Factors Predicting Long-term Outcome and the Need for Surgery in Graves Orbitopathy: Extended Follow-up From the CIRTED Trial

Peter Taylor,[1] Rathie Rajendram,[2] Stephanie Hanna,[1] Victoria Wilson,[3] Julie Pell,[1] Chunhei Li,[1] Anne Cook,[4] Rao Gattamaneni,[5] Nicholas Plowman,[6] Sue Jackson,[7] Robert Hills,[8] Robert French,[1] Jimmy M. Uddin,[2] Richard W. J. Lee,[2,3,9,*] and Colin M. Dayan;[1,*] on behalf of the CIRTED investigators

[1]Thyroid Research Group, Cardiff University School of Medicine, Cardiff CF14 4XN, UK
[2]Moorfields Eye Hospital NHS Foundation Trust, London, EC1V 2PD, UK
[3]Bristol Medical School, University of Bristol, UK
[4]Manchester Royal Eye Hospital, Manchester, M13 9WH, UK
[5]Christie Hospital, Manchester, M20 4BX, UK
[6]Bristol Haematology and Oncology Centre, Bristol, BS2 8ED, UK
[7]Centre for Appearance Research, University of the West of England, Bristol, BS16 2JP, UK
[8]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
[9]Institute of Ophthalmology, Faculty of Brain Sciences, University College London, London EC1V 9EL, UK
Correspondence: Peter Taylor, PhD, Thyroid Research Group, C2 Link Corridor, UHW, Heath Park, Cardiff CF14 4XN, UK. Email: taylorpn@cardiff.ac.uk.

*Joint senior authors.

Abstract

Graves orbitopathy is both disabling and disfiguring. Medical therapies to reduce inflammation are widely used, but there is limited trial data beyond 18 months of follow-up.
Methods: Three-year follow-up of a subset of the CIRTED trial (N = 68), which randomized patients to receive high-dose oral steroid with azathioprine/placebo and radiotherapy/sham radiotherapy.
Results: Data were available at 3 years from 68 of 126 randomized subjects (54%). No additional benefit was seen at 3 years for patients randomized to azathioprine or radiotherapy with regard to a binary clinical composite outcome measure (BCCOM), modified European Group on Graves' Orbitopathy score, or Ophthalmopathy Index.
  Clinical Activity Score (CAS), Ophthalmopathy Index, and Total Eye Score improved over 3 years ($P < .001$). However, quality of life at 3 years remained poor. Of 64 individuals with available surgical outcome data, 24 of 64 (37.5%) required surgical intervention. Disease duration of greater than 6 months before treatment was associated with increased need for surgery [odds ratio (OR) 16.8; 95% CI 2.95, 95.0; $P = .001$]. Higher baseline levels of CAS, Ophthalmopathy Index, and Total Eye Score but not early improvement in CAS were associated with increased requirement for surgery.
Conclusion: In this long-term follow-up from a clinical trial, 3-year outcomes remained suboptimal with ongoing poor quality of life and high numbers requiring surgery. Importantly, reduction in CAS in the first year, a commonly used surrogate outcome measure, was not associated with improved long-term outcomes.
Key Words: graves orbitopathy, CAS, GOQOL, azathioprine, radiotherapy, CIRTED

Graves' disease is common throughout the world (1). Active moderate to severe Graves orbitopathy (GO; also known as thyroid eye disease or thyroid-associated orbitopathy) occurs in approximately 5% to 10% of cases of Graves' disease (2, 3). It can be both visually disabling and cosmetically disfiguring and substantially impairs quality of life (2, 4-6). Middle-aged women are predominantly affected, and the negative impact of GO can be prolonged.

Current medical therapeutic strategies have been directed toward suppression of orbital inflammation in the hope of reducing tissue remodeling in the extraocular muscles, orbital fat, and other periocular soft tissues (4, 7). Immunosuppressive therapies, particularly corticosteroids, are the mainstay of treatment (3, 8, 9). More recently, randomized clinical trials have indicated substantial potential benefit from novel/additional agents such as azathioprine (10), teprotumumab (11), mycophenolate, (12) and tocilizumab (13). However, these clinical trials have largely been of modest duration, and follow-up has been at most 18 months and often less. One exception is the study of Leo et al that followed patients up for 4 to 10 years. This

showed continued improvement and a requirement for surgery in 27% of patients, but limited quality of life data was reported (14). Alongside these advances, surgery continues to be required to reduce the degree of diplopia or proptosis (3) and in some cases to preserve vision, as well as to improve quality of life (15). Frequently, surgical intervention is performed a year or more after disease onset and so is rarely included in trial outcomes. At present there is limited data on what proportion of individuals with moderate to severe GO require surgery despite extensive medical intervention. Taken together, there is a pressing need for longer term data to establish which of the characteristics of GO correlate with poor clinical outcomes once the inflammatory phase of the disease has resolved. These may be key in assessing who will potentially benefit the most from new and costly interventions and provide insight into disease trajectory while awaiting for trials with longer duration of follow-up.

There are several well-established tools for assessing GO. Typically disease activity is measured using Clinical Activity Score (CAS) or by the inflammatory component of the vision, inflammation, strabismus, and appearance scale (3, 16, 17). Other scores objectively assess the severity of GO including European Group on Graves' Orbitopathy (EUGOGO) (18, 19), Ophthalmopathy Index (OI) (20), Thyroid Eye Score, NOSPECS (21, 22), Graves Ophthalmopathy Quality of Life (GOQOL) (23, 24), and the vision, inflammation, and appearance component of the vision, inflammation, strabismus, and appearance scale (17). In addition to these measures, a greater understanding of disease trajectory is essential for identifying which patients should be followed more closely in clinic. Potential variables that might predict the disease trajectory and likely need for additional interventions such as surgery include patient characteristics such as baseline GO severity scores as well as age, sex, and duration of disease as well as potent risk factors such as smoking status.

The CIRTED trial was a multicenter, factorial design, double-masked, randomized controlled trial based in the UK. Results from the primary outcome at 1 year have been reported (10), but participants were also invited to attend an optional follow-up assessment at 3 years. All patients in the study received high-dose oral corticosteroids and were randomized to receive either azathioprine or placebo and orbital radiotherapy or sham radiotherapy. These planned medical interventions were completed within 6 months of study entry. The 3-year data in CIRTED enables us to identify important predictors of disease outcome and treatment response, especially with regard to quality of life and the need for surgery.

Methods

Study Details and Procedures

The CIRTED trial was a randomized multifactorial trial and has been described in detail previously (10, 25). Baseline characteristics were obtained on age at enrollment, disease duration, ethnicity, sex thyroid status, study center, CAS, and Thyroid Eye Score. In brief, all patients received high-dose oral prednisolone in a tapering regime for 24 weeks (80 mg per day, reduced to 20 mg per day by 6 weeks, 10 mg per day by 15 weeks, and 5 mg per day by 21 weeks). Those randomized to radiotherapy received 20 Gy of radiation administered to the retrobulbar compartment in 10 to 12 fractions over 2 to 3 weeks. Those randomized to receive azathioprine received this treatment for 48 weeks with doses adjusted according to weight and safety blood monitoring. The majority of patients were recruited from 2 centers in the UK (Bristol and Moorfields; n = 105) (10). The CIRTED trial identified potential benefit at 48 weeks for azathioprine, but no clear benefit was observed for radiotherapy (10). Following week 48 (completion of the original CIRTED study) follow-up was optional and patients were managed by their local teams.

