HHS Public Access

Author manuscript

Ophthal Plast Reconstr Surg. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Ophthal Plast Reconstr Surg. 2016 ; 32(2): 106–112. doi:10.1097/IOP.0000000000000437.

Axial Globe Position Measurement: A Prospective Multi-center Study by the International Thyroid Eye Disease Society

Chad M. Bingham, M.D.[1], **Jennifer A. Sivak-Callcott, M.D.**[1], **Matthew J. Gurka, Ph.D.**[2], **John Nguyen, M.D.**[1], **Jeffery P. Hogg, M.D.**[3], **Steve E. Feldon, M.D.**[4], **Aaron Fay, MD**[5], **Lay-Leng Seah, M.D.**[6], **Diego Strianese, M.D.**[7], **Vikram D. Durairaj, M.D.**[8], **Jimmy Uddin, M.D.**[9], **Martin H. Devoto, M.D.**[10], **Matheson Harris, M.D.**[1], **Justin Saunders, M.D.**[1], **Tammy H. Osaki, M.D.**[5], **Audrey Looi, M.D.**[6], **Livia Teo, M.D.**[6], **Brett W. Davies, M.D.**[8], **Andrea Elefante, M.D.**[11], **Sunny Shen, M.D.**[6], **Tony Realini, M.D., M.P.H.**[1], **William Fischer**[4], and **Michael Kazim, M.D.**[12]

[1]West Virginia University Eye Institute, West Virginia University Department of Ophthalmology, Morgantown, WV, USA [2]West Virginia University Department of Biostatistics, Morgantown, WV, USA [3]West Virginia University Department of Radiology, Morgantown, WV, USA [4]David and Ilene Flaum Eye Institute, University of Rochester Department of Ophthalmology, Rochester, NY, USA [5]Harvard University Department of Ophthalmology, Boston, MA, USA [6]Singapore National Eye Centre, Singapore [7]Department of Neuroscience, Reproductive Science and Odontostomatology University Federico II, Naples, Italy [8]University of Colorado, Department of Ophthalmology, Denver, CO, Texas Oculoplastic Consultants, Austin, TX, USA [9]Moorfields Eye Hospital, London, England [10]Consultores Oftalmológicos, Buenos Aires, Argentina [11]Department of Advanced Biomedical Sciences, University Federico II, Naples Italy [12]Harkness Eye Institute, Columbia University Department of Ophthalmology, New York, NY, USA

Abstract

Purpose—Identify a reproducible measure of axial globe position (AGP) for multicenter studies of patients with thyroid eye disease (TED).

Methods—This is a prospective, international, multicenter, observational study in which 3 types of AGP evaluation were examined: radiologic, clinical, and photographic. In this study, computed tomography (CT) was the modality to which all other methods were compared. CT AGP was measured from an orthogonal line between the anterior lateral orbital rims to the cornea. All CT measurements were made at a single institution by 3 individual clinicians. Clinical evaluation was performed with exophthalmometry. Three clinicians from each clinical site assessed AGP with 3 different exophthalmometers and horizontal palpebral width using a ruler. Each physician made 3 separate measurements with each type of exophthalmometer, not in succession. All photographic

Please direct all correspondence to Jennifer A. Sivak-Callcott, M.D., West Virginia University Eye Institute, One Medical Center Drive, Morgantown, WV 26505, USA; Telephone: (304)598-6944; Facsimile: (304)598-6928; sivakj@wvuhealthcare.com.

Previous Presentation:

American Society of Ophthalmic Plastic and Reconstructive Surgery Fall Meeting. Chicago, IL, October 2014. International Orbit Symposium. Buenos Aires, Argentina. September 2012.

Financial Disclosures: The authors have no proprietary or commercial interest in any of the materials discussed in this article.

measurements were made at a single institution. AGP was measured from lateral photographs in which a standard marker was placed at the anterior lateral orbital rim. Horizontal and vertical palpebral fissure were measured from frontal photographs. Three trained readers measured 3 separate times, not in succession.

Exophthalmometry and photography method validity was assessed by agreement with CT (mean differences calculation, ICC's, Bland-Altman figures). Correlation between palpebral fissure and CT AGP was assessed with Pearson correlation. Intraclinician and interclinician reliability was evaluated using intraclass correlation coefficients (ICC).

Results—Sixty-eight patients from 7 centers participated. CT mean AGP was 21.37mm (15.96 – 28.90mm) right, 21.22mm (15.87 – 28.70mm) left (ICC 0.996 and 0.995). Exophthalmometry AGP fell between 18mm and 25mm. Intraclinician agreement across exophthalmometers was ideal (ICC 0.948 – 0.983). Agreement between clinicians was greater than 0.85 for all upright exophthalmometry measurements. Photographic mean AGP was 20.47mm (10.92 – 30.88mm) right, 20.30mm (8.61 – 28.72mm) left. Intrareader and interreader agreement was ideal (ICC 0.991 – 0.989). All exophthalmometers' mean differences from CT ranged between −0.06mm (+/− 1.36mm) and 0.54mm (+/− 1.61mm); 95% CI fell within 1mm. Magnitude of AGP did not affect exophthalmometry validity. Oculus best estimated CT AGP but differences form other exophthalmometers were not clinically meaningful in upright measurements. Photographic AGP (right ICC=0.575, left ICC=0.355) and palpebral fissure do not agree with CT.

