

Schippling, S; Balk, L; Costelo, F; Albrecht, P; Balcer, L; Calabresi, P; Fredriksen, J; (2015) Quality control for retinal OT in multiple sclerosis validation of the OSCAR-IB criteria. **Multiple Sclerosis Journal**, 21 (2) pp. 163-170. 10.1177/1352458514538110. Downloaded from UCL Discovery: <http://discovery.ucl.ac.uk/1433670>

ARTICLE

Quality control for retinal OCT in Multiple Sclerosis: Validation of the OSCAR-IB criteria

Schippling S, MD^{1,}, Balk LJ, MD^{2,*}, Costello F, MD, FRCP^{3,*}, Albrecht P, MD⁴, Balcer L, MD, MSCE⁵, Calabresi PA, MD⁶, Frederiksen JL, MD DMSci⁷, Frohman E, MD, PhD⁸, Green AJ, MD⁹, Klistorner A, MD, PhD¹⁰, Outteryck O, MD¹¹, Paul F, MD,¹² Plant GT, MD, FRCP, FRCOph¹³, Traber G, MD¹⁴, Vermersch P, MD¹¹, Villoslada P, MD¹⁵, Wolf S, MD¹⁶, Petzold A, MD, PhD^{2*&}*

¹ Neuroimmunology and Multiple Sclerosis Research Section, Department of Neurology, University Hospital Zurich, Zurich, Switzerland; ² MS Centre Amsterdam, VU University Medical Centre, Department of Neurology, Amsterdam, The Netherlands; ³ University of Calgary, Departments of Clinical Neurosciences and Surgery, Calgary, Alberta, Canada; ⁴ Department of Neurology, Medical Faculty, Heinrich-Heine University Düsseldorf, Germany; ⁵ Department of Ophthalmology, New York, USA; ⁶ Johns Hopkins Hospital, Department of Neurology, Baltimore, MD, USA;

⁷ Department of Neurology, University of Copenhagen, Denmark; ⁸ Departments of Neurology and Ophthalmology, University of Texas Southwestern Medical Center at Dallas, Dallas, USA;

⁹ Multiple Sclerosis Center, Department of Neurology, University of California San Francisco, San Francisco, CA, USA; ¹⁰ Department of Ophthalmology, Save Sight Institute, University of Sydney, P.O. Box 4337, Sydney, NSW 2001, Australia; ¹¹ University of Lille Nord de France, Department of Neurology, Lille, France; ¹² NeuroCure Clinical Research Center (NCRC) and Clinical and Experimental Multiple Sclerosis Research Center, Department of Neurology, Charite, Universitätsmedizin Berlin, Berlin, Germany; ¹³ Moorfields Eye Hospital, The National Hospital for Neurology and Neurosurgery & St. Thomas' Hospital, London, United Kingdom; ¹⁴ Department of Ophthalmology, University Hospital Zurich, Zurich, Switzerland; ¹⁵ Center for Neuroimmunology, Institute of Biomedical Research August Pi Sunyer (IDIBAPS), Hospital Clinic of Barcelona, Spain;

¹⁶ Department of Ophthalmology, Bern University Hospital and University of Bern, Switzerland

* These authors contributed equally to the manuscript

Abstract

Background: Retinal optical coherence tomography (OCT) permits quantification of retinal layer atrophy relevant to assessment of neurodegeneration in multiple sclerosis (MS). Measurement artefacts may limit the use of OCT to MS research.

Objective: An expert task force convened with the aim to provide guidance on the use of validated quality control (QC) criteria for the use of OCT in MS research and clinical trials.

Methods: A prospective multi-centre (n=13) study. Peripapillary ring scan QC rating of a OCT training set (n=50) was followed by a test set (n=50). Inter-rater agreement was calculated using kappa statistics. Results were discussed at a round table after the assessment had taken place.

