Some studies on the natural history of Graves’ orbitopathy: increase in orbital fat is a rather late phenomenon

Peter W Potgieser¹, Wilmar M Wiersinga², Noortje I Regensburg¹ and Maarten P Mourits¹

Departments of ¹Ophthalmology, D2-431 and ²Endocrinology and Metabolism, Academic Medical Centre, Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

Correspondence should be addressed to P W Potgieser
Email p.w.potgieser@amc.uva.nl

Abstract

Aim: To describe volumes of orbital fat (FV) and extraocular muscles (MV) in Graves’ orbitopathy (GO) as a function of the duration of GO.

Patients: i) Cross-sectional survey among 95 consecutive patients with untreated GO who had been referred to the combined thyroid–eye clinics of our university hospital. ii) Longitudinal survey among 39 of the 95 patients who did not receive any therapeutic intervention and were followed for 1 year.

Methods: A computed tomography (CT)-based and well-validated method for calculating orbital soft tissues. In order to neutralize sex differences, results are expressed as ratios of FV:orbital volume (OV) and MV:OV.

Results: i) Patients with GO duration of >1 year had greater FV:OV (0.65 vs 0.55, p=0.004), similar MV:OV (0.22 vs 0.21, not significant (NS)), and more proptosis (22 mm vs 21 mm, p=0.03) as compared to those with shorter duration. ii) As compared to baseline, after 1 year, FV:OV had increased (0.56 vs 0.63, p=0.000), MV:OV had not changed (0.19 vs 0.19, NS), proptosis was higher (20 mm vs 21 mm, p=0.003), and clinical activity scores had become lower (2 vs 1, p=0.02) (median values).

Conclusion: CT images show that a longer duration of GO is associated with a higher orbital FV. Extraocular MV, however, is not associated with GO duration; rather, it is related to the severity of GO.

Introduction

Enlargement of orbital fat and extraocular muscles are the hallmark of Graves’ orbitopathy (GO), as is evident from surgical pathology and magnetic resonance imaging or CT imaging (1, 2, 3). Whereas a large amount of progress has been made in understanding the swelling of retrobulbar tissues in terms of molecular immunology, less attention has been paid to the differential involvement of orbital fat and extraocular muscles. It has been suggested that subsets of GO patients exist: exophthalmos and lid retraction are predominant in some patients, but motility impairment and diplopia predominate in others (4). The former are associated with preferentially a higher orbital fat volume (FV), and the latter are mainly associated with a higher extraocular muscle volume (MV).

Quantitative measurements of orbital volumes (OVs) to substantiate these claims are rare. We have previously reported on a series of 95 consecutive patients with untreated GO: relative to age-specific reference ranges, a higher MV (but normal FV) was observed in 61%, a higher MV and a higher FV in 9%, an isolated higher FV in 5%, and normal MV and normal FV in 25% (5). Patients with increased FV had greater lid aperture and more proptosis as compared to patients with normal FV. Patients with increased MV were older, had more proptosis, had more motility disturbances, and had higher serum thyrotropin (TSH) receptor antibodies as compared to patients with normal MV. No group differences were observed with respect to clinical activity.
score (CAS) or thyroid function. It was unexpected that the vast majority of these newly diagnosed patients with untreated GO had no enlargement of the orbital fat compartment. We hypothesized that a longer duration of the orbitopathy could be related to a higher orbital FV. Therefore, we evaluated OVs of untreated GO patients as a function of GO duration, both cross-sectionally (at referral) and longitudinally (after 1 year in patients who did not require medical or surgical intervention).