Conclusions—Upright clinical exophthalmometry accurately estimates CT AGP in TED. AGP measurement was reliably reproduced by the same clinician and between clinicians at multiple institutions using the protocol in this study. These findings allow reliable measurement of AGP that will be of considerable value in future outcome studies.

Introduction

There is no gold standard for measuring axial globe position (AGP). This makes globe position outcomes difficult to compare between institutions and published studies. Although ideal, identifying a gold standard is beyond the scope of this study. Rather, we aim to identify a clinically reasonable, reproducible method of measurement that ultimately will allow valid comparison between patient visits, different observers, and multiple centers in future outcome studies.

In this study, CT was the modality to which other methods were compared.[1] Our primary hypothesis is that computed tomography (CT) and exophthalmometry readings of AGP exhibit excellent agreement, such that exophthalmometry can be considered a clinical standard for assessment of proptosis. Investigation of a previously unpublished method using external photography was also performed and analyzed for agreement with CT data. Correlation between axial globe position and both horizontal and vertical palpebral fissure measurements was studied.

Methods

This is an international, multi-center, observational study of methods used to measure axial globe position in patients with thyroid eye disease (TED). Each site had institutional review board approval. The research adhered to the tenets of the Declaration of Helsinki, and was HIPAA-compliant. Informed consent was obtained from all participants.

Three types of AGP evaluation were examined: radiologic, clinical, and photographic. CT scan was considered the standard to which other methods were compared.[1] Participants had to be 18 years or older, with TED, a medical need for orbital CT scan (related to their TED), and no history of orbital decompression, eye muscle surgery, or facial trauma. Patients under age 18, pregnant, or with a history of orbital or eye muscle surgery were excluded. All clinicians had completed or were in fellowship training.

Radiologic Methods

Non-contrast orbital CT scans were acquired on each subject within one week of clinical measurements using a multidetector CT (MDCT) technique with a field of view selected to adequately contain the orbital structures. Patients were instructed to keep their eyes open and fixed on a target in the gantry. MDCT with a single volume of data was acquired in the axial plane at submillimeter thickness and displayed at 1–2mm thickness. All images were archived on compact disc and sent to a single institution (WVU) for measurement.

Measurements were made with an open source DICOM viewer (OsiriX; Geneva, Switzerland) at 400% magnification in the axial plane. A single investigator (CMB) chose the axial image which included the thickest intraocular lens and drew a line between the right and left anterior orbital rims as described by Segni et al.[2] To compensate for any tilt of the subject's head, images for right and left orbits were determined independently of each other. Three clinicians (JASC, JN, CMB) independently measured the AGP between the reference line and the anterior cornea with an orthogonal line and recorded the length (Figure 3).

Clinical Methods

At each institution three clinicians (oculoplastic faculty or fellows) performed three exophthalmometry measurements, not in rapid succession, on each participant in both the upright and supine positions. All clinicians were instructed in the standard technique (Figure 1).[3,4,5] The base (at the most narrow position) of the exophthalmometer was determined by the lead clinician with the first upright measurement.[4] The measurements were repeated with three different types of exophthalmometers: single mirror, single prism, curved footplate (Hertel, Inami & Co. Tokyo, Japan), double mirror, straight footplate (Oculus Inc., Dutenhofen, Germany), and double prism, straight footplate (Mourits, Medical Workshop, Groningen, The Netherlands) (Figure 2). Measurements were performed as described by Frueh et al. with the clinician seated to the patient's right for the upright position and standing at the head of the patient for the supine position.[6, 25] In addition to exophthalmometry, each clinician made three horizontal fissure measurements using a standard ruler (Figure 1).

Photographic Methods

Right profile, left profile, and full-face photographs were taken on the day of clinical measurements. An arrow sticker was placed on the anterior edge of the lateral orbital rim at the level of the lateral canthus as a reference (Figure 4). The camera (Canon digital Elph SD1400, Tokyo Japan), settings (Portrait mode, ISO 200, full megapixels, flash, maximum zoom and focal length), distance (19 inches) and the arrow stickers were standardized between all centers. The images were uploaded to a website for measurement at a single institution (URMC). All measurements were performed 3 times, not in succession, by 3 different observers who were blind to the results of the other observers. The horizontal and vertical palpebral fissures were measured on the full face view. In the right and left profile view, a vertical line was drawn at the position of the lateral orbital rim, using the sticker as a guide, with an orthogonal line drawn to the anterior surface of the cornea to produce the AGP value (Figure 4).

