Interobserver and intraobserver variability of measurements of uveal melanomas using standardised echography

C Haritoglou, A S Neubauer, H Herzum, W R Freeman, A J Mueller

Br J Ophthalmol 2002;86:1390–1394

See end of article for authors' affiliations

Correspondence to: Christos Haritoglou, MD, Department of Ophthalmology, Ludwig-Maximilians-Universität, Mathildenstrasse 8, 80336 Muenchen, Germany; Christos.Haritoglou@ak-i.med.uni-muenchen.de

Accepted for publication 26 June 2002

Aim: To report on the intraindividual and interindividual variability of tumour size (height and base diameter) measurements using standardised echography in a masked prospective study.

Methods: 20 consecutive eyes of 20 patients were examined on four different visits by three experienced examiners using standardised echography. As common in standardised echography, tumour height was evaluated with A-scan technique, while transverse and longitudinal base diameter were calculated with B-scan.

Results: Tumour height measurements using A-scan were more accurate than base diameter measurements using B-scan. The standard deviation for tumour height over all visits/measurements was 0.18 mm (A-scan), 0.79 mm for transverse, and 0.69 mm for longitudinal base diameters (B-scan). The interclass correlation coefficient (ICC) was much higher for tumour height measurements with A-scan (0.7735 for three examiners on one visit) than for transverse (0.6563) or longitudinal (0.4522) base diameter measurements with B-scan techniques.

Conclusions: A-scan techniques for tumour height measurements provide very reproducible results with little intraindividual and interobserver variability. As B-scan techniques for tumour base evaluation are less accurate they should be used for topographic and morphological examinations.

Several approaches in the management of uveal melanoma exist, mainly depending on the size of the tumour.^{1–6} In the past 20 years an increasing percentage of patients has been treated with globe preserving therapies. Thus, it is important to have reliable parameters indicating whether the treated lesions decrease in size over time or remain unchanged. In addition, most ophthalmic-oncologists prefer to follow small melanocytic lesions until tumour growth is observed before a treatment is initiated. In these cases standardised A-scan echography is the most commonly used technique for biometry of the eye. It is also essential to establish the diagnosis of uveal melanoma using specific criteria.⁷

Standardised echography was introduced in the 1960s by Ossoinig for the purpose of ophthalmic tissue differentiation. The term standardised echography refers to a special examination technique which is based on the use of a standardised A-scan instrument especially developed for tissue differentiation. It is complemented by a real time contact B-scan.^{8–12} Standardised A-scan is characterised by special signal processing through defined parameters, the so called "internal standardisation" which is provided by the manufacturer (narrowband receiver, special S-shaped type of amplification with a well defined dynamic range, high frequency filtering, etc).¹³ The "external standardisation" is performed by the examiner and includes the ascertainment of the "tissue sensitivity setting," which is optimal for tissue diagnosis.¹⁴ However, besides specially designed equipment, the term standardised echography also implies a standardisation of the A-scan and B-scan examination of the globe and orbit. Meanwhile, standardised echography is established as a very helpful and reliable diagnostic tool in ophthalmology for the evaluation of ocular as well as orbital diseases.^{15–19} In standardised echography A-scan is used for tumour height measurement and B-scan for tumour base measurement as well as topographic and morphological evaluation.

However, follow up examinations may not always be performed by the same examiner, particularly in larger hospitals. Consequently, it is important to know about the intraindividual and interobserver variability of tumour size measurements obtained by standardised echography in order to estimate whether differences in tumour height and base diameter are the result of individual variations of the examination technique or real variations in tumour size. The knowledge about the reliability of tumour measurements may also be an important parameter in prospective clinical studies following tumour patients over a longer period of time.^{6, 20, 21}

METHODS

Twenty eyes of 20 consecutive patients of different age and sex were included in this prospective study. Criteria for inclusion were uveal melanomas not involving the iris or ciliary body. Each patient was examined by three independent and experienced examiners on four occasions during a period of less than 2 weeks. This short period of time was chosen to assure that possible variations of tumour height were not due to real tumour growth. Patients after any radiation treatment within the last 6 months were excluded for the same reason.