**Subjects and methods**

Consecutive adult GO patients referred to the combined thyroid-eye clinic of the Academic Medical Center at the University of Amsterdam were studied as described previously (5). The tenets of the Helsinki Declaration were followed. Although our ethical committee considered the present study not subject to consent, all participants were asked to sign informed consents at baseline and at the second visit. The only inclusion criterion was definite GO, as judged by attending consultants in endocrinology and ophthalmology based on thyroid function, thyroid antibodies, an assessment of GO severity (according to the NOSPECS classification), an assessment of GO activity (according to a CAS), and an orbital CT scan (6). The exclusion criteria included: overt hypo- or hyperthyroidism (defined as abnormal TSH in combination with abnormal free thyroxine); previous medical, radiological, or surgical treatment for GO apart from local lubricants; non-Caucasia (to minimize bias by racial differences in orbital anatomy); pregnancy; drug abuse; an incomplete set of CT images; and an uncorrected Gantry tilt on CT. Out of the 143 consecutive referrals that met the inclusion criterion, 48 had an exclusion criterion, which left 95 patients with untreated GO for use in our previous study (5) and in the present study. First, in a cross-sectional survey, we calculated orbital FV and MV from the CT scan performed at referral, and we related these volumes to the duration of GO as ascertained from the patient’s history. The duration of GO was defined as the interval between the time of referral and the first time the patient recalled that something was wrong with his or her eyes. In order to diminish recollection bias, we dichotomized GO duration into <1 and >1 year. Second, in a longitudinal survey, we repeated an orbital CT scan as a routine clinical course after 1 year in those patients from the original cohort who did not require medical or surgical intervention, had remained euthyroid by using a stable dose of thyroid supplementation, and had been followed. Changes over time in orbital FV or MV in this group could be considered to reflect the natural history of GO.

The applied methods of orbital volumetry have been described in detail previously (7). In short, axial CT images were burned on to a CD-ROM and imported in to a work station (DC7900 Small Form Factor, Hewlett-Packard (Palo Alto, CA, USA)) using Windows XP (Microsoft) and Mimics 13.1 (Materialise, Leuven, Belgium). Mimics Software converts the axial images into coronal and sagittal images. In Mimics, a soft tissue window (grayscale) setting was used to discriminate between orbital muscle, fat, and bone tissue. Orbital soft tissue CT numbers were set at −200 to +100 Hounsfield units (HU) for bony OV, −200 to −30 HU for fat tissue, and −30 to +100 HU for muscle tissue. As a result, the tissues of interest were highlighted on the display in specific colors. Region growing (the computer-assisted separation of different tissues) and the manual deletion of tissues were used.

The bony OV was measured by removing all of the selected tissues outside of the orbit. Because the orbital aperture does not have an existing border, we used the following landmarks: frontal bone, frontozygomatic suture, inferior orbital rim, and anterior lacrimal crest. All of the selected pixels in front of these landmarks were removed. The orbital septum was used for the anterior border of the orbital fat.

FV and MV were measured by manually removing all of the tissues that were not of interest. Orbital FV and extraocular MV were calculated in mm³ and were selected on the masks with region growing and manual segmentation and differentiation, as described previously (7). Intra-observer variation was <5% for FV and MV. The ratios FV:OV and MV:OV, our primary outcome measures, were used to eliminate anatomical differences between sexes (8). Age-specific reference ranges were obtained (8). The FV:OV and MV:OV ratios are not normally distributed, and they are therefore presented as medians and interquartile ranges. The secondary outcome measures were disease activity, which was measured with a CAS, and disease severity, which was measured with the NOSPECS classification.

Statistical analysis was performed using SPSS for Mac version 21.0 (SPSS, Inc.). The Mann–Whitney U test was used for evaluating the differences between two unrelated samples, and the Wilcoxon’s test was used for differences in time within one sample. The Spearman’s test was used for correlations. A P value of <0.05 was considered to be statistically significant.
Results

Cross-sectional survey

At referral of the 95 patients, the duration of GO was <1 year in 50 patients and >1 year in 39 patients. In six of the patients, the duration of complaints and smoking habits were unknown. Of the in the group with GO <1 year included in Table 1, 25 patients never smoked, 11 patients were current smokers, and 14 did smoke but had stopped. In the group of patients with a GO duration longer than 1 year, 15 patients did not smoke, 16 were current smokers, and eight were ex-smokers. There was no difference in smoking habits between the two groups (P=0.693). Patients with a GO duration of longer than 1 year had significantly higher FV (17.4 ml vs 14.9 ml, P=0.061) and FV:OV (0.65 vs 0.55, P=0.004) but similar MV (5.9 ml vs 5.2 ml, P=0.197) and MV:OV (0.22 vs 0.21, P=0.194) as compared to patients with shorter duration (Table 1). The higher FV was accompanied by a greater degree of exophthalmos, but otherwise, the severity and activity of GO were not different between the two groups.