Statistical Methods

Analyses were performed utilizing SAS® Version 9.3 (SAS Institute Inc., Cary, NC). For clinical and photographic methods, intraclass correlation coefficients (ICC's) were computed to quantify agreement, both interclinician and intraclinician, in an effort to assess reliability of these measurements. A value of 1 indicates perfect agreement; values above 0.9 are considered ideal in measuring inter-rater and intra-rater reliability to be used in clinical settings.[7] Agreement of clinical and photographic AGP measurement methods with the CT scan values was also assessed after averaging across multiple measurements (within and across raters) by patient. To assess agreement between mean measures from these methods and the CT scans, an initial graphical examination was performed, plotting the individual pairs of estimates and visually assessing the overall proximity to the identity line. Bland-Altman figures were also produced, in which differences between the two estimates were plotted against the values of the CT scan estimates.[8] Means and standard deviations of the difference values between each measurement and CT were computed as an estimate of overall bias and precision of the measurement. The 95% confidence intervals (CI's) were also computed for differences between the measurement and CT, with 95% CI's not containing 0 indicating statistically significant differences at α=0.05. Paired t-tests were used to compare mean differences from CT between supine and upright measures of the exophthalmometers. Finally, ICC's again were computed to get an overall statistic quantifying agreement (i.e., proximity to the identity line) to the CT measure (validity). For those measurement methods that were not direct measures of axial globe position (horizontal and vertical eyelid fisher distances), Pearson correlations were calculated between those measurements and CT. All analyses were performed separately for right and left sides.

Results

Radiologic Results (Table 1)

Most right eye AGP measurements on CT across all patients fell between 18mm and 25mm with a range of 15.96mm to 28.90mm, mean=21.37. Similarly, most left eye AGP fell between 18mm and 24mm with a range of 15.87mm to 28.70mm, mean=21.22mm. Interclinician agreement for all CT measurements was near perfect, ICC right=0.996 and ICC left=0.995.

Clinical Results (Tables 1–2)

Sixty-eight patients from 7 centers were enrolled; 8,484 measurements were analyzed. The majority of AGP measurements made with all types of exophthalmometers fell between 18mm and 25mm. The mean measurement, range and standard deviation for each exophthalmometer, for each side and position are listed in Table 1. Intraclinician agreement was ideal across all exophthalmometers, positions, and patient sides (right and left), (ICC ranging from 0.948 to 0.983). Agreement between clinicians was ideal for Hertel and Oculus measurements (Hertel ICC range 0.907 – 0.947; Oculus ICC range 0.898–0.933). Although interclinician agreement was nearly ideal in right-sided Mourits measurements (ICC range 0.897–0.918), agreement was poor on the left side and worst in the upright position, ICC=0.794 (Table 1). Across all types of exophthalmometers and positions, right sided measurements had slightly better intraclinician and interclinician agreement than left sided measurements.

Clinical measurement of horizontal palpebral fissure had ideal intraclinician agreement, ICC range 0.943–0.951, but agreement between clinicians was poor for both right and left sides, ICC range 0.541–0.577 (Table 2).

Photographic Results (Tables 1–2)

The range and standard deviation of the mean in photographic measurement values was broader than that seen in both clinical and radiographic methods. Mean right eye AGP was 20.47mm +/− 4.99mm, range 10.92mm–30.88mm. Mean left eye AGP was 20.30mm +/− 3.72mm, range 8.61mm–28.72mm. Both intrareader and interreader agreement was ideal, ICC range 0.991–0.989 (Table 1). Photographic measurement of horizontal palpebral fissure demonstrated ideal intrareader and interreader agreement. Vertical palpebral fissure measurement had slightly worse agreement both in the same reader and between readers, ICC range 0.840–0.872 (Table 2).

Clinical and Photographic Agreement with CT (Tables 3–4)

All exophthalmometers' mean differences from CT ranged between −0.06mm +/− 1.36mm and 0.54mm +/− 1.61mm and are visually demonstrated in the Bland-Altman plots of Table 4. The 95% CI for all exophthalmometers fell within 1mm. The magnitude of proptosis (AGP) did not appear to affect the validity of any exophthalmometer measurement. Right sided measurements were more closely correlated with CT measurements compared to left-sided measurements and this was most apparent with the Mourits exophthalmometer in the supine position, right ICC=0.905, left ICC=0.836 (Table 3). The clinical AGP measurement method that best estimated CT AGP measurement was Oculus exophthalmometry although the Hertel estimates were also nearly ideal. ICC ranged from 0.855 to 0.916. The Oculus was most precise with mean difference ranging from −0.22mm +/−1.36mm (95% CI −0.56mm, 0.11mm) to 0.12mm +/−1.53 (95% CI −0.25mm to 0.50mm) for both sides and positions. Right sided Hertel measurements were also very precise (mean difference range from 0.11mm +/− 1.58mm (95% CI −0.28mm, 0.50mm) to 0.14 +/− 1.45(95% CI −0.21mm, 0.50mm)). Left sided supine Hertel measurements, and right and left supine Mourits measurements were least precise with statistically significant mean differences from CT ($p<0.05$) (Table 3). Although paired t-test showed a statistically significant difference with patient position for left sided Mourits measurements, there were no clinically meaningful differences in exophthalmometry and CT based on patient position (Table 3).

Photographic AGP measurement was a poor estimate of CT AGP measurement, right side ICC=0.575, left side ICC=0.355. In addition, the photographic measurement method was imprecise, right side mean difference from CT = 0.95mm +/− 3.85mm (95% CI −1.90, −0.01), left side mean difference from CT=0.88mm +/− 3.73mm (95% CI −1.80, 0.04) which is visually demonstrated in the Bland-Altman plots in Table 4.

