

Evaluation of optical coherence tomography findings in age-related macular degeneration: a reproducibility study of two independent reading centres

Markus Ritter,¹ Julee Elledge,² Christian Simader,¹ Gabor G Deak,¹ Thomas Benesch,³ Barbara A Blodi,² Ursula M Schmidt-Erfurth¹

¹Department of Ophthalmology, Medical University of Vienna, Vienna, Austria

²Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Wisconsin, USA

³Department of Medical Statistics, Medical University of Vienna, Vienna, Austria

Correspondence to

Ursula M Schmidt-Erfurth, Department of Ophthalmology, Medical University of Vienna, Vienna, Austria; ursula.schmidt-erfurth@meduniwien.ac.at

Accepted 18 April 2010
Published Online First
30 August 2010

ABSTRACT

Background/aims To determine the reproducibility among readers of two independent certified centres, the Vienna Reading Center (VRC) and the University of Wisconsin-Madison Reading Center (UW-FPRC) for optical coherence tomography (OCT) images in age-related macular degeneration (AMD).

Methods Fast macular thickness scans and 6 mm cross hair scans were obtained from 100 eyes with all subtypes of AMD using Stratus OCT. Consensus readings were performed by two certified OCT readers of each reading center using their grading protocol. Common variables of both grading protocols, such as presence of cystoid spaces, subretinal fluid, vitreomacular traction and retinal pigment epithelial detachment, were compared using κ statistics. In addition, the intraclass correlation coefficient (ICC) was calculated for centre point thickness (CPT) of values re-measured manually in the presence of alignment errors.

Results The reproducibility was dependent on the variable measured with a κ value of 0.81 for the presence of cystoid spaces, 0.78 for the presence of subretinal fluid and 0.795 for the presence of vitreomacular traction. The lowest reproducibility was found for the presence of retinal pigment epithelial detachment with a κ value of 0.51. The CPT was re-measured in 29 out of 100 scans at both sites with an ICC of the re-measured thicknesses of 0.92.

Conclusion OCT scan data are crucial in monitoring treatment efficacy in AMD clinical trials. For comparison of results obtained by different reading centers, the inter-reading center reproducibility is essential. Although the reproducibility is generally high, the reliability depends on the selected morphological parameters.

INTRODUCTION

Optical coherence tomography (OCT) is a rapid, non-invasive imaging method capable of generating high-resolution optical cross-sections of the macula. A number of studies have elucidated the usefulness of OCT in the evaluation and follow-up of retinal conditions such as age-related macular degeneration (AMD).^{1,2}

OCT has also become a critical tool for assessing the morphological response of the retina to therapeutic interventions. OCT-derived measurements of the central retinal thickness are important outcome measures in clinical trials for treatment of macular oedema^{3–5} and choroidal neovascularisation.^{6,7} In addition to quantitative metrics, OCT is

widely used for qualitative assessment to establish the presence of cystoid spaces, pigment epithelium detachments (PEDs), subretinal fluid (SRF) or vitreomacular pathologies such as vitreomacular traction (VMT). Recently, some of these variables have been used to determine the need for initial treatment and subsequent re-treatment in clinical trials evaluating novel treatment modalities in the management of neovascular AMD.⁸ In current clinical trials of neovascular AMD, the most commonly used OCT system is the Stratus OCT (Carl Zeiss Meditec, Dublin, California, USA) with the fast macular thickness map (FMTM) scan mode. However, the automated algorithms frequently misidentify the inner and outer retinal boundaries. This leads to an inaccurate thickness measurement requiring a manual re-measurement. For the evaluation of the various morphological characteristics of exudative AMD, typically the cross hair line scans have been obtained.

In multicentre AMD clinical trials, OCT scans are being sent to a central reading center (RC) for evaluation of morphology and accuracy of automated thickness measurements. Certified readers interpret retinal morphology according to standardised protocols and the centre point thickness (CPT) is re-measured manually when the automated measurement is incorrect. Controlled RC procedures provide the most reliable method to obtain realistic and solid OCT data and to conclude on the value and benefit of OCT imaging in intravitreal treatment of AMD. Therefore, certified RCs play an important role for defining OCT strategies.

Although data have been published regarding reproducibility within OCT RCs,⁹ little information is available regarding reproducibility between RCs. However, Food and Drug Administration (FDA) approved trials are usually performed as two independent trials in US and non-US sites with individual RCs responsible for each trial. Also, a comparison of strategies and results from different centres allows a solid evaluation of reliability and reproducibility.