Outcomes

In the original study, the coprimary outcomes were a BCCOM (Box 1) and OI. Secondary outcome measures included Total Eye Score (TES) as an additional assessment of disease severity and the patient-reported GOQOL score (6, 23, 24), which has 2 components: Visual Function (GOQOL-VF) and Appearance (GOQOL-AP). CAS was also measured throughout. In this 3-year report we additionally calculated the EUGOGO score from the individual components of the trial dataset (18, 19). As this did not include a measurement of lid retraction, we inferred the presence of lid retraction if there was a >2 mm difference in palpebral aperture between an individual patient's eyes. On this basis, moderate to severe GO was classified as having 2 or more of ≥ 2 mm difference in palpebral aperture, moderate or severe soft tissue involvement, exophthalmos ≥ 3 mm above normal for race and gender, and inconstant or constant diplopia. Hence, the indices we used to assess longitudinal features and stability of disease over 3 years or follow-up were the OI, TES, CAS, EUGOGO, and GOQOL scores. In addition, whether ophthalmic surgery was required, which was entirely at the discretion of the patients and their local ophthalmologists, was also recorded.

Box 1 Calculation of the BCCOM

BCCOM Major Criteria
- An improvement of ≥1 grade in diplopia score
- An improvement of >8 degrees of eye movement in any direction
- A reduction of ≥2 mm in proptosis

BCCOM Minor Criteria
- A reduction of ≥2 mm in lid aperture
- An improvement of ≥1 grade in soft tissue involvement
- An improvement in best-corrected visual acuity of ≥1 line on the Snellen chart
- Subjective improvement

All items refer to the worst eye.

Response to treatment is calculated as follows:

Improved = improvement in ≥1 major criteria or ≥2 minor criteria

No Change = improvement or deterioration in ≤1 minor criterion

Worse = deterioration in ≥1 major or ≥2 minor criteria (even if other criteria improve) or requiring rescue therapy (radiotherapy, intravenous steroid, or orbital decompression) at any point

Statistical Analysis

The impact of being randomized to azathioprine and radiotherapy against the original primary outcome measures of disease severity BCCOM and OI were studied first. CAS was a coprimary outcome, although we anticipated all participants would have a significant improvement in CAS well before 3 years in accordance with the natural history of the disease (26) and the EUGOGO score was utilized as an additional outcome. Analysis was undertaken adjusting for the same confounders as in the initial CIRTED analysis; these included smoking status at the time of GO diagnosis, thyroid status on enrollment, previous corticosteroid use, sex, disease severity, disease duration, age greater than 60 years, and disease activity (baseline CAS). Additional analyses were then undertaken to explore the effect of being randomized to azathioprine or radiotherapy on TES, quality of life scores, and need for ophthalmological surgery.

Multivariable linear and logistic regression models were used, adjusting for minimization variables, the factorial design, and the value of the outcome variable at baseline. Where appropriate, to allow easy comparison variables were standardized; these analyses are therefore presented as per standard deviation. Comparison between nonnormally distributed outcomes between groups was undertaken using the Wilcoxon rank sum test. The original trial was a factorial design, and, as patients who followed up at 3 years were broadly representative of the whole trial population, with no striking benefit observed for either azathioprine or radiotherapy and no evidence of interaction between them, we combined treatment groups when exploring baseline ophthalmic and clinical assessments and their relationship to subsequent clinical outcomes to increase power. In these analyses we explored the relationship between the baseline clinical assessments of CAS, OI, TES, GOQOL-VF, and GOQOL-AP and their respective scores at 3 years. We also explored the association between baseline CAS (as a continuous measure), OI (in quartiles) TES, and EUGOGO scores as well as baseline characteristics on the need for ophthalmological surgery over the 3 years of follow-up.

Statistical significance was defined in advance as a *P* value of <.05. All statistical analyses were undertaken using STATA version 16 (STATACORP, College Station, TX, USA).

Results

Data were available after 3 years of follow-up from 68 of 126 (54%) individuals initially randomized into the CIRTED study. Of these 68 individuals, 64 provided data on their requirement for ophthalmological surgery, 58 on CAS score, 52 on BCCOM, 52 on GOQOL, and 49 on OI. The EUGOGO score could be calculated from the individual measurements recorded in our dataset (as set out previously) for 47 trial subjects. The 68 individuals with 3-year data were fairly evenly distributed across the intervention groups: 17 were in the placebo/sham radiotherapy group, 16 were in the placebo/radiotherapy group, 16 were in the azathioprine/sham radiotherapy group, and 19 were in the azathioprine/radiotherapy group. The derivation of study participants is shown in Figure 1.

Baseline characteristics of key risk factors for GO, potential confounders, and disease levels at baseline (CAS/TES) for trial subjects who provided 3-year data, compared to those who did not, did not show any substantial differences (Table 1)

aside from ethnicity, where non-Caucasians were less likely to provide 3-year data (OR = 0.30; 95% CI .11, .79; *P* = .02). Baseline characteristics of the 3-year study group vs the remainder of the CIRTED cohort are shown in Supplementary Table 1 (27).

Analysis of the Effect of Azathioprine and Radiotherapy on Primary Outcomes at 3 Years

At the 3-year assessment, we observed no clear benefit in terms of improvement in BCCOM between individuals randomized to azathioprine vs placebo. The adjusted OR for improvement in BCCOM in individuals randomized to azathioprine was $OR_{(adj)}$ = 0.51 (95% CI .08, 3.13; *P* = .47) (Table 2). There was also no clear benefit of improvement with orbital radiotherapy vs sham radiotherapy in terms of improvement in BCCOM. The adjusted OR for improvement was $OR_{(adj)}$ = 2.17 (95% CI .43, 10.9; *P* = .35) (Table 2). There was no evidence of interaction between azathioprine and radiotherapy (p_{int} = 0.20).

No additional benefits above oral steroid were seen with either azathioprine or orbital radiotherapy on OI. Individuals randomized to azathioprine had an adjusted beta $[(B)_{(adj)}]$ of 1.33 (95% CI −.68, 3.34; *P* = .19) on OI, and for those randomized to orbital radiotherapy, $B_{(adj)}$ was 0.11 (95% CI −1.93, 2.13; *P* = .92) on OI (Table 2). There was also no evidence of an interaction between azathioprine and radiotherapy in their effect on OI (p_{int} = 0.85).

No differences in change in CAS were observed between individuals who received treatment with azathioprine vs placebo ($B_{(adj)}$ = 0.58; 95% CI −.20, 1.36; *P* = .14) or those who received radiotherapy vs sham radiotherapy ($B_{(adj)}$ = 0.20; 95% CI −.58, .98; *P* = .61) (Table 2). There was no evidence of interaction between azathioprine and radiotherapy in their effect on CAS (p_{int} = 0.76).

We also observed no differences in terms of improvement in EUGOGO score between individuals randomized to azathioprine vs placebo or radiotherapy vs sham radiotherapy. The adjusted OR for having odds of moderate to severe GO as opposed to mild GO in individuals randomized to azathioprine was $OR_{(adj)}$ = 2.48 (95% CI .14, 44.7; *P* = .54) and for radiotherapy was $OR_{(adj)}$ = 0.58 (95% CI .03, 10.97; *P* = .72) (Table 1).