Results: The inter-rater QC agreement was substantial (kappa=0.7). Disagreement was found highest for judging signal strength (kappa=0.40). Future steps to resolve these issues were discussed.

Conclusion: Substantial agreement for QC assessment was achieved with aid of the OSCAR-IB criteria. The task force has developed a website for free online training and QC certification. The criteria may prove useful for future research and trials in MS using OCT as a secondary outcome measure in a multi-centre setting.

Introduction

Retinal OCT is increasingly used in multiple sclerosis (MS) research as a tool to quantify changes of retinal layer thickness [1]. Following MS associated optic neuritis (MSON) there may be severe atrophy of the peripapillary retinal nerve fibre layer (pRNFL). Importantly, a small degree of pRNFL atrophy can also be observed in patients with MS who never experienced a clinical episode of MSON [1]. In addition to these cross-sectional data, longitudinal data from time-domain OCT suggests that the annual loss of pRNFL in MS accounts for as little as 1-2 μm compared to about 0.1 μm among healthy individuals [2]. This degree of pRNFL atrophy is within the measurement range of recently introduced spectral-domain OCT devices from a number of manufactures. Therefore longitudinal spectral-domain OCT assessments of the retina in patients with MS may emerge as a surrogate for neurodegeneration, even in the absence of MSON. In this context it is important to recognise that the magnitude of localised measurement artifacts introduced by a poor OCT scanning technique can be significant, measuring up to 40 μm when caused by off-centre placement of the measurement beam [3]; over 3.4 μm when due to displacement of the peripapillary ring-scan [outside an ellipse of 200 μm (horizontal shift) and 600 μm (vertical shift)]; [4] and up to 10 μm when arising from poor signal quality alone [5]. Therefore, OCT scans containing such measurement artifacts should be excluded from clinical studies or trials as they may obscure the detection and interpretation of the much smaller degree of retinal layer atrophy expected from longitudinal studies [2].

In addition to these quantitative considerations there are qualitative reasons to reject peripapillary OCT scans which may contain inherent inaccuracies. These comprise boundary line errors or algorithm failures and poorly centered ring scans [6]. Poor illumination and obvious protocol violations are other reasons for OCT scans to be deemed poor in quality. In a previous study we developed a set of quality control (QC) criteria which address all of these issues [7]. In a multi-centre validation approach we found a substantial inter-rater agreement (kappa 0.61) for the seven criteria that we named the OSCAR-IB QC criteria [7].

The present study aimed to test the practicability and reliability of the OSCAR-IB criteria in a world-wide multi-centre setting. To this end we created a training- (n=50) and a test set (n=50) of retinal OCTs from healthy subjects and patients with MS. Trained OCT readers from 13 centres participated. The group of participating experts from MS clinics, ophthalmology, and statistics convened in Lyon, France, to discuss the data and make recommendations for future OCT QC reading in MS research.

Methods

Retinal images were obtained using a Spectral Domain (SD)-OCT device (Heidelberg Spectralis, Software version 1.1.6.3) with the eye tracking function (EBF) enabled. All OCT scans were recorded from MS patients and healthy subjects who gave written informed consent for participation at the MS Centre in Amsterdam. In all subjects a peripapillary ring scan (diameter 12°) around the optic nerve head (ONH) was recorded. All scans were anonymised, randomised and split into a training (n=50) and test (n=50) set.

The training set was rated individually by two trained OCT readers (LB, AP) with an inter-rater kappa of 1.0. A slide presentation was prepared which first showed the scans and then the dichotomised QC results: accept or reject. In case of rejection the reason for rejection was detailed. This could be one single out of the seven OSCAR-IB criteria [7] or, more frequently, a combination of several criteria (e.g. a poor signal strength can also lead to an algorithm failure). The training set was sent to each of the participating OCT readers.