Longitudinal survey

In 39 patients (37 women and two men) out of the 95 patients in the cross-sectional survey, therapeutic intervention for GO was not deemed necessary in view of the rather mild nature of GO or they were put on the waiting list for rehabilitative surgery. They did not receive any further treatment apart from local measures, such as artificial tears, lubricants, or sunglasses; rather, they were just observed. After a follow-up of 1 year, ten of the patients were still smoking, two had recently stopped smoking, and 27 had never smoked. Euthyroidism was maintained. A new CT scan was made, and OVs were calculated. During the 1-year period, the FV (16.2 ml vs 14.7 ml, P=0.000) and FV:OV (0.56 vs 0.63, P=0.000) but not the MV (4.9 ml vs 4.9 ml, P=0.919) or MV:OV (0.19 vs 0.19) increased significantly (Table 2). Individual analysis of the FV:OV showed an increase in 80% of the patients, whereas 10% remained stable and 10% had a decrease. The MV:OV was increased in 32% of patients, whereas 17% remained stable and 51% had a mild decrease (data not shown). The increase in fat mass was correlated with a small but significant increase in exophthalmos (20 mm vs 21 mm, P=0.000) and also a small but significant decrease in CAS (2 vs 1, P=0.020). The lowest measured Hertel at diagnosis was 14 mm, and the highest was 28 mm; the CAS ranged from 0 to 5. At the 1-year follow-up, the lowest and highest measured Hertel were 13 and 30 mm, and the CAS ranged from 0 to 3.

Discussion

We conclude that a longer duration of GO is associated with a larger orbital fat mass. There are several underpinnings of this conclusion.

First, an increase in FV over time was observed both in the cross-sectional survey (after dichotomizing patients into those with a GO duration of <1 year and those with a GO duration of >1 year).

Table 1 Characteristics of 89 patients with untreated Graves’ orbitopathy (GO) according to GO duration of <1 and >1 year.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>&lt;1 year</th>
<th>&gt;1 year</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>52 (45–60)</td>
<td>50 (36–55)</td>
<td>0.087</td>
</tr>
<tr>
<td>Fat volume (ml)</td>
<td>14.8 (12.9–17.5)</td>
<td>17.4 (13.8–19.3)</td>
<td>0.061</td>
</tr>
<tr>
<td>Muscle volume (ml)</td>
<td>5.2 (4.3–6.8)</td>
<td>5.9 (4.8–7.7)</td>
<td>0.197</td>
</tr>
<tr>
<td>Orbital volume (ml)</td>
<td>25.7 (24.2–27.7)</td>
<td>25.6 (23.8–28.1)</td>
<td>0.676</td>
</tr>
<tr>
<td>FV:OV</td>
<td>0.55 (0.50–0.65)</td>
<td>0.65 (0.57–0.70)</td>
<td>0.004</td>
</tr>
<tr>
<td>MV:OV</td>
<td>0.21 (0.17–0.25)</td>
<td>0.22 (0.19–0.27)</td>
<td>0.194</td>
</tr>
<tr>
<td>Clinical activity score</td>
<td>2 (1–3)</td>
<td>2 (1–3)</td>
<td>0.185</td>
</tr>
<tr>
<td>Lid aperture (mm)</td>
<td>12.00 (11–13)</td>
<td>12 (11–14)</td>
<td>0.522</td>
</tr>
<tr>
<td>Hertel (mm)</td>
<td>21 (19–23)</td>
<td>22 (21–24)</td>
<td>0.030</td>
</tr>
<tr>
<td>Elevation (degrees)</td>
<td>42 (39–47)</td>
<td>40 (30–46)</td>
<td>0.903</td>
</tr>
<tr>
<td>Diplopia grade 1</td>
<td>0.00 (0–2)</td>
<td>1 (0–1)</td>
<td>0.721</td>
</tr>
<tr>
<td>TSH (mU/l)</td>
<td>2.20 (0.02–4.65)</td>
<td>0.61 (0.09–2.10)</td>
<td>0.258</td>
</tr>
<tr>
<td>FT4 (pmol/l)</td>
<td>15.2 (11.8–17.1)</td>
<td>16.0 (12.7–18.7)</td>
<td>0.183</td>
</tr>
<tr>
<td>TPO-Ab (kU/l)</td>
<td>160 (40–860)</td>
<td>100 (35–680)</td>
<td>0.417</td>
</tr>
<tr>
<td>TBII (U/l)</td>
<td>7.60 (3.00–15.0)</td>
<td>5.20 (2.45–20.35)</td>
<td>0.762</td>
</tr>
</tbody>
</table>