Analysis of key individual components of primary outcomes showed no clear benefits for azathioprine or radiotherapy on individual components apart from radiotherapy being associated with reduced incidence of diplopia at 3 years (B = −0.56; 95% CI −1.08, −.06; *P* = .03) (Supplementary Table 2) (27).

Analysis of the Effect of Azathioprine and Radiotherapy on Additional Outcomes at 3 Years

Individuals randomized to azathioprine had a higher TES after adjustment than those who received placebo [TES $(B)_{(adj)}$ of 3.11; 95% CI .04, 6.19; *P* = .05), and in those randomized to orbital radiotherapy the $B_{(adj)}$ was −0.61 (95% CI −3.87, 2·65; *P* = .71) (Table 2). There was also no evidence of an interaction between azathioprine and radiotherapy in their effect on OI (p_{int} = .31).

Individuals who were randomized to receive azathioprine did not have higher GOQOL-VF than those who did not (*P* = .32), nor did those randomized to receive radiotherapy (*P* = .36). A potential late benefit of radiotherapy was seen with regard to GOQOL-AP [median 81.3 (interquartile range [IQR] 62.5-93.8) vs median 68.9 (IQR 31.2-87.5); *P* = .03]. In
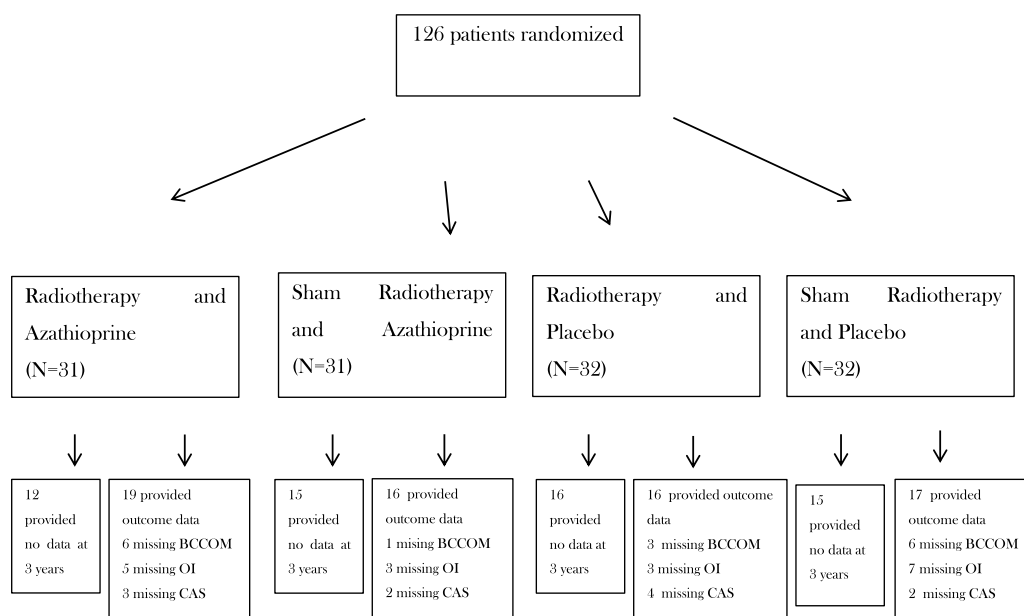
Figure 1. Derivation of study participants.

contrast, no late benefit was observed for GOQOL-AP in those who received azathioprine [median 68.9 (IQR 31.3-87.5) vs median 84.4 (IQR 68.8-93.8) for placebo; $P = .06$].

No clear benefit on reducing the odds of requiring any surgery was observed for either azathioprine (OR = 0.39; 95% CI .09, 1.69; $P = .21$) or radiotherapy (OR = .34; 95% CI .08, 1.52; $P = .16$) (Table 3).

Individuals randomized to receive azathioprine did not see a reduction in the need for lid surgery (OR = 0.60; 95% CI .12, 3.04; $P = .54$); however, a borderline significant result was observed for radiotherapy (OR = 0.17; 95% CI .03, 1.03; $P = .05$) with the 95% CI crossing equality (Supplementary Table 3) (27). No benefit was seen for either azathioprine (OR = 0.39; 95% CI .06, 2.68) or radiotherapy (OR = 0.71; 95% CI .09. 5.74; $P = .75$) in reducing the need for orbital decompression (Supplementary Table 4) (27).

Changes in Clinical Assessments of GO Over the Study Period

OI fell over the study period from a mean at 12 weeks 9.45 (SD 3.95) to 6.02 (SD 2.99) ($P = <.001$ at 3 years) (Fig. 2A). OI at 12 weeks correlated with OI at 3 years (B_{(std)} = 0.59; 95% CI .34, .84; $P < .001$; $r^2 = 0.35$). OI at 48 weeks was also correlated with OI at 3 years (B_{(std)} = 0.42; 95% CI .21, .63; $P < .001$; $r^2 = 0.28$).

TES fell over the study period from a baseline mean of 14.9 (SD 6.27) to 6.33 (SD 4.65) ($P < .001$ at 3 years) (Fig. 2B). Baseline TES was not associated with TES at 3 years (B_{(std)} = −0.13; 95% CI −0.42, .16; $P = .38$; $r^2 = 0.01$) but was associated with TES at week 48 (B_{(std)} = 0.50; 95% CI .32, .68; $P < .001$; $r^2 = 0.25$). Week 48 TES was not associated with TES at 3 years (B_{(std)} = −0.05; 95% CI −.33, .23; $P = .71$; $r^2 = 0.002$).

CAS score also fell progressively over the study period from a median baseline of 5 (IQR 4-5) to a 3-year median of 1 (IQR 0-1; $P = <.001$) (Fig. 2C). By week 12, 17.8% of individuals had a CAS of 0 or 1, rising to 23.8% by 24 weeks 40.8% by week 48 and 58.6% by 3 years. Baseline CAS was not associated with CAS at 3 years (B_{(std)} = 0.07; 95% CI −.20, .33;

$P = .62$; $r^2 = 0.005$). However, baseline CAS was associated with CAS at 12 weeks (B_{(std)} = 0.37; 95% CI .18, .55; $P < .001$; $r^2 = 0.14$) and to a lesser extent CAS at 48 weeks (B_{(std)} = 0.24; 95% CI .04, .44; $P = .02$; $r^2 = 0.05$). Not even CAS at week 48 was associated with CAS at 3 years (B_{(std)} = 0.09; 95% CI −.18, .37; $P = .49$; $r^2 = 0.01$).

EUGOGO score fell progressively over the study from 70.8% of patients having moderate to severe GO at baseline to 38.9% having moderate to severe GO at 12 weeks, falling further to 32.1% at 48 weeks and 6.4% at 3 years. Baseline EUGOGO score was associated with 12-week EUGOGO score (B = 0.53; 95% CI .26, .79; $P < .001$; $r^2 = 0.24$) but not 48-week score (B = 0.26; 95% CI −.02, .54; $P = .07$; $r^2 = 0.05$) or 3-year score ($P = .64$).

GOQOL-VF improved progressively over the study period from a baseline median of 68.7 (IQR 43.3-87.5) to 93.8 (IQR 76.7-100) at 3 years ($P < .001$) (Fig. 2D). GOQOL-VF at baseline was associated with GOQOL-VF at 3 years (B_{(std)} = 0.62; 95% CI .39, .85; $P < .001$; $r^2 = 0.35$). GOQOL-AP also rose over the study period, although to a lesser degree than visual function, rising from a baseline median of 56.2 (IQR 37.5 −75) to 75 (IQR 56.3-93.8; $P = .002$) (Fig. 2E). GOQOL-AP at baseline was associated with GOQOL-AP at 3 years (B_{(std)} = 0.61; 95% CI .38, .83; $P < .001$; $r^2 = 0.35$).

