of TSH receptor antibodies, or markers of inflammation (e.g., cytokines). But they may also include assessment of the effects of treatment on disease-related psychosocial well-being and public health relevance [33].

Adverse Events

Adverse events of IVGC, the current first-line treatment for active, moderate-to-severe GO, are well known and can be short-term or long-term [86, 87]. Less is known about novel immunosuppressive agents, such as rituximab, tocilizumab, and teprotumumab, particularly in the long-term. Therefore, adverse events should be carefully reported, as an integral part of the evaluation of the effects of treatment. In our opinion, they should be mentioned among secondary outcomes to allow a precise cost/benefit

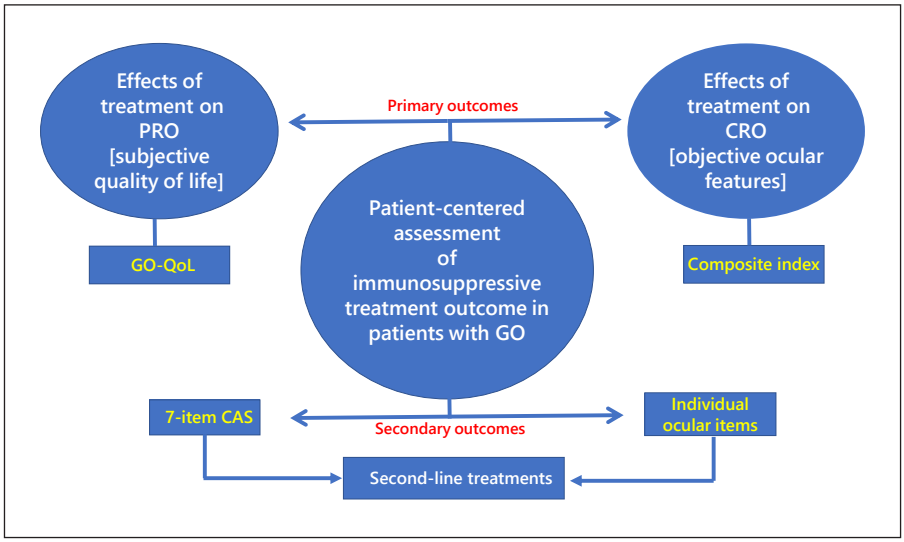

Fig. 2. Proposal for assessment of treatment outcome in randomized clinical trials for active, moderate-to-severe Graves' orbitopathy.

evaluation. A standardized tool to report adverse events is represented by the medical dictionary for regulatory affairs (MedDRA), recommended by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) [88]. By this tool, both the number and severity of adverse events should be reported in detail.

When Should Treatment Outcomes Be Assessed?

Treatment outcome assessment also needs to be standardized. Some studies report assessment of the primary outcome and response to treatment at the end of drug therapy. But this time point is extremely variable because duration of treatment varies from 3 months for IVGC [54] and tocilizumab [7], 6 months for mycophenolate [55] and teprotumumab [6, 81], or 1 day to 2 weeks for rituximab [4, 5]. Response to IVGC may observed early but also be delayed [89]. Relapse of GO or its progression to sight-threatening forms may occur after a response to treatment has been considered positive at a too early evaluation [54]. For these reasons, it seems reasonable to propose that in future RCTs for active, moderate-to-severe GO assessment of response by primary outcomes and secondary outcomes be carried out 3 months after treatment has been completed.

Discussion

In the last 20 years, the field of management of active, moderate-to-severe GO has witnessed a renewed interest driven by application of old drugs to this disease, combination therapies, or development of novel, biological agents providing promising results. Comparison of different studies and different drugs is extremely difficult due to the heterogeneity of reporting results. Therefore, we aimed at introducing a standardized way of evaluating treatment outcomes to overcome the above difficulties.

