# RESEARCH

**BMC Ophthalmology** 

# **Open Access**



# Magnetic resonance imaging parameters on lacrimal gland in thyroid eye disease: a systematic review and meta-analysis

Nicole Tsz Yan Wong<sup>1</sup>, Ka Fai Kevin Yuen<sup>1</sup>, Fatema Mohamed Ali Abdulla Aljufairi<sup>1,2,3</sup>, Kenneth Ka Hei Lai<sup>1,4</sup>, Zhichao Hu<sup>1</sup>, Karen Kar Wun Chan<sup>1,2</sup>, Clement Chee Yung Tham<sup>1,2,5</sup>, Chi Pui Pang<sup>1</sup> and Kelvin Kam Lung Chong<sup>1,2,5\*</sup>

## Abstract

**Background** Thyroid eye disease is an extrathyroidal manifestation of Graves' disease and is associated with dry eye disease. This is the first systematic review and meta-analysis to evaluate the role of magnetic resonance imaging lacrimal gland parameters in thyroid eye disease diagnosis, activity grading, and therapeutic responses prediction.

**Methods** Up to 23 August, 2022, 504 studies from PubMed and Cochrane Library were analyzed. After removing duplicates and imposing selection criteria, nine eligible studies were included. Risk of bias assessment was done. Meta-analyses were performed using random-effect model if heterogeneity was significant. Otherwise, fixed-effect model was used. Main outcome measures include seven structural magnetic resonance imaging parameters (lacrimal gland herniation, maximum axial area, maximum coronal area, maximum axial length, maximum coronal length, maximum axial width, maximum coronal width), and three functional magnetic resonance imaging parameters (diffusion tensor imaging-fractional anisotropy, diffusion tensor imaging-apparent diffusion coefficient or mean diffusivity, diffusion-weighted imaging-apparent diffusion coefficient).

**Results** Thyroid eye disease showed larger maximum axial area, maximum coronal area, maximum axial length, maximum axial width, maximum coronal width, diffusion tensor imaging-apparent diffusion coefficient/ mean diffusivity, and lower diffusion tensor imaging-fractional anisotropy than controls. Active thyroid eye disease showed larger lacrimal gland herniation, maximum coronal area, diffusion-weighted imaging-apparent diffusion coefficient than inactive. Lacrimal gland dimensional (maximum axial area, maximum coronal area, maximum axial length, maximum axial width, maximum coronal width) and functional parameters (diffusion tensor imaging-apparent diffusion coefficient, diffusion tensor imaging-apparent diffusion coefficient) could be used for diagnosing thyroid eye disease; lacrimal gland herniation, maximum coronal area, and diffusion-weighted imaging-apparent diffusion coefficient for differentiating active from inactive thyroid eye disease; diffusion tensor imaging parameters (diffusion tensor imaging-fractional anisotropy, diffusion tensor imaging-mean diffusivity) and lacrimal gland herniation for helping grading and therapeutic responses prediction respectively.

**Conclusions** Magnetic resonance imaging lacrimal gland parameters can detect active thyroid eye disease and differentiate thyroid eye disease from controls. Maximum coronal area is the most effective indicator for thyroid eye

\*Correspondence: Kelvin Kam Lung Chong chongkamlung@cuhk.edu.hk Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/ficenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

disease diagnosis and activity grading. There are inconclusive results showing whether structural or functional lacrimal gland parameters have diagnostic superiority. Future studies are warranted to determine the use of magnetic resonance imaging lacrimal gland parameters in thyroid eye disease.

**Keywords** Thyroid eye disease, Thyroid-associated ophthalmopathy, Graves' ophthalmopathy, Graves' orbitopathy, lacrimal gland, Magnetic resonance imaging, Systematic review, Meta-analysis

## Introduction

Thyroid eye disease (TED), also known as thyroid-associated ophthalmopathy (TAO), Graves' ophthalmopathy (GO), or Graves' orbitopathy, is an autoimmune disorder involving the orbital soft tissues, namely the extraocular muscles (EOMs) and orbital fat (OF) [1]. It is an extrathyroidal manifestation of Graves' disease (GD) [2]. In GD, thyrotropin receptor (TSHr) autoantibodies (TRAbs) and insulin-like growth factor-1 (IGF-1) receptor (IGF-1r) autoantibodies attack their respective receptors on orbital fibroblasts and EOMs, stimulating adipogenesis and inflammation. The resultant increased volume of retrobulbar soft tissues within the limited space contributes to various thyroid eye signs [3].

TED is a biphasic disease that begins with an active phase with progressive inflammation, followed by an inactive phase with stable fibrosis of orbital soft tissues [4]. After the current clinical assessment of TED, its differentiation of disease activity and severity is commonly based on the European Group of Graves' Orbitopathy (EUGOGO) [5] classification system, in which the disease activity is assessed by the modified Clinical Activity Score (CAS), including spontaneous and gaze-evoked orbital pain, eyelid swelling and erythema, conjunctival erythema and chemosis, as well as inflammation of caruncle. A cutoff score of three or above out of seven items is defined as active ophthalmopathy; while less than three is inactive. TED is then classified into mild, moderate-to-severe, and sight-threatening [6]. This classification often guides TED management. However, the different clinical presentations in Asian population would raise questions whether implementing this Caucasian based classification may delay and underestimate TED diagnosis in this group, and whether there are any better potential parameters to help in early TED diagnosis.

There is currently an increasing number of studies supporting the correlation between LG dysfunction and TED progress, ranging from LG enlargement clinically [7], TSHr on LG and the involvement of TRAbs immunologically [8], LG inflammation pathologically [9], to increased proinflammatory cytokines and proteomic changes in tear films molecularly [10–13]. Recently, imaging has been acting as an adjunct to the clinicalendocrinological assessment for the diagnosis, grading, treatment, and monitoring of TED. With the established role of LG in TED, imaging studies have been changing their focus from the traditional retrobulbar soft tissues to the structural and functional changes of LG. Previously, computed tomography (CT) studies reported an increase in LG dimensions and volume in TED patients. When compared to CT, magnetic resonance imaging (MRI) has higher soft tissue resolution without radiation.

Apart from the quantitative measurements of structural parameters like dimensions [14–16] and LG herniations [17–19], functional parameters like signal intensity ratio (SIR) in T2-weighted imaging (T2WI) [16, 19], apparent diffusion coefficient (ADC) in diffusion-weighted imaging (DWI) [15, 20], and fractional anisotropy (FA) [21, 22] and ADC [21] or mean diffusivity (MD) [22] in diffusion tensor imaging (DTI) were investigated in multiple studies. To the best of our knowledge, there is no review on MRI parameters of LG in TED patients. We would like to explore whether this newer imaging modality (i.e. MRI), combined with LG parameters could better aid the clinical management of TED. Herein, this systematic review and meta-analysis review and report the outcomes and clinical implications of different MRI parameters on LG in TED patients.

## Methodology

Our systematic review and meta-analysis followed the PRISMA 2020 guidelines [23]. (PROSPERO registration number: CRD42022335591).

## Search strategy

On 23 August, 2022, we performed our literature search on the following electronic bibliographic databases: Pub-Med and the Cochrane Central Register of Controlled Trials (issue 7 of 12, July 2022). We formulated sensitive search strategies using keywords and Medical Subject Heading (MeSH) terms stated in Table 1. No language restrictions nor limitations on publication years were applied. A total of 504 results were yielded (468 from Pub-Med, 36 from the Cochrane Library). 15 duplicates were identified, and 489 results were left for screening (Fig. 1).

## Selection criteria

The search mainly focused on mapping existing literature on MRI parameters on LG in TED. From the 489 results, we included studies based on the following inclusion criteria: 1) comparative studies including case–control

Table 1 Search strategies used in PubMed and Cochrane Library

Search number	Search terms
1	Graves Ophthalmopathy (MeSH Terms)
2	Graves' Ophthalmopathy (Title/Abstract)
3	Thyroid eye disease (Title/Abstract)
4	Thyroid Associated Ophthalmopathy (Title/Abstract)
5	Thyroid Associated Orbitopathy (Title/Abstract)
6	Graves' orbitopathy (Title/Abstract)
7	Graves' eye disease (Title/Abstract)
8	#1 or #2 or #3 or #4 or #5 or #6 or #7
9	Magnetic Resonance Imaging (MeSH Terms)
10	MRI (Title/Abstract)
11	Echo-Planar Imaging (Title/Abstract)
12	Magnetic Resonance (Title/Abstract)
13	Diffusion Tensor Imaging (Title/Abstract)
14	#9 or #10 or #11 or #12 or #13
15	#8 and #14

and cohort studies, 2) cases were TED patients based on clinical diagnosis, 3) controls were healthy subjects or GD patients without TED or patients with inactive TED based on clinical diagnosis, 4) study focuses were on LG findings on MRI, 5) study subjects were unrelated individuals from clearly defined populations, 6) clear MRI LG results (or existing data adequate for calculation) in both case and control groups were provided. Animal studies, case reports, case series, reviews, abstracts, studies without or with incomplete original data were excluded.