Outcomes: Factors Predicting the Need for Surgery Over the Study Period (3 Years).

Requirement for any surgical intervention

Twenty-four of the 64 individuals (37.5%) who provided surgical outcome data required surgery over 3 years of follow-up. The decision to perform surgery was left to the local clinical team and was not protocolized. The majority of interventions involved lid surgery (n = 16) or orbital decompression (n = 12), with strabismus correction (squint surgery) being less common (n = 2). Five individuals required both lid surgery and decompression surgery. Eight of 49 respondents (16.33%) had undergone thyroidectomy over the 3-year period.

Table 1. Odds of providing 3 year data

Reason	OR	95% CI	P	OR[a,b]	95% CI[a]	P[a]
Age						
<60 years	1			1		
≥60 years	2.25	(0.80, 6.28)	.12	1.22	(0.39, 3.83)	.74
Gender						
Male	1			1		
Female	1.35	(0.61, 2.99)	.46	1.56	(0.63, 3.83)	.33
Ethnicity						
Caucasian	1			1		
Non-Caucasian	0.39	(0.17, 0.86)	.02	0.30	(0.11, 0.79)	.02
Thyroid state						
Hypothyroid/euthyroid	1			1		
Hyperthyroid	1.01	(0.42, 2.39)	.99	0.77	(0.29, 2.02)	.59
Recent use of steroids						
No recent use	1			1		
Use in last 6 months	0.83	(0.29, 2.38)	.73	0.78	(0.29, 2.07)	0.61
Disease duration						
<6 months	1			1		
≥6 months	1.26	(0.58, 2.71)	.56	1.40	(0.58, 3.41)	.46
Smoking status						
Nonsmoker	1			1		
Smoker	0.57	(0.28, 1.17)	.12	0.46	(0.20, 1.07)	.07
Study center						
Other	1			1		
Bristol/Moorfields	0.68	(0.26, 1.77)	.43	0.85	(0.27, 2.74)	.79
Baseline CAS						
2-3	1			1		
4-5	0.49	(0.18, 1.38)	.18	0.41	(0.13, 1.31)	.13
6-7				0.64		
Baseline TES	0.96	(0.28, 3.27)	.94		(0.15, 2.77)	.55
<22	1			1		
≥22	0.58	(0.23, 1.51)	.27	0.49	(0.16, 1.48)	.21
Received Aza	1.07	(0.53, 2.16)	.85	0.96	(0.43, 2.15)	.92
Received RT	1.47	(0.73, 2.97)	.28	1.50	(0.68, 3.29)	.32

Sixty-eight individuals provided data; 58 did not.
P = P value against the null hypothesis of no association.
Abbreviations: Aza, azathioprine; CAS, Clinical Activity Score; OR, odds ratio; RT, radiotherapy; TES, Total Eye Score.
[a]Adjusted for age group, ethnicity, smoking status, gender, thyroid state, disease duration, study center, recent steroid use, baseline CAS, baseline TES.
[b]Four individuals with incomplete baseline CAS excluded from analysis.

Disease duration of greater than 6 months at baseline was associated with the need for surgery, although with a wide confidence interval (OR = 16.8; 95% CI 2.95, 95.0; $P = .001$) (Table 3). CAS at baseline assessed as a continuous measure was associated with the need for surgery (B = 0.11; 95% CI .02, .20; $P = .02$) (Table 4). No other baseline variables including age, gender, ethnicity, thyroid state, recent use of steroids, smoking status, or baseline TES and CAS (split high vs low) predicted the need for surgery (Table 3) (Supplementary Figs. 1 and 2) (27).

Although CAS at baseline predicted the need for surgery, prompt reduction in CAS to 0 or 1 by 12 weeks was not associated with a reduction in the need for surgery overall (OR = 0.47; 95% CI .09, 2.55; $P = .38$) or lid surgery (OR = 0.90; 95% CI .16, 5.02; $P = .90$) or orbital decompression (OR = 0.54; 95% CI .06, 4.94; $P = .59$).

Using quartiles of OI we found no evidence of its association with needing surgery (OR = 1.12; 95% CI .71, 1.77; $P = .63$); we also saw no clear evidence with TES (OR = 1.35; 95% CI .87, 2.09; $P = .19$) (Table 4). Components of OI including presence of diplopia at baseline was also not associated with need for surgery (OR = 1.32; 95% CI .78, 2.44; $P = .26$), nor was proptosis (OR = 1.07; 95% CI .69, 1.65; $P = .76$) or palpebral aperture (OR = 1.34; 95% CI .85, 2.12; $P = .20$). EUGOGO score at baseline did not predict the need for surgery overall (Table 4).

Requirement for lid or decompression surgery

Of the 64 individuals who provided data, 16 needed lid surgery during the 3 years. The needs for lid and decompression

Table 2. Effect of azathioprine and radiotherapy on the binary composite clinical outcome, Ophthalmopathy Index, Clinical Activity Score, and Total Eye Score at 3 years follow-up

Reason	OR/B	95% CI	*P*	OR/B[a],[b]	95% CI[a]	*P*[a]
Binary clinical composite outcome improvement (N = 52)						
Azathioprine	0.83	(0.21, 3.29)	.70	0.51	(0.08, 3.13)	.47
Radiotherapy	2.82	(0.64, 12.4)	.17	2.17	(0.43, 10.9)	.35
Ophthalmopathy Index (N = 49)						
Azathioprine	1.36	(−0.35 3.06)	.12	1.33	(−0.68, 3.34)	.19
Radiotherapy	−0.53	(−2.26, 1.21))	.54	0.11	(−1.92, 2.14()	.92
Clinical Activity Score (N = 58)						
Azathioprine	0.24	(−0.48, 0.96)	.51	0.58	(−0.20, 1.36)	.14
Radiotherapy	0.16	(−0.57, 0.88)	.66	0.19	(−0.58, 0.98)	.61
Total Eye Score (N = 46)						
Azathioprine	2.97	(0.31, 5.62)	.03	3.11	(0.04, 6.19)	.05
Radiotherapy	−1.20	(−3.97, 1.57)	.39	−0.61	(−3.87, 2.65)	.71
EUGOGO classification[a] (N = 49)						
Azathioprine	1.67	(0.14, 19.8)	.69	2.48	(0.14, 44.7)	.54
Radiotherapy	0.42	(0.04, 4.94)	.49	0.58	(0.03, 10.97)	.72

P = *P* value against the null hypothesis of no association.
Adjusted for age group, ethnicity, smoking status, gender, thyroid state, disease duration, study center, recent steroid use, and baseline Clinical Activity Score.
Abbreviations: B, beta coefficient; EUGOGO, European Group on Graves' Orbitopathy; OR, odds ratio.
[a]Odds of having moderate to severe vs mild on EUGOGO classification.

surgery were also separately assessed. CAS at baseline as a continuous measure was associated with the need for lid surgery ($P = .02$), as was duration of disease prior to enrollment (OR = 7.18; 95% CI 1.24, 41.4; $P = .03$) (Supplementary Table 2) (27). OI, TES, and EUGOGO score at baseline were not predictive of the need for lid surgery (Table 4). No other baseline factors were predictive of the need for lid surgery including diplopia at baseline (OR = 0.94; 95% CI .51, 1.73; $P = .84$).