We attempted to perform this study with as much masking as was feasible. Thus, the three examiners performed echographic evaluations of each patient independently from each other. Every patient was measured at four different visits by all three examiners. The results of each examination were documented by the examiner. The three examiners were masked to the results of the other examiners. It was not possible, however, to completely mask the examiners to their own previous results, as each examiner had to choose the most accurate measurement documenting the maximum tumour height and so got to know about the numerical value of the measurement.

In all patients the same instrumentation was used for A-scan and B-scan examination (ultrasound B-scan-S V-plus,

Table 1 Standard deviations of tumour height, transverse and longitudinal base diameter measurements between different examiners on four visits as well as over all visits

Mean (SD) over all patients	Examiner			SD between different examiners
	A	B	C	
Visit 1				
Height	4.20 (2.37)	4.34 (2.57)	4.38 (2.70)	0.30 (n=3)*
Transverse	9.25 (2.67)	9.66 (2.07)	9.37 (2.79)	1.01
Longitudinal	9.76 (2.50)	10.05 (2.67)	9.59 (2.81)	0.72
Visit 2				
Height	4.34 (2.54)	4.44 (2.57)	4.32 (2.56)	0.15 (n=3)*
Transverse	8.90 (2.34)	9.40 (2.28)	9.12 (2.54)	0.81
Longitudinal	9.75 (2.49)	10.04 (2.20)	9.74 (2.51)	0.58
Visit 3				
Height	4.37 (2.58)	4.36 (2.55)	4.35 (2.56)	0.13 (n=3)*
Transverse	9.11 (2.38)	9.59 (2.31)	9.20 (2.58)	0.76
Longitudinal	9.63 (2.63)	9.94 (2.52)	9.48 (2.72)	0.78
Visit 4				
Height	4.30 (2.51)	4.34 (2.53)	4.33 (2.53)	0.10 (n=3)*
Transverse	9.08 (2.35)	9.62 (2.19)	9.42 (2.46)	0.68
Longitudinal	9.47 (2.25)	9.94 (2.30)	9.70 (2.58)	0.63
Visits 1–4				
Height	4.30 (2.45)	4.37 (2.50)	4.35 (2.54)	
Transverse	9.09 (2.40)	9.57 (2.17)	9.28 (2.55)	
Longitudinal	9.65 (2.43)	9.99 (2.38)	9.63 (2.60)	
Mean SD within one patient over all 4 visits (within each examiner)				
Height	0.18 (n=4)†	0.16 (n=4)†	0.15 (n=4)†	0.18 (n=12)*
Transverse	0.56	0.61	0.50	0.79
Longitudinal	0.63	0.72	0.47	0.69

*Number of different measurements over all three examiners.

†Number of different measurements within one examiner.

Memory Card Version S 2.07, all by Biovision/Schwind, Kleinostheim, Germany). We followed the specific criteria described by Ossoinig⁷ that are used to evaluate an intraocular tumour with standardised echography.

The eye was open during examination. Each examination consisted of a preliminary topographic B-scan evaluation. B-scan was also used to evaluate the shape of the lesion and to measure both the maximum transverse and longitudinal extension of the tumour base with calipers.⁷ A relatively low gain setting was used in order to obtain the best measurement. Measurements were taken to the inner sclera, which was identified as the first distinct line at the tumour base that was continuous with the surrounding fundus. Gain settings for B-scan evaluation were not specified. Tumour height measurements were then obtained with the A-scan.⁷ B-scan was not used to evaluate the tumour height. The standardised A-scan instrument was first set at tissue sensitivity with the probe placed to the opposite side of the tumour. Great attention was paid to the perpendicular orientation of the sound beam with respect to the point of maximal tumour elevation and the inner sclera. Once the tumour surface spike and the scleral spike were displayed with their maximum height, the gain was lowered while continuously monitoring the screen until the peaks were distinct and clear. Measurements were obtained by placing the calipers on the peak of the tumour surface spike and the inner scleral spike. At least three images of high quality were taken. One of these photographs representing the most accurate measurement was finally chosen by the examiner and the tumour height was documented. Other documented parameters were shape, internal reflectivity (low, medium, high), internal structure (homogeneous-regular, heterogeneous-regular, irregular), vascularity (positive, negative), retinal detachment (positive, negative), and location (posterior, equator, anterior).