Objective ocular changes are extremely important, and it is fundamental to know which components of GO (inflammatory changes, exophthalmos, eye motility, and visual acuity), if not all, are influenced by therapy. In this regard, an overall evaluation of the objective response to treatment is permitted by the revised composite Index (Table 4), which is completely deprived of subjective elements. This composite index should, therefore, represent the primary CRO. Changes in the composite index, while providing a reliable overview of the general effectiveness of a given treatment, do not unveil the effects of treatment on the individual components of GO. For this reason, we propose that details on the individual features of the disease be provided as secondary outcomes, because this is helpful to orientate subsequent medical and/or surgical interventions.

If objective assessment is essential, we feel that subjective evaluation by the patient himself/herself is even more important. We physicians can tell the patient that exophthalmos has decreased by 2–3 mm, eye motility has improved by 8–10°, or the CAS has decreased by 2 points, and we consider this a success. But, from the patient's point of view, this may still be a failure, if it does not translate into a substantial improvement in the quality of life [33, 90]. Accordingly, we believe that, much more

than in the past, assessment of outcome of treatment for active, moderate-to-severe GO should be patient-centered. Therefore, we propose, as illustrated in Figure 2, that primary outcomes should include both PRO (GO-QoL changes) and CRO (changes in the revised composite index). Secondary outcomes may include changes in individual objective features, safety (information on adverse events), and information on relapses after treatment (durability) and on the need of additional medical or surgical treatments. Changes in the 7-item CAS remain an important secondary outcome to facilitate the shared decision on the need of further medical or surgical treatments.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The authors did not receive any funding.

Author Contributions

Both authors equally contributed to design the study, to search literature, to write the paper, and to revise it, and are fully responsible for its content.

References

1 Brain R, Aterman K, Cope CL, Dunlop DM, Russel Fraser T, Greene R, et al. CORTISONE in exophthalmos: report on a therapeutic trial of cortisone and corticotrophin (A.C.T.H.) in exophthalmos and exophthalmic ophthalmoplegia by a panel appointed by the medical research council. Lancet. 1955;268(6853):6–9.

2 van Geest RJ, Sasim IV, Koppeschaar HP, Kalmann R, Stravers SN, Bijlsma WR, et al. Methylprednisolone pulse therapy for patients with moderately severe Graves orbitopathy: a prospective, randomized, placebo-controlled study. Eur J Endocrinol. 2008;158(2):229–37.

3 Bartalena L, Baldeschi L, Boboridis K, Eckstein A, Kahaly GJ, Marcocci C, et al. The 2016 European thyroid association/European group on Graves' orbitopathy guidelines for the management of Graves' orbitopathy. Eur Thyroid J. 2016;5(1):9–26.

4 Salvi M, Vannucchi G, Currò N, Campi I, Covelli D, Dazzi D, et al. Efficacy of B-cell targeted therapy with rituximab in patients with active moderate to severe Graves' orbitopathy: a randomized controlled study. J Clin Endocrinol Metab. 2015;100(2):422–31.

5 Stan MN, Garrity JA, Carranza Leon BG, Prabin T, Bradley EA, Bahn RS. Randomized controlled trial of rituximab in patients with Graves' orbitopathy. J Clin Endocrinol Metab. 2015;100(2):432–41.

6 Smith TJ, Kahaly GJ, Ezra DG, Fleming JC, Dailey RA, Tang RA, et al. Teprotumumab for thyroid-associated ophthalmopathy. N Engl J Med. 2017;376(18):1748–61.

7 Perez-Moreiras JV, Gomez-Reino JJ, Maneiro JR, Perez-Pampin E, Romo Lopez A, Rodríguez Alvarez FM, et al. Efficacy of tocilizumab in patients with moderate-to-severe corticosteroid-resistant Graves orbitopathy: a randomized clinical trial. Am J Ophthalmol. 2018;195:181–90.

8 Available from: https://www.fda.gov/drugs/development-approval-process-drugs/drug-development-tool-qualification-programs.

9 Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims.