Twelve of 64 individuals underwent decompression surgery within 3 years. Similar to lid surgery, disease duration prior to enrollment of greater than 6 months was predictive of the need for orbital decompression (OR = 52.5; 95% CI 4.10, 672; $P = .002$), but there was no significant association with baseline CAS score ($P = .39$) (Table 4). Baseline diplopia was also associated with need for orbital decompression (OR = 6.00; 95% CI 1.78, 20.3; $P = .004$), as was palpebral aperture (OR = 3.18; 95% CI 1.54, 6.57; $P = .002$) and OI (OR = 2.70; 95% CI 1.33, 5.47; $P = .006$), with a possible association with TES, with 5% requiring surgery in the lowest quartile but 37% requiring surgery in the highest quartile ($P = .05$), (Table 4), but not degree of proptosis (OR = 0.06; 95% CI .03, .15; $P = .16$) (Supplementary Table 4) (27). No individuals with mild GO on baseline EUGOGO score required orbital decompression over 3 years. No other baseline variables predicted the need for decompression surgery including age, gender, ethnicity, thyroid state, recent use of steroids, smoking status, or baseline TES (Supplementary Table 4) (27).

Baseline GOQOL-VF was lower in those who required orbital decompression over the 3-year follow-up [median 58.3 (IQR 28.6-78.6) vs those who did not, median 75 (IQR 63.4-100); $P = .02$]. However no clear association was seen between GOQOL-VF and the need for any surgery or lid surgery. Baseline GOQOL-AP was not predictive of individuals undergoing any form of surgery (Supplementary Table 5) (27).

Factors Predicting Improvement in Quality of Life (GOQOL)

There was improvement overall after 3 years of follow-up in patients' perception of their visual function (GOQOL-VF; $P < .001$) and visual appearance (GOQOL-AP; $P = .002$) (Fig. 2D and 2E); however, individual trajectories were variable (Fig. 3A). Some individuals showed minimal improvement or even deterioration in GOQOL-AP (Fig. 3A), especially among those who did not undergo surgery (Fig. 3B). Overall, 35% of individuals failed to improve their GOQOL-VF and 42.4% failed to improve GOQOL-AP at 3 years (Fig. 3A). These percentages were even higher in those who did not have surgery, with 42.1% failing to improve GOQOL-VF and 50% failing to improve GOQOL-AP (Figs. 3B and 3C).

After 3 years, 15 of the 60 people (25%) who provided GOQOL-VF scores and 32 of the 60 people (54.2%) who provided GOQOL-AP at 3 years had a score of 75 or less, indicating ongoing substantial impairment of quality of life. No baseline characteristics, potential confounders, or disease levels at baseline (CAS/TES) were associated with having increased odds of having a GOQOL score less than 75 at 3 years (Supplementary Tables 5 and 6) (27).

Discussion

There is limited outcome data for GO following medical interventions beyond 2 years. This 3-year report on prospectively acquired subjective and objective assessments in just over half of the participants in the CIRTED trial, who were broadly representative of the overall study population, therefore adds substantially to the current evidence base for long-term treatment outcomes in this important cause of visual disability, disfigurement, and impaired quality of life. In summary, we observed no clear benefit beyond the effect of high-dose oral steroids in individuals randomized to receive azathioprine or radiotherapy on BCCOM, OI, CAS, and EUGOGO score at

Table 3. Odds of requiring any surgery by 3 years

Reason	OR	95% CI	*P*	OR[a],[b]	95% CI[a]	*P*[a]
Age						
<60 years	1			1		
≥60 years	0.70	(0.26, 3.11)	.88	0.64	(0.14, 2.93)	.56
Gender						
Male	1			1		
Female	1.0	(0.31, 3.22)	.99	1.92	(0.31, 11.7)	.48
Ethnicity						
Caucasian	1			1		
Non-Caucasian	0.27	(0.05, 1.37)	.12	0.08	(0.01, 1.05)	.05
Thyroid state						
Hypothyroid/euthyroid				1		
Hyperthyroid	10.81	(0.22, 2.90)	.74	0.44	(0.08, 2.40)	.34
Recent use of steroids						
No recent use	1			1		
Use in last 6 months	0.82	(0.14, 4.84)	.83	0.45	(0.03, 6.61)	.56
Disease duration						
<6 months	1			1		
≥6 months	4.71	(1.50, 14.8)	.008	16.8	(2.95, 95.0)	.001
Smoking status						
Nonsmoker	1			1		
Smoker	0.8	(0.31, 2.43)	.80	0.76	(0.16, 3.57)	.73
Study center						
Other	1			1		
Bristol/Moorfields	0.71	(0.17, 2.97)	.64	0.74	(0.09, 6.02)	.11
Baseline CAS						
2-3	1			1		
4-5	5.72	(0.65, 50.3)	.12	6.48	(0.41, 103)	.19
6-7	11.6	(1.17, 114)	.04	13.6	(0.78, 237)	.07
Baseline TES						
<22	1			1		
≥22	0.81	(0.19, 3.59)	.78	1.40	(0.18, 11.0)	.75
Received Aza	0.90	(0.33, 2.49)	.85	0.39	(0.09, 1.69)	.21
Received RT	0.52	(0.19, 1.47)	.22	0.34	(0.08, 1.52)	.16

Sixty-four individuals provided data—24 had surgery; 40 did not.
Sixteen individuals had lid surgery; 12 individuals required orbital decompression; 2 required squint surgery. Five individuals required both lid surgery and decompression surgery.
P = *P* value against the null hypothesis of no association.
Abbreviations: Aza, azathioprine; CAS, Clinical Activity Score; OR, odds ratio; RT, radiotherapy; TES, Total Eye Score.
[a]Adjusted for age group, ethnicity, smoking status, gender, thyroid state, disease duration, study center, recent steroid use, baseline CAS, baseline TES.
[b]Two individuals with incomplete baseline CAS excluded from analysis.

3 years. Furthermore, for azathioprine we also observed no additional benefit with regard to need for subsequent surgical intervention or quality of life measures. This is in contrast to our original findings at 1 year, which showed a modest potential benefit from azathioprine (10). The results for radiotherapy are less clear. In our original study we found no evidence of benefit with radiotherapy (10), but in this longer-term follow-up we did observe that those randomized to receive radiotherapy had a higher GOQOL-AP score [median 81.3 (IQR 62.5-93.8) vs median 68.9 (IQR 31.2-87.5); *P* = .03] than those who did not. We also observed a potential reduction in the need for lid surgery (OR = 0.17; 95% CI .03, 1.03; *P* = .05), although it should be highlighted that the 95% CI does cross equality. Analysis of key individual components

of outcome measures did suggest radiotherapy reduced diplopia (Supplementary Table 2) (27). Taken together this does raise the possibility of late benefits from radiotherapy; however, caution is needed over interpretation of these marginal benefits given our incomplete patient follow-up and the number of statistical tests being performed, which raises the possibility of a type 1 error.