Next, another slide presentation was prepared, containing the test set. This time only the OCT scan was shown without giving any more information on QC rating. For this new technology expert centres were defined as those who have published in the field (see references in [1]) and were either of ophthalmological or neurological background. First and corresponding authors of these papers were approached either by email or personal contact. In addition, contact was made with established OCT reading centres. The raters from those 13 centres who had agreed to participate were asked to indicate if a scan was accepted or rejected in a spreadsheet document. In case of rejection the raters were also asked to fill in which one(s) of the OSCAR-IB criteria was (were) missed. This information was then sent back to the MS Centre in Amsterdam for statistical evaluation.

Data analysis

Kappa statistics for multiple raters were calculated to assess the inter-centre agreement using the magree macro in SAS software (V9.3) [8]. The level of agreement was rated as slight (0-0.2), fair (0.2-0.4), moderate (0.4-0.6), substantial (0.6-0.8) or almost perfect (0.8-1) [9]. Results were presented at a joint meeting of the participating centres at the 2012 ECTRIMS meeting in Lyon, France. Sources of disagreement were identified and discussed.

Results

A total of 13 centres participated in the study (Table 2). Most centres (n=9) provided data for a single rater. Data from multiple raters were provided by 4 centres (Table 2).

Equipped with the OSCAR-IB criteria (Table 1), all raters participated in a training set (n=50) prior to performing their own rating (n=50) subject to statistical analysis (Figure 1).

Two raters had to be excluded because the order of scans rated got irretrievably mixed up. The inter-rater level of agreement for the overall accept-reject classification was substantial (kappa 0.7, $p < 0.0001$).

The average rejection rate was 55% of scans (range 48-74%; see Table 2).

Raters had the right to accept-reject based on one or more of the OSCAR-IB criteria. In order to evaluate the relative utility of individual features individual inter-centre kappas were calculated. The individual kappa values were lower compared to the overall accept-reject kappa (Table 3). The data demonstrates that a scan poorly centered at the optic disc was a clear reject for most (highest kappa 0.579). A comparable level of agreement was found for the R- (kappa 0.543), I- (kappa 0.565) and B-criteria (kappa 0.549). Disagreement between raters was largest for the S- (kappa 0.404), O- (kappa 0.434) and A-criteria (kappa 0.482).

The consensus discussion at the round table meeting included scans like the one exemplified in Figure 2. This scan of a patient with MSON showing severe thinning of the pRNFL was rejected by a number of readers for the I-criterion. The shadow seen on the bottom right of the infrared fundus image was cast by a floater. However, the illumination of the scan is acceptable, the contrast of the pRNFL in the OCT B-scan is good and there is no algorithm failure. Such an image should be accepted because the quantitative data is reliable. In contrast, Figure 3 shows a scan with truly poor illumination that should be rejected based on the I-criterion.

Another point of discussion related to the B-criterion that was not intuitively clear to each reader. The two scans in the supplementary Figure illustrate the problem of an off-centered beam-placement in different directions for a baseline and follow-up scan in the same subject. It can be readily seen that in the first scan the signal intensity is lowest in the ONL/OPL inferonasally, highest at the opposite part of the ring scan (located clockwise to sectors 7 o'clock and 1 o'clock). The localised measurement error is illustrated by the red/green shaded areas in the summary image. Such a scan should thus be rejected.

Discussion

OCT is a promising new imaging technique that allows rapid and reliable quantification of retinal structures [1; 2]. To date, pRNFL thickness has been the OCT measurement most commonly used in tracking longitudinal changes in optic neuritis (ON) and MS patients. A recent meta-analysis of time domain OCT studies (14 studies [2,063 eyes]) demonstrated that RNFL values are reduced between 5 to 40 μm in MSON eyes [1]. Based upon the wealth of information derived from the OCT studies to date, pRNFL thickness changes may open a real window of opportunity for patient friendly assessment of neuro-axonal degeneration in ON and MS [1,2,13,14].