In our first review, the titles and abstracts were screened by two independent reviewers (K.Y., and N.W.,) after applying search strategy and eligibility criteria. Disagreements were resolved after discussions with four senior reviewers (K.C., F.A., K.L., and Z.H.,). 475 irrelevant results were removed. Full text screening was then performed on the remaining 14 eligible articles by two independent reviewers (K.Y., and N.W.,). Disagreements were resolved after discussions with four senior reviewers (K.C., F.A., K.L., and Z.H.,). After ensuring eligibility,



Fig. 1 PRISMA flow diagram of literature search and selection process

a total of nine qualified studies were included in our review and meta-analysis (Fig. 1).

#### Data extraction

We adopted a pre-designed form to collect all the extracted data, including the name of first author, year of publication, country of study, ethnicity, definition of case and control groups (thyroid status), CAS, age, sex, sample size, MRI parameters and their respective findings (expressed as mean ± standard deviation (SD) or median (interquartile range (IQR)), and the purposes of the parameters used (including disease activity, severity, or therapeutic responses). We extracted and analysed data on an eye-basis instead of a patient-basis. If results were reported on a patient-basis, we estimated and converted to eye-basis based on the assumption that all subjects had two eyes. If no extractable MRI results were obtained from an eligible study, or if there were confusions with the data reported, we emailed the authors for the missing data and for verification. Two independent reviewers (K.Y., and N.W.,) extracted data, and discrepancies were resolved after discussions with four senior reviewers (K.C., F.A., K.L., and Z.H.,).

## Risk of bias (quality) assessment

We adopted a modified Newcastle–Ottawa Scale (NOS) for cross-sectional studies in the assessment of the quality of each of the selected studies [24, 25]. We gave scores based on the selection, comparability, and outcome to each study. The number of scores were marked in the same form for collecting extracted data. Any study obtaining less than or equal to six out of ten scores was considered as a high risk of inducing bias. Two independent reviewers (K.Y., and N.W.,) were involved in the quality assessment. Disagreements were resolved after discussions with four senior reviewers (K.C., F.A., K.L, and Z.H.,).

#### Statistical analysis

Review Manager (RevMan, Version 5.4. The Cochrane Collaboration, 2020.) was used to perform the metaanalysis for outcome measures which were included in two or more studies. We analyzed LG herniation (LGH), LG dimensions (maximum axial area (MAA), maximum coronal area (MCA), maximum axial length (MAL), maximum coronal length (MCL), maximum axial width (MAW), maximum coronal width (MCW)), DTI-FA, DTI-ADC/MD and DWI-ADC as continuous variables. As all outcomes in the included studies were measured in the same scale, we used mean difference as the summary effect measure for all variables. Mean difference is the difference between the mean of two groups [26]. It is interpreted with *P*-value and 95% confidence interval (CI). If median and interquartile range were used in the studies, mean and standard deviation (SD) were estimated respectively as suggested by Wan et al. [27]. (See *Formula (1)* and *(2)*, Additional file 2) If the measurement of right eye and left eye were grouped and reported separately in the studies, we combined the two subgroups [26] (See *Formula (3)*, Additional file 2).

Heterogeneity was tested using Cochran's Q-statistics chi-square test and I<sup>2</sup>-statistic. If significant heterogeneity was found between the studies (P < 0.1 or  $I^2 \ge 50\%$ ), a random-effect model was used for meta-analysis. Otherwise, a fixed-effect model was used.

## Results

From our literature search, we identified a total of 504 titles and abstracts, and retrieved 14 full texts for review. We finally included nine studies in our systematic review and meta-analyses [14-22].

## **Characteristics of included studies**

Table 2 summarizes the characteristics of the nine included studies. A total of 1012 eyes were included in the nine studies, in which 693 were cases and 319 were controls. Seven studies were conducted in China recruiting Chinese subjects [14-16, 18, 19, 21], while the remaining two studies were conducted in Italy and Egypt recruiting Italians [17] and North Africans subjects [22] respectively. The age ranged from 33.5 [20] to 54.1 [18], while the sample sizes ranged from 64 [17] to 222 eyes [15]. Given that the CAS of a study [15, 17] could not be retrieved, the CAS ranged from 1 [14, 16, 17] to 4.6 [21]. Around 78% (seven out of nine) studies discussed diagnostic purposes [14–18, 20, 21], in which four studies compared both active and inactive TED with healthy controls (HCs) [14, 16, 20, 21], two studies compared active TED with inactive TED [17, 18], while one study used GD as control to make comparisons with both active and inactive TED [15]. The other two studies focused on grading [22] and therapeutic purposes [19] respectively; whilst the former compared mild and moderate-to-severe TED with HCs, the latter compared responsive to unresponsive group to glucocorticoid (GC) therapy in patients with active and moderate-to-severe TED. In terms of the MRI parameters used, three studies [14-16] looked into LG dimensions, among which all three studies reported MCA, MCL, and MCW, while only two studies [14, 16] reported MAA, MAL and MAW. Two studies investigated LG herniation in T2WIfat suppression (T2WI-FS) [17, 18], two studies explored SIR in T2WI [16, 19], and two studies studied DWI-ADC [15, 20]. For the two studies that studied DTI-FA [21, 22], one study reported DTI-ADC [21], while the other reported DTI-MD [22].

No.	Author (year)	Country	Ethnicity	Definition		CAS (Mean≟ (Median [IQI	ESD) or R])	Age (Mean±SL [IQR])	0) or (Median	Gender (M/F		Sample size <sup>a</sup>	MRI parameters	Purpose
				Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	(cases/ controls)	used	
	D. Huang (2014)	China	Chinese	TED (active+inac- tive)	HCS	Active: 4 (3–7) Inac- tive: 1 (0–2)		Active RE: 42.69±11.593 Active LE: 42.71±1.2238 Inactive RE: 37.91±10.406 Inactive LE: 37.40±8.947	40.75 ± 11.947	Active RE: 4/9^ Active LE: 5/9^ Inactive RE: 4/7^ Inactive LE: 3/7 ^	8/8	48/32 (27 active + 21 inactive)/32	Dimensions (MAA, MCA, MAL, MCL, MAW, MCW)	Diagnostic
7	C. Gagliardo (2020)	Italy	Italians	TED (active)	TED (inac- tive)	Active: 4.2±1.93 Inactive: 1±0.91	~	52 [35—68]^	43 [30—64]^	5/11	5/11	32/32	T2WI-FS LGH	Diagnostic
m	L. Chen (2021)	China	Chinese	TED (active + inac- tive)	HCs	Active: 4.6±0.8 Inactive: 1.3±0.8	~	46.4±13.4 Active: 47.8±9.9 Inac- tive: 44.9±16.4	46.5±13.5	12/18 Active: 7/8 Inactive: 5/10	6/9	60/30 (30 active + 30 inactive)/30	DTI – FA DTI—ADC	Diagnostic
4	D. Wu (2021)	China	Chinese	TED (active + inac- tive)	GD with- out TED	~	~	46.98 ± 12.45	36.67 ± 11.16	39/60	4/8	198/24 (142 active + 56 inactive)/24	Dimen- sions (MCA, MCL, MCW) DWI—ADC	Diagnostic
ى ك	H. Hu (2016)	China	Chinese	TED (active+inac- tive)	HCs	Active: 4±1 Inactive: 1 ±1	~	48.4±13.9	50.0±12.9	12/21	8/16	66/48 (26 active + 40 inactive)/48	Dimensions (MAA, MCA, MAL, MCL, MAW, MCW) T2WI-FS SIR	Diagnostic
9	Y. Gao (2022)	China	Chinese	TED (active)	TED (inac- tive)	<b>4.21</b> ± 0.80	1.28±0.69	54.1 ± 8.4	51.3±14.0	13/16^	19/24^	29/43	T2WI-FS LGH	Diagnostic
$\sim$	A. A. Razek (2019)	Egypt	Africans	TED (active + inac- tive)	HCs	Active: 4.4±0.4 Inactive: 1.42±0.3	~	38±12.6	33.5±1.3	17/27	8/12	88/40 (48 active + 40 inactive)/80	DWI-ADC	Diagnostic
00	H. Hu (2020)	China	Chinese	Acitve and moder- ate-severe TED responsive to GC therapy	Active and mod- erate- severe TED unrespon- svie to GC therapy	3.9±0.8	3.8±0.9	47.9±13.1	44.2±12.5	9/20	10/8	58/36	T2WI-FS - LGH T2WI- FSSIR	Therapeutic

Table 2 Characteristics of included studies in our systematic review and meta-analysis

No.	Author (year)	Country	Ethnicity	Definition		CAS (Mean: (Median [IQ	±SD) or įR])	Age (Mean± [IQR])	SD) or (Median	Gender (M/	(F)	Sample size <sup>a</sup>	MRI parameters	Purpose
				Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	(cases/ controls)	used	
6	L. Rui (2021)	China	Chinese	TED	HCS	Mild: 2.00		Mild:	35.4±12.3	Mild: 9/13	7/10	114/34 (44	DTI – FA	Grading
				(mild + moder-		(1.00, 2.00)		39.1±12.3		Moderate-		mild+70	DTIMD	)
				ate-severe)		Moderate-		Moderate-		severe:		moderate-		
						severe: 4.00		severe:		15/20		severe)/34		
						(3.00, 5.00)		47.1±14.5						

controls, LE left eye, LGH Lacrimal gland herniation, M Male, MAA Maximum axial area, MCA Maximum coronal area, MAL Maximum axial length, MCL Maximum coronal length, MAW Maximum axial width, MCW Maximum coronal length, MAW Maximum axial width, MCW Maximum coronal with, MD Mean diffusivity, RE Right eye, SD Standard deviation, T2W+F5 T2-weighted imaging-fat suppression, TED Thyroid eye disease