The most striking finding from this 3-year follow-up study was that quality of life outcomes in GO remain poor, as 35% of individuals failed to improve their GOQOL-VF and 42.4% failed to improve their GOQOL-AP at 3 years (Fig. 3). This figure was even higher in those patients who did not undergo surgery. Overall, after 3 years, 25% of patients had a GOQOL-VF of 75 or lower and 54.2% had a GOQOL-AP of 75 or lower,
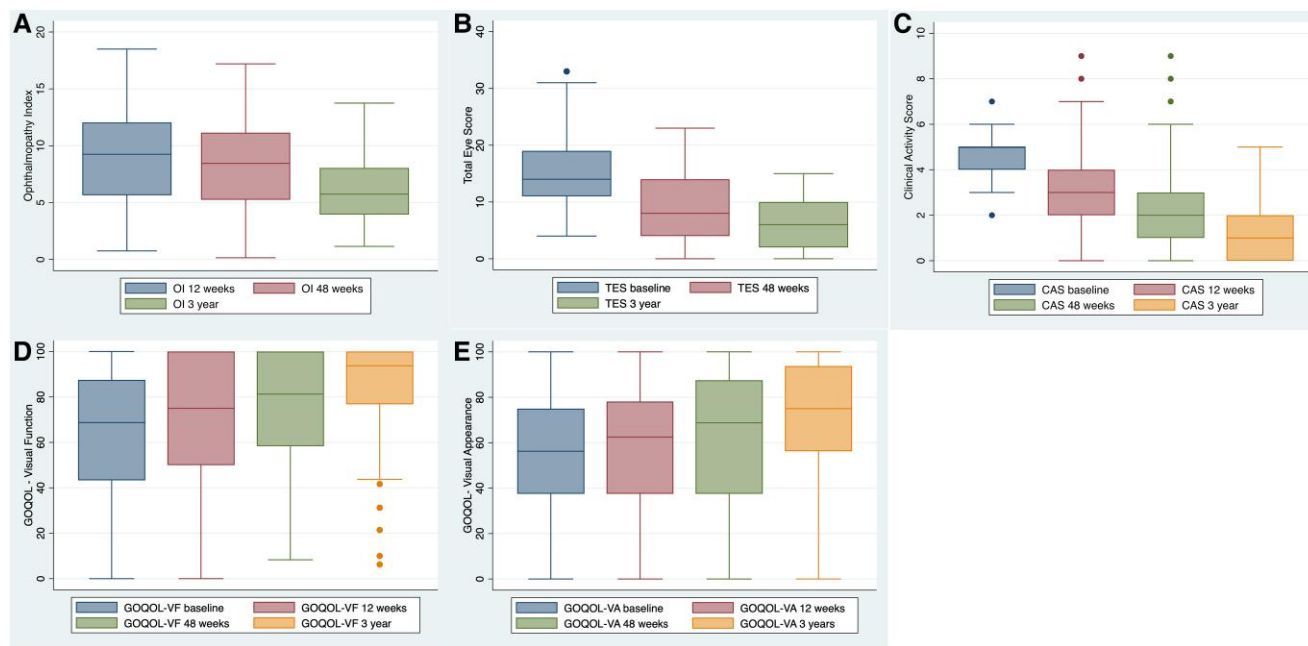
Figure 2. (A) Ophthalmopathy Index at 12 weeks, 48 weeks, and 3 years. (B) TES at baseline, 12 weeks, 48 weeks, and 3 years. (C) CAS at baseline, 12 weeks, 48 weeks, and 3 years. (D) GOQOL Visual Function at baseline, 12 weeks, 48 weeks, and 3 years. (E) GOQOL Visual Appearance at baseline, 12 weeks, 48 weeks, and 3 years. Abbreviations: CAS, Clinical Activity Score; GOQOL, Graves Ophthalmopathy Quality of Life; TES, Total Eye Score.

Table 4. Need for surgery over 3-year follow-up by baseline CAS, OI, and TES

	Any Surgery				Lid				Decompression			
	No	Yes	%	P	No	Yes	%	P	No	Yes	%	P
CAS				.02				.02				.39
2	5	0	0		5	0	0		5	0	0	
3	4	1	20		4	1	20		4	1	20	
4	8	7	46.6		11	4	26.7		12	3	20	
5	14	7	33.3		18	3	14.3		16	5	23.8	
6	6	4	45.5		7	4	36.4		10	1	9.1	
7	1	4	80		1	4	80		3	2	40	
OI				.63				.66				.006
Q1	8	5	38.4		18	5	38.5		13	0	0	
Q2	12	5	29.4		14	3	17.7		15	2	11.7	
Q3	7	5	41.6		9	3	25.0		8	4	33.3	
Q4	11	8	42.1		14	5	26.3		14	5	26.3	
TES				0.19				0.46				0.05
Q1	14	6	30		15	5	25		19	1	5	
Q2	11	5	31.3		14	2	12.5		12	4	25	
Q3	7	5	41.7		8	4	33.3		11	1	8.3	
Q4	8	8	50		11	5	31.2		10	6	37.5	
EUGOGO classification												
Mild	9	6	40		3	3	50		15	33	68.8	
Moderate to severe	27	18	40	.99	10	4	28.6	0.36	0	12	100	

Abbreviations: CAS, Clinical Activity Score; EUGOGO, European Group on Graves' Orbitopathy; OI, Ophthalmopathy Index; Q1, quartile 1; Q2, quartile 2; Q3, quartile 3; Q4, quartile 4; TES, Total Eye Score.