As with each new technique entering the clinical arena, quality control becomes an issue. In absence of any consensus guidelines on QC rating of retinal OCT scans we recently developed and validated a set of seven criteria (OSCAR-IB) addressing major sources of artifacts causing inaccurate measurements. The intent was to guarantee high quality image acquisition and interpretation for multicenter OCT studies in which reading was to be done at a central site. The aim of the current study was to build on this experience and extend the validation of the OSCAR-IB criteria to a multi-reading centre setting. Twenty independent raters from thirteen international OCT expert centres participated in the trial. Results were discussed during a round table meeting in Lyon in order to provide further guidance for the use of the OSCAR-IB criteria in clinical trial settings.

In a set of test scans consisting of 50 optic nerve head ring scans (12°), the OSCAR-IB criteria proved to work better than in the pilot study with a substantial level of agreement reaching an inter-centre kappa of 0.7 with an average rejection rate of 55% of scans (range 48-74%). This seeming high rejection rate is due to the fact that the dataset was put together with the intention to have a balanced number of good versus poor scans. It needs to be highlighted that none of the readers was aware that the intended rejection rate was 48%. In line with our previous experience, OCT scans were rejected for several reasons.

For the first of the OSCAR-IB criteria, “Obvious” [O], scans were rejected for highly apparent issues including severe lens opacities or vitreous hemorrhage. Likewise, rejection also occurred based on a protocol violation such as deliberately not averaging several B-scans by turning off the ART mode. As the rejection rate based on this criterion was overall low our expert panel expressed no new concerns regarding the use of this criterion.

For the second trial criterion, “Signal strength” [S], we arbitrarily defined a cut off of 15 dB in the pilot study [7] based on our experience and published evidence suggesting that poor signal strength decreases the signal-to-noise ratio of RNFL measures [10-12]. However, scans with poor signal strength (defined as < 15 dB) but relatively good contrast between layers allowing for proper image post-processing and segmentation do exist. This might also be reflected by the fact that the S-criterion had the lowest inter-rater agreement in our study (inter-centre kappa 0.404). In the absence of any FDA approved retinal layer segmentation algorithm and corresponding normal control databases it is currently not possible to systematically test the relative contribution of signal strength and number of averaged A-scans to the final B-scan image quality. The contrast level between the retinal layer interfaces is likely to become a major criterion to judge the quality of OCT scans in ON and MS patients. In the future, this warrants further systematic studies. We opted to leave the cut off of 15 dB unchanged until such data is being provided.

The third criterion, a correctly “Centered” ring scan [C] remains an important issue [4]. More recent scanning algorithms from the Heidelberg Spectralis device are centering the ring scans after defining the borders of the Bruch’s membrane opening (new Nsite protocol; e.g. used in the PASSOS trial;

NCT01705236). Preliminary experience with new protocols suggests that a perfectly centered ring-scan may not be required. The C-criterion is therefore not fully applicable to scans obtained using the Nsite protocol. However, as a change of position (vertical or horizontal) at follow-up scans may introduce relevant measurement artifacts [4] it is crucial to set the baseline scans as reference and use the device's follow-up setting to assure consistency of scan placement during longitudinal investigations.

Admittedly, "Algorithm" failures [A] remain a problem. In selected cases it may be possible to correct such errors by image post-processing techniques such as manual correction. However, this procedure is time-consuming when analyzing large collectives. A decision may be made on individual scans. Whilst this criterion should be handled rather strictly on the baseline scan in order to ensure optimal settings for longitudinal studies, we accept that a minor, manually correctable algorithm failure on a valuable follow-up scan should not necessarily lead to immediate rejection.