<sup>a</sup> Sample size is expressed in an eye-basis

Gender is expressed in an eye-basis

The definitions of the MRI parameters are consistent among the included studies. For structural parameters, MCA is defined as LG area in the coronal image in which the LG is the largest, as shown in Fig. 2. MCL is defined as the distance between the superior tip and the inferior tip of LG in the coronal cut where MCA is obtained. MCW is defined as the widest distance perpendicular to the length (MCL) within the LG. The same principle applies to the axial parameters. MAA is defined as LG area in the axial image in which the LG is the largest, as shown in Fig. 3. MAL is defined as the distance between the anterior tip and the posterior tip of LG in the axial cut where MAA is obtained. MAW is defined as the widest distance perpendicular to the length (MAL) within the LG. LGH is defined as the distance between the anterior tip of LG and the interzygomatic line as shown in Fig. 4. For functional parameters, DWI-ADC, DTI-ADC (or MD) and DTI-FA were obtained by first placing a region of interest in the LG which has the largest cross-sectional area, and then measuring the value of ADC, MA or FA of that region of interest in DTI or DWI scan. In our paper, we combine the findings of both DTI-ADC and MD together since they both reflect the magnitude of water diffusion.

## **Risks of bias in included studies**

Table 3 summarized the risk of bias assessment using a modified scale adapted from the Newcastle–Ottawa Scale (NOS) for cohort studies for all our nine selected cross-sectional studies. For more in-depth details, see Supplementary Table 1, Additional File 1.

Among the nine studies, except for one which scored six [14], all of them scored seven or above, indicating that they have a low risk of inducing bias [15-22]. Six of them (67%) adopted convenience sampling by choosing consecutive patients with TED [15, 17, 18, 20-22]. Two of them (22%) lacked detailed descriptions on the recruitment method of subjects [14, 22]. In terms of sample size, all studies did not justify nor show relevant sample size calculation. They all lacked an explanation on the expected sample size to provide a statistically significant information [14-22]. For outcome assessment, five of them (56%) involved more than one assessor who were blinded to the clinical condition of subjects to independently evaluate the MRI results. The intra- or inter-observer variability were appropriately adjusted using relevant statistical methods [15, 18, 19, 21, 22]. Two of them did not mention whether the assessors were blinded or not [16, 17]; while one of them involved only one blinded assessor and did not mention the correction of intra-observer variability [20]. Otherwise, all studies had satisfactory response rates and established characteristics of the subjects, included CAS to ascertain the exposure of subjects (*i.e.* the status of active or inactive TED),

adjusted age and sex as confounders, and clearly stated the appropriate statistical test for data analysis [14-22].

#### Outcome measures

#### i. Active vs inactive

Table 4 summarized the MRI parameters used in the included studies to compare between active TED group and inactive TED group.

We conducted a meta-analysis on eight MRI measurements of active TED patients with inactive TED patients as control group, including LG herniation, LG dimensional parameters (MAA, MCA, MAL, MCL, MAW, MCW) and DWI-ADC. The results are shown in Fig. 5, and the summary is shown in Table 5. Two to three studies were included in each outcome measures. In MAA, MAW and DWI-ADC, there were statistically significant heterogeneity.

The active TED group showed a significant larger LG herniation than the inactive TED group by 3.37 mm (Fig. 5a). For LG dimensions, there was significant difference between the two groups only in MCA by 8.1 mm<sup>2</sup> (Fig. 5c). In contrast, there were no significant differences in MAA (pooled mean difference: 8.3 mm<sup>2</sup>; Fig. 5b), MAL (pooled mean difference: 0.55 mm; Fig. 5d), MCL (pooled mean difference: 0.19 mm; Fig. 5e), MAW (pooled mean difference: 0.05 mm; Fig. 5f) and MCW (pooled mean difference: 0.22 mm; Fig. 5g). The active TED group was also associated with higher DWI-ADC than the inactive TED group by  $0.1 \times 10^{-3}$  mm<sup>2</sup>/s (Fig. 5h).

## ii. TED vs control

Table 6 summarized the MRI parameters used in the included studies to compare between TED group and healthy control group.

We also conducted a meta-analysis on nine MRI measurements of TED patients compared to control group, including LG dimensional parameters (MAA, MCA, MAL, MCL, MAW, MCW), DTI-FA, DTI-ADC/MD, and DWI-ADC. All included studies used healthy subjects as control group except Wu, who used Grave's disease patients without TED as control [15]. The results are shown in Fig. 6, and the summary is shown in Table 7. Two to three studies were included in each outcome measures. In MCA, MCL, MCW, DTI-FA and DWI-ADC, there were statistically significant heterogeneity.

For LG dimensions, there was significant difference between the TED group and the control group in MAA (pooled mean difference: 23.28 mm<sup>2</sup>; Fig. 6a), MCA (pooled mean difference: 14.44 mm<sup>2</sup>; Fig. 6b), MAL (pooled mean difference: 1.88 mm; Fig. 6c), MAW (pooled mean difference: 1.45 mm; Fig. 6e), and MCW (pooled mean difference: 1.00 mm; Fig. 6f). There was no significant difference in MCL (pooled mean difference: 0.37 mm; Fig. 6d). The TED group was associated with lower DTI-FA than the control group



Fig. 2 Definition of MCA, MCL and MCW. MCA = maximum coronal area; MCL = maximum coronal length; MCW = maximum coronal width



Fig. 3 Definition of MAA, MAL and MAW. MAA = maximum axial area; MAL = maximum axial length; MAW = maximum axial width



Fig. 4 Definition of LGH. LGH = lacrimal gland herniation

No.	Author (year)	Selection (max 5)				Comparability (max 2)	Outcome (max 3	3)	Total no. of
		(1) Representativeness of the sample	(2) Sample size	(3) Non- respondents	(4) Ascertainment of the exposure (risk factor)	(1) The subjects in different outcome groups are comparable, based on the study design or analysis. Confounding factors are controlled	(1) Assessment of the outcome	(2) Statistical test	
1	D. Huang (2014)	-	-	1	2	2	-	1	6
2	C. Gagliardo (2020)	-	-	1	2	2	1	1	7
3	L. Chen (2021)	-	-	1	2	2	2	1	8
4	D. Wu (2021)	-	-	1	2	2	2	1	8
5	H. Hu (2016)	1	-	1	2	2	1	1	8
6	Y. Gao (2022)	-	-	1	2	2	2	1	8
7	A. A. Razek (2019)	-	-	1	2	2	1	1	7
8	H. Hu (2020)	-	-	1	2	2	2	1	8
9	L. Rui (2021)	-	-	1	2	2	2	1	8

**Table 3** Risk of bias summary

Details of criteria of each item could be found in supplementary materials. Max Maximum

by 0.04 (Fig. 6g), and higher DTI-ADC/MD by  $0.05 \times 10^{-3}$  mm<sup>2</sup>/s (Fig. 6h). No significant difference in DWI-ADC was found (pooled mean difference:  $0.12 \times 10^{-3}$  mm<sup>2</sup>/s; Fig. 6i).

## iii. Other MRI parameters

Table 8 summarized other MRI parameters used in the included studies for grading or therapeutic purposes.

## Discussion

This systematic review and meta-analysis focused on the MRI measurement on LG of TED patients. Two to three studies were included in the meta-analyses. Active TED

patient group has significantly larger LGH, larger MCA and larger DWI-ADC value than inactive TED patients. TED patient group was significantly larger in five dimensional parameters (MAA, MCA, MAL, MAW, MCW) and DTI-ADC/MD, and was significantly lower in DTI-FA than health controls.

