Reproducibility of the unaided subjective assessment of orbital computed X-ray tomographic features in thyroid eye disease

Isobel Landray, B.Sc.[1], James Carpenter, D.Phil.[1,2], Kaveh Vahdani, F.R.C.Ophth.[3], Katherine Miszkiel, F.R.C.R.[3], Lakshmi A. Ratnam, F.R.C.R.[4], Geoffrey E. Rose, D.Sc.[3]

[1] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, U.K

[2] MRC Clinical Trials Unit at UCL, 90 High Holborn, London WC1V 6LJ, U.K.

[3] Moorfields Eye Hospital NHS Foundation Trust, City Road, London EC1V 2PD, U.K.

[4] Department of Radiology, St George's Hospital, Blackshaw Road, London SW17 0QT, U.K.

Correspondence: Professor Geoffrey E. Rose, D.Sc., F.R.C.Ophth.,
Adnexal Service, Moorfields Eye Hospital NHS Foundation Trust,
City Road, London EVC1V 2PD, U.K.

E-mail: geoff.rose1@nhs.net

Conflicts of interest: The authors declare no conflicts of interest.

Key words: thyroid eye disease; thyroid orbitopathy; computed X-ray tomography; orbital imaging; reproducibility; subjective image assessment; CT scan

Précis Without use of any specialist research software, the subjective estimation of eight bone and soft-tissue changes from orbital CT scans of patients with thyroid eye disease shows relatively high intra-observer and inter-observer variability.

Abstract

Objectives: To assess the reproducibility of subjective interpretation of CT for eight features associated with thyroid eye disease (TED).

Methods: Patients with confirmed TED had three distinct orbital CT sections presented as anonymised montages to three masked observers ((#1 orbital radiologist, #2 general radiologist, #3 orbital surgeon). Eight features were graded: Superior orbital fissure (SOF) clarity, degree of orbital fat prolapse through the SOF, loss of fat space at the apex, muscle enlargement, increase in orbital fat volume, vascular congestion, superior ophthalmic vein (SOV) size, and lamina papyracea bowing. Thirty montages were randomly triplicated within the completed image-testing-file.

Results: Each observer provided 3296 assessments of montages from 146 patients (68% female). Observer #2 had the highest rate of "indeterminate" gradings (13.3%), while #1 had the lowest (6.7%). For ***intra-observer*** agreement, the Kappa statistics were 'substantial' to 'almost perfect' for apical crowding, muscular enlargement, and medial bowing, whereas orbital fat expansion and vascular congestion showed only 'slight' to 'moderate' agreement. Excluding SOV size (where indeterminacy was too great for statistical analysis), there was a wide and statistically-significant ***inter-observer*** variation for the other seven features, with no consistent ranking of observer scores.

Conclusions: subjective interpretation of CT images for TED patients has high variability, particularly for inter-observer comparisons. Only the assessment of apical crowding, muscular enlargement, and bowing of the lamina papyracea showed fairly consistent intra-observer gradings. The results suggest that variability in interpretation of such images might only be improved with the use of objective measures applied to the CT images.

Introduction

Orbital imaging is frequently used in the diagnosis and management of thyroid eye disease (TED), particularly for surgical planning, to exclude other pathologies, and for monitoring disease progression and its complications (such as dysthyroid optic neuropathy; 'DON').[1] Whereas magnetic resonance imaging (MRI) might be preferred for assessing tissue oedema (and, by inference, disease activity), computed X-ray tomography (CT) has distinct advantages due to its ability to clearly image bone and soft tissue, its rapid acquisition time (effectively eliminating movement artefacts), and its ready availability and lower cost.[1]

Early CT investigations showed fusiform enlargement of extraocular muscles, particularly the inferior and medial recti, to characterise TED.[2-6] These changes have later been quantified -- both linearly and volumetrically -- and include the widely-used 'Barrett's muscle index' (the proportion of orbit occupied by recti), albeit with variability in diagnostic thresholds, sensitivity, and specificity.[7-13] Other imaging parameters for assessment of TED severity include fat volume and prolapse through the superior orbital fissure, apical crowding, osseous morphology, optic nerve size and configuration, superior ophthalmic vein size, and vascular congestion.[8-16] Assessment of these parameters often requires manual delineation in multiple image-planes, specialist imaging segmentation programmes, or advanced automated multiparametric analysis -- and these complex methods precludes day-to-day usage in clinical care. Their use may be further limited by intricate and cumbersome data acquisition, variation in observer reliability, and complexities of analysis. The few studies that used subjective methods were limited by relatively small sample sizes, or by scans with relatively low resolution.[2-5,8]