As a fifth criterion, "Retinal pathology" [R] unrelated to MS can independently influence RNFL measures and should thus be taken into account. Previously, we listed a number of pathologies that should be considered when it comes to OCT scan interpretation [7] (supplementary Table). Such a list will naturally grow with further experience generated by the use of OCT in different disease models as with the improvement of the technology per se. Of note, a number of recent OCT studies have provided evidence that retinal pathology other than RNFL thinning exists in MS [13-16], including microcystic macular oedema (MMO) predominantly affecting the inner nuclear layer (INL) of the retina. In the studies to date, MMO (and INL thickening) appear to be associated with higher degrees of MS-related disability [14, 17-18]. Hence, we do not recommend rejecting OCT scans showing evidence of MMO because this may be a clinically relevant finding in the disease and the impact of MMO on RNFL quantification is low.

Poor "Illumination" of scans [I] raised some discussion between experts. Scans displaying shadow casts on the retina by large vitreous floaters were rejected by some raters based on the I-criterion whereas these scans were accepted by others. Although floaters were identified as potential cause of shadowing, they were not listed as a general exclusion criterion in the original OSCAR-IB manuscript. In most cases the contrast of retinal layers underneath the shadow cast was expected to remain of sufficient quality to enable accurate retinal layer segmentation by automated algorithms.

In contrast, poor illumination by a sub-optimal placement of the laser beam causes partial illumination of the retina with just about acceptable B-scan quality in some areas and very dark B-scans from poorly illuminated areas. Typically, the poorly illuminated areas are located at the border of the infrared image. Clearly, in poorly illuminated areas the signal may drop to such low levels that no image post-processing can be performed. In summary, causes for shadows such as floaters, cataracts, long eye lashes, etc. should not be considered a general reason for exclusion by the I-criterion in cases where A- and B-scan quality is good.

The relevance of the "B" (beam placement) criterion was not intuitively obvious to all raters. It is important to keep the OCT scan in the live window horizontally orientated whenever possible. Problems arise when a scan is tilted in one direction at baseline and to the opposite direction at follow-up, because, as a consequence, measurement errors in the range of 9-40 μm can be introduced [3]. Scans with sub-optimal beam placement as described, should be rejected. In some cases the anatomy of the retina or refraction anomalies such as astigmatism may make it difficult to achieve a perfectly horizontally aligned OCT. In this context, the QC rating of the follow up scan will ensure that the same direction of tilting is repeated, thus ensuring consistency. The principle of the

B-criterion is that the direction of tilt causes a characteristic, reproducible, robust signal pattern in the outer plexiform- (OPL) and outer nuclear layers (ONL). This sign allows tracing back possible errors due to off-centre beam placement. Notably, a post-hoc analysis of the present data excluding all cases that were rejected based on the B-criterion resulted in a comparable kappa as the original analysis.

A limitation of the OSCAR-IB criteria is that they are entirely based on the pRNFL measures from the Spectralis device (Heidelberg Engineering). There is a need to test the criteria also on other machines on the market. As segmentation algorithms will certainly be used in multi-centre OCT-studies in MS in the near future as soon as the necessary post-processing software becomes more broadly available, we cannot be sure about limitations to such measures' quality that in turn impact on the inter-rater agreement of segmentation analyses. While data on inter-rater- and agreement between different segmentation algorithms has been provided [20], sources of disagreement due to poor quality have not yet been identified. As for peripapillary ring scans, a sharp and good contrast between layers is likely to be the core pre-requisite for the segmentation of macular volume scans as well. This is covered by the O-, S-, R-, I- and B-criteria of the OSCAR-IB criteria. For a reading center working with a large number of sites and photographers the criteria may be rather strict. In our experience the initial rejection rate may be as high as 43% [7]. It is therefore recommended that OCT technicians are specifically trained to comply with the OSCAR-OB criteria, and pass a certification procedure before starting to include patients into clinical trials. In order to evaluate the influence of the QC criteria for the "real" world one would need to test their impact by power calculations, ideally from ongoing studies. Probably some pragmatism will be needed to permit for reasonable sample size estimates in a "real" world multi-centre setting.