## **MRI LG parameters comparisons**

## i. Active TED vs. inactive TED

In the comparison between active TED patients and inactive TED patients, we found that active TED patient

Parameter	No.	Author (year)	Sample	e size (eye)		Outcome (Mean ± SD OR Medi	ian (IQR))
			Active	Inactive	Control	Active	Inactive
LG herniation (mm)	#2	Gagliardo 2020	32	32	/	RE: 10.1 (7.3–17) LE: 8.5 (6.6–13)	RE: 7 (0-13.4) LE: 5.8 (0-12)
	#6	Gao 2022	29	43	/	15.5±3.25	12.21±2.09
T2WI-FS—SIR	#5	Hu 2016	26	40	48	$2.72 \pm 0.51$	$2.34 \pm 0.36$
Dimension							
Maximum axial area (mm²)	#1	Huang 2014	27	21	32	94.29±17.29	79.92±14.86
	#5	Hu 2016	26	40	48	68.9±31.49	68.63±20.97
Maximum coronal area (mm <sup>2</sup> )	#1	Huang 2014	27	21	32	$103.93 \pm 20.44$	89.32±17.36
	#4	Wu 2021	142	56	24	75 (64, 92)	66 (57, 86)
	#5	Hu 2016	26	40	48	70.36±29.74	68.42±17.6
Maximum axial length (mm)	#1	Huang 2014	27	21	32	18.64±3.09	18.28±2.35
	#5	Hu 2016	26	40	48	16.7±3.7	$15.9 \pm 3.4$
Maximum coronal length (mm)	#1	Huang 2014	27	21	32	21.67±4.04	20.96±3.35
	#4	Wu 2021	142	56	24	18.8 (17.4, 20.4)	18.9 (16.1, 21.0)
	#5	Hu 2016	26	40	48	16.0±4.5	16.3±2.8
Maximum axial width (mm)	#1	Huang 2014	27	21	32	7.21±1.39	$6.44 \pm 1.41$
	#5	Hu 2016	26	40	48	5.1±1.1	6.1 ± 3.8
Maximum coronal width (mm)	#1	Huang 2014	27	21	32	6.77±1.48	6.61±1.3
	#4	Wu 2021	142	56	24	5.8 (4.9, 6.5)	5.6 (4.7, 5.9)
	#5	Hu 2016	26	40	48	$5.4 \pm 1.2$	$5.4 \pm 0.9$
DTI-FA	#3	Chen 2021	30	30	30	0.277±0.04	0.349±0.071
DTI-ADC (10 <sup>-3</sup> mm <sup>2</sup> /s)						1.477±0.342	1.329±0.164
DWI-ADC (10 <sup>-3</sup> mm <sup>2</sup> /s)	#4	Wu 2021	142	56	24	1007.08 (939.71, 1074.83)	955.89 (863.26, 1043.70)
	#7	Razek 2019	48	40	40	1.83±0.08	$1.69 \pm 0.04$

Table 4 Summa	ary of MRI parar	neters of Active TE	D group vs. Inacti	ve TED group ir	n included studies
			- /		

group has significantly larger LGH, larger MCA and larger DWI-ADC value. This implies that these three parameters are potential parameters to differentiate active TED patients from the inactive ones. Out of the seven structural parameters, only LGH and MCA have significant difference. In contrast to functional MRI parameters, structural parameters measure physical LG characteristics to indirectly reflect the degree of inflammation, which is generally more severe in active TED. As a result, structural MRI parameters may not be superior to differentiate active TED patients from inactive ones, comparing with functional MRI parameters. Out of the three parameters showing significant differences, LGH ( $I^2 = 0\%$ ; Fig. 5a) and MCA ( $I^2 = 16\%$ ; Fig. 5c) have insignificant heterogeneity, while DWI-ADC ( $I^2 = 93\%$ ; Fig. 5h) has substantial heterogeneity. The high heterogeneity may affect the validity of the result.

## ii. TED vs. control

In the comparison between TED patients and control, TED patient group was significantly larger in five dimensional parameters (MAA, MCA, MAL, MAW, MCW) and DTI-ADC/MD, and was significantly lower in DTI-FA. This implies that these seven parameters are potential parameters to differentiate TED patients from healthy subjects. Five out of six structural parameters (MAA, MCA, MAL, MAW, MCW) show significant differences. As the difference in the severity of inflammation between TED patients and control is larger than that between active TED patients and the inactive ones, structural parameters can also differentiate TED patients from healthy subjects.

For dimensional parameters, all coronal parameters, *i.e.*, MCA ( $I^2 = 85\%$ ; Fig. 6b), MCL ( $I^2 = 81\%$ ; Fig. 6d) and MCW ( $I^2 = 59\%$ ; Fig. 6f), show significant heterogeneity. In contrast, all axial parameters, *i.e.*, MAA ( $I^2 = 0\%$ ; Fig. 6a), MAL ( $I^2 = 0\%$ ; Fig. 6c) and MAW ( $I^2 = 48\%$ ; Fig. 6e), show insignificant heterogeneity. The study by Wu only measured coronal parameters [15]. It is observed that the result of Wu's study showed lower mean differences consistently, accounting to the high heterogeneity in coronal parameters.

## iii. Structural parameters

Among the dimensional parameters, area parameters perform better at differentiating active TED patients from the inactive ones and TED patients from healthy subjects than length and width parameters. Only MCA

## a. LG Herniation (in mm)

	A	ctive		In	active			Mean Difference				Mean Di	fference		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	Year			IV, Fixed	, 95% CI		
Gagliardo 2020	10.42	6.08	32	6.37	9.28	32	10.8%	4.05 [0.21, 7.89]	2020						
Gao 2022	15.5	3.25	29	12.21	2.09	43	89.2%	3.29 [1.95, 4.63]	2022						
													-		
Total (95% CI)			61			75	100.0%	3.37 [2.11, 4.64]					-		
Heterogeneity: Chi <sup>2</sup> =	0.13, df	= 1 (P	= 0.71)	); I <sup>2</sup> = 09	6					+		5			40
Test for overall effect:	Z = 5.23	(P < (	0.0000	0						-10	Larger	in inactive	Largerin	activo	10
											Larger	mmacuve	Largerin	active	

#### b. Maximum Axial Area (in mm<sup>2</sup>)

		Active		Ir	active			Mean Difference		Mean I	ifference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Rand	om, 95% Cl	
Huang 2014	94.29	17.29	27	79.92	14.86	21	56.9%	14.37 [5.26, 23.48]	2014		-	
Hu 2016	68.9	31.49	26	68.63	20.97	40	43.1%	0.27 [-13.47, 14.01]	2016		<b>*</b>	
Total (95% CI)			53			61	100.0%	8.30 [-5.39, 21.98]				
Heterogeneity: Tau² = Test for overall effect:	64.05; 0 Z = 1.19	Chi <sup>2</sup> = 2 I (P = 0.	.81,df= 23)	= 1 (P =	0.09); P	<sup>e</sup> = 64%				-50 -25 Larger in inactive	0 25 Larger in active	50

#### C. Maximum Coronal Area (in mm<sup>2</sup>)

	A	ctive		In	active			Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% Cl	Year	IV, Fixed, 95% CI
Huang 2014	103.93	20.44	27	89.32	17.36	21	22.9%	14.61 [3.91, 25.31]	2014	
Hu 2016	70.36	29.74	26	68.42	17.6	40	16.4%	1.94 [-10.73, 14.61]	2016	
Wu 2021	77	20.74	142	69.7	21.48	56	60.7%	7.30 [0.72, 13.88]	2021	<b>⊢∎</b> -
Total (95% CI)			195			117	100.0%	8.10 [2.97, 13.22]		
Heterogeneity: Chi² = Test for overall effect	2.39, df = Z = 3.10	2 (P = ) (P = 0.0	0.30); F 02)	²=16%						-50 -25 0 25 50 Larger in inactive

## d. Maximum Axial Length (in mm)

	A	ctive		In	active			Mean Difference		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	Year	
Huang 2014	18.64	3.09	27	18.28	2.35	21	56.9%	0.36 [-1.18, 1.90]	2014	
Hu 2016	16.7	3.7	26	15.9	3.4	40	43.1%	0.80 [-0.97, 2.57]	2016	
Total (95% CI)			53			61	100.0%	0.55 [-0.61, 1.71]		
Heterogeneity: Chi <sup>2</sup> =	0.14, df	= 1 (P	= 0.71	; l² = 09	6					- 4
Test for overall effect:	Z = 0.93	P = 0	1.35)							La



#### e. Maximum Coronal Length (in mm)

	A	ctive		In	active			Mean Difference		Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% Cl	Year	IV, Fixed, 95% CI	
Huang 2014	21.67	4.04	27	20.96	3.35	21	15.5%	0.71 [-1.38, 2.80]	2014		
Hu 2016	16	4.5	26	16.3	2.8	40	18.1%	-0.30 [-2.24, 1.64]	2016		
Wu 2021	18.9	2.2	142	18.7	3.6	56	66.4%	0.20 [-0.81, 1.21]	2021	<mark>_</mark>	
Total (95% CI)			195			117	100.0%	0.19 [-0.63, 1.01]			
Heterogeneity: Chi² = Test for overall effect	= 0.48, df ::Z = 0.45	= 2 (P 5 (P = 0	= 0.79) 0.65)	); I² = 09	6				_	-4 -2 0 2 Larger in inactive Larger in active	4

## f. Maximum Axial Width (in mm)

	A	ctive		In	active			Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year	IV, Random, 95% Cl
Huang 2014	7.21	1.39	27	6.44	1.41	21	53.9%	0.77 [-0.03, 1.57]	2014	
Hu 2016	5.1	1.1	26	6.1	3.8	40	46.1%	-1.00 [-2.25, 0.25]	2016	
Total (95% CI)			53			61	100.0%	-0.05 [-1.78, 1.68]		· · · · · · · · · · · · · · · · · · ·
Heterogeneity: Tau <sup>2</sup> =	: 1.28; Ch	ni² = 5.	.46, df=	= 1 (P =	0.02);	<sup>2</sup> = 829	%			
Test for overall effect:	Z= 0.05	(P = 0	).96)							Larger in inactive Larger in active