10 Wiersinga WM. Quality of life. In: Wiersinga WM, Kahaly GJ, editors. Graves' orbitopathy. A multidisciplinary approach – questions and answers. 3rd, revised and expanded ed. Basel: Karger; 2017. p. 260–74.

11 Terwee CB, Dekker FW, Prummel MF, Wiersinga WM. Graves' ophthalmopathy through the eyes of the patient: a state of the art on health-related quality of life assessment. Orbit. 2001;20(4):281–90.

12 Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132.

13 Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set" – a practical guideline. Trials. 2016;17(1):449.

14 Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based standards for the selection of health measurement instruments. BMC Med Res Methodol. 2006;6:2.

15 Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45.

16 Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147–57.

17 Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. Br J Ophthalmol. 1998;82(7):773–9.

18 Tehrani M, Krummenauer F, Mann WJ, Pitz S, Dick HB, Kahaly GJ. Disease-specific assessment of quality of life after decompression surgery for Graves' ophthalmopathy. Eur J Ophthalmol. 2004;14(3):193–9.

19 Yeats RP. Quality of life in patients with Graves ophthalmopathy. Trans Am Ophthalmol Soc. 2005;103:368–411.

20 Bradley EA, Sloan JA, Novotny PJ, Garrity JA, Woog JJ, West SK. Evaluation of the National eye institute visual function questionnaire in Graves' ophthalmopathy. Ophthalmology. 2006;113(8):1450–4.

21 Watt T, Hegedüs L, Groenvold M, Bjorner JB, Rasmussen AK, Bonnema SJ, et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. Eur J Endocrinol. 2010;162(1):161–7.

22 Fayers T, Dolman PJ. Validity and reliability of the TED-QoL: a new three-item questionnaire to assess quality of life in thyroid eye disease. Br J Ophthalmol. 2011;95(12):1670–4.

23 Insull EA, Sipkova Z, David J, Turner HE, Norris JH. Early low-dose rituximab for active thyroid eye disease: an effective and well-tolerated treatment. Clin Endocrinol. 2019;91(1):179–86.

24 Terwee CB, Gerding MN, Dekker FW, Prummel MF, van der Pol JP, Wiersinga WM. Test-retest reliability of the GO-QOL: a disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy. J Clin Epidemiol. 1999;52(9):875–84.

25 Terwee CB, Dekker FW, Mourits MP, Gerding MN, Baldeschi L, Kalmann R, et al. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. Clin Endocrinol. 2001;54(3):391–8.

26 Terwee C, Wakelkamp I, Tan S, Dekker F, Prummel MF, Wiersinga W. Long-term effects of Graves' ophthalmopathy on health-related quality of life. Eur J Endocrinol. 2002;146(6):751–7.