This study assessed the reproducibility of non-assisted (subjective) image interpretation for eight commonly-used imaging features associated with TED, the aim being to determine reliability of clinical assessment and to identify which features (if any) might be regarded as a practical guide for general radiologists and oculoplastic specialists.

Methods

Patients with a well-established diagnosis of stable TED (but of variable severity) were identified from the orbital database at Moorfields Eye Hospital and a cohort with suitable electronic orbital CT imaging selected; images with previous orbital fracture or surgery were excluded. Three images were extracted from each scan: (a) an axial section in the plane of the apical optic nerve (Figure 1A), (b) a coronal section ~1cm anterior to the Annulus of Zinn (Figure 1B), and (c) an axial section with the greatest size for the superior orbital fissure (SOF) fat-pad (Figure 1C). For each patient, the images were anonymised and assembled onto a single montage within a 'Master' image-file, with each 'three-image' montage (and, thereby, patient) being assigned a unique number.

An arbitrary selection of 30 montages was triplicated within the Master image-file, with each of the two *copy* images being assigned a new, arbitrary and unique number. To create an unstructured order, the slide-montages were shuffled multiple times within the Master image-file. Finally, to reduce possible bias from observer 'pattern-recognition' (of repeated images), a number of montages were 'side-to-side' reversed (although retaining the same unique number); naturally, during later analysis, the gradings for these 'reversed' montages were assigned to the appropriate side.

Imaging features and grading

The Master image-file, with anonymously-numbered montages for 146 patients, was presented to each of three masked observers – an orbital radiologist (KM, >25 years' experience; Observer 1), a general radiologist (LAR, >15 years' experience; Observer 2), and an orbital surgeon (GER, >30 years' experience; Observer 3).

Montages were subjectively graded for eight imaging features associated with TED (*On-line material*: Table 1). Six changes were graded as being 'normal' (grade 1), or having 'mild' (grade 2), 'moderate' (grade 3) or 'marked' (grade 4) change – namely, (i) loss of soft-tissue clarity at the superior orbital fissure (SOF) ('Clarity SOF'); (ii) degree of orbital fat prolapse through the SOF ('Bulge SOF'); (iii) loss of fat space at the orbital apex ('Apex'); (iv) enlargement of extraocular muscles (EOMs) ('Muscle'); (v) increase in orbital fat ('Fat'); and (vi) degree of orbital vascular congestion ('Congestion'). The two other features were graded as 'normal' (grade 1), 'mild' (grade 2) or 'moderate/marked' (grade 3) – namely, (vii) increased size of the superior ophthalmic vein (SOV) ('SOV size') and (viii) bowing of the lamina papyracea ('Bowing'). If the observer considered a feature unquantifiable or 'indeterminate', this was designated as 'grade 5' [for features (i)-(vi)] or 'grade 4' [for features (vii) and (viii)] (*On-line material*: Table 1).

The study received Institutional Review Board approval (MEH CA#1370) and adhered to the tenets of the Declaration of Helsinki.

Statistical analysis

For the whole series of 146 montages, the number and proportion of 'indeterminate' gradings assigned by each observer was evaluated. For the 30 triplicated montages, the number rated 'indeterminate' (whether once, twice, thrice, or not at all) was tabulated by observer -- with the distribution of counts being compared across observers by a chi-squared test. For each feature and each right or left orbit, inter-observer disagreement on the number of montages with indeterminacy was tallied separately from the number with disagreement between three repeated reviews by the same observer (intra-observer disagreement); 'disagreement' meant either one or two of the three scans being rated as 'indeterminate'. For inter-observer comparison using the whole series of 146 patients, only the first montage was utilised (the second and third montages of the triplicated series being omitted). Non-quantifiable scan features rated as 'indeterminate' were excluded from further analyses.