#### g. Maximum Coronal Width (in mm)

0.					,									
	P	ctive		In	active			Mean Difference		Mean Difference				
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% Cl	Year		IV, Fix	ed, 95% (	1	
Huang 2014	6.77	1.48	27	6.61	1.3	21	10.2%	0.16 [-0.63, 0.95]	2014			•	_	
Hu 2016	5.4	1.2	26	5.4	0.9	40	21.8%	0.00 [-0.54, 0.54]	2016			1_		
Wu 2021	5.7	1.2	142	5.4	0.89	56	68.0%	0.30 [-0.01, 0.61]	2021					
Total (95% CI)			195			117	100.0%	0.22 [-0.03, 0.47]				٠		
Heterogeneity: Chi <sup>2</sup> = Test for overall effect	: 0.93, df : Z = 1.71	= 2 (P   (P = (	= 0.63) ).09)	); I² = 09	6				-	-2 Larg	-1 er in inactiv	0 /e Large	1 r in active	2

#### h. DWI-ADC (in 10<sup>-3</sup> mm<sup>2</sup>/s)

	Active Inactive							Mean Difference		Mean Difference		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Random, 95% Cl		
Razek 2019	1.83	0.08	48	1.69	0.04	40	51.4%	0.14 [0.11, 0.17]	2019	-		
Wu 2021	1.007	0.1	142	0.954	0.134	56	48.6%	0.05 [0.01, 0.09]	2021			
Total (95% CI)			190			96	100.0%	0.10 [0.01, 0.18]		-		
Heterogeneity: Tau <sup>2</sup> Test for overall effect	= 0.00; C t: Z = 2.25	hi² = 1 5 (P = 0	-0.2 -0.1 0 0.1 0.2 Larger in inactive Larger in active									



Parameter	Pooled mean difference	95% Cl	P-value	Significant difference?	<b>I</b> <sup>2</sup>	
LG herniation	3.37 mm	[2.11, 4.64]	< 0.00001	Yes	0%	
Dimension						
Maximum axial area	8.3 mm <sup>2</sup>	[-5.39, 21.98]	0.23	No	64%	
Maximum coronal area	8.1 mm <sup>2</sup>	[2.97, 13.22]	0.002	Yes	16%	
Maximum axial length	0.55 mm	[-0.61, 1.71]	0.35	No	0%	
Maximum coronal length	0.19 mm	[-0.63, 1.01]	0.65	No	0%	
Maximum axial width	-0.05 mm	[-1.78, 1.68]	0.96	No	82%	
Maximum coronal width	0.22 mm	[-0.03, 0.47]	0.09	No	0%	
DWI-ADC	$0.1 \times 10^{-3} \text{ mm}^2/\text{s}$	[0.01, 0.18]	0.02	Yes	93%	

Table 5	Meta-analyses of	of MRI paramete	rs of Active TED	group vs. Inactive	TED group

CI Confidence interval

Table 6 Summary of MRI parameters of TED group vs. Healthy control group in included studies

Parameter	No.	Author (Year)	Sample	size (eye)		Outcome (Mean ± SD OR Median (IQR))			
			Active	Inactive	Control	TED	Control		
T2WI-FS—SIR	#5	Hu 2016	26	40	48	2.49±0.46	2.17±0.32		
Dimension									
Maximum axial area (mm <sup>2</sup> )	#1	Huang 2014	27	21	32	88±17.64	65.8±11.56		
	#5	Hu 2016	26	40	48	68.74±25.40	43.73±18.57		
Maximum coronal area (mm <sup>2</sup> )	#1	Huang 2014	27	21	32	97.54±20.33	73.49±12.61		
	#4	Wu 2021	142	56	24	74 (60, 87)	67 (58, 81)		
	#5	Hu 2016	26	40	48	69.18±22.97	54.96±19.11		
Maximum axial length (mm)	#1	Huang 2014	27	21	32	18.48±2.77	16.97±2.12		
	#5	Hu 2016	26	40	48	$16.2 \pm 3.5$	$14.0 \pm 1.9$		
Maximum coronal length (mm)	#1	Huang 2014	27	21	32	21.36±3.74	$20.06 \pm 3.05$		
	#4	Wu 2021	142	56	24	18.8 (17.2, 20.7)	19.7 (18.8, 20.5)		
	#5	Hu 2016	26	40	48	16.2±3.5	$15.3 \pm 2.7$		
Maximum axial width (mm)	#1	Huang 2014	27	21	32	6.87±1.44	$5.65 \pm 1.12$		
	#5	Hu 2016	26	40	48	5.7±3.1	3.8±0.8		
Maximum coronal width (mm)	#1	Huang 2014	27	21	32	6.7±1.39	$5.39 \pm 1.01$		
	#4	Wu 2021	142	56	24	5.7 (4.9, 6.3)	4.9 (4.4, 5.8)		
	#5	Hu 2016	26	40	48	$5.4 \pm 1.0$	4.3±0.7		
DTI-FA	#3	Chen 2021	30	30	30	0.313±0.067	0.375±0.082		
	#9	Rui 2021	56	58	34	0.376±0.031	0.399±0.049		
DTI-ADC (10 <sup>-3</sup> mm <sup>2</sup> /s)	#3	Chen 2021	30	30	30	1.403±0.276	1.289±0.172		
DTI-MD ( $10^{-3} \text{ mm}^2/\text{s}$ )	#9	Rui 2021	56	58	34	1.39±0.08	1.34±0.04		
DWI-ADC (10 <sup>-3</sup> mm <sup>2</sup> /s)	#4	Wu 2021	142	56	24	989.81 (915.00, 1063.27)	965.56 (881.40, 1014.81)		
	#7	Razek 2019	48	40	40	$1.73 \pm 0.08$	$1.52 \pm 0.04$		

can differentiate active TED patients from the inactive ones (Fig. 5c), while both MCA and MAA can differentiate TED patients from healthy subjects (Fig. 6a and 6b). Length and width parameters cannot differentiate active TED patients from the inactive ones. The possible reason is that areas are two-dimensional entities. The differences between the groups are more prominent. Area parameters are also more accurate as they reflect changes in two dimensions. LG volume may be an even better dimensional parameter because it is three-dimensional. Among the nine included studies, only the study by Hu (2016) measured LG volume [16]. Hu's method for measuring LG volume requires delineating LG in all slides to obtain the areas, and then multiplying the sum of area and slice interval to compute the volume [16]. It is much more labour-intensive to measure volume than area, as it

## a) Maximum Axial Area (in mm<sup>2</sup>)

		TED		0	control			Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	Year	IV, Fixed, 95% CI
Huang 2014	88	17.64	48	65.8	11.56	32	61.4%	22.20 [15.80, 28.60]	2014	
Hu 2016	68.74	25.4	66	43.73	18.57	48	38.6%	25.01 [16.94, 33.08]	2016	
<b>Total (95% CI)</b> Heterogeneity: Chi <sup>2</sup> = Test for overall effect	: 0.29, df : Z = 9.10	= 1 (P = I (P < 0.	<b>114</b> : 0.59); 00001)	l² = 0%		80	100.0%	23.28 [18.27, 28.30]		-50 -25 0 25 50 Larger in control Larger in TED

#### b) Maximum Coronal Area (in mm<sup>2</sup>)

		TED		(	ontrol			Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Random, 95% Cl
Huang 2014	97.54	20.33	48	73.49	12.61	32	33.6%	24.05 [16.83, 31.27]	2014	
Hu 2016	69.18	22.97	66	54.96	19.11	48	32.9%	14.22 [6.48, 21.96]	2016	
Wu 2021	73.67	20	198	68.67	17.04	24	33.4%	5.00 [-2.36, 12.36]	2021	+=
Total (95% CI)			312			104	100.0%	14.44 [3.44, 25.44]		-
Heterogeneity: Tau <sup>2</sup> =	= 80.06; •	Chi <sup>2</sup> = 1	3.11, đ	f= 2 (P :	= 0.001)	; l² = 8	5%			-50 -25 0 25 50
Test for overall effect	: Z = 2.57	r (P = 0.	01)							Larger in control Larger in TED

#### c) Maximum Axial Length (in mm)

	TED Control							Mean Difference		Mean Difference			
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	Year	IV, Fixed, 95% Cl			
Huang 2014	18.48	2.77	48	16.97	2.12	32	46.5%	1.51 [0.44, 2.58]	2014				
Hu 2016	16.2	3.5	66	14	1.9	48	53.5%	2.20 [1.20, 3.20]	2016				
Total (95% Cl)			114			80	100.0%	1.88 [1.15, 2.61]		•			
Heterogeneity: Chi² = Test for overall effect	0.85, df Z = 5.03	= 1 (P 3 (P < 0	= 0.36	); I² = 09 1)	6					-4 -2 0 2 4			

#### d) Maximum Coronal Length (in mm)

	TED Control							Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Random, 95% CI
Huang 2014	21.36	3.74	48	20.06	3.05	32	28.4%	1.30 [-0.20, 2.80]	2014	
Hu 2016	16.2	3.5	66	15.3	2.7	48	32.9%	0.90 [-0.24, 2.04]	2016	
Wu 2021	18.9	2.59	198	19.67	1.26	24	38.7%	-0.77 [-1.39, -0.15]	2021	
Total (95% CI)			312			104	100.0%	0.37 [-1.04, 1.77]		-
Heterogeneity: Tau <sup>2</sup> = Test for overall effect:	1.23; Cl Z = 0.51	hi <sup>2</sup> = 1 (P = 0	0.70, d 0.61)	f= 2 (P :	= 0.005	5); <b>I</b> ² = 8	31%			-4 -2 0 2 4 Larger in control Larger in TED