27 Jellema HM, Merckel-Timmer E, Kloos R, Saeed P, Mourits MP. Quality of life improves after strabismus surgery in patients with Graves' orbitopathy. Eur J Endocrinol. 2014; 170(5):785–9.
28 Terwee CB, Dekker FW, Prummel MF, Wiersinga WM. Chapter 7. English translation of the GO-QoL. In: Terwee CB, editor. Graves' ophthalmopathy through the eyes of the patient. Assessment of health-related quality of life. PhD thesis. Netherlands: University of Amsterdam; 2000. p. 131–43.: http://hdl.handle.net/11245/2.9088.
29 Dietrich A, Taylor P, White P, Wilson V, Uddin J, Lee RWJ, et al. CIRTED investigators: Establishing the usefullness of the GO-QoL in a UK hospital-treated population with thyroid eye disease in the CIRTED trial. Psychol Health Med. 2018;23(Suppl 1):1341–65.
30 Lin IC, Lee CC, Liao SL. Assessing quality of life in Taiwanese patients with Graves' ophthalmopathy. J Formos Med Assoc. 2015; 114(11):1047–54.
31 Shen L, Zhu H, Li XX, Liu DD, Chen XY, Liu H. [The influence factors of quality-of-life in patients with thyroid-associated ophthalmopathy]. Zhonghua Yan Ke Za Zhi. 2017; 53(8):575–82.
32 Zeng P, Fan S-X, Li Z-J, Peng Y-Y, Hu Y-X, Xu M-T, et al. Evaluation of the Graves' orbitopathy-specific quality of life questionnaire in the mainland Chinese population. J Ophthalmol. 2019;2019:7602419.
33 Ponto KA, Hommel G, Pitz S, Elflein H, Pfeiffer N, Kahaly GJ. Quality of life in a German Graves orbitopathy population. Am J Ophthalmol. 2011;152(3):483–90.e1.
34 Choi YJ, Lim HT, Lee SJ, Lee SY, Yoon JS. Assessing Graves' ophthalmopathy-specific quality of life in Korean patients. Eye. 2012; 26(4):544–51.
35 Kashkouli MB, Karimi N, Aghamirsalim M, Abtahi MB, Nojomi M, Shahrad-Bejestani H, et al. Measurement properties of the Persian translated version of Graves orbitopathy quality of life questionnaire: a validation study. Ophthalmic Epidemiol. 2017;24(1): 3–10.
36 Sawicka-Gutaj N, Bednarczuk T, Daroszewski J, Waligórska-Stachura J, Miśkiewicz P, Sowiński J, et al. GO-QOL: disease-specific quality of life questionnaire in Graves' orbitopathy. Endokrynol Pol. 2015;66(4):362–6.
37 Villagelin D, Romaldini J, Andrade J, Santos R, Milkos A, Teixeira PFDS, et al. Evaluation of quality of life in the Brazilian Graves' disease population: focus on mild and moderate Graves' orbitopathy patients. Front Endocrinol. 2019 Apr 4;10:192.
38 Delfino LC, Zunino A, Sapia V, Croome MdCS, Ilera V, Gauna AT. Related quality of life questionnaire specific to dysthyroid ophthalmopathy evaluated in a population of patients with Graves' disease. Arch Endocrinol Metab. 2017;61(4):374–81.
39 Watt T, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedüs L, Bjorner JB, et al. Quality of life in patients with benign thyroid disorders. A review. Eur J Endocrinol. 2006; 154(4):501–10.
40 Du Y, Ye H, Li K, Xiao X, Chen R, He JF, et al. Vision-related quality of life tends to be more severely impaired in patients with dysthyroid optic neuropathy. Curr Eye Res. 2014;39(5): 532–6.
41 Terwee CB, Dekker FW, Prummel MF, Wiersinga WM. Graves' ophthalmopathy through the eyes of the patient: a state of the art on health-related quality of life assessment. Orbit. 2007;20(4):281–90.
42 Watt T, Cramon P, Hegedüs L, Bjorner JB, Bonnema SJ, Rasmussen ÅK, et al. The thyroid-related quality of life measure ThyPRO has good responsiveness and ability to detect relevant treatment effects. J Clin Endocrinol Metab. 2014;99(10):3708–17.
43 Watt T, Bjorner JB, Groenvold M, Rasmussen AK, Bonnema SJ, Hegedüs L, et al. Establishing construct validity for the thyroid-specific patient reported outcome measure (ThyPRO): an initial examination. Qual Life Res. 2009;18(4):483–96.
44 Watt T, Groenvold M, Deng N, Gandek B, Feldt-Rasmussen U, Rasmussen ÅK, et al. Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. Health Qual Life Outcomes. 2014;12:126.