Three measures of reproducibility were estimated for each feature:

(i) Intra-observer reproducibility was estimated by observer, and left/right orbit, using the 30 triplicated montages (that is, on 90 montages) with the percentage agreement and Cohen's Kappa statistic.[17,18] These estimates required each triplicated assessment to have *no* cases of 'indeterminate' grading for the feature-orbit side under consideration; any montage with even a single 'indeterminate' rating was, therefore, omitted.

The Kappa statistic estimates percentage agreement after adjusting for the agreement that would be expected by chance.[17,18] The Kappa statistic for agreement between all three observers was calculated as the average of the Kappa statistics for each of the three pair-wise observer comparisons. As image-grading is entirely subjective, the Kappa statistic was weighted such that disagreements between 'normal' and 'mild', or between 'moderate' and 'marked' were considered of less clinical importance than disagreements between 'mild' and 'moderate' (*On-line material*: Table 2): This weighting recognizes that a feature read as 'mild' might be 'normal', and that one read as 'moderate' might be 'marked -- in other words, that some harder-to-interpret features might, in fact, be tending towards a binary measure. Likewise, for features with three possible grades, disagreements between 'normal' (Gde.1) and 'mild' (Gde.2) were considered of less clinical importance than disagreements between 'mild' (Gde.2) and 'moderate/marked' (Gde.3) (*On-line material*: Table 2). Confidence intervals for Kappa statistics were calculated through the jack-knife procedure,[19] and, following convention, the Kappa statistic (κ) was interpreted as "poor agreement" (for $\kappa < 0$), "slight" ($0 \leq \kappa < 0.2$), "fair" ($0.2 \leq \kappa < 0.4$), "moderate" ($0.4 \leq \kappa < 0.6$), "substantial" ($0.6 \leq \kappa < 0.8$), and "almost perfect" agreement ($0.8 \leq \kappa \leq 1$).[20]

(ii) Inter-observer reproducibility was analysed using only a single montage (the first) for each of the 146 patients in the study: for each feature, a mixed model (*On-line material*: Appendix A) was fitted to these data, with fixed effects for the side (left/right) and observers, and random intercepts for each patient. These models allowed for comparisons between ratings given by different observers to the same montage and side for each feature.

(iii) Right-left bias for each feature was formally tested from the mixed model fit to assess inter-observer variability (*On-line material*: Appendix A). A Wald test was applied to the coefficient of the fixed effect of orbit side and, if significantly different from zero, provided evidence to suggest a difference in the gradings for right and left orbits.

(iv) An analysis of variance (ANOVA) model with observer, subject, side and observer-subject interaction terms was used to compare intra-observer and inter-observer variability for each feature.[21,22] For this model, all 30 repeated montages were included, and an 'indeterminate' grading was treated as missing. The ratio of inter-observer to intra-observer variance was estimated for each feature.

STATA 18.0 was used to conduct all the analyses and R 4.2.1 was used to produce graphics. Being a study of the reproducibility of clinical imaging techniques, statistical levels of significance were not set for this investigation.

Results

Two-hundred-and-six montages were prepared from 146 patients (99 female; 68%) and, of the 30 triplicated montages, 17 (57%) were from female patients. TED was clinically unilateral in 34/146 (23%) patients, with 14/34 (41%) having DON at the time of imaging; conversely, 112 (77%) had bilateral TED, with 5/112 (4.5%) having unilateral DON and 51/112 (46%) having bilateral DON. Of the triplicated montage subset (90 montages in all), 23/90 were randomly reversed in 21/30 patients (that is, 19 patients had 1/3 scans reversed and 2 patients had 2/3 scans reversed).

Indeterminate ratings

Each observer made 3296 assessments from the Master image-file ('8 features' x '2 sides' x '206 montages') and, of the 3296 assessments, the number classed as 'indeterminate' varied from 220/3296 (6.7%) for Observer 1, to 439/3296 (13.3%) for Observer 2 ($p<0.0001$) (Table 3).

For the 'repeated-test' scenario, each observer reviewed the 30 triplicated scans for '8 features' on '2 sides'; that is, each observer made three repeated judgements on 480 'feature-side-montage' combinations. Most montages were free of indeterminacy, but the proportion of repeated montages with 0, 1, 2 or 3 repeats classed as 'indeterminate' differed significantly by observer (χ_6^2=72.5, $p<0.001$) (Table 3).