The data were collected and analysed using SPSS 10.0 for Windows (SPSS Inc, Chicago, IL, USA), $p < 0.05$ was considered statistically significant. Intraclass correlation coefficients (ICC) were calculated for the intraindividual and intra-

observer variation. The ICC is commonly used as a measure of reliability, with a value of 1 representing a perfect correlation. It was calculated on mean values according to Portney²² based on a program by A Chang.²³

RESULTS

Demographics

Twenty patients, 10 men and 10 women, were included in this study. The mean age was 64 years (range 43–82 years). Seventeen patients had undergone radiation therapy in the past (more than 6 months ago). Tumour height varied from 1.5 to 11.5 mm, mean tumour height was 4.35 mm. Standard deviations are analysed in detail below. Six tumours were located at the posterior pole close to the optic nerve, 11 at the equator, and three were located anteriorly with the borders still being distinguishable. In six cases a shallow retinal detachment overlying the tumour was noted. There were 16 dome-shaped tumours (five of them with an irregular surface contour) and four tumours with a collar button shape.

Intraobserver and interobserver variability

The main results of the obtained measurements are summarised in Table 1. Comparing tumour evaluation with standardised A-scan and B-scan, tumour height measurements using A-scan technique were approximately three times more reproducible than transverse or longitudinal base diameter measurement using B-scan (Fig 1). The standard deviation over all three examiners and all patients was 0.18 mm for tumour height (A-scan), 0.79 mm for transverse and 0.69 mm for longitudinal base diameter (both B-scan).

With respect to tumour height measurements with standardised A-scan, the mean coefficient of variation (CV) within the same examiner (intraobserver CV) was in the range of 20–35%. Comparing the three different examiners, there were only small differences regarding the intraobserver CV, meaning that different observers measured with comparable precision. We observed no statistically significant difference between small (<5 mm) and large melanomas in terms of CV.

Figure 1 Differences between examiner A, B, and C concerning measurements of tumour height (A) on A-scan and on transverse (B) and longitudinal (C) base diameter on B-scan. The three examiners are labelled A, B, and C.

Table 3 Interclass correlation coefficients (ICC) for other ultrasonic parameters between three examiners for 20 patients

	Correlation between the 3 raters on the first visit	
	ICC	F value
Localisation	0.8855	1.3571
Vascularity	0.9678	0.4873
Structure	0.7344	2.0496
Reflectivity	0.6795	2.9835
Retinal detachment	0.9317	1.0000

There was only a trend towards larger differences in tumour height measurement using A-scan in larger tumours (Fig 1A). As shown in Figure 1B and C large differences were observed in both transverse and longitudinal base diameter measurements using B-scan techniques.

The mean deviation of tumour height measurements between different examiners decreased when measuring the same subject repeatedly (standard deviation over different examiners of first measurement was 0.30 and 0.10 of fourth measurement, see Table 1). Because of very anterior location, the longitudinal base diameter could not be obtained by all three examiners in two patients.

Interclass correlation coefficients (ICC)

In Table 2 maximum tumour height is the parameter with the highest correlation coefficient between the three raters on their first visit (ICC 0.7735). Tumour height was also the parameter which shows highest correlation when the same examiner re-examines the patient on different visits (ICC 0.9334 for examiner A, 0.7970 for examiner B, and 0.6879 for examiner C). Correlation coefficients for transverse and longitudinal measurements were much lower. For all ICCs no significant F value was found between the raters or between visits within the same rater.

For the other morphological and quantitative parameters such as localisation, possible retinal detachment, vascularisation, internal structure, and reflectivity (shown in Table 3), which were only assessed on the first visit, very high ICCs were found indicating good concordance between the different raters.

DISCUSSION

To our knowledge, the present study is the first to evaluate both interobserver and intraobserver variabilities of tumour measurements within one institution in a masked prospective setting.