However, the comparability and reproducibility of OCT gradings of readers from independent RCs using different grading protocols have not been evaluated to our knowledge. Herein, we determine the inter-reading centre reproducibility for OCT image evaluation between an established European and an established US Reading centre, the Vienna Reading Center (VRC) and the Reading Center of the University of Wisconsin-Madison (UW-FPRC), respectively.



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://bj.o.bmj.com/site/about/unlocked.xhtml>

METHODS**Data collection**

In this retrospective study, cases were collected sequentially from a list of patients with the clinical diagnosis of AMD who underwent Stratus-OCT imaging of the macula at the retina unit of the Department of Ophthalmology of the Medical University of Vienna. Approval for the analysis of OCT images was obtained from the Institutional Review Board of the Medical University of Vienna and of the University of Wisconsin-Madison. The research adhered to the tenets set forth in the Declaration of Helsinki.

All scans were acquired by certified OCT operators following pupil dilatation using the FMTM scan mode for quantitative measures of the CPT and the 6 mm cross hair scan mode for qualitative assessment of retinal morphology.

The FMTM protocol acquires six 6 mm radial lines consisting of 128 A-scans per line in 1.92 s of scanning, whereas the cross hair protocol includes two 6 mm lines (6–12 to 9–3 o'clock) at 512 scan resolution.

Scans were analysed with the version 4.0 software (Carl Zeiss Meditec), all patient identifying data were removed and a consecutive patient number was given. Data for each case were sent as printouts to the UW-FPRC and as digital data files to the VRC. After completion of the grading procedure, the data files with the results were independently sent by the two RCs to the Department of Medical Statistics of the Medical University of Vienna for data analysis keeping each of the RCs blinded from the results of the other RC.

Grading procedure

The OCT image sets were examined as printouts at the UW-FPRC. Standard grid templates for identification of the centre point and the central mm, printed on a transparent stock, were overlaid on the scans and used for grading of location and extent of the variables.

A validated computer assisted grading software was used at the VRC. This software imports OCT scan data exported from the Stratus-OCT and allows the grader to electronically measure various parameters in the 6 mm cross hair and FMTM scans. The distances are calculated in pixels and using the dimensions of the B scan image, the pixels are converted into micrometers. A superimposed grid indicating the centre point and the central mm was used for grading of location and extent of parameters.

A consensus reading was performed by two certified readers at each RC. The readers were trained according to the individual protocol of the related RC and certification was awarded on the successful completion of all requirements. All readers are actively participating in the grading of Stratus OCT scans of clinical trials evaluating treatments for AMD.

The 6 mm cross hair scans were graded for presence and location of cysts, subretinal fluid (SRF), pigment epithelial detachment (PED) and vitreomacular traction (VMT). FMTM scans were graded for presence of alignment errors at the centre point. Scans with alignment errors of more than 25% of the retinal thickness at the centre point in at least one of the six FMTM scans and with a standard deviation of CPT at >10% were manually re-measured using a calibrated ruler at the UW-FPRC and digitally at the VRC.

Definition of morphological parameters

Before the start of the study, staff from the two RCs discussed the grading parameters but did not thereafter interact. Cysts were defined as round, minimally reflective spaces within the

neurosensory retina. SRF was identified as a non-reflective space between the posterior boundary of the neurosensory retina and the retinal pigment epithelium/choriocapillaris reflection. Localisation of cysts and SRF was categorised: (1) being outside the central mm and (2) involving the central mm. In addition, central cysts were categorised as small (<200 µm), medium (<400 µm) or large (>400 µm).

PED was described as a focal elevation of the reflective retinal pigment epithelium (RPE) band over an optically clear or moderately reflective space. A PED was graded if the elevation was greater than 400 µm at the base and/or greater than 200 µm from the surface of the RPE band to the surface of the choriocapillaris. Location of PED was described as (1) being outside the central mm and (2) involving the central mm.

VMT was identified if a thickened posterior hyaloid was observed to insert in the foveal/perifoveal area. No grading regarding extent of VMT was performed.

Both RCs followed their own grading protocols for the selected parameters. To compare gradings between the two RCs, a comparison key was developed.