In summary, the OSCAR-IB QC have shown substantial inter-rater agreement in hands of experienced OCT raters in this multi-centre validation study. Experts participating in the meeting agreed that making the training set more broadly available would be helpful in order to guarantee rigorous QC in future clinical OCT trials in MS and other diseases. Rigorous QC will be particularly important for longitudinal clinical studies where only a small degree of retinal layer atrophy can be expected. The training and test set of this manuscript can be accessed for free on www.oscar-ib.org. Likewise, revisions to the criteria might become necessary as the use of OCT technology grows and scanning protocols continue to evolve. We would like to reach out and invite other centres in the field to join forces.

Conflicts of interests

The authors declare no conflicts of interest related to this study.

Disclosures

Philipp Albrecht received research grants from Novartis, Biogen Idec, Teva, Merz Pharmaceuticals and travel/accommodations/meeting expenses by Novartis, Teva, Biogen Idec, Merz Pharmaceuticals, Ipsen, Esai and Glaxo Smith Kline.

Pablo Villoslada has received consultancy fees from Roche, Novartis, MedImmune, TFS, Heidelberg Engineering, Digna Biotech, Neurotec Farma and is a shareholder in Bionure Farma .

Peter Calabresi has received personal compensation for consulting and serving on scientific advisory boards from; Vertex, Vaccinex, Prothena, and Abbott; and has received research funding from companies; Biogen-IDEC and Novartis and is supported by NIH, R01 NS082347.

Fiona Costello has nothing to disclose.

Friedemann Paul is supported by the German Research Council (DFG Exc 257) and by the German Ministry of Education and Research (BMBF "Competence Network Multiple Sclerosis").

Olivier Outteryck has nothing to disclose.

Jette Lautrup Frederiksen has served on scientific advisory boards for and received funding for travel related to these activities and honoraria from Biogen Idec, Sanofi-Aventis, Genzyme, Teva, Novartis and Almirall. Jette Frederiksen has received speaker honoraria from Biogen Idec, Merck Serono and Teva. She has served as advisor on preclinical development for Takeda.

Patrick Vermesch has nothing to disclose.

Axel Petzold has nothing to disclose.

Lisanne Balk has nothing to disclose.

Sebastian Wolf is on advisory boards of, and has served as consultant and speaker for Allergan, Bayer, Heidelberg Engineering, Molecular Partners, Novartis, Roche, and Optos.

Elliot Frohman has nothing to disclose.

Alexander Klistorner has nothing to disclose.

Ghislaine Traber has nothing to disclose.

Authors contributions

SS, FC, LB and AP contributed to designing the study concept. All authors performed OCT QC assessments at their centre. LB, AP performed data collection and statistical analyses. All authors participated either in person or by proxy at the ECTRIMS 2012 consensus meeting. FC, SS and AP prepared the first draft of the manuscript. All authors revised the manuscript. All authors gave final approval of the version submitted.

Acknowledgments

The consensus meeting at ECTRIMS was privately sponsored (AP).

Table 1. The OSCAR-IB quality control criteria for retinal OCT scans.

Item	criteria
O	Obvious problems not covered by items below. Please document for discussion + consensus agreement
S	Is the OCT signal sufficient? Signal strength > 15 (ring and volume scans) with appropriate averaging of multiple scans (ART activated).
C	Is the ring scan correctly centred? for circular discs: ONH must not cross more than two colours of the RAF logo (outer ring of RAF adjusted to outer ring of scan either by paper or electronically). In contrast to the ONH ring scan, post-hoc readjustment is possible for the macular volume scan.
A	Is there an algorithm failure? Red lines correctly identify the superior and inferior RNFL border (ring scan); Red lines correctly identify the retinal borders (volume scan)
R	Is there visible retinal pathology which may potentially impair the RNFL reading? See Table 2 (note some of these conditions are also exclusion criteria for OCT studies in MS)
I	Is the fundus well illuminated? Retinal structures visible (ring and volume scans)
B	Is the measurement beam placed centrally? Homogeneous ONL reflectivity (ring and volume scans)

Table 2: Participating centers in alphabetical order, specifying the respective clinical specialty, number of raters and overall QC acceptance rate.