In our institution standardised echography is routinely used to evaluate maximum tumour height as well as

Table 2 Interclass correlation coefficients (ICC) within and between three examiners for 20 patients

	Within rater A on 4 different visits		Within rater B on 4 different visits		Within rater C on 4 different visits		Between the 3 raters on the first visit	
	ICC over all days (4 days)	F value	ICC over all days (4 days)	F value	ICC over all days (4 days)	F value	ICC for the first measurement	F value
20 patients								
Height	0.9334	2.4281	0.7970	0.5132	0.6879	1.8146	0.7735	1.9750
Transverse	0.2156	0.5024	0.4484	2.7872	0.5745	1.6656	0.6563	1.1423
Longitudinal	0.1063	0.8687	0.4063	0.6453	0.6801	0.7203	0.4522	1.2342

transverse and longitudinal base diameters of uveal melanomas. The goal of this study was to evaluate the accuracy of the examination technique. In standardised echography the A-scan technique is commonly used for tumour height evaluation. As there is no specific gain setting for B-scan evaluation, B-scan is only used for gross tumour height measurements.²⁴ B-scan is useful for measuring the basal diameter of a tumour, although in some cases these measurements might not represent the true maximal diameter of a lesion. However, it still provides reproducible measurements for follow up evaluations. B-scan may also be not accurate in some cases because of difficulties in precise localisation of the tumour borders, especially in tumours with gradually sloping borders. Therefore B-scan measurements are not considered a reliable method to determine tumour growth or regression. However, B-scan is the only method other than ophthalmoscopy that can be used to evaluate the basal diameter of a tumour. It is therefore useful as a second source of measurement and can be especially helpful when the shape of the tumour prevents an accurate assessment with the ophthalmoscope. In addition, B-scan measurement of the tumour base is useful if radiation is being considered for treatment. Sizing of the plaque is important in order that the tumour is appropriately covered. Main domain of A-scan is the measurement of tumour height and evaluation of internal structure, reflectivity, and vascularisation.⁷ The standardisation of A-scan ultrasound technique in terms of a "tissue sensitivity setting"¹⁴ allows easy reproduction of measurements in contrast with non-standardised B-scan technique. While B-scan provides an easily understandable topographic and morphological information, no statement can be made concerning internal structure, reflectivity or vascularisation of a lesion. This information can only be obtained from amplitudes, shapes, and motions of A-scan spikes. However, optimal results can only be obtained, when all the acoustic information of A- and B-scan methods is put together.¹²

One of the main questions when following tumour patients is either whether the tumour decreases in size—for example, after radiation treatment, or whether tumour growth is observed in smaller melanocytic lesions in order to initiate an appropriate therapy. It is therefore important to know about the accuracy of the examination technique in terms of interobserver and intraobserver variabilities of measurements.

According to our data the following statistical consideration should be taken in account when deciding on tumour growth: measurements falling outside a 95% confidence interval usually are considered statistically reasonably safe to indicate a change. Given normal distribution of the measurements this means that the measurements have to fall outside plus or minus two standard deviations (2 SD) from the measured value. Our overall standard deviation (Table 1) was 0.18 mm for maximum height, 0.79 mm for transverse, and 0.69 for longitudinal base diameters. Thus, a difference in measurements of more than 0.36 mm (that is, 2 SD) may be considered a true change in tumour height, regardless of the absolute tumour height. Tumour shape or location did not influence the results of the measurements obtained in this study.

Nevertheless, as shown in our data, repeated measurement reduces the variance between the examiners. This should be due to a training effect of the examiners (and patients), as the examinations were performed in a masked manner. Therefore, the echographic estimation of tumour height should always be based on several measurements in clinical routine.

Interestingly we observed only small interobserver differences and a high ICC of tumour height measurements (mean SD 0.18, 0.16, and 0.15 mm, Table 1). It is therefore possible to have the patient examined by different examiners and still make valid conclusions on changes of tumour height, but attention should be paid to compare only measurements obtained by examiners having approximately the same level of experience. This relative independence of the examiner is a clear advantage of standardised A-scan echography.