Palpebral fissure correlation with AGP

Neither horizontal nor vertical palpebral fissure meaningfully correlate with AGP. Right sided clinical horizontal correlation coefficient was r=0.58 (p<0.01), left side was r=0.47 (p<0.01). Right sided photographic horizontal correlation coefficient was r=0.54 (p<0.01), left was r=0.35 (p<0.01). Vertical palpebral fissure correlation coefficients were lower than horizontal, right r=0.38 (p<0.01), left r=0.19 (p=0.14).

Discussion

The axial position of the eye relative to the bony orbit (AGP) is used to diagnose, characterize, and follow orbital disease, including TED. Although CT, MRI, and eyelid correlation measurements have been used to assess AGP, exophthalmometry is most commonly reported and most familiar to clinicians.[2,9,10,11,12,13,14] Exophthalmometer design and user technique affect accuracy and reproducibility.[3,16, 25] Interobserver variability is common.[17,18]

Because there are many different types of exophthalmometers and multiple variables affect measured values, assessment of AGP between patient visits, clinicians and centers is difficult. Nevertheless, AGP, and change in AGP, is an important parameter in the management of TED. This is the first prospective study undertaken to specifically identify a reproducible, reliable AGP measurement method at multiple institutions.

Radiologic measurement of AGP was the modality to which all other methods were compared in this study. We demonstrated nearly perfect interobserver agreement for CT measurement using our protocol, but radiation exposure, cost and the frequent need for imaging make use of this method for all AGP assessment impractical. However, a strong correlation between CT and exophthalmometry has been reported. Segni et al. measured AGP on 42 TED patients with a Krahn exophthalmometer and compared results to AGP measurements obtained by CT in a manner similar to our study; a correlation coefficient of r=0.91 was found.[2] Hauck et al reported an ICC of 0.95 with an average difference of 0.03mm between Hertel and CT measuremtns in 53 patients.[26] In our study of 68 TED patients, the mean difference between all exophthalmometers and CT ranged from −0.06mm (+/− 1.36mm) to 0.54mm (+/− 1.61mm) with 95% CI falling within 1mm. This confirms not just a strong correlation between CT and exophthalmometry measurement, but strong agreement between the measures.

Given the strong agreement between CT and exophthalmometry, we hypothesized that the latter could be the standard for proptosis measurement, and studied 3 different types of exophthalmometers. Intra-clinician correlation was ideal with all instruments, showing good reproducibility between measurements by the same clinician. Inter-clinician correlation was greater than 0.85 for all upright exophthalmometry measurements. When comparing the exophthalmometry measurements to CT, the Oculus was found to be most accurate, but the differences between the exophthalmometers were very small. Given the 95% confidence interval within 1mm, the true differences between all upright exophthalmometry measurements and CT values are likely minimal and not clinically significant.

For the present study, each institution used the same set of exophthalmometers (one of each type) on every patient. Of note, Sleep et al. studied 10 Hertel-type exophthalmometers and found variation between manufacturers, as well as between instruments provided by the same company.[17] This variation was up to 2.9mm, which is much greater than the maximum variation for any of the exophthalmometers in this study. Because of the marked variability between instruments of the same make and model, we recommend the same instrument be used for all measurements on the individual patient at every visit.

Statistical analysis found the Oculus to have the best agreement with CT (mean difference not statistically significant). A similar conclusion was reached by Vardizer et al. based on their study of 8 different exophthalmometers using a mechanical model; when used by experienced orbital surgeons, the single mirror, straight footplate design was more accurate than others tested.[15] However, the only statistically significant exophthalmometry differences were seen with supine measurements; the upright measurements in all 3 types of exophthalmometers were not statistically different from CT.

Technique affects exophthalmometry readings.[3,4,5,20] All clinicians in the present study were instructed in published standard exophthalmometry procedure: including positioning the examiner as far from the patient as possible, measuring both eyes without removing the instrument, aligning the examiner’s eyes in the same plane as those of the patient, and sitting to the right of the patient. [3,4,5,20] As in other published studies, the base was determined by the first clinician, at the narrowest point, and then held constant for all subsequent measurements.[4] As indicated by the exceptional intra-clinician agreement for all measurements, all clinicians in the study were consistent in their technique with all 3 exophthalmometers. Training in proper exophthalmometry technique is clearly essential for all clinicians involved in future studies, regardless of prior experience. Additionally, intra-clinician agreement analysis should be used to verify consistency.

The accuracy of AGP measurement is influenced by clinician experience. Musch et al. studied inter-observer variability among 4 clinicians (oculoplastic surgeon, fellow, resident, technician) using a Hertel exophthalmometer.[4] They found a 61–80% agreement with the senior clinician among observers (within +/− 1mm), with the technician showing the least agreement. In the study by Kashkouli et al., exophthalmometry measurements were made by an experienced oculoplastic surgeon and by a third-year ophthalmology resident in 1063 patients, with 60% agreement (within +/− 1 mm) between the observers.[23] In our study design, we chose clinicians who routinely use exophthalmometry in practice. All clinicians were orbital surgeons or orbital surgery fellows, and as previously noted, all were consistent in their particular measurements across all three exophthalmometers. Although agreement was best with Hertel and Oculus instruments, the clinicians in this study did not routinely use the Mourits exophthalmometer in practice. Based on results of the present and previous studies, we recommend that participants in future AGP outcome research be at the fellowship level of experience or higher.