## e) Maximum Axial Width (in mm)

	TED Control							Mean Difference		Mean Difference			
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	Year	IV, Fixed, 95% Cl			
Huang 2014	6.87	1.44	48	5.65	1.12	32	65.9%	1.22 [0.66, 1.78]	2014				
Hu 2016	5.7	3.1	66	3.8	0.8	48	34.1%	1.90 [1.12, 2.68]	2016				
Total (95% CI)			114			80	100.0%	1.45 [1.00, 1.91]		•			
Heterogeneity: Chi <sup>2</sup> =	1.92, df	= 1 (P	-4 -2 0 2 4										
l est for overall effect	Z = 0.23	(P < I	1.0000	0						Larger in control Larger in TED			

#### f) Maximum Coronal Width (in mm)

	TED Control							Mean Difference		Mean Difference		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Random, 95% Cl		
Huang 2014	6.7	1.39	48	5.39	1.01	32	27.0%	1.31 [0.78, 1.84]	2014			
Hu 2016	5.4	1	66	4.3	0.7	48	40.9%	1.10 [0.79, 1.41]	2016			
Wu 2021	5.63	1.04	198	5.03	1.04	24	32.1%	0.60 [0.16, 1.04]	2021			
Total (95% CI)			312			104	100.0%	1.00 [0.62, 1.38]		•		
Heterogeneity: Tau <sup>2</sup> :	= 0.07; C	hi² = 4										
Test for overall effect	:Z=5.16	6 (P < 6		-z -i u 1 2								

#### g) DTI-FA

		TED		0	ontrol			Mean Difference		Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	Year	IV, Random, 95% Cl
Chen 2021	0.313	0.067	60	0.375	0.082	30	42.8%	-0.06 [-0.10, -0.03]	2021	
Rui 2021	0.376	0.031	114	0.399	0.049	34	57.2%	-0.02 [-0.04, -0.01]	2021	-
Total (95% CI)			174			64	100.0%	-0.04 [-0.08, -0.00]		•
Heterogeneity: Tau <sup>2</sup> =	0.00; C	hi² = 4.0	12, df =	1 (P = 0	.04); l²=	75%				
Test for overall effect:	Z = 2.06	(P = 0.	04)							Larger in control Larger in TED

## h) DTI-ADC/MD (in 10<sup>-3</sup> mm<sup>2</sup>/s)



## i) DWI-ADC (in 10<sup>-3</sup> mm<sup>2</sup>/s)



Fig. 6 Lacrimal gland parameters in TED and control groups. SD=standard deviation; IV=inverse variance; CI=confidence interval

Parameter	Pooled mean difference	95% Cl	P-value	Significant difference?	l <sup>2</sup>	
Dimension						
Maximum axial area	23.28 mm <sup>2</sup>	[18.27, 28.30]	< 0.00001	Yes	0%	
Maximum coronal area	14.44 mm <sup>2</sup>	[3.44, 25.44]	0.01	Yes	85%	
Maximum axial length 1.88 mm		[1.15, 2.61]	< 0.00001	Yes	0%	
Maximum coronal length	0.37 mm	[-1.04, 1.77]	0.61	No	81%	
Maximum axial width	1.45 mm	[1.00, 1.91]	< 0.00001	Yes	48%	
Maximum coronal width	1.00 mm	[0.62, 1.38]	< 0.00001	Yes	59%	
DTI-FA	-0.04	[0.00, 0.08]	0.04	Yes	75%	
DTI-ADC/MD	$0.05 \times 10^{-3} \text{ mm}^2/\text{s}$	[0.03, 0.07]	< 0.00001	Yes	42%	
DWI-ADC 0.12 × 10 <sup>-3</sup> mm <sup>2</sup> /s		[-0.05, 0.30]	0.16	No	98%	

**Table 7** Meta-analyses of MRI parameters of TED group vs. Healthy control group

Table 8 Summary of other MRI parameters for grading or therapeutic purposes

Sequence— Parameter	No.	Author (year)	Sample size (eye)		Purpose (diagnostic/	Outcome (Mean ± SD)	
			Group 1	Group 2	grading/ therapeutic)	Group 1	Group 2
T2WI-FS—LG hernia- tion (mm)	#8	Hu 2020	Responsive: 58	Unresponsive: 36	Therapeutic: respon- sive vs unrespon- sive to IV steroid after 6 months	11.12±2.15	9.89±2.54
T2WI-FS—SIR	#8	Hu 2020	Responsive: 58	Unresponsive: 36	Therapeutic: respon- sive vs unrespon- sive to IV steroid after 6 months	SIR-max: 3.35±0.88 SIR-mean: 2.42±0.55 SIR-min: 1.43±0.44	3.44±0.94 2.42±0.76 1.48±0.53
DTI-FA DTI-MD (10^-3 mm^2/s)	#9	Rui 2021	Moderate to Severe: 35	Mild: 22	Grading (EUGOGO): moderate to severe vs mild	0.370±0.03 1.41±0.08	0.386±0.031 1.37±0.08

is required to delineate LG in all slides. In clinical practice, it is more difficult to manually measure LG volume for all patients. Measuring maximum area is easier and more practical.

Among the six dimensional parameters, the best parameter is MCA and the worst parameter is MCL. MCA can differentiate active TED patients from the inactive ones (Fig. 5c), and TED patients from healthy subjects (Fig. 6b). In contrast, MCL cannot differentiate both (Figs. 5e and 6d).

## iv. Functional parameters

Substantial heterogeneity was observed in DWI-ADC of both comparisons ( $I^2$ =93% in active TED vs. inactive TED, Fig. 5h;  $I^2$ =98% in TED vs control, Fig. 6i). It is observed that DWI-ADC value in Wu's study is generally lower than that in Razek's study [15, 20]. Wu's method of measuring DWI-ADC value involved delineating the largest coronal area in T2 weighted sections, and then measuring the ADC value of that area in DWI sequence. The most hyperintense spot, which represents the area of most severe inflammation, may not be hit. In contrast,

Razek's method involved placing region of interest directly in DWI sequence and measuring the ADC value. The difference in the method of measuring DWI-ADC value is a possible reason for the generally low DWI-ADC values in Wu's study, and thus the high heterogeneity in the meta-analyses. Another possible reason is ethnicity difference. Wu's study recruited Chinese subjects while Razek's study recruited Egyptians [15, 20].

## xxii. Structural vs. functional parameters

In evaluating the inflammatory activity, functional MRI parameters may be better than structural MRI parameters as functional parameters reflect directly on the level of metabolic activity. However, in both comparisons (*i.e.*, active TED vs inactive TED and TED vs healthy control), structural and functional MRI parameters show comparable results in differentiating between two groups. Two (*i.e.*, LGH and MCA) out of seven structural parameters, and one (*i.e.*, DWI-ADC) out of one functional parameter can differentiate active TED patients from the inactive ones. Five (*i.e.*, MAA, MCA, MAL, MAW and MCW) out of six structural parameters, and two (*i.e.*, DTI-FA

and DTI-ADC/MD) out of three functional parameters can differentiate TED patients from healthy subjects. The result of this meta-analyses showed that functional MRI parameters has no superiority than structural MRI parameters, and vice versa. Computed tomography (CT) is another imaging modality that can measure structural parameters. With comparable results between functional and structural MRI parameters, CT may be comparable to functional MRI in diagnosing TED. However, a study by Lee showed the sensitivity of CT and MRI for detecting active inflammation in TED is 50% and 100% respectively, despite limited validation of the ability of MRI in the study [28]. As a result, there was an inconclusive result in evaluating the superiority between structural and functional MRI parameters.

## **Clinical implications**

## i. Clinical diagnosis of disease activity

Out of the nine included studies, seven studies (78%) were conducted on Chinese patients (Table 2). This could possibly be accounted by the more difficult diagnosis and management based on the clinical manifestations of Chinese patients with TED, where Lim et al. concluded that East Asians generally had fewer exophthalmos, upper eyelid retractions and edema than Caucasian patients, leading to more research interests in finding alternatives (e.g. imaging modalities) for earlier detection or diagnosis of TED [29]. While five [14–16, 18, 21] out of these seven studies, plus two studies done in Italy [17] and Egypt [20], looked into the different LG parameters to aid the diagnosis of TED, we are the first study to do a meta-analysis on the data provided in all these studies. We have found that the LG parameters, both structural and functional, generally provide a more significant diagnostic value in differentiating TED from disease-free patients than in differentiating active from inactive TED patients. To be more precise, LG dimensional parameters including MAA, MCA, MAL, MAW, MCW (Fig. 6a-c, e, f), as well as LG functional parameters including DTI-FA and DTA-ADC/MD (Fig. 6g, h) could possibly be used in clinical practices for differentiating TED from disease-free patients. This is compared to the fewer parameters, i.e., LG herniation, MCA, and DWI-ADC (Fig. 5a, c, h), that could possibly be used to differentiate active from inactive TED patients. This implies that we might take these LG MRI parameters into account when diagnosing TED in the future along with the traditional modified CAS. The possibility of creating a new scoring system for TED activity diagnosis incorporating LG MRI parameters may also be considered, especially among the Asian population.