45 Watt T, Barbesino G, Bjorner JB, Bonnema SJ, Bukvic B, Drummond R, et al. Cross-cultural validity of the thyroid-specific quality-of-life patient-reported outcome measure, ThyPRO. Qual Life Res. 2015;24(3):769–80.
46 Watt T, Bjorner JB, Groenvold M, Cramon P, Winther KH, Hegedüs L, et al. Development of a short version of the thyroid-related patient-reported outcome ThyPRO. Thyroid. 2015;25(10):1069–79.
47 Rasmussen SL, Rejnmark L, Ebbehøj E, Feldt-Rasmussen U, Rasmussen ÅK, Bjorner JB, et al. High level of agreement between electronic and paper mode of administration of a thyroid-specific patient-reported outcome, ThyPRO. Eur Thyroid J. 2016;5(1):65–72.
48 Fayers T, Fayers PM, Dolman PJ. Sensitivity and responsiveness of the patient-reported TED-QoL to rehabilitative surgery in thyroid eye disease. Orbit. 2016;35(6):328–34.
49 de Boer MR, Moll AC, de Vet HC, Terwee CB, Völker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt. 2004;24(4):257–73.
50 Wong CK, Lang BH, Lam CL. A systematic review of quality of thyroid-specific health-related quality-of-life instruments recommends ThyPro for patients with benign thyroid diseases. J Clin Epidemiol. 2016;78:63–72.
51 The European Group on Graves' Orbitopathy; Wiersinga WM, Wiersinga WM, Perros P, Kahaly GJ, Mourits MP, 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.
52 Marcocci C, Kahaly GJ, Krassas GE, Bartalena L, Prummel M, Stahl M, et al. Selenium and the course of mild Graves' orbitopathy. N Engl J Med. 2011;364(20):1920–31.
53 Aktaran S, Akarsu E, Erbağci I, Araz M, Okumuş S, Kartal M. Comparison of intravenous methylprednisolone therapy vs. oral methylprednisolone therapy in patients with Graves' ophthalmopathy. Int J Clin Pract. 2007;61(1):45–51.
54 Bartalena L, Krassas GE, Wiersinga W, Marcocci C, Salvi M, Daumerie C, et al. Efficacy and safety of three different cumulative doses of intravenous methylprednisolone for moderate to severe and active Graves' orbitopathy. J Clin Endocrinol Metab. 2012;97(12):4454–63.
55 Kahaly GJ, Riedl M, König J, Pitz S, Ponto K, Diana T, et al. Mycophenolate plus methylprednisolone versus methylprednisolone alone in active, moderate-to-severe Graves' orbitopathy (MINGO): a randomised, observer-masked, multicentre trial. Lancet Diabetes Endocrinol. 2018;6(4):287–98.
56 Swiglo BA, Murad MH, Schünemann HJ, Kunz R, Vigersky RA, Guyatt GH, et al. A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the grading of recommendations, assessment, development, and evaluation system. J Clin Endocrinol Metab. 2008; 93(3):666–73.
57 EUGOGO; Mourits MP, Mourits MP, Bijl H, Altea MA, Baldeschi L, et al. Outcome of orbital decompression for disfiguring proptosis in patients with Graves' orbitopathy using various surgical procedures. Br J Ophthalmol. 2009;93(11):1518–23.
58 Wickwar S, McBain H, Ezra DG, Hirani SP, Rose GE, Newman SP. The psychosocial and clinical outcomes of orbital decompression surgery for thyroid eye disease and predictors of change in quality of life. Ophthalmology. 2015;122(12):2568–76.e1.
59 Iacobaeus L, Sahlin S. Evaluation of quality of life in patients with Graves' ophthalmopathy, before and after orbital decompression. Orbit. 2016;35:121–5.
60 Cheng AM, Wei YH, Tighe S, Sheha H, Liao SL. Long-term outcomes of orbital fat decompression in Graves' orbitopathy. Br J Ophthalmol. 2018;102:69–73.
61 Fichter N, Krentz H, Guthoff RF. Functional and esthetic outcome after bony lateral wall decompression with orbital rim removal and additional fat resection in Graves' orbitopathy with regard to the configuration of the lateral canthal region. Orbit. 2013;32(4):239–46.
62 Jellema HM, Saeed P, Mombaerts I, Dolman PJ, Garrity J, Kazim M, et al. Objective and subjective outcomes of strabismus surgery in Graves' orbitopathy: a prospective multicentre study. Acta Ophthalmol. 2017;95(4):386–91.