There are conflicting reports on the effect of patient position, supine versus upright, on AGP measurement.[6,21] Some studies suggest there is an increase in AGP when measured with the patient supine. [6,24] We found no clinically meaningful effect of patient position. The only statistically significant difference was with left-sided measurements made with the Mourits instrument. Although the patient is supine when a CT is performed, the least precise exophthalmometry measurements compared to CT were with the patient supine. Consequently, we recommend exophthalmometry be performed with the patient upright.

Right-sided measurements in our study were slightly more reliable and better correlated with CT, both for the individual clinician and between clinicians across all exophthalmometers. All measurements were taken from the patient’s right side as previously described. [3,4,5,20] Clinician position (sitting directly in front versus to the right or left of the patient), handedness, and ocular dominance may contribute to the right-left differences, and future studies could focus on these effects. It is possible the Mourits agreement may have been better if the clinician were positioned in front of the patient. [25] However, given that measurements made with the Hertel and Oculus instruments (patient upright) showed ideal intra-clinician, inter-clinician agreement, and ideal or nearly ideal agreement with CT for both right and left sides, we recommend that clinicians continue to sit to the right of the patient.

Some studies have found a greater level of error in exophthalmometry measurement as the amount of proptosis increases.[3,15,16] In our study, the magnitude of proptosis did not appear to affect accuracy of measurement with any of the exophthalmometers, but few patients had proptosis greater than 25mm. The mean values and ranges of exophthalmos in our study were very similar to those of previous reports. Frueh et al. measured AGP on 84 patients with TED using a Hertel exophthalmometer, and reported mean values of 22.5 mm (SD=3.5) on the right and 21.9 mm (SD=3.8) on the left.[19] In the study by Segni et al. mean clinical AGP was 22.6 mm (SD=2.9), and 21.3 mm (SD=2.8) by CT.[2]

Our hypothesis, that exophthalmometry provides the best estimate of AGP, was confirmed by the results of the present study. The photographic method used in this study was not a good estimate of CT AGP. The photographic measurements had a much broader range and standard deviation, much lower ICCs, and broader range of mean differences, compared to the clinical and radiologic measurements. This is likely due to imprecision in marking the lateral rim as both intra- and inter-reader agreement was ideal. Additionally, horizontal and vertical palpebral fissure measurements in our study do not correlate well enough with CT AGP to be clinically useful. Others have investigated the correlation of eyelid position, photography, and clinical exophthalmometry. Edwards et al. demonstrated a strong correlation between clinical and photographic measurement of eyelid position, including vertical palpebral fissure determinations.[22] However, these were not compared to AGP measurements. Miot et al. obtained multiple photographic measurements of eyelid position relative to points on the eye, and compared these to clinical exophthalmometry.[9] The only measurements for which there was a correlation were the distance of the lateral limbus to the lateral canthus, and the distance from the superior limbus to the lateral canthus. [9] Based on available data, vertical and horizontal palpebral fissure measurements should not be used to estimate AGP.

In summary, we found that clinical exophthalmometry provides an accurate estimate of AGP as determined by CT in patients with TED. Base on our findings we recommend at least fellowship level experience, setting the base (as narrow as possible) at initial examination, use of the same instrument for every measurement, sitting to the right of the upright patient, and averaging 3 non-sequential readings. We also recommend using an exophthalmometer with which the clinicians have significant experience and familiarity. In this study, the values for AGP obtained with exophthalmometry are reliable and reproducible for the individual clinician and between clinicians at multiple institutions. This standardized method could be used for future multi-center outcome studies.

Acknowledgments

Grant Support:

Dr. Gurka’s work on this study was supported by NIGMS grant U54GM104942.

Drs. Sivak, Bingham, Nguyen, Saunders, and Realini’s work on this study was supported by Research to Prevent Blindness, Unrestricted Grant, West Virginia University.

Dr. Feldon’s work on this study was supported by Research to Prevent Blindness, Unrestricted Grant, University of Rochester.

Dr. Realini’s work on this study was supported by NEI grant K23EY18859.

ITEDS support by an ASOPRS Foundation Grant

References

1. Nkenke E, Maier T, Benz M, Wiltfang J, Holbach LM, Kramer M, Hausler G, Neukam FW. Hertel exophthalmometry versus computed tomography and optical 3D imaging for the determination of the globe position in zygomatic fractures. Int J Oral Maxillofac Surg. 2004 Mar; 33(2):125–33. [PubMed: 15050067]
2. Segni M, Bartley GB, Garrity JA, Bergstralh EJ, Gorman CA. Comparability of proptosis measurements by different techniques. Am J Ophthalmol. 2002 Jun; 133(6):813–8. [PubMed: 12036674]
3. Frueh WT, Freuh BR. Error of single-mirror or prism Hertel exophthalmometers and recommendations for minimizing the errors. Ophthal Plast Reconstr Surg. 2007 May-Jun;23(3): 197–201.
4. Musch DC, Frueh BR, Landis JR. The reliability of Hertel exophthalmometry. Observer variation between physician and lay readers. Ophthalmology. 1985 Sep; 92(9):1177–80. [PubMed: 4058879]
5. Lam AK, Lam CF, Leung WK, Hung PK. Intra-observer and inter-observer variation of Hertel exophthalmometry. 2009 Jul; 29(4):472–6.
6. Frueh BR, Garber F, Grill R, Musch DC. Positional effects on exophthalmometer readings in Graves’ eye disease. Arch Ophthalmol. 1985; 103:1355–1356. [PubMed: 3840014]