With respect to the great differences of B-scan evaluation of a tumour (Fig 1B and C), B-scan as a non-standardised technique should not be used for the definition of tumour growth. However, it is still suitable for morphological and topographic examinations. Our standard deviation of 0.18 mm for all examiners is comparable to the data reported by other authors. Nicholson²⁵ described that approximately 90% of independent measurements by two technicians made from photographic records of 53 previously performed scans were within 0.4 mm of each other with a standard deviation of 0.22 mm. A standard deviation of 0.20 mm for tumour height was also reported by Char in a prospective series of 26 patients.²⁶ However, this study focused only on intraobserver variabilities of echographic measurements and confirmed ultrasound to be the most accurate method of measuring tumour height. Furthermore, in a retrospective review of 32 uveal melanomas Char²⁷ reported a difference in ultrasound measurements of thickness of 0.64 mm between two institutions (interobserver variability, range 0–2.2 mm, SD 0.60 mm), both in small and large tumours. The correlation between the two institutions was 0.89. According to Char several factors affect these correlations like the experience in ultrasound measurements, the location of the tumour or the instrumentation used.

However, as those two studies on interobserver variability were performed in a retrospective manner, they did not address the precision of independent measurements during separate dynamic echographic examinations as in our series, but relied on photographic evidence.^{25–27} In addition, the authors did not describe the intraobserver variability of echographic measurements that plays an important part if a patient is seen by the same examiner on different occasions.

In conclusion, standardised echography is a very helpful tool for tumour height measurements. The interobserver variability of A-scan measurements is low. Therefore, the results of the examinations obtained on different occasions during follow up may be compared and conclusions concerning tumour growth may be drawn with little limitations, even if the examinations are performed by different, yet experienced, examiners. Tumour growth may be suspected if the difference between measurements is at least two standard deviations, in our study equalling 0.36 mm. Nevertheless, further studies will be needed to compare the technique described in this report with recent developments in ophthalmic ultrasound such as three dimensional imaging of intraocular lesions and measurement of tumour volume.²⁸ Compared to A-scan technique, B-scan is less accurate, but remains a helpful method for morphological and topographic evaluation and tumour base measurements. However, with the latter larger standard deviations of the measurements have to be taken into account.

Authors' affiliations

C Haritoglou, A S Neubauer, H Herzum, A J Mueller, Department of Ophthalmology, Ludwig-Maximilians-University, Munich, Germany
W R Freeman, UCSD Shiley Eye Center, La Jolla, CA 92093, USA

The authors do not have any commercial interest in any of the materials and methods used in this study.

REFERENCES

- 1 Bornfeld N. Diagnose und Behandlung maligner Melanome der Uvea (Aderhaut und Ziliarkörper). *Ophthalmologie* 1992;**89**:61–78.
- 2 Shields J. Management of uveal melanoma. A continuing dilemma. *Cancer* 1993;**72**:2067–8.
- 3 Müller AJ, Talies S, Schaller UC, et al. Die stereotaktische Präzisionsbestrahlung großer uvealer Melanome mit dem Gamma-Knife. *Ophthalmologie* 2000;**97**:537–545.
- 4 Char DH, Kroll SM, Castro J. Ten-year follow-up of helium ion therapy for uveal melanoma. *Am J Ophthalmol* 1998;**125**:81–9.
- 5 Foerster M, Bornfeld N, Wessing A, et al. Die Behandlung von malignen Melanomen der Uvea mit 106-Ruthenium-Applikatoren. *Klin Monatsbl Augenheilkd* 1986;**185**:490–5.
- 6 Mueller AJ, Talies S, Schaller UC, et al. Stereotactic radiosurgery of large uveal melanomas with the gamma-knife. *Ophthalmology* 2000;**107**:1381–138.