When considering 'indeterminate' assessments for each of the 8 features, the number of 'within observer' disagreements across the 30 'repeated-test' sequences varied between nil (for most measures) and 11/30 (37%); all three observers showed intra-observer disagreements whilst assessing

SOV size (between 3.3% and 37%) (Table 3). With the exception of SOV size, there was generally a good inter-observer agreement when judging indeterminacy in the series of 146 single montages (Table 3). When considering the complete series of 206 montages, SOV size was regarded as ‘indeterminate’ in 49%, 98% and 79% of 412 judgements (Observers 1, 2 and 3, respectively).

Intra-observer variation

Figure 2A shows the Kappa statistics and percentages of intra-observer variation, classified by feature, observer and side; due to a high rate of indeterminacy, the results for SOV size cannot be interpreted: Observer 1 had ‘moderate’ agreement (or better) across repeated testing for all features, Observer 3 had ‘fair’ to ‘substantial’ agreement for all features, and Observer 2 had more varied levels of agreement (Figure 2A). All three observers had ‘substantial’ to ‘almost perfect’ agreement for apical crowding (‘Apex’), ‘substantial’ for muscular enlargement (‘Muscle’), ‘moderate’ to ‘substantial’ for SOF clarity, SOF bulging and bowing of the lamina papyracea (‘Clarity SOF’, ‘Bulge SOF’ & ‘Bowing’). Agreement was ‘moderate’ for vascular congestion (‘Congestion’) and between ‘slight’ and ‘moderate’ for degree of orbital fat expansion (‘Fat’) (Figure 2A).

Inter-observer variation

The model fitted to assess inter-observer variation (*On-line material*: Appendix A) provided an estimate of the mean inter-observer difference for each feature (Table 4) (Figure 2B). Inter-observer differences for all features except SOV size (with its high indeterminacy) showed high statistical significance and no consistent ranking of observer scores across each feature. While the average differences between grades for some observers and features is less than 0.5 (e.g. between observers 1 and 2 for ‘Clarity SOF’ the average difference is -0.22), for others the average differences are larger than an integer (e.g. for ‘Fat’, the average difference between observers 1 and 2 is -1.37). Figures 3A and 3B show the Kappa statistics and percentages of inter-observer agreement overall and pairwise, classified by feature and side.

Right-left bias

Estimated from the mixed model fit to assess inter-observer variation (*On-line material*: Appendix A), the differences between right and left orbits are very small, and just reach 5% significance (with no adjustment for multiple testing) for ‘Bulge SOF’ and ‘Bowing’ (Table 4) (Figure 2B).

Comparison of intra-observer and inter-observer variation

Table 5 shows, for each feature, the intra- and inter-observer variation from the ANOVA model; due to indeterminacy in the triplicated montages, the model could not be fitted for ‘SOV size’. Except for bulging of the SOF (‘Bulge SOF’), intra-observer variation for all features explains less than a half of the variation in gradings attributable to inter-observer variation. For all seven measures, the variability between observers is greater than the variability within observers. The model residuals were checked using quantile-quantile plots for approximate normality and there was no evidence of major violations.

Discussion

Subjective assessment of CT in a clinical setting is both practical and rapid, especially as imaging is routinely used by oculoplastic physicians to assess the presence and severity of TED, and to evaluate optic nerve jeopardy.

This study investigated variation in the subjective interpretation of CT images for eight imaging features associated with TED, it therefore questioning the widely-held view that there is relatively little ambiguity in interpreting commonly-used TED imaging features. Whilst comparison of the reproducibility of CT imaging with measures of TED activity (and with normal orbits) would be of interest, it was outside the limitations of this large study. The results reveal considerable variation in assessment of all such features, with SOV size, orbital vascular congestion, and expansion of orbital fat showing the greatest variation (and therefore least reliability for clinical usage). Assessing size of the SOV from the single montages was often not possible and, as such, had the highest rate of ‘indeterminacy’; whilst use of single montages was an unforeseen limitation with our assessment of this feature, use of the whole image sequence during normal clinical practice might actually allow assessment of SOV size. The study also examined whether grading (quantitation) of each imaging feature was observer-dependent and/or side-dependent – the latter being an estimate of any inherent bias of laterality within the scanner or observer. Observer 2 (a general radiologist) had more ‘indeterminate’ assessments as compared to Observers 1 (an orbital radiologist) and 3 (an orbital surgeon). A mixed model, fitted to assess inter-observer variability, did not show any significant bias of laterality.