#### ii. Grading of disease severity

Among all the nine included studies, only a Chinese study by Rui et al. [22] took a step further to compare mild to moderate-severe TED patients and investigate the use of DTI parameters for grading TED severity. Thus, we could not perform a meta-analysis regarding this perspective. Based on the findings by Rui et al. [22], moderate-severe TED group had significantly lower DTI-FA, especially of medial rectus (MR) (P=0.017), and higher DTI-MD (P=0.021) than mild TED group. It also concluded that DTI parameters, especially FA, of MR were sensitive indicators that could help in the differentiation between mild and moderate-severe TED. From this result, we could see the potential role of LG DTI parameters in guiding the grading of TED severity and hence the management plan of TED patients more accurately. However, it is obvious that more studies need to be carried out to draw a more statistically significant conclusion.

## iii. Prediction of therapeutic responses and prognosis

Similar to the above, among all the nine included studies, only a Chinese study by Hu et al. [19] compared the LG parameters, i.e., LG herniation and SIR (SIR-max, SIR-mean, SIR-min), of active and moderatesevere TED patients responsive to intravenous (IV) steroidal therapy after six months to those unresponsive patients. Thus, a meta-analysis regarding the therapeutic responses was not performed. Based on the sole results by Hu et al. [19], it is found that those responsive to IV steroids had a significantly larger LG herniation than those unresponsive (P=0.019), while there were no statistically significant differences in SIR (SIR-max, SIR-mean, SIR-min) between the two groups (P = 0.514, 0.776 and 0.642 respectively). It summarized that the larger LG herniation could possibly be used to distinguish treatment responsive and unresponsive group. This could then possibly allow a wiser allocation of treatment plans, i.e., glucocorticoid therapy for responsive patients, and immunotherapies for unresponsive patients. With more studies investigating this aspect, a more accurate conclusion could then be drawn, and hence more targeted treatment plans could be made for patients to improve their disease prognosis.

## Use of (LG) imaging in managing other orbital/ inflammatory diseases

Apart from the LG parameters studied in our systematic review and meta-analysis for the diagnosis, grading, and prediction of therapeutic responses in TED, in fact, different LG parameters have gained an emerging role in the assessment of other orbital or inflammatory diseases. For instance, one of the differential diagnoses of dry eyes is primary Sjogren's syndrome (pSS), which is an autoimmune disease affecting the salivary and LGs, causing dryness of mouth and eyes [30]. However, its clinical diagnosis is difficult due to its non-specific signs and symptoms [31]. While the current diagnostic criteria related to the orbit involves Schirmer's test and ocular dry scores [32, 33], discomfort to patients and uneasy interpretation arises respectively [31]. Hence, different studies have tried to look for different LG parameters by non-invasive imaging to aid pSS diagnosis. For example, the change in LG size and enhanced signal intensity with accelerated fat deposition in MRI could predict pSS stages [34]; the significantly lower DWI-ADC of LG may suggest LG abnormalities in pSS patients [35]; the lower 11C-MET uptake by LG in PET-CT scan has found to have positive correlation with reduced tear flow [36]. Another example of LG-related orbital disease is IgG4-related disease (IgG4-RD), which is a fibroinflammatory disease with lymphoplasmacytic IgG4-positve plasma cells infiltration to multiple organ tissues that could involve the orbit, in case of IgG4-related ophthalmic disease (IgG4-ROD) [37]. Its current diagnosis is based on the typical organ dysfunctions or structural changes (*i.e.*, swelling), high serum IgG4 titer, and histopathological results from biopsy which is invasive, while imaging could also serve as a non-invasive tool to aid the diagnosis [38]. For instance, it is found that the hypointense and enlarged LG on T2W MRI [39], and the higher uptake of 68 Ga-FAPI in PET-CT by LG could aid the diagnosis and assessment of IgG4-RD [40]. While both pSS and IgG4-RD could result in enlarged LG, the infraorbital nerve enlargement (IONE) in MRI could act as a specific MRI sign of IgG4-ROD [41]. Besides, IONE could also help differentiate IgG4-ROD from other lymphoproliferative orbital diseases, including lymphoma, reactive lymphoid hyperplasia, and idiopathic or other orbital inflammation [42]. From these examples, we could see an increasingly important role of non-invasive imaging techniques, as well as the rising role of LG parameters on different imaging to aid the diagnosis of various orbital or inflammatory diseases in which ophthalmologists are of particular interests.

## Limitations

There were few limitations in our systematic review. First, only two to three studies were included under each outcome measures. The pooled sample sizes may not be large enough to draw a clinically significant conclusion due to random sampling error. Secondly, if heterogeneity was found to be significant by Cochran's Q-statistics chi-square test and  $I^2$ -statistic, subgroup analysis cannot be performed as each subgroup would consist of one to two studies only. As a result, heterogeneity may be significant and may affect the validity of the result of meta-analysis. Further investigations may be needed to explore the reasons behind the high heterogeneity, such as ethnicity differences and scanner dependent differences. Thirdly, the few numbers of included studies also reflects that the field of MRI on lacrimal gland in thyroid eye disease requires further studies.

Out of the nine included studies, seven studies were written by Chinese authors and recruited Chinese subjects [14–16, 18, 19, 21, 22]. Care in application of results to other ethnicities should be considered. Previous studies have demonstrated that differences in clinical manifestations of TED exist between East Asian and Caucasian patients [29, 43]. Radiological differences can arise between different ethnicities, affecting the representativeness of this meta-analysis as most included patients are Chinese.

#### **Future insights**

This is the first systematic review and meta-analysis on the use of MRI LG parameters in TED patients. MRI is a non-invasive imaging modality that can effectively guide the management of TED patients. While the current number of studies on MRI LG parameters is limited, where only two to three studies focus on each parameter, more studies with larger sample sizes and from a wider range of ethnicities would be warranted. The potential LG imaging markers for TED, especially in the aspects of disease grading and therapeutic responses prediction are still under investigations. The use of MRI, which is non-invasive, safe, highly sensitive, could possibly be a rising trend for the diagnosis of TED or other orbital diseases.

## Conclusions

The systematic review and meta-analyses suggest that lacrimal gland herniation, maximum coronal area, and DWI-ADC are able to detect TED patients with active diseases. Maximum axial area, maximum coronal area, maximum axial length, maximum axial width, maximum coronal width, DTI-FA, and DTI-ADC/MD are able to differentiate TED patients from healthy controls. Further studies on the use of MRI on lacrimal gland in the field of thyroid eye disease are warranted to confirm our results.

#### Abbreviations

TED	Thyroid eye disease
GD	Graves' disease
DED	Dry eye disease
MRI	Magnetic resonance imaging
LG	Lacrimal gland
LGH	Lacrimal gland herniation
MAA	Maximum axial area
MCA	Maximum coronal area
MAL	Maximum axial length
MCL	Maximum coronal length
MAW	Maximum axial width
MCW	Maximum coronal width
SIR	Signal intensity ratio
T2WI	T2-weighted imaging
ADC	Apparent diffusion coefficient
DWI	Diffusion-weighted imaging
FA	Fractional anisotropy
MD	Mean diffusivity
DTI	Diffusion tensor imaging
TAO	Thyroid-associated ophthalmopathy
GO	Graves' ophthalmopathy
EOMs	Extraocular muscles
OF	Orbital fat
TSHr	Thyrotropin receptor
TRAbs	Thyrotropin receptor autoantibodies
IGF-1	Insulin-like growth factor-1
IGF-1r	Insulin-like growth factor-1 receptor
EUGOGO	European Group of Graves' Orbitopathy
CAS	Clinical Activity Score
DON	Dysthyroid optic neuropathy
CT	Computed tomography,
MeSH	Medical Subject Heading
SD	Standard deviation
IQR	Interquartile range
NOS	Newcastle–Ottawa Scale
CI	Confidence interval
HCs	Healthy controls
GC	Glucocorticoid
pSS	Primary Sjogren's syndrome
lgG4-RD	lgG4-related disease
lgG4-ROD	lgG4-related ophthalmic disease
IONE	Infraorbital nerve enlargement

## **Supplementary Information**

The online version contains supplementary material available at https://doi. org/10.1186/s12886-023-03008-x.

## Additional file 1.

Additional file 2.

## Acknowledgements

Not applicable.

#### **Financial support**

No conflicting relationship exists for any author.

#### Author's contributions

K.Y., and N.W., reviewed the literature, analyzed the data, and wrote the manuscript. All co-authors have contributed to the work, read and approved the final manuscript.

#### Funding

Not applicable.

#### Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

#### Declarations

#### **Ethics approval and consent to participate** Not applicable.

tot applicable.

**Consent for publication** Not applicable.

#### **Competing interests**

The authors declared that they do not have any competing interests.

#### Author details

<sup>1</sup>Department of Ophthalmology and Visual Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, Hong Kong, China. <sup>2</sup>Department of Ophthalmology and Visual Sciences, Prince of Wales Hospital, Hong Kong Special Administrative Region, Hong Kong, China. <sup>3</sup>Department of Ophthalmology, Salmaniya Medical Complex, Government Hospitals, Manama, Bahrain. <sup>4</sup>Department of Ophthalmology, Tung Wah Eastern Hospital, Hong Kong Special Administrative Region, Hong Kong, China. <sup>5</sup>Hong Kong Eye Hospital, Hong Kong Special Administrative Region, Hong Rong, China.