Centre	Speciality	Number of raters	QC accept (%)
Amsterdam	Neurology	3	52
Baltimore	Neurology	1	39
Berlin	Neurology	1	38
Bern	Ophthalmology	4	52
Calgary	Neurology	1	38
Copenhagen	Neurology	1	44
Düsseldorf	Neurology	1	52
Lille	Neurology	1	50
London	Neurology	1	52
Philadelphia	Ophthalmology	1	48
Sydney	Ophthalmology	1	26
UTSW	Neurology	2	39
Zurich	Neurology	1	50
	Ophthalmology	1	

Table 3: Inter-centre kappas for each of the individual OSCAR-IB QC criteria. Note that individual kappas are lower compared to the overall kappa of 0.7 because raters had the right to reject a scan based on one OSCAR-IB criterion alone, also several may apply simultaneously.

QC criterion	kappa
O	0.434
S	0.404
C	0.579
A	0.482
R	0.543
I	0.565
B	0.549

References

- [1] Petzold A, de Boer JF, Schippling S, et al. Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol* 2010; 9:921-932.
- [2] Talman L S, Bisker ER, Sackel DJ, et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. *Ann Neurol* 2010 67:749-760.
- [3] Balk LJ, de Vries-Knoppert WAEJ, Petzold A. A Simple Sign for Recognizing Off-Axis OCT Measurement Beam Placement in the Context of Multicentre Studies. *PLoS One* 2012; 7:e48222.
- [4] Gabriele M L, Ishikawa H, Wollstein G, et al. Optical coherence tomography scan circle location and mean retinal nerve fiber layer measurement variability. *Invest Ophthalmol Vis Sci* 2008; 49:2315-2321.
- [5] Balasubramanian M, Bowd C, Vizzeri G, Weinreb RN, Zangwill LM. Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography. *Opt Express* 2009; 17:4019-4036.
- [6] Domalpally A, Danis RP, Zhang B, Myers D, Kruse CN. Quality issues in interpretation of optical coherence tomograms in macular diseases. *Retina* 2009; 29:775-781.
- [7] Tewarie P, Balk LJ, Costello F, et al. The OSCAR-IB Consensus Criteria for Retinal OCT Quality Assessment. *PLoS One* 2012; 7:e34823.
- [8] Fleiss J. Measuring nominal scale agreement among many raters. *Psych Bull* 1971; 76:378-382.
- [9] Landis J and Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.
- [10] Vizzeri G, Bowd C, Medeiros FA, Weinreb RN, Zangwill LM. Effect of Signal Strength and Improper Alignment on the Variability of Stratus Optical Coherence Tomography Retinal Nerve Fiber Layer Thickness Measurements. *Am J Ophthalmol* 2009; 148:245–255.
- [11] Yim Lui Cheung C, Kai Shun Leung C, Lin D, Pang CP, Shun Chiu Lam D. Relationship between Retinal Nerve Fiber Layer Measurement and Signal Strength in Optical Coherence Tomography. *Ophthalmol* 2008; 115:1347–1351.
- [12] Huang J, Liu X, Wu Z, Sadda S. Image quality affects macular and retinal nerve fiber layer thickness measurements on fourier-domain optical coherence tomography. *Ophthalmic Surg Lasers Imaging* 2011; 42:216–21.
- [13] Saidha S, Syc SB, Ibrahim MA, et al. Primary retinal pathology in multiple sclerosis as detected by optical coherence tomography. *Brain* 2011; 134:518-33.
- [14] Brandt AU, Oberwahrenbrock T, Ringelstein M, et al. Primary retinal pathology in multiple sclerosis as detected by optical coherence tomography. *Brain* 2011; 134:e193.