Comparing variances for subjective clinical assessment of the eight imaging features revealed greater variability between observers than within observers. Intra-observer agreement was 'substantial' to 'almost perfect' for gradings of apical crowding, muscular enlargement, and medial wall bowing (Figure 2A). In contrast, orbital fat expansion and vascular congestion showed only 'slight' to 'moderate' agreement.

Several studies, utilizing research devices and software, have reported the assessment of apical crowding, or muscle and/or fat volumes in TED. With specialist imaging software, a study of 60 TED patients found over 90% concordance for measures of muscle diameter and apical crowding.[16] With interclass correlation coefficients (ICCs) from 0.81 to 0.99, other automated or semi-automated studies also report high inter-observer and intra-observer reliability – this indicating high reproducibility for assisted techniques.[11, 13, 23-25] Likewise, quantitative bone angle measurements in TED show excellent interobserver reliability (ICC 0.87 to 0.97) and very good intra-observer reliability (ICC 0.84 to 0.98).[11] Whilst a few studies have assessed fat prolapse through the SOF in TED, none appear to have evaluated the clarity of the SOF on imaging – with the latter possibly having a significant impact on the assessment of fat prolapse. Birchall and colleagues reported intracranial fat herniation ranging from 2 to 4 mm in 50 TED patients, with excellent inter-observer agreement (K = 0.96),[14] while Cheng et al.[16] found over 90% concordance, and Chan et al.[11] reported "high reliability".

In summary, our investigation highlights the high variability in subjective interpretation of CT images for TED patients. Based on intra-observer variation, the study offers a clinically-useful ranking of the eight features, with substantial intra-observer consistency for estimates of apical crowding, muscular enlargement, and bowing of the lamina papyracea, whilst expansion of orbital fat and orbital vascular congestion showed significant variability. Future research might focus on determining whether variation in clinical image-interpretation applied to patients with active and inactive disease (or to normal patients), on improving the reproducibility through standardised assessment, and on the training of assessors (both in radiology and oculoplastics). Incorporating Artificial Intelligence algorithms might also enhance diagnostic consistency of TED imaging.[26]

References

1) Siakallis LC, Uddin JM, Miszkiel KA. Imaging investigation of thyroid eye disease. *Ophthalmic Plast Reconstr Surg*. 2018;34(4S Suppl 1):S41-S51.
2) Enzmann D, Marshall WH, Rosenthal AR, Kriss JP. Computed tomography in Graves ophthalmopathy. *Radiology* 1976;118:615-620.
3) Trokel SL, Jakobiec FA. Correlation of CT scanning and pathologic features of ophthalmic Graves' disease. *Ophthalmology* 1981;88:553-564.
4) Kennerdell JS, Rosenbaum AE, El-Hoshy MH. Apical optic nerve compression of dysthyroid optic neuropathy on computed tomography. *Arch Ophthalmol* 1981;99:807-809.
5) Neigel JM, Rootman J, Belkin RI, et al. Dysthyroid optic neuropathy. The crowded orbital apex syndrome. *Ophthalmology*. 1988;95:1515-1521.
6) Barrett L, Glatt HJ, Burde RM, Gado MH. Optic nerve dysfunction in thyroid eye disease: CT. *Radiology* 1988;167:503–507.

7) Monteiro ML, Gonçalves AC, Silva CT, Moura JP, Ribeiro CS, Gebrim EM. Diagnostic ability of Barrett's index to detect dysthyroid optic neuropathy using multidetector computed tomography. *Clinics (Sao Paulo)*. 2008;63:301-306.

8) Nugent RA, Belkin RI, Neigel JM, et al. Graves orbitopathy: correlation of CT and clinical findings. *Radiology*. 1990;177:675-682.

9) Giaconi JA, Kazim M, Rho T, Pfaff C. CT scan evidence of dysthyroid optic neuropathy. *Ophthalmic Plast Reconstr Surg*. 2002;18:177-182.