- 7 **Ossoinig KC**, Bigar F, Kaefring SL. Malignant melanoma of the choroid and the ciliary body. A differential diagnosis in clinical echography. *Ultrasonography in ophthalmology*. Basle: Karger, 1975;**83**:141–54.
- 8 **Ossoinig KC**, Seher K. Echographische Untersuchungen feingeweblicher Strukturen zur Klärung des Echoursprungs in Tumoren. I:Theoretische Überlegungen. *Graefes Arch Clin Exp Ophthalmol* 1966;**171**:17.
- 9 **Ossoinig KC**, Steiner H. Zum Problem der Normung in der Ultraschalldiagnostik des Auges. *Graefes Arch Clin Exp Ophthalmol* 1966;**169**:241.
- 10 **Ossoinig KC**. The first standardized system for echo-ophthalmography. In: Massin M, Poujol J, eds. *Diagnostica ultrasonica in ophthalmologia*. Paris: Centre National d'Ophthalmologie des Quinze-Vingts, 1973:131.
- 11 **Ossoinig KC**. Quantitative echography-the basis of tissue differentiation. *J Clin Ultrasound* 1974;**2**:33.
- 12 **Ossoinig KC**. Standardized echography:basic principles, clinical applications, and results. *Int Ophthalmol Clin* 1979;**19**:127–210.
- 13 **Stefani FH**, Hasenfratz G. *Macroscopic ocular pathology. An atlas including correlations with standardized echography*. Heidelberg: Springer, 1987:1.
- 14 **Till P**. Solid tissue model for the standardization of the echo-ophthalmograph 7200 MA (Kretztechnik). *Doc Ophthalmol* 1976;**41**:205.
- 15 **Sacca S**, Polizzi A, Macri A, et al. Echographic study of extraocular muscle thickness in children and adults. *Eye* 2000;**14**:765–9.
- 16 **Dick AD**, Nangia V, Alta H. Standardized echography in the differential diagnosis of extraocular muscle enlargement. *Eye* 1992;**6**:610–17.
- 17 **Pulido JS**, Byrne SF, Clarkson JG, et al. Evaluation of eyes with advanced stages of retinopathy of prematurity using standardized echography. *Ophthalmology* 1991;**98**:1099–104.
- 18 **Ossoinig KC**. The role of standardized ophthalmic echography in the management of Graves' ophthalmopathy. *Dev Ophthalmol* 1989;**20**:28–37.
- 19 **Doro D**. Optic neuropathies:diagnostic role of standardized echography. *Metab Pediatr Syst Ophthalmol* 1990;**13**:67–71.
- 20 **Mueller AJ**, Freeman W, Folberg R, et al. Die Muenchen/San Diego/Iowa City Collaboration (MuSIC):Design, Charakterisierung des Kollektives und erste Ergebnisse. MuSIC-Report I. *Ophthalmologie* 2002;**99**:193–9.
- 21 **Mueller AJ**, Schaller U, Freeman W, et al, for the Munich/San Diego/Iowa City Collaboration (MuSIC) Complex microcirculation patterns detected by confocal Indocyanine green angiography predict time to growth of small choroidal melanocytic tumors. MuSIC-Report II. *Ophthalmology* (in press).
- 22 **Portney LG**, Watkins MP. *Foundations of clinical research. Applications and practice*. Norwalk, CT: Appleton and Lange, 1993:509–16.
- 23 **Chang A**. Interclass correlation. <http://www.obg.cuhk.edu.hk/researchsupport/statement.asp>.
- 24 **Frazier-Byrne S**, Green, RL. *Ultrasound of the eye and the orbit*. Chapter 5. *Intraocular tumors*. St Louis: Mosby Year Book, 1992:133.
- 25 **Nicholson DH**, Frazier-Byrne S, Chiu MT, et al. Echographic and histologic tumor height measurements in uveal melanoma. *Am J Ophthalmol* 1985;**100**:454–7.
- 26 **Char DH**, Stone RD, Irvine AR, et al. Diagnostic modalities in chorioidal melanoma. *Am J Ophthalmol* 1980;**89**:223–30.
- 27 **Char DH**, Kroll S, Stone RD, et al. Ultrasonographic measurement of uveal melanoma:interobserver variability. *Br J Ophthalmol* 1990;**74**:183–5.
- 28 **Grasbon T**, Schriever S, Hoops JP, et al. 3D-Ultraschall. Erste Erfahrungen bei verschiedenen Augenerkrankungen. *Ophthalmologie* 2001;**98**:88–93.

Is your paper being cited?

CiteTrack service

CiteTrack will alert you by email whenever new content in the *British Journal of Ophthalmology* or a participating journal is published that matches criteria you want to track

Topics: Tell CiteTrack which words or subjects to watch for in new content

Authors: Be alerted whenever key authors you are following publish a new paper

Articles: Know whenever a paper of interest to you is referenced by another paper

www.bjophthalmol.com