- [15] Gelfand JM, Nolan R, Schwartz D M, Graves J, Green AJ. Microcystic macular oedema in multiple sclerosis is associated with disease severity. *Brain* 2012; 135:1786-93.
- [16] Saidha, S Sotirchos ES, Ibrahim MA, et al. Microcystic macular oedema, thickness of the inner nuclear layer of the retina, and disease characteristics in multiple sclerosis: a retrospective study. *Lancet Neurol* 2012; 11:963-72.
- [17] Balk LJ, Killestein J, Polman C H, Uitdehaag BM, Petzold A. Microcystic macular oedema confirmed, but not specific for multiple sclerosis. *Brain* 2012; 135:e226.
- [18] Sotirchos ES, Saidha S, Byraiah G, et al. In vivo identification of morphologic retinal abnormalities in neuromyelitis optica. *Neurology*. 2013; 80:1406-1414.
- [19] Wolff B, Basdekidou C, Vasseur V, Mauget-Faÿsse M, Sahel JA, Vignal C. RETINAL INNER NUCLEAR LAYER MICROCYSTIC CHANGES IN OPTIC NERVE ATROPHY: A Novel Spectral-Domain OCT Finding. *Retina* 2013; May 2. [Epub ahead of print].
- [20] Seigo MA, Sotirchos ES, Newsome S, et al. In vivo assessment of retinal neuronal layers in multiple sclerosis with manual and automated optical coherence tomography segmentation techniques. *J Neurol* 2012; 259:2119-30.

Figure Legends

Figure 1. Study design.

Figure 2. Example of an OCT scan rejected by some raters for the I-criterion. Shadow seen on the bottom right of the fundus image is cast by a floater.

Figure 3. Example of a scan to be rejected due to truly poor illumination (I-criterion).

Supplementary Figure. Two scans originating from the same healthy control subject scanned with EBF activated and use of the follow-up function within 1 minute which illustrates the problem of a *systematic error* introduced by off-centre beam-placement (failed B-criterion), for a ring scan well centered on the ONH (passed C-criterion) [3]. **(A)** shows the IR image OD. The blue line indicates the position of the ONH and foveola. The green circle indicates the area of the OCT B-scan shown in **(B)** the pRNFL is the bright layer between the red and green lines determined by automated layer segmentation. There is a bright signal for the OPL/ONL for the papilomacular bundle (PMB) indicated by a yellow arrow, which turns darker in for the more nasally located OPL/ONL indicated by the magenta arrow. The corresponding quantitative data are shown in **(C)** with a global average pRNFL of 106 μm , a PMB RNFL of 51 μm and a nasal RNFL of 96 μm . **(D)** A change of the beam-placement, whilst keeping the ring scan on precisely the same position causes a change of the OCT B-scan in the live image which is only visible to the OCT operator and lost in the summary image uploaded to an OCT reading center (for more detailed see serial live images in reference [3]). However, what remains clearly visible is the change of the OPL/ONL signal in **(E)**. Compared to the earlier OCT B-scan shown in **(B)** the OPL/ONL appear relatively darker (magenta arrow) below the PMB and relative brighter (yellow arrow) below the nasal RNFL. Closer inspections shows that subsequent automated layer segmentation performs different. Consequently the data shown in **(F)** is different compared to the data in **(C)** with an averaged pRNFL of 102 μm , a PMB of 54 μm and a nasal RNFL of 87 μm . To better appreciate the local differences in automated layer segmentation caused by off-center beam placement on overlay of both RNFL profiles are shown in **(G)** with the blue line corresponding to the OCT B-scan in **(B)** and the magenta line corresponding to the OCT B-scan in **(E)**. Please note that in addition to the *systematic error* introduced by off-center beam placement there are also non systematic segmentation artifacts related to the void signal caused by the retinal vasculature.