Statistical methods

Analyses were conducted using the software package SPSS V.17.0. To quantify the reproducibility of manual re-measurements of the CPT in cases of alignment errors, the intraclass correlation coefficient (ICC) was calculated. In addition, the mean, SD, standard error, Pearson ρ and Bias correction factor C_b was determined. For CPT re-measurements, absence/presence, location and extent of cysts, SRF, PED and VMT, the percentage of agreement and the κ statistic were calculated. κ Statistics were interpreted using the ranges suggested by Landis and Koch¹⁰: 0–0.20, slight agreement; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and more than 0.80, almost perfect agreement.

RESULTS

The study included 100 eyes of 89 patients with an average age (SD) of 67±8 y (59 female, 41 male). The OCT scans showed a wide range of morphological changes associated with AMD, including drusenoid changes, serous, haemorrhagic and fibrovascular PED, SRF, intraretinal cysts, VMT or macular oedema.

Manually re-measured CPT

The CPT was re-measured in 29 out of 100 FMTM scans at both sites (in 35 scans at the VRC and in 37 scans at the UW-FPRC). Table 1 shows the reproducibility as to whether or not a re-measurement was necessary and shows the mean, minimum and maximum CPT, SD, SE, ICC, Pearson ρ and Bias correction factor C_b for the re-measured CPTs at both sites.

Figure 1A shows the Bland-Altman plot of the differences in the re-measured CPTs (x axis: the average of the CPT values of the two RCs; y axis: the differences of the CPT measurements between the two RCs). Figure 1B shows the scatter plot of the same measurements.

Presence of cysts, SRF, PED and VMT

Results of gradings for presence of cysts, PED, SRF and vitreomacular traction are summarised in table 2. Bold values indicate consistent gradings. The highest reproducibility was found for presence of cysts with a κ value of 0.81 (almost perfect) and the lowest for PED with a κ value of 0.51 (moderate).

The number of detected VMT was low with only two cases identified at both sites and one additional at the UW-FPRC.

Table 1 Scans requiring re-measurements and re-measured values

UW-FPRC	VRC		Total
	No	Yes	
No	57	6	63
Yes	8	29	37
Total number	65	35	100
κ	0.70		
SE	0.07		
95% CI	0.55 to 0.84		

Manually re-measured centre point thickness	
No of re-measured scans at both sites	29
Intraclass correlation coefficient	0.92
95% CI	0.845 to 0.96
Pearson ρ (precision)	0.94
Bias correction factor C_b (accuracy)	0.9824

	VRC	UW-FPRC
Mean (μm)	298	276
SE	21.3	22.5
SD	114.9	121.2
Minimum	139	115
Maximum	501	544

Values in bold indicate consistent grading.

FMTM, fast macular thickness map; UW-FPRC, University of Wisconsin-Madison Reading Center; VRC, Vienna Reading Center.

Location and extend of cysts, PED and SRF

Results of gradings for location and extent of cysts, PED and SRF are summarised in tables 3 and 4. Table 3 shows the reproducibility of thickness gradings of cysts at the centre point. Bold values indicate consistent gradings.

DISCUSSION

The introduction of a novel diagnostic method such as OCT of the retina into clinical practice, and the acceptance/approval of used parameters by the responsible institution, requires solid proof of the value and relevance of the acquired data. In order to assess treatment efficacy and to evaluate treatment indications, it is important to gather robust data sets in a controlled manner. Few prospective clinical trials have included OCT measurements in a rigorous way using a standardised protocol. Due to the limited availability of solid data, the technique, although widely used in clinical routine, has not yet obtained general acceptance as a gold standard method for monitoring patients with AMD, by institutions and the community. The data available from trials such as PRONTO (Prospective Optical Coherence Tomography Imaging of Patients with Neovascular Age-Related Macular Degeneration Treated with intraocular Ranibizumab),¹¹ EXCITE (Efficacy and Safety of Ranibizumab in Patients

With Subfoveal Choroidal Neovascularization Secondary to Age-Related Macular Degeneration)¹² and SUSTAIN (Study of Ranibizumab in Patients With Subfoveal Choroidal Neovascularization Secondary to Age-Related Macular Degeneration)¹³ are uniformly based on conventional time domain (TD) OCT using the Stratus-OCT.