Diplopia, grade 1, intermittent (only when tired or upon awakening); grade 2, inconstant (only at extremes of gaze); grade 3, constant (in primary gaze or reading position); TSH, thyrotropin; FT4, free thyroxine; TPO-Ab, thyroid peroxidase antibodies; TBII, TSH binding inhibitory immunoglobulins.

*Median values with interquartile range in brackets.
a duration of >1 year) and in the longitudinal survey with a fixed follow-up of 1 year.

Second, the error in recollecting the date of onset of eye changes was rather small because the choice was purposefully limited to either ‘<1 year ago’ or ‘more than 1 year ago’ in the cross-sectional survey.

Third, the difference in FV:OV between patients with a GO duration of <1 year and those with a duration of more than 1 year was 0.10, and in the longitudinal study, the change after 1 year was 0.07. These figures denote substantial changes that far exceed any changes that can be attributed to intra- or inter-observer variation in the applied volumetric method.

Fourth, the greater orbital fat mass with a longer GO duration in both surveys was associated with a higher degree of proptosis, as was evident from Hertel exophthalmometry (P=0.000). The imaging results thus had a direct correlation in a clinical parameter.

Fifth, in contrast to orbital fat, extraocular MV was not dependent on the GO duration in either survey, which lends biological plausibility to the specificity of the observed relationship between fat mass and GO duration.

Sixth, smoking behavior might have been a confounder, seeing as we previously demonstrated that current smokers had significantly higher MV than did ex-smokers and never smokers but also that smoking did not influence FV (9). We believe it highly unlikely that smoking was a confounder in the present series: in the longitudinal survey, the increase in FV:OV after 1 year was observed both in the 12 current smokers and in the 27 ex-smokers and never smokers, and the change was not different between both groups (data not shown).

The prevalence of enlarged extraocular muscles was 70% in our series of consecutive patients with untreated GO (5). Increased MV was related to older age, more proptosis, more diplopia, more impaired ductions, higher, TSH binding inhibitory immunoglobulin levels, and currently smoking (5, 10). Older age, high serum concentrations of TSH receptor antibodies, and current smoking have all been identified in the past as being factors that are associated with more severe GO. The prevalence of enlarged orbital fat mass in this series was only 14% and was only related to proptosis values. The data suggest that muscle enlargement is directly related to the severity of GO (5) and that the enlargement of orbital fat is a later phenomenon. These findings may have relevance for our understanding of the pathogenesis of GO, but more studies on active moderate-to-severe GO patients are still needed in order for the present results to be extrapolated.

The present longitudinal survey provides proof, for the first time, that peak activity precedes peak severity in the natural history of GO. Studies on the natural history of GO have been rare. The most informative ones were carried out by F F Rundle in the 1940s and 1950s at a time when there were no antithyroid drugs, no radioactive iodine, and no steroids. Rundle described a dynamic phase that involved first the aggravation of eye signs and then incomplete recovery, followed by a static phase in which any remaining eye changes remained stable. Sequential measurements of exophthalmos, the degree of elevation, and upper scleral show allowed the construction of curves that depict the natural history of each of these items, and these were recently reproduced in a paper by Bartley (10). The course of these separate eye changes can be summarized in a single curve that reflects the natural history of disease severity; nowadays, it is known as Rundle’s curve.