7. Kottner J, Audige L, Brorson S, Donner A, Gaiewski BJ, Hrobiartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan; 64(1):96–106. [PubMed: 21130355]
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; i:307–310. [PubMed: 2868172]
9. Miot HA, Fernandes LPS, Jorge EN, Pivotto DR, Nogueira CR, Mazeto GMFA. Comparative evaluation of oculometric variables In Graves' ophthalmopathy. Clinics. 2009; 64(9):885–9. [PubMed: 19759882]
10. Drews LC. Exophthalmometry and a new exophthalmometer. Trans Am Ophth. 1956; 54:215–250.
11. Bogren HG, Schermer MJ, Franti C, Elfstrom G, Tengroth B. Radiographic exophthalmometry. Tr Am Acad Ophth & Otol. 1975:298–304.
12. Krimsky E. Photorecording Exophthalmometer. Ophthalmology. 1978; 85:1186–1189. [PubMed: 733171]
13. Detorakis ET, Drakonaki E, Papadaki E, Pallikaris IG, Tsilimbaris MK. Effective orbital volume and eyeball position: an MRI study. Orbit. 2010 Oct; 29(5):244–9. [PubMed: 20812829]
14. Campi I, Vannucchi GM, Minetti AM, Dazzi D, Ayignone S, Covelli D, Curro N, Ratiglia R, Guastella C, Pignataro L, Beck-Peccoz P, Salvi M. A quantitative method for assessing the degree of axial proptosis in relation to orbital tissue involvement in Graves' orbitopathy. Ophthalmology. 2013 May; 120(5):1092–8. [PubMed: 23399378]
15. Vardizer Y, Berendschot TT, Mourits MP. Effect of exophthalmometer design on its accuracy. Ophthal Plast Reconstr Surg. 2005 Nov; 21(6):427–30.
16. Davanger M. Principles and Sources of Error in Exophthalmometry. A New exophthalmometer. Acta Ophthalmologica. 1970; 48:625–633. [PubMed: 5536709]
17. Sleep TJ, Manners RM. Interinstrument variability in Hertel-type exophthalmometers. Ophthal Plast Reconstr Surg. 2002 Jul; 18(4):254–7.
18. O'Donnell NP, Virdi M, Kemp EG. Hertel exophthalmometry; the most appropriate measuring technique. Br J Ophthalmol. 1999 Sep.83(9):1096B.
19. Frueh BR, Musch DC, Garber FW. Exophthalmometer readings in patients with Graves' eye disease. Ophthalmic Surg. 1986 Jan; 17(1):37–40. [PubMed: 3754041]
20. Ameri H, Fenton S. Comparison of unilateral and simultaneous bilateral measurement of the globe position, using Hertel exophthalmometer. Ophthal Plast Reconstr Surg. 2004 Nov; 20(6):448–51.
21. Asad R, Tewari HK, Ahuja MM, Mithal A. Postural variation in Graves' ophthalmopathy. Indian J Ophthalmol. 1990; 38:166–168. [PubMed: 2086467]
22. Edwards DT, Bartley GB, Hodge DO, Gorman CA, Bradley EA. Eyelid position measurement in Graves' ophthalmopathy: reliability of a photographic technique and comparison with a clinical technique. Ophthalmology. 2004 May; 111(5):1029–34. [PubMed: 15121384]
23. Kashkouli MB, Beigi B, Noorani MM, Nojoomi M. Hertel exophthalmometry: reliability and interobserver variation. Orbit. 2003 Dec; 22(4):239–45. [PubMed: 14685897]
24. Duke-Elder, S.; MacFaul, PA. Lacrimal, Orbital, and Periorbital Diseases. In: Duke-Elder, editor. System of Ophthalmology. St. Louis MO: CV Mosby; 1974. p. 781
25. Genders SW, Mourits DL, Jasem M, Kloos RJHM, Saeed P, Mourits MP. Parallax-free Exophthalmometry: A Comprehensive Review of the Literature on Clincial Exophthalmometry and the Introduction of the First Parallax-Free Exophthalmometer. Orbit. 2014 Oct.:1–7. Early Online.
26. Hauck M, Tao J, Burgett R. Computed tomography exophthalmometry: a comparison with Hertel. Ophthalmic Surgery, Lasers & Imaging. 2010 Mar.22:1–4.10.3928/15428877-20100210-82