Standardised assessment of morphological changes in OCT scans by one central RC is becoming a widely used method of determining outcomes in large clinical trials. Benefits of using a central RC include standardised data acquisition procedures at the clinical sites, standardised grading procedures by the use of certified readers and masking of readers to patient data regarding treatment assignment. An additional advantage is that the reproducibility of the readings can be easily assessed.⁹ Measurement of retinal thickness and grading for presence, location and extent of intraretinal cysts, SRF and PED are currently defined as the primary morphological outcomes in recent clinical trials evaluating novel treatments for neovascular AMD.¹¹ Although these morphological parameters have been similarly applied by several RCs, there are no studies comparing OCT gradings between them despite the fact that RCs often cooperate in clinical trials. RCs developed individual grading protocols, software tools and measuring devices. There is only one study comparing results from two fundus photograph RCs in grading cytomegalovirus retinitis progression.¹⁴ Gradings were performed on light boxes using standard grid templates and measuring circles by an ophthalmologist grader at one RC and by a group of four certified graders at the other RC. Another study investigated agreement between clinician and RC gradings of fundus photographs of diabetic retinopathy.¹⁵ The agreement between diabetic retinopathy severity classification as determined by ophthalmoscopy performed by retina specialists and gradings of stereoscopic fundus photographs at a RC was determined. However, to our knowledge no study has been published on the inter-reading centre reproducibility of OCT gradings.

As more competing trials are being performed worldwide with increasing numbers of study participants, more RCs are being involved to satisfy the need for certified data analysis.

The comparison of consensus readings of different OCT parameters in AMD showed a high level of agreement between the VRC and the UW-FPRC. However, the reproducibility was dependent on the specific morphological parameters measured with a κ value of 0.81 for the presence of cystoid spaces, 0.78 for the presence of SRF and 0.795 for the presence of VMT. The lowest reproducibility was found for the presence of PED with a κ value of 0.51. Reproducibility of gradings for localisation and extent of the selected parameters was lower, but still 'moderate' with κ values of 0.54 for SRF and 0.51 for PED and 'substantial' for intraretinal cysts with a κ value of 0.78. These findings, based on the capability

Figure 1 (A) Blant-Altman plot comparing mean and difference of manually re-measured centre point thickness (CPT) values between the two reading centres. (B.) Scatterplot of manually re-measured CPT values from the two reading centres. UW-FPRC, University of Wisconsin-Madison Reading center; VRC, Vienna Reading Center.

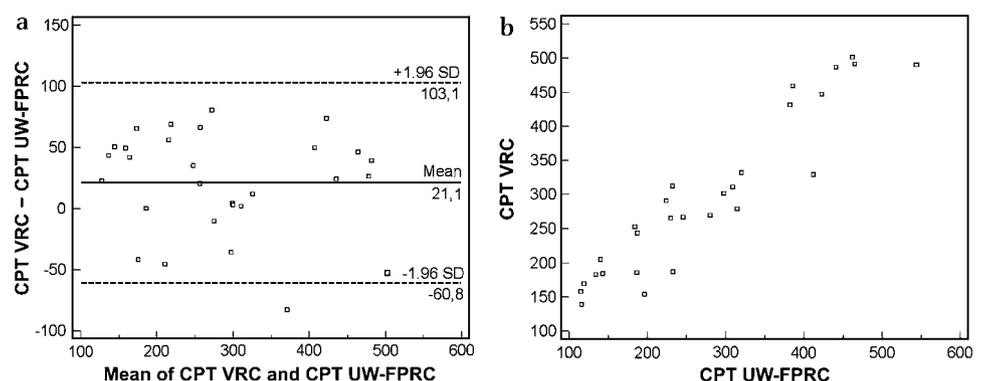


Table 2 Presence of cystoid spaces, SRF, PED and vitreomacular traction

UW-FPRC	VRC					
	Cystoid spaces			SRF		
	No	Yes	Total	No	Yes	Total
No	65	8	73	52	8	60
Yes	0	27	27	3	37	40
Total no	65	35	100	55	45	100
κ	0.81			0.78		
SE	0.06			0.06		
95% CI	0.69 to 0.94			0.65 to 0.90		

UW-FPRC	VRC					
	PED			Vitreomacular traction		
	No	Yes	Total	No	Yes	Total
No	32	26	58	97	0	97
Yes	0	42	42	1	2	3
Total no	32	68	100	98	2	100
κ	0.51			0.795		
SE	0.07			0.20		
95% CI	0.37 to 0.65			0.40 to 1.0		

Values in bold indicate consistent gradings. FMTM, fast macular thickness map; PED, pigment epithelium detachments; SRF, subretinal fluid; UW-FPRC, University of Wisconsin-Madison Reading Center; VRC, Vienna Reading Center.

of trained analysts in a standardised setting also highlight the difficulties in identifying specific morphologic pathologies in general—for example, lesions of the RPE clearly represent a challenge for diagnosis and grading, while the identification of intraretinal cysts appears to be more reliable. These observations from a certified centre condition allow an estimation of the value of such data from studies not using certified procedures.