63 Kashkouli MB, Heidari I, Pakdel F, Jam S, Honarbakhsh Y, Mirarmandehi B. Change in quality of life after medical and surgical treatment of Graves' ophthalmopathy. Middle East Afr J Ophthalmol. 2011;18(1):42–7.
64 Dickinson AJ, Hintschich C. Clinical manifestations. In: Wiersinga WM, Kahaly GJ, editors. Graves' orbitopathy. A multidisciplinary approach – questions and answers. 3rd, revised and expanded ed. Basel: Karger; 2017. p. 1–25.
65 Bahn RS, Gorman CA. Choice of therapy and criteria for assessing treatment outcome in thyroid-associated ophthalmopathy. Endocrinol Metab Clin North Am. 1987;16(2): 391–407.
66 Kahaly GJ, Rösler HP, Pitz S, Hommel G. Low- versus high-dose radiotherapy for Graves' ophthalmopathy: a randomized, single blind trial. J Clin Endocrinol Metab. 2000; 85(1):102–8.
67 Marcocci C, Bartalena L, Tanda ML, Manetti L, Dell'Unto E, Rocchi R, et al. Comparison of the effectiveness and tolerability of intravenous or oral glucocorticoids associated with orbital radiotherapy in the management of severe Graves' ophthalmopathy: results of a prospective, single-blind, randomized study. J Clin Endocrinol Metab. 2001;86(8):3562–7.
68 Macchia PE, Bagattini M, Lupoli G, Vitale M, Vitale G, Fenzi G. High-dose intravenous corticosteroid therapy for Graves' ophthalmopathy. J Endocrinol Invest. 2001;24(3):152–8.
69 Kauppinen-Mäkelin R, Karma A, Leinonen E, Löyttyniemi E, Salonen O, Sane T, et al. High dose intravenous methylprednisolone pulse therapy versus oral prednisone for thyroid-associated ophthalmopathy. Acta Ophthalmol Scand. 2002;80(3):316–21.
70 Dickinson AJ, Vaidya B, Miller M, Coulthard A, Perros P, Baister E, et al. Double-blind, placebo-controlled trial of octreotide long-acting repeatable (LAR) in thyroid-associated ophthalmopathy. J Clin Endocrinol Metab. 2004; 89(12):5910–5.
71 Wémeau JL, Caron P, Beckers A, Rohmer V, Orgiazzi J, Borson-Chazot F, et al. Octreotide (long-acting release formulation) treatment in patients with Graves' orbitopathy: clinical results of a four-month, randomized, placebo-controlled, double-blind study. J Clin Endocrinol Metab. 2005;90(2):841–8.
72 Ng CM, Yuen HK, Choi KL, Chan MK, Yuen KT, Ng YW, et al. Combined orbital irradiation and systemic steroids compared with systemic steroids alone in the management of moderate-to-severe Graves' ophthalmopathy: a preliminary study. Hong Kong Med J. 2005;11(5):322–30.
73 Kahaly GJ, Pitz S, Hommel G, Dittmar M. Randomized, single blind trial of intravenous versus oral steroid monotherapy in Graves' orbitopathy. J Clin Endocrinol Metab. 2005; 90(9):5234–40.
74 Chang TC, Liao SL. Slow-release lanreotide in Graves' ophthalmopathy: a double-blind randomized, placebo-controlled clinical trial. J Endocrinol Invest. 2006;29(5):413–22.
75 Stan MN, Garrity JA, Bradley EA, Woog JJ, Bahn MM, Brennan MD, et al. Randomized, double-blind, placebo-controlled trial of long-acting release octreotide for treatment of Graves' ophthalmopathy. J Clin Endocrinol Metab. 2006;91(12):4817–24.