#### Received: 7 September 2022 Accepted: 31 May 2023 Published online: 08 August 2023

#### References

- 1. Weiler DL. Thyroid eye disease: a review. Clin Exp Optom. 2017;100(1):20–5.
- Eckstein A, Esser J. Graves' orbitopathy. Klin Monbl Augenheilkd. 2011;228(5):432–8.
- Khoo TK, Bahn RS. Pathogenesis of Graves' ophthalmopathy: the role of autoantibodies. Thyroid. 2007;17(10):1013–8.
- Dolman PJ. Grading Severity and Activity in Thyroid Eye Disease. Ophthalmic Plast Reconstr Surg. 2018;34(4S):S34–40.
- Bartalena L, Baldeschi L, Dickinson A, Eckstein A, Kendall-Taylor P, Marcocci C, et al. Consensus statement of the European Group on Graves' orbitopathy (EUGOGO) on management of GO. Eur J Endocrinol. 2008;158(3):273–85.
- Barrio-Barrio J, Sabater AL, Bonet-Farriol E, Velázquez-Villoria Á, Galofré JC. Graves' Ophthalmopathy: VISA versus EUGOGO Classification, Assessment, and Management. J Ophthalmol. 2015;2015: 249125.
- Khu J, Freedman KA. Lacrimal gland enlargement as an early clinical or radiological sign in thyroid orbitopathy. Am J Ophthalmol Case Rep. 2017;5:1–3.
- Eckstein AK, Finkenrath A, Heiligenhaus A, Renzing-Köhler K, Esser J, Krüger C, et al. Dry eye syndrome in thyroid-associated ophthalmopathy: lacrimal expression of TSH receptor suggests involvement of TSHR-specific autoantibodies. Acta Ophthalmol Scand. 2004;82(3 Pt 1):291–7.
- 9. Jacobson DH, Gorman CA. Endocrine ophthalmopathy: current ideas concerning etiology, pathogenesis, and treatment. Endocr Rev. 1984;5(2):200–20.
- Zoukhri D. Effect of inflammation on lacrimal gland function. Exp Eye Res. 2006;82(5):885–98.
- 11. Huang D, Xu N, Song Y, Wang P, Yang H. Inflammatory cytokine profiles in the tears of thyroid-associated ophthalmopathy. Graefes Arch Clin Exp Ophthalmol. 2012;250(4):619–25.
- 12. Matheis N, Okrojek R, Grus FH, Kahaly GJ. Proteomics of tear fluid in thyroidassociated orbitopathy. Thyroid. 2012;22(10):1039–45.
- Khalil HA, De Keizer RJ, Bodelier VM, Kijlstra A. Secretory IgA and Iysozyme in tears of patients with Graves' ophthalmopathy. Doc Ophthalmol. 1989;72(3–4):329–34.
- Huang D, Luo Q, Yang H, Mao Y. Changes of lacrimal gland and tear inflammatory cytokines in thyroid-associated ophthalmopathy. Invest Ophthalmol Vis Sci. 2014;55(8):4935–43.
- Wu D, Zhu H, Hong S, Li B, Zou M, Ma X, et al. Utility of multi-parametric quantitative magnetic resonance imaging of the lacrimal gland for diagnosing and staging Graves' ophthalmopathy. Eur J Radiol. 2021;141: 109815.
- Hu H, Xu XQ, Wu FY, Chen HH, Su GY, Shen J, et al. Diagnosis and stage of Graves' ophthalmopathy: Efficacy of quantitative measurements of the lacrimal gland based on 3-T magnetic resonance imaging. Exp Ther Med. 2016;12(2):725–9.

- Gagliardo C, Radellini S, Morreale Bubella R, Falanga G, Richiusa P, Vadalà M, et al. Lacrimal gland herniation in Graves ophthalmopathy: a simple and useful MRI biomarker of disease activity. Eur Radiol. 2020;30(4):2138–41.
- Gao Y, Chang Q, Li Y, Zhang H, Hou Z, Zhang Z, et al. Correlation between extent of lacrimal gland prolapse and clinical features of thyroid-associated ophthalmopathy: a retrospective observational study. BMC Ophthalmol. 2022;22(1):66.
- Hu H, Xu XQ, Chen L, Chen W, Wu Q, Chen HH, et al. Predicting the response to glucocorticoid therapy in thyroid-associated ophthalmopathy: mobilizing structural MRI-based quantitative measurements of orbital tissues. Endocrine. 2020;70(2):372–9.
- Razek AA, El-Hadidy EM, Moawad ME, El-Metwaly N, El-Said AAE. Assessment of lacrimal glands in thyroid eye disease with diffusion-weighted magnetic resonance imaging. Pol J Radiol. 2019;84:e142–6.
- Chen L, Hu H, Chen W, Wu Q, Zhou J, Chen HH, et al. Usefulness of readoutsegmented EPI-based diffusion tensor imaging of lacrimal gland for detection and disease staging in thyroid-associated ophthalmopathy. BMC Ophthalmol. 2021;21(1):281.
- Rui L, Jing L, Zhenchang W. Diffusion Tensor Imaging Technology to Quantitatively Assess Abnormal Changes in Patients With Thyroid-Associated Ophthalmopathy. Front Hum Neurosci. 2021;15: 805945.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.
- Wells GA, Wells G, Shea B, Shea B, O'Connell D, Peterson J, et al., editors. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses2014.
- Herzog R, Álvarez-Pasquin MJ, Díaz C, Del Barrio JL, Estrada JM, Gil Á. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? a systematic review. BMC Public Health. 2013;13(1):154.
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3: Cochrane; 2022 [updated February 2022. Available from: www.training. cochrane.org/handbook.
- Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.
- Lee MJ, Hamilton BE, Pettersson D, Ogle K, Murdock J, Dailey RA, et al. Radiologic imaging shows variable accuracy in diagnosing orbital inflammatory disease and assessing its activity. Sci Rep. 2020;10(1):21875.
- 29. Lim NC, Sundar G, Amrith S, Lee KO. Thyroid eye disease: a Southeast Asian experience. Br J Ophthalmol. 2015;99(4):512–8.
- Jonsson R, Brokstad KA, Jonsson MV, Delaleu N, Skarstein K. Current concepts on Sjögren's syndrome - classification criteria and biomarkers. Eur J Oral Sci. 2018;126 Suppl 1(Suppl Suppl 1):37–48.
- Bjordal O, Norheim KB, Rødahl E, Jonsson R, Omdal R. Primary Sjögren's syndrome and the eye. Surv Ophthalmol. 2020;65(2):119–32.
- 32. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. Ann Rheum Dis. 2017;76(1):9–16.
- 33. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 2002;61(6):554–8.
- Izumi M, Eguchi K, Uetani M, Nakamura H, Takagi Y, Hayashi K, et al. MR features of the lacrimal gland in Sjögren's syndrome. AJR Am J Roentgenol. 1998;170(6):1661–6.
- Kawai Y, Sumi M, Kitamori H, Takagi Y, Nakamura T. Diffusion-weighted MR microimaging of the lacrimal glands in patients with Sjogren's syndrome. AJR Am J Roentgenol. 2005;184(4):1320–5.
- Jimenez-Royo P, Bombardieri M, Ciurtin C, Kostapanos M, Tappuni AR, Jordan N, et al. Advanced imaging for quantification of abnormalities in the salivary glands of patients with primary Sjögren's syndrome. Rheumatology (Oxford). 2021;60(5):2396–408.
- 37. McNab AA, McKelvie P. IgG4-related ophthalmic disease. Part I: background and pathology. Ophthalmic Plast Reconstr Surg. 2015;31(2):83–8.

- Mod Rheumatol. 2012;22(1):21–30.
  Toyoda K, Oba H, Kutomi K, Furui S, Oohara A, Mori H, et al. MR imaging of IgG4-related disease in the head and neck and brain. AJNR Am J Neuroradiol. 2012;33(11):2136–9.
- Luo Y, Pan Q, Yang H, Peng L, Zhang W, Li F. Fibroblast Activation Protein-Targeted PET/CT with (68)Ga-FAPI for Imaging IgG4-Related Disease: Comparison to (18)F-FDG PET/CT. J Nucl Med. 2021;62(2):266–71.
- Soussan JB, Deschamps R, Sadik JC, Savatovsky J, Deschamps L, Puttermann M, et al. Infraorbital nerve involvement on magnetic resonance imaging in European patients with IgG4-related ophthalmic disease: a specific sign. Eur Radiol. 2017;27(4):1335–43.
- 42. Ohshima K, Sogabe Y, Sato Y. The usefulness of infraorbital nerve enlargement on MRI imaging in clinical diagnosis of IgG4-related orbital disease. Jpn J Ophthalmol. 2012;56(4):380–2.
- Chng CL, Seah LL, Khoo DH. Ethnic differences in the clinical presentation of Graves' ophthalmopathy. Best Pract Res Clin Endocrinol Metab. 2012;26(3):249–58.

## **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

#### At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