indicating ongoing prolonged substantial impairment in the quality of life of these patients despite aggressive medical management. This appears to contrast with the substantial improvement in physician-conducted GO assessments, in particular measures of inflammation (CAS) and overall orbital effects/deformity (eg, OI and TES) (Fig. 2). Taken together, this
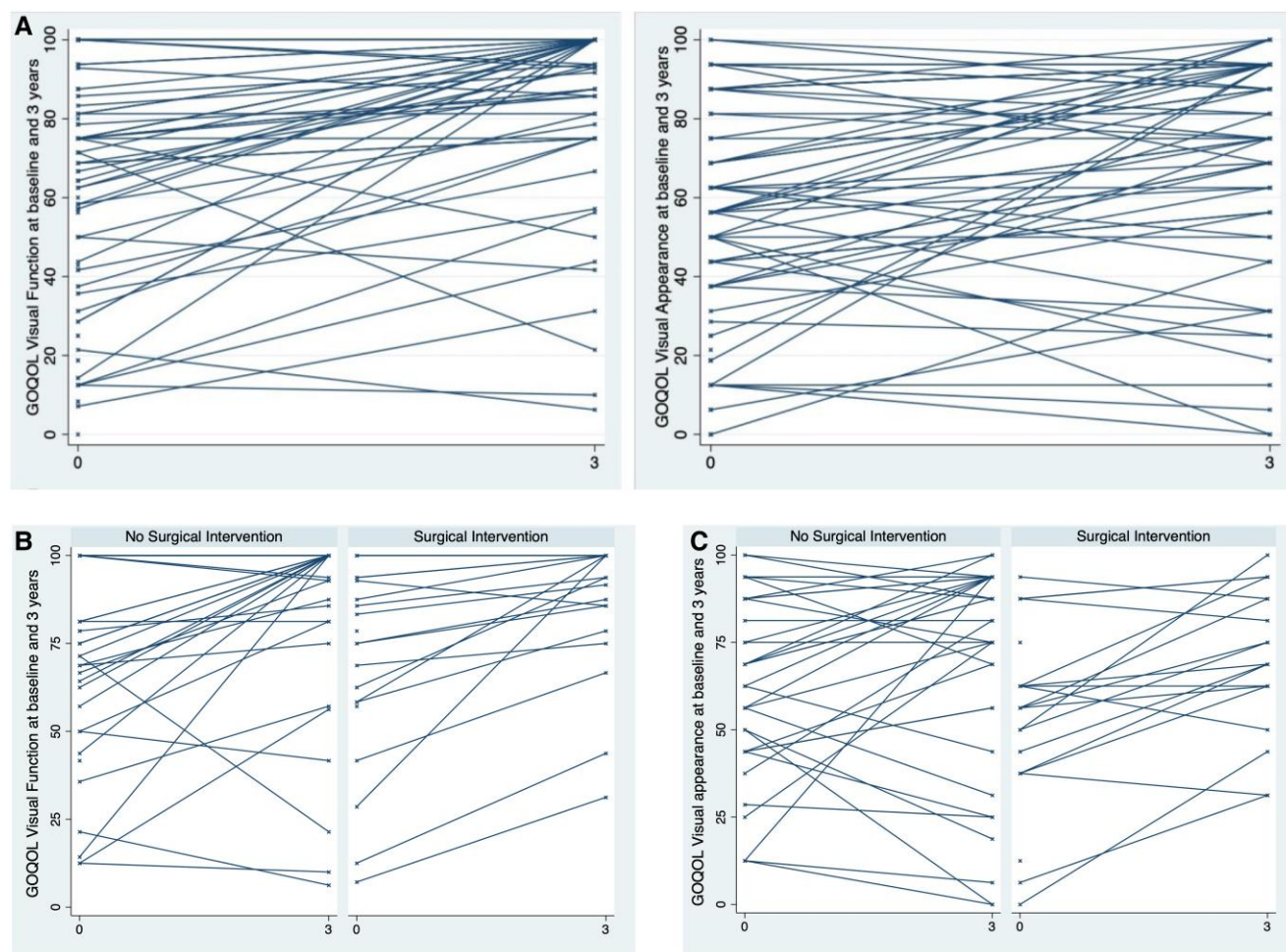
Figure 3. (A) GOQOL Visual Function and Appearance at baseline and 3 years—overall. (B) GOQOL Visual Function at baseline and 3 years by surgical intervention. (C) GOQOL Visual Appearance at baseline and 3 years by surgical intervention. Abbreviations: GOQOL, Graves Ophthalmopathy Quality of Life.

indicates there is a substantial unmet need for improving outcomes in this visually disabling and cosmetically disfiguring condition. In keeping with this, we confirm that although CAS is a valuable baseline measure (3), change in CAS did not correlate with key surgical and quality of life outcomes.

In addition, we have been able to quantify the need for surgery beyond the inflammatory phase of GO. Despite extensive medical therapy, including high-dose oral (but not intravenous) steroids in all individuals, 37.5% of patients still required surgical intervention, predominantly lid surgery or decompression. This is comparable but somewhat higher than the surgery rates in the study of Leo et al, although it is noted that some patients in their study declined surgery (14). The high requirement for surgery also indirectly highlights the poor efficacy of oral steroids and/or azathioprine and radiotherapy in treating thyroid eye disease. It is possible that more of our patients would also benefit from surgical intervention, especially since a recent meta-analysis has highlighted the patient-reported benefits of surgery (15).

Whereas baseline CAS was predictive of the need for surgery overall (particularly lid surgery but not orbital decompression), it was not associated with quality of life at 3 years. Of note, baseline CAS was not associated with CAS later on in the study, and early suppression of inflammation as shown by a reduction in

CAS to 0 or 1 in the first 12 weeks was not associated with improved outcomes with regard to needing surgery or improving quality of life, raising the possibility that early control of inflammation—a key part of current treatment approaches for GO—may not have the impact on long-term orbital remodeling that have been predicted (7). This contrasts with the improvements in proptosis as well as quality of life seen with disease-modifying therapies such as teprotumumab, although this also improves CAS substantially (28, 29) and emphazies the need for longer-term follow-up and assessments of quality of life and responses other than CAS, to ensure benefits translate into improved outcomes for patients.

An additional finding from our study was the consistent observation that the need for surgery over 3 years was higher in those who had thyroid eye disease for more than 6 months before enrolment in CIRTED (OR = 16.8; 95% CI 2.95, 95.0; $P = .001$), in particular orbital decompression (OR = 52.5; 95% CI 4.10, 672; $P = .002$). This emphazies the importance of surveillance and early case finding in GO, as in the UK TEAMeD 5 approach (http://www.btf-thyroid.org/TEAMeD-5) (30), which supports patients and endocrinologists to detect GO early. More than 80% of cases of GO arise at the same time or after the diagnosis of thyrotoxicosis, a period in which patients are normally under endocrine follow-up (31).

Our study also highlights that OI and TES are associated with need for decompression surgery over 3 years (Table 4), and OI at baseline is still associated with OI 3 years later. This perhaps reflects the robustness of these assessments but also provides evidence of incomplete treatment efficacy in our study. Whereas baseline CAS, OI, and TES are associated with surgical outcomes, they do not appear to be able to clearly identify those patients who do not need surgery; they are better at identifying those at highest risk of surgery and identifying individuals who need the closest attention (Table 4). Although baseline OI and TES are predictive of identifying those who need surgery, they are not associated with quality of life at 3 years. Perhaps unsurprisingly, quality of life at baseline is predictive of quality of life at 3 years: baseline GOQOL-VF was lower in those who required orbital decompression over the 3 years [follow-up median score 58.3 (IQR 28.6-78.6)] vs those who did not [median score 75 (IQR 63.4-100) $P = .02$]. A positive change in GOQOL-VF was observed in those who received orbital decompression [median 28.1 (IQR 18.3-36.5) vs 7.13 (IQR 0—25); $P = .05$] (Supplementary Table 6) (27). This suggests that orbital decompression in particular improves GOQOL-VF in people with thyroid eye disease. Taken together, to improve GOQOL outcomes in GO we need to take more account of the baseline GOQOL-VF and GOQOL-AP scores and use these in conjunction to perhaps better define who might benefit from surgery or consider novel agents such as teprotumumab.

Although our study is one of the largest in thyroid eye disease with longer follow-up than other studies, there are several limitations. The sample size is still modest, which means we are potentially underpowered to detect important effects such as identifying clear benefits from surgery in terms of quality of life scores. The number of statistical tests performed also raises the possibility of a type 1 error. As the 3-year assessment was optional, almost half the study participants were also not evaluated at this time point, although a majority did provide data and are representative of the original cohort.

In conclusion, we have identified no clear evidence of benefit for the addition of azathioprine to high-dose oral steroid in individuals with moderate to severe GO at 3 years. The results for orbital radiotherapy were less clear, but there may have been possible benefit in the visual appearance domain for quality of life and a potential reduction in the need for lid surgery. Furthermore, we have identified that 3-year outcomes in GO despite the medical interventions we used are heterogenous and poor, particularly for GOQOL-AP. A large proportion of patients still need surgery, and in those who receive surgery there are potential benefits, particularly for GOQOL-VF. Despite considerable effectiveness of therapy in improving GO indices such as CAS, this is not reflected in quality of life scores, and future trials of treatment and treatment strategies should reflect this. Our results also strongly suggest that intervening in GO early will likely reduce the need for future surgical intervention, although rapid reduction in CAS was not associated with extra benefit. Longer-term studies, particularly for promising new therapies, are also needed, with an emphasis on quality of life measures and not just objective clinical assessments of GO.