Developments in the 1990s saw the introduction of the concept of disease activity. It was realized that about one-third of patients with moderate-to-severe GO did not respond to steroids, and it was hypothesized that a response to steroids was more likely when the disease was in its early, ‘active’ stage with a lot of inflammatory edema rather than in its late, ‘inactive’ stage with more fibrosis. A CAS based on the classical signs of inflammation was introduced by our group in 1989 in order to predict whether the orbitopathy would respond to steroids (11). Assessment of GO activity by the CAS was recommended internationally in 1992 in view of its relevance to the choice of therapy (12). Remarkably, the CAS has stood the test of time, and it is nowadays widely used not only to assess disease activity but also as a primary or secondary outcome parameter of treatment (13). The CAS led us to realize that the natural history of GO could be described not only in terms of disease

Table 2  Characteristics of 39 patients with untreated GO at referral and after 1 year follow-up without any therapeutic interventiona.

<table>
<thead>
<tr>
<th></th>
<th>At referral</th>
<th>After 1 year follow-up</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>49.60 (43.80–59.40)</td>
<td>50.75 (45.05–60.27)</td>
<td>0.000</td>
</tr>
<tr>
<td>Fat volume (ml)</td>
<td>14.7 (12.5–17.5)</td>
<td>16.2 (14.3–18.8)</td>
<td>0.919</td>
</tr>
<tr>
<td>Muscle volume (ml)</td>
<td>4.9 (4.2–6.2)</td>
<td>4.9 (4.0–5.9)</td>
<td>0.000</td>
</tr>
<tr>
<td>FV:OV</td>
<td>0.56 (0.50–0.65)</td>
<td>0.63 (0.56–0.68)</td>
<td>0.000</td>
</tr>
<tr>
<td>MV:OV</td>
<td>0.19 (0.16–0.22)</td>
<td>0.19 (0.16–0.22)</td>
<td>0.220</td>
</tr>
<tr>
<td>Clinical activity score</td>
<td>2 (1.00–2.00)</td>
<td>1 (0.00–2.00)</td>
<td>0.020</td>
</tr>
<tr>
<td>Lid aperture</td>
<td>12 (11.00–13.00)</td>
<td>12 (11.00–14.00)</td>
<td>0.150</td>
</tr>
<tr>
<td>Hertel (mm)</td>
<td>20 (18–22)</td>
<td>21 (18–23)</td>
<td>0.003</td>
</tr>
<tr>
<td>Diplopia</td>
<td>0 (0.00–1.00)</td>
<td>0 (0.00–1.00)</td>
<td>0.354</td>
</tr>
</tbody>
</table>

*Median values with interquartile range in brackets.
severity but also in terms of disease activity. After some early attempts to incorporate the concept of activity into Rundle’s curve of severity (14, 15), we proposed the two curves that reflect the natural history of disease activity and disease severity and first published them in 1995 (16). We hypothesized at that time that the activity curve should be located to the left of the severity curve, because it is reasonable to assume that peak activity precedes peak severity in GO. The obtained prospective data from the present longitudinal survey support this hypothesis: disease activity decreases at the same time that disease severity increases. The few other studies that have been conducted on the natural history of GO have been unable to make this observation, either because CAS was not measured (17) or because the intervals between measurements were not optimal (18). Indeed, it can be expected that with decreasing CAS, GO severity will eventually also decrease.

It would be interesting to follow GO patients over longer periods of time to see what happens to FV and MV. The increase in FV will undoubtedly stop, but will there be a spontaneous reduction of FV (and MV) with extended follow-up? Measurements of FV and MV in patients who are treated with steroids or other modalities could also provide better insight into the nature of patient response to those treatments.

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

Funding
This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References