10) Al-Bakri M, Rasmussen AK, Thomsen C, Toft PB. Orbital volumetry in Graves' orbitopathy: Muscle and fat involvement in relation to dysthyroid optic neuropathy. *ISRN Ophthalmol*. 2014:435276.

11) Chan LL, Tan HE, Fook-Chong S, Teo TH, Lim LH, Seah LL. Graves ophthalmopathy: the bony orbit in optic neuropathy, its apical angular capacity, and impact on prediction of risk. *Am J Neuroradiol*. 2009;30:597-602.

12) Chaganti S, Mundy K, DeLisi MP, et al. Assessment of orbital computed tomography (CT) imaging biomarkers in patients with thyroid eye disease. *J Digit Imaging*. 2019;32:987-994.

13) Gonçalves AC, Silva LN, Gebrim EM, Monteiro ML. Quantification of orbital apex crowding for screening of dysthyroid optic neuropathy using multidetector CT. *Am J Neuroradiol*. 2012;33:1602-1607.

14) Birchall D, Goodall KL, Noble JL, Jackson A. Graves ophthalmopathy: intracranial fat prolapse on CT images as an indicator of optic nerve compression. *Radiology*. 1996;200:123-127.

15) Rose GE, Vahdani K. Optic nerve stretch is unlikely to be a significant causative factor in dysthyroid optic neuropathy. *Ophthalmic Plast Reconstr Surg*. 2020;36:157-163.

16) Cheng S, Ming Y, Hu M, et al. Risk prediction of dysthyroid optic neuropathy based on CT imaging features combined the bony orbit with the soft tissue structures. *Front Med (Lausanne)*. 2022;9:936819.

17) Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20:37-46.

18) Abraira V, Perez de Vargas, A. Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs. *Qüestiió*. 1999;23:561-571.

19) Quenouille M. Notes on bias in estimation. *Biometrika*. 1956;43;353-360.

20) Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

21) Bland J. How do I analyse observer variation studies? Available at: https://www-users.york.ac.uk/~mb55/meas/observer.pdf (Accessed 21 August 2024).
22) Bland J, Altman, D. Measurement error. *BMJ*. 1996;313;7059.
23) Zhang T, Chen R, Ye H, Xiao W, Yang H. Orbital MRI 3D reconstruction based on volume rendering in evaluating dysthyroid optic neuropathy. *Curr Eye Res*. 2022;47:1179-1185.
24) Bontzos G, Papadaki E, Mazonakis M, et al. Extraocular muscle volumetry for assessment of thyroid eye disease. J *Neuroophthalmol*. 2022;42:e274-e280.
25) Ugradar S, Goldberg RA, Rootman DB. Bony orbital volume expansion in thyroid eye disease. *Ophthalmic Plast Reconstr Surg*. 2019;35:434-437.
26) Yi C, Niu G, Zhang Y, et al. Advances in artificial intelligence in thyroid-associated ophthalmopathy. *Front Endocrinol (Lausanne)*. 2024;15:1356055.

Legends

Figure 1

Representative CT imaging sections for assessment of thyroid eye disease. **(A)** Axial section in the plane of the apical optic nerve, **(B)** coronal section approximately 1 cm anterior to the Annulus of Zinn, and **(C)** axial section at level of the greatest size for the superior orbital fissure fat-pad.

Figure 2

(**A**) Kappa statistics for intra-observer agreement, classified by radiological feature and left/right orbit. **(B)** Mean ratings for eight radiological features across all 206 assessments (116 single and 30 triplicated montages), classified by imaging feature, observer and side (with 95% confidence intervals).

"Poor" agreement on Cohen-Kappa convention is denoted by "-", "Slight" by "+/-", "Fair" by "+", "Moderate" by "++", "Substantial" by "+++" and "Almost perfect" by "++++". "NA" denotes "not applicable" (where too many image-gradings were classed as 'indeterminate') and "%" denotes the percent absolute agreement.

Figure 3

(A) Kappa statistics for inter-observer agreement, by feature and left/right orbits. **(B)** Kappa statistics for pair-wise inter-observer agreement, classified by feature; for this analysis the two orbits are assumed independent and analysed together. "NA" denotes "not applicable", where too many montages were graded as 'indeterminate'.