These data appear even more relevant as a digital grading method at the VRC was compared to a paper print grading method at the UW-FPRC. In general, all reading parameters were discussed by members of the two RCs and defined prior to the

Table 3 Location, extent and thickness at the centre point of cystoid spaces

UW-FPRC	VRC			
	0	1	2	Total
0	65	0	0	73%
1	2	3	0	5%
2	6	2	22	30%
Total	73%	5%	12%	100%
0—Absent				κ 0.78
1—Definite, outside central 1 mm				SE 0.07
2—Definite, central 1 mm involved				95% CI 0.53 to 0.82

UW-FPRC	VRC				
	0	1	2	3	Total
0	83	2	1	0	86%
1	0	7	3	0	10%
2	0	0	2	1	3%
3	0	0	0	1	1%
Total	83%	9%	6%	2%	100%
0—Not applicable					κ 0.75
1—Small, <200 μ m					SE 0.09
2—Medium, <400 μ m					95% CI 0.56 to 0.93
3—Large, >400 μ m					

Values in bold indicate consistent gradings. UW-FPRC, University of Wisconsin-Madison Reading Center; VRC, Vienna Reading Center.

Table 4 Location and extent of subretinal fluid and pigment epithelial detachment

UW-FPRC	VRC				Total
	0	1	2	3	
0	52	5	7	1	65%
1	2	6	2	0	10%
2	5	3	17	0	25%
3	1	0	0	0	0%
Total	59%	14%	26%	1%	100%
0—Absent					κ 0.54
1—Definite, outside central 1 mm					SE 0.08
2—Definite, central 1 mm involved					95% CI 0.37 to 0.69
3—Cannot grade					

UW-FPRC	VRC				Total
	0	1	2	3	
0	32	2	23	0	57%
1	0	4	2	0	6%
2	0	1	35	0	36%
3	0	0	1	0	1%
Total	32%	7%	61%	0%	100%
0—Absent					κ 0.51
1—Definite, outside central mm					SE 0.08
2—Definite, central mm involved					95% CI 0.36 to 0.66
3—Cannot grade					

Values in bold indicate consistent gradings. UW-FPRC, University of Wisconsin-Madison Reading Center; VRC, Vienna Reading Center.

study but there was no interaction between the two RCs thereafter offering completely independent analysis. Limitations of this analysis include the restricted range of morphological parameters included and the relatively small number of study cases. In addition, this study is based solely on Stratus OCT data, which is still the most commonly used OCT in research protocols to date; however, it will be probably replaced soon by Fourier domain OCT (FD-OCT) devices. Stratus OCT imaging is based on six radial, cross-sectional scans and the information, therefore, is limited to a few randomly selected locations and an overall low resolution of structural details. The FD-OCT uses a fast spectral-domain technique and performs scans in a raster pattern throughout the entire macular area at a superior resolution of 5 μ m in axial and 20 μ m in transverse directions. As a result, the retinal morphological features such as RPE or vitreomacular interface abnormalities can be imaged at all locations and in more details.

In AMD, there may be up to a 92% rate of inner or outer retinal boundary errors with subsequent incorrect segmentation for retinal thickness.¹⁶ Recently, studies have shown the superiority of the central subfield for analysis of the central retinal thickness when compared to the centre point thickness because of reduced variability.¹⁷ Although the majority of the FD-OCT platforms will allow recalculation of central subfield thickness measurements, only the centre point thickness can be re-measured manually on Stratus OCT when the automated algorithm fails. The CPT measurements in case of alignment errors were highly reproducible between the two RCs with a κ value of 0.70 and an ICC of 0.92. However the mean CPT re-measured by readers of the VRC was 22 μ m higher than the mean of the re-measured CPT at the UW-FPRC. The agreement on thickness gradings of central cysts was 'substantial' with a κ value of 0.75. This high degree of measurement agreement is

relevant for clinical trials in which lesion thickness is a primary or secondary study endpoint.