Acknowledgments

Disclosures

Role of the Funding Source

Data Availability

Restrictions apply to the availability of some or all data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

References

1. Taylor PN, Albrecht D, Scholz A, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*. 2018;14(5):301-316. https://doi.org/10.1038/nrendo.2018.18
2. Bartalena L, Pinchera A, Marcocci C. Management of Graves' ophthalmopathy: reality and perspectives. *Endocr Rev*. 2000;21(2): 168-199. https://doi.org/10.1210/edrv.21.2.0393
3. Taylor PN, Zhang L, Lee RWJ, *et al*. New insights into the pathogenesis and nonsurgical management of Graves orbitopathy. *Nat Rev ndocrinol*. 2020;16(2):104-116. https://doi.org/10.1038/s41574-019-0305-4
4. Wiersinga WM. Advances in treatment of active, moderate-to-severe Graves' ophthalmopathy. *Lancet Diabetes Endocrinol*. 2017;5(2):134-142. https://doi.org/10.1016/S2213-8587(16)30046-8
5. Terwee C, Wakelkamp I, Tan S, Dekker F, Prummel MF, Wiersinga W. Long-term effects of Graves' ophthalmopathy on health-related quality of life. *Euro J Endocrinol*. 2002;146(6):751-757. https://doi.org/10.1530/eje.0.1460751
6. Dietrich A, Taylor P, White P, *et al*. Establishing the usefulness 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):1341-1355. https://doi.org/10.1080/13548506.2018.1503693
7. Salvi M, Campi I. Medical treatment of Graves' orbitopathy. *Horm Metab Res*. 2015;47(10):779-788. https://doi.org/10.1055/s-0035-1554721
8. Dosiou C, Kossler AL. Thyroid eye disease: navigating the new treatment landscape. *J Endocr Soc*. 2021;5(5):bvab034. https://doi.org/10.1210/jendso/bvab034
9. Bartalena L, Kahaly GJ, Baldeschi L, *et al*. The 2021 European Group on Graves' Orbitopathy (EUGOGO) clinical practice guidelines for the medical management of Graves' orbitopathy. *Euro J Endocrinol*. 2021;185(4):G43-g67. https://doi.org/10.1530/EJE-21-0479
10. Rajendram R, Taylor PN, Wilson VJ, *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. https://doi.org/10.1016/S2213-8587(18)30021-4

11. Smith TJ, Kahaly GJ, Ezra DG, *et al*. Teprotumumab for thyroid-associated ophthalmopathy. *N Engl J Med*. 2017;376-(18):1748-1761. https://doi.org/10.1056/NEJMoa1614949

12. Kahaly GJ, Riedl M, Konig J, *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):287-298. https://doi.org/10.1016/S2213-8587(18)30020-2

13. Perez-Moreiras JV, Gomez-Reino JJ, Maneiro JR, *et al*. Efficacy of tocilizumab in patients with moderate-to-severe corticosteroid-resistant Graves orbitopathy: a randomized clinical trial. *Am J Ophthalmol*. 2018;195:181-190. https://doi.org/10.1016/j.ajo.2018.07.038

14. Leo M, Marcocci C, Pinchera A, *et al*. Outcome of Graves' orbitopathy after total thyroid ablation and glucocorticoid treatment: follow-up of a randomized clinical trial. *J Clin Endocrinol Metab*. 2012;97(1):E44-E48. https://doi.org/10.1210/jc.2011-2077

15. Woo T, Li C, Ganesananthan S, *et al*. The effect of ophthalmic surgery for Graves' orbitopathy on quality of life: a systematic review and meta-analysis. *Thyroid*. 2022;32(2):177-187. https://doi.org/10.1089/thy.2021.0411

16. Mourits MP, Prummel MF, Wiersinga WM, Koornneef L. Clinical activity score as a guide in the management of patients with Graves' ophthalmopathy. *Clin Endocrinol (Oxf)*. 1997;47(1):9-14. https://doi.org/10.1046/j.1365-2265.1997.2331047.x

17. Dolman PJ, Rootman J. VISA classification for Graves orbitopathy. *Ophthalmic Plast Reconstr Surg*. 2006;22(5):319-324. https://doi.org/10.1097/01.iop.0000235499.34867.85

18. Bartalena L, Baldeschi L, Dickinson A, *et al*. Consensus statement of the European Group on Graves' Orbitopathy (EUGOGO) on management of GO. *Euro J Endocrinol*. 2008;158(3):273-285. https://doi.org/10.1530/EJE-07-0666

19. 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):333-346. https://doi.org/10.1089/thy.2007.0315

20. Perros P, Crombie AL, Matthews JN, Kendall-Taylor P. Age and gender influence the severity of thyroid-associated ophthalmopathy: a study of 101 patients attending a combined thyroid-eye clinic. *Clin Endocrinol (Oxf)*. 1993;38(4):367-372. https://doi.org/10.1111/j.1365-2265.1993.tb00516.x

21. Werner SC. Classification of the eye changes of Graves' disease. *Am J Ophthalmol*. 1969;68(4):646-648. https://doi.org/10.1210/jcem-29-7-982

22. Werner SC. Modification of the classification of the eye changes of Graves' disease: recommendations of the Ad Hoc Committee of the American Thyroid Association. *J Clin Endocrinol Metab*. 1977;44(1):203-204. https://doi.org/10.1210/jcem-44-1-203

23. Terwee CB, Dekker FW, Mourits MP, *et al*. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. *Clin Endocrinol (Oxf)*. 2001;54(3):391-398. https://doi.org/10.1046/j.1365-2265.2001.01241.x

24. 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):773-779. https://doi.org/10.1136/bjo.82.7.773

25. Rajendram R, Lee RWJ, Potts MJ, *et al*. Protocol for the combined immunosuppression & radiotherapy in thyroid eye disease (CIRTED) trial: A multi-centre, double-masked, factorial randomised controlled trial. *Trials*. 2008;9:6-6. https://doi.org/10.1186/1745-6215-9-6

26. Rundle FF, Wilson CW. Development and course of exophthalmos and ophthalmoplegia in Graves' disease with special reference to the effect of thyroidectomy. *Clin Sci*. 1945;5(3-4):177-194. https://doi.org/10.1136/bjo.82.7.773

27. Taylor PN. Data from: CIRTED 3 Year Data Supplementary Online Material. Deposited December 26, 2022. https://doi.org/10.5281/zenodo.7483891.

28. Kahaly GJ. Immunotherapies for thyroid eye disease. *Curr Opin Endocrinol Diabetes Obes*. 2019;26(5):250-255. https://doi.org/10.1097/MED.0000000000000493

29. Douglas RS, Kahaly GJ, Patel A, *et al*. Teprotumumab for the treatment of active thyroid eye disease. *N Engl J Med*. 2020;382(4):341-352. https://doi.org/10.1056/NEJMoa1910434

30. BTF. TEAMeD-5. 2018.

31. Marcocci C, Bartalena L, Bogazzi F, Panicucci M, Pinchera A. Studies on the occurrence of ophthalmopathy in Graves' disease. *Acta Endocrinol (Copenh)*. 1989;120(4):473-478. https://doi.org/10.1530/acta.0.1200473