76 Menconi F, Marinò M, Pinchera A, Rocchi R, Mazzi B, Nardi M, et al. Effects of total thyroid ablation versus near-total thyroidectomy alone on mild to moderate Graves' orbitopathy treated with intravenous glucocorticoids. J Clin Endocrinol Metab. 2007;92(5):1653–8.
77 Zhu W, Ye L, Shen L, Jiao Q, Huang F, Han R, et al. A prospective, randomized trial of intravenous glucocorticoids therapy with different protocols for patients with Graves' ophthalmopathy. J Clin Endocrinol Metab. 2014; 99(6):1999–2007.
78 He Y, Mu K, Liu R, Zhang J, Xiang N. Comparison of two different regimens of intravenous methylprednisolone for patients with moderate to severe and active Graves' ophthalmopathy: a prospective, randomized controlled trial. Endocr J. 2017;64(2):141–9.
79 Ye X, Bo X, Hu X, Cui H, Lu B, Shao J, et al. Efficacy and safety of mycophenolate mofetil in patients with active moderate-to-severe Graves' orbitopathy. Clin Endocrinol. 2017; 86(2):247–55.
80 Rajendram R, Taylor PN, Wilson VJ, Harris N, Morris OC, Tomlinson M, et al. Combined immunosuppression and radiotherapy in thyroid eye disease (CIRTED): a multicentre, 2 × 2 factorial, double-blind, randomised controlled trial. Lancet Diabetes Endocrinol. 2018;6(4):299–309.
81 Douglas RS, Kahaly GJ, Patel A, Sile S, Thompson EHZ, Perdok R, et al. Teprotumumab for the treatment of active thyroid eye disease. N Engl J Med. 2020;382(4):341–52.
82 Piantanida E, Tanda ML, Lai A, Sassi L, Bartalena L. Prevalence and natural history of Graves' orbitopathy in the XXI century. J Endocrinol Invest. 2013;36(6):444–9.
83 Mourits MP, Koornneef L, Wiersinga WM, Prummel MF, Berghout A, van der Gaag R. Clinical criteria for the assessment of disease activity in Graves' ophthalmopathy: a novel approach. Br J Ophthalmol. 1989;73(8):639–44.
84 Mourits MP, Prummel MF, Wiersinga WM, Koornneef L. Clinical activity score as a guide in the management of patients with Graves' ophthalmopathy. Clin Endocrinol. 1997; 47(1):9–14.
85 Yang M, Wiersinga WM, Soeters MR, Mourits MP. What is the aim of immunosuppressive treatment in patients with Graves' orbitopathy? Ophthalmic Plast Reconstr Surg. 2014;30(2):157–61.
86 Zang S, Ponto KA, Kahaly GJ. Clinical review: intravenous glucocorticoids for Graves' orbitopathy: efficacy and morbidity. J Clin Endocrinol Metab. 2011;96(2):320–32.
87 Marcocci C, Watt T, Altea MA, Rasmussen AK, Feldt-Rasmussen U, Orgiazzi J, et al. Fatal and non-fatal adverse events of glucocorticoid therapy for Graves' orbitopathy: a questionnaire survey among members of the European thyroid association. Eur J Endocrinol. 2012;166(2):247–53.
88 ICH. Data elements for transmission of individual case safety reports E2B. London: International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2001.
89 Bartalena L, Veronesi G, Krassas GE, Wiersinga WM, Marcocci C, Marinò M, et al. Does early response to intravenous glucocorticoids predict the final outcome in patients with moderate-to-severe and active Graves' orbitopathy? J Endocrinol Invest. 2017;40(5): 547–53.
90 Ponto KA, Merkesdal S, Hommel G, Pitz S, Pfeiffer N, Kahaly GJ. Public health relevance of Graves' orbitopathy. J Clin Endocrinol Metab. 2013;98(1):145–52.