A constant high level of agreement between RCs probably requires the development of a quality control programme and RCs may need to consider establishing and maintaining cooperative programs to harmonise their measuring and grading procedures and settings. Otherwise, a comparison of results from trials analysed by different RCs cannot be certain of the impact of possible differences between RCs. However, optimal standardisation and reliability is essential to identify the value of new surrogate parameters in the development of novel treatment strategies, which also introduce appropriate strategies for patient management and follow-up.

Competing interests None.

Ethics approval This study was conducted with the approval of the Medical University of Vienna, Austria, and Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Wisconsin, USA.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Puliafito CA**, Hee MR, Lin CP, *et al.* Imaging of macular diseases with optical coherence tomography. *Ophthalmology* 1995;**102**:217–29.
2. **Baumal CR**. Clinical applications of optical coherence tomography. *Curr Opin Ophthalmol* 1999;**10**:182–8.
3. **Rivellese M**, George A, Sulkes D, *et al.* Optical coherence tomography after laser photocoagulation for clinically significant macular edema. *Ophthalmic Surg Lasers* 2000;**31**:192–7.
4. **Panozzo G**, Gusson E, Parolini B, *et al.* Role of OCT in the diagnosis and follow up of diabetic macular edema. *Semin Ophthalmol* 2003;**18**:74–81.
5. **Massin P**, Duguid G, Erginay A, *et al.* Optical coherence tomography for evaluating diabetic macular edema before and after vitrectomy. *Am J Ophthalmol* 2003;**135**:169–77.
6. **Hee MR**, Baumal CR, Puliafito CA, *et al.* Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology* 1996;**103**:1260–70.
7. **Rogers AH**, Martidis A, Greenberg PB, *et al.* Optical coherence tomography findings following photodynamic therapy of choroidal neovascularization. *Am J Ophthalmol* 2002;**134**:566–76.
8. **Schmidt-Erfurth UM**, Richard G, Augustin A, *et al.* European Society for Retina Specialists' Guidelines Committee (EURETINA). Guidance for the treatment of neovascular age-related macular degeneration. *Acta Ophthalmol Scand* 2007;**85**:486–94.
9. **Zhang N**, Hoffmeyer GC, Young ES, *et al.* Optical coherence tomography reader agreement in neovascular age-related macular degeneration. *Am J Ophthalmol* 2007;**144**:37–44.
10. **Landis RJ**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;159–74.
11. **Fung AE**, Lalwani GA, Rosenfeld PJ, *et al.* An optical coherence tomography-guided, variable-dosing regimen with intravitreal ranibizumab (Lucentis) for neovascular age-related macular degeneration. *Am J Ophthalmol* 2007;**143**:566–83.
12. **Schlingemann RO**, Schmidt-Erfurth U, Eldem B. and on behalf of the EXCITE study group. Safety and efficacy of quarterly vs. monthly ranibizumab injections in patients with neovascular age-related macular degeneration: 12-months results of the EXCITE study. *Invest Ophthalmol Vis Sci* 2009;**50**:E-Abstract 2382.
13. **Holz FG**, Meyer C, Eter N. and on behalf of the SUSTAIN study group. Safety and efficacy of ranibizumab treatment in patients with neovascular age-related macular degeneration: 12-month results of the SUSTAIN study. *Invest Ophthalmol Vis Sci* 2009;**50**:E-Abstract 3095.
14. **Hubbard LD**, Ricks MO, Martin BK. Comparability of two fundus photograph reading centers in grading cytomegalovirus retinitis progression. *Am J Ophthalmol* 2004;**137**:426–34.
15. **Scott IU**, Bressler NM, Bressler SB, *et al.* Agreement between clinician and reading center gradings of diabetic retinopathy severity level at baseline in a phase 2 study of intravitreal bevacizumab for diabetic macular edema. *Retina* 2008;**28**:36–40.
16. **Domalpally A**, Danis RP, Zhang B, *et al.* Quality issues in interpretation of optical coherence tomograms in macular diseases. *Retina* 2009;**29**:775–81.
17. **Diabetic Retinopathy Clinical Research Network**, Krzystolik MG, Strauber SF, Aiello LP, *et al.* Reproducibility of macular thickness and volume using Zeiss optical coherence tomography in patients with diabetic macular edema. *Ophthalmology* 2007;**114**:1520–5.