

Software updates of OCT segmentation algorithms influence longitudinal assessment of retinal atrophy

Danko Coric^{a,b}, MD; Axel Petzold^{a,b,c,d}, MD, PhD; Bernard MJ. Uitdehaag^a, MD, PhD;
Lisanne J. Balk^{a,b}, PhD

- a. Department of Neurology, VU University Medical Center, Amsterdam, the Netherlands
- b. Dutch Expertise Center for Neuro-ophthalmology, VU University Medical Center, Amsterdam, the Netherlands
- c. Moorfields Eye Hospital, City Road, London, UK
- d. The National Hospital for Neurology and Neurosurgery, Queen Square, London, UK

Corresponding author

Danko Coric, MD

VU University Medical Center, department of Neurology

PO Box 7057, 1007 MB Amsterdam, the Netherlands

Phone: 0031 20 4440717 Fax: 0031 20 4440715

Email: d.coric@vumc.nl

Word count manuscript: ...

Key words: multiple sclerosis, optical coherence tomography, retinal nerve fiber layer, neuro-ophthalmology, retinal layer segmentation, Bland-Altman analysis

ABSTRACT

Objective – To investigate whether there is a systematic difference in peripapillary retinal nerve fiber layer (pRNFL) thickness measurements between subsequent updates of pRNFL segmentation software provided by Heidelberg Spectralis optical coherence tomography (OCT).

Methods – In total, 838 pRNFL scans from 213 multiple sclerosis (MS) patients and 61 healthy controls were analysed. All ~~pRNFL~~ scans were performed on the