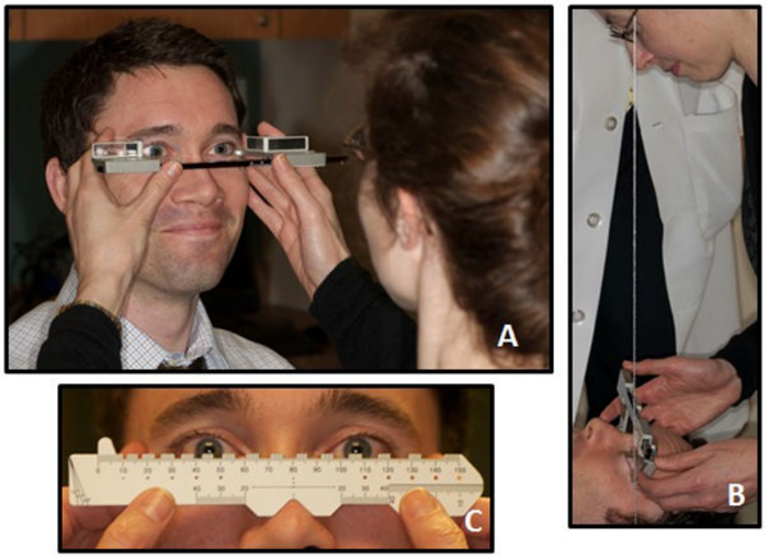


Figure 1.
A: Clinical exophthalmometry, patient upright
B. Clinical exophthalmometry, patient supine
C: Clinical horizontal eyelid fissure measurement

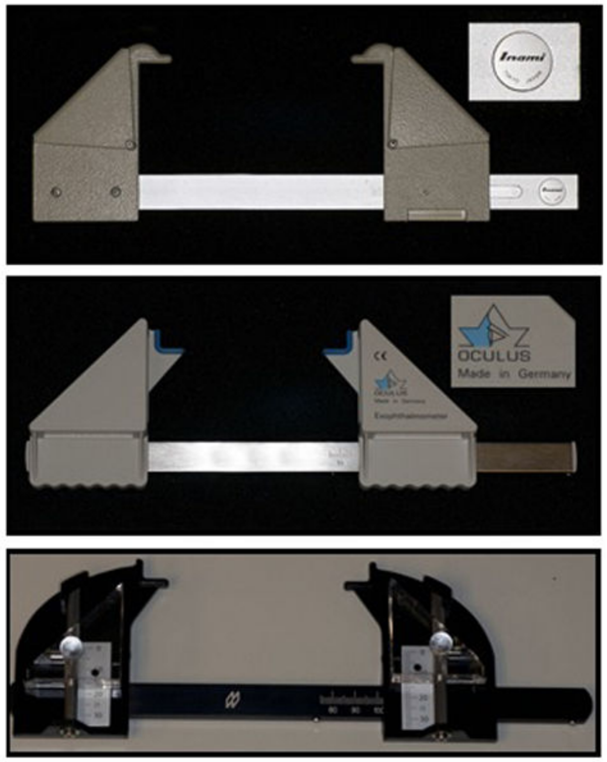


Figure 2. Three types of exophthalmometers utilized in this study

A: Single mirror, single prism (Hertel)

B: Double mirror, no prism (Oculus)

C: Double prism (Mourits)

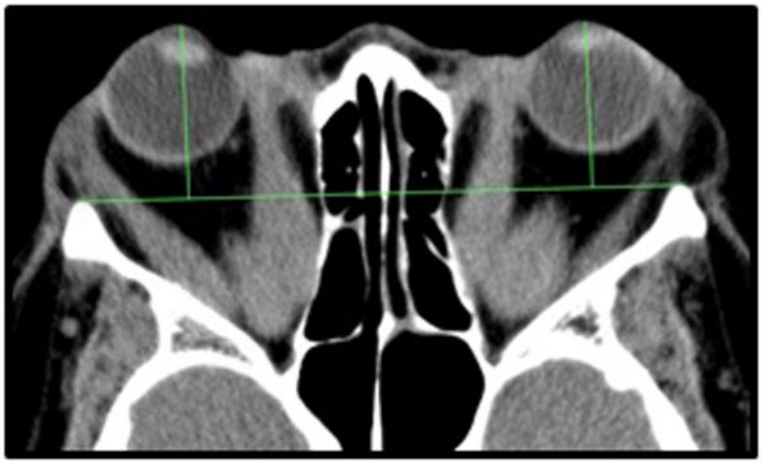


Figure 3.
CT Measurement of axial globe position

Figure 4. Photographic Measurements

A: Lateral measurement of axial globe position

B: Horizontal and Vertical eyelid fissure measurement

Table 1

Results of all Three Types of AGP Measurement: Means, SD's, Range, and ICC's (Interclinician, Intraclinician, Interreader, Intrareader).

Measurement Method	Position	Side	Mean (mm)	SD	Range	ICC (Agreement)	
						Inter	Intra
Radiologic							
Axial CT scan	Supine	Right	21.37	3.35	(15.96, 28.90)	0.996	--
		Left	21.22	2.88	(15.87, 28.70)	0.995	--
Clinical							
Hertel Exophthalmometer	Supine	Right	21.45	3.35	(15.44, 30.78)	0.940	0.983
		Left	21.59	2.81	(16.33, 28.56)	0.907	0.979
	Upright	Right	21.40	3.49	(14.22, 30.67)	0.947	0.980
		Left	21.44	2.88	(16.22, 29.22)	0.924	0.973
Oculus Exophthalmometer	Supine	Right	21.20	3.40	(15.00, 29.78)	0.933	0.983
		Left	21.33	2.83	(16.00, 28.78)	0.898	0.975
	Upright	Right	21.10	3.30	(14.00, 28.67)	0.928	0.982
		Left	21.14	2.82	(15.33, 28.56)	0.904	0.978
Mourits Exophthalmometer	Supine	Right	21.71	3.55	(14.67, 31.11)	0.918	0.972
		Left	21.71	3.00	(15.56, 30.67)	0.863	0.965
	Upright	Right	21.59	3.50	(15.11, 30.22)	0.897	0.976
		Left	21.34	2.86	(15.78, 29.11)	0.794	0.948
Photographic							
Lateral AGP Measurement	Upright	Right	20.47	4.99	(10.92, 30.88)	0.987	0.989
		Left	20.30	3.72	(8.61, 28.72)	0.991	0.992

ICC=0.850 –0.900 highlighted in grey (nearly ideal)

ICC<0.850 highlighted in yellow (not ideal)

Table 2

Palpebral Fissure Measurements: Means, SD's, Range, and ICC's (Interclinician, Intraclinician, Interreader, Intrareader).

Measurement Method	Position	Side	Mean (mm)	SD	Range	ICC (Agreement)	
						Inter	Intra
Clinical							
Horizontal Palpebral Fissure	Upright	Right	29.73	1.98	(15.96, 28.90)	0.577	0.951
		Left	29.74	1.80	(25.89, 34.33)	0.541	0.943
Photographic							
Horizontal Palpebral Fissure	Upright	Right	30.45	2.77	(22.90, 38.21)	0.968	0.977
		Left	29.81	2.44	(25.83, 36.31)	0.959	0.969
Vertical Palpebral Fissure	Upright	Right	12.18	2.18	(7.64, 17.38)	0.872	0.872
		Left	12.07	2.13	(7.65, 18.14)	0.840	0.875

ICC=0.850 –0.900 highlighted in grey (nearly ideal)

ICC<0.850 highlighted in yellow (not ideal)

Table 3

Clinical and Photographic AGP Measurement Agreement with CT scan

Measurement Method	Position	Difference from CT Scan Value (millimeters)						ICC (Agreement)	
		Right			Left				
		Mean (mm)	SD	95% CI	Mean (mm)	SD	95% CI	Right	Left
Clinical									
Hertel Exophthalmometer	Supine	0.14	1.45	(−0.21, 0.50)	**0.43**	**1.48**	**(0.07, 0.80)**	0.908	0.855
	Upright	0.11	1.58	(−0.28, 0.50)	0.28	1.52	(−0.10, 0.65)	0.895	0.858
	*p-value**	*0.79*			*0.14*				
Oculus Exophthalmometer	Supine	−0.14	1.43	(−0.49, 0.21)	0.12	1.53	(−0.25, 0.50)	0.913	0.858
	Upright	−0.22	1.36	(−0.56, 0.11)	−0.06	1.36	(−0.40, 0.27)	0.916	0.887
	*p-value**	*0.32*			*0.04*				
Mourits Exophthalmometer	Supine	**0.38**	**1.48**	**(0.02, 0.75)**	**0.54**	**1.61**	**(0.15, 0.94)**	0.905	0.836
	Upright	0.27	1.57	(−0.12, 0.65)	0.14	1.51	(−0.23, 0.51)	0.896	0.863
	*p-value**	*0.21*			*< 0.01*				
Photographic									
	Upright	**−0.95**	**3.85**	**(−1.90, −0.01)**	−0.88	3.73	(−1.80, 0.04)	0.575	0.355

ICC=0.850 –0.900 highlighted in grey (nearly ideal)

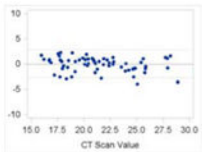
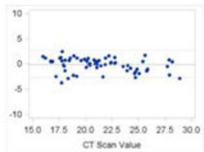
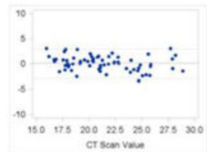
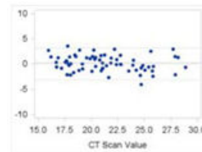
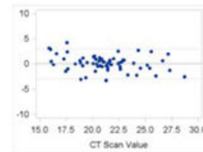
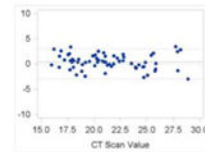
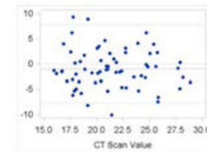
ICC<0.850 highlighted in yellow (not ideal)

Numbers in bold indicate statistically significant ($p < 0.05$) mean differences from the CT scan values

* paired t-test comparing supine vs. upright differences from the CT scan value

Table 4

Bland Altman Plots, Clinical and Photographic AGP versus CT AGP measurement

Right-Sided Measurements: Differences between Clinical and Photographic vs. CT						
Oculus Supine Mean Diff=−0.1mm	Oculus Upright Mean Diff=−0.2mm	Hertel Supine Mean Diff=0.2mm	Hertel Upright Mean Diff=0.2mm	Mourits Upright Mean Diff=0.3mm	Mourits Supine Mean Diff=0.4mm	Photography Mean Diff=−1.0mm
						
Left-Sided Measurements: Differences between Clinical and Photographic vs. CT						
Oculus Supine Mean Diff=0.1mm	Oculus Upright Mean Diff=−0.1mm	Hertel Supine Mean Diff=0.4mm	Hertel Upright Mean Diff=0.3mm	Mourits Upright Mean Diff=0.1mm	Mourits Supine Mean Diff=0.5mm	Photography Mean Diff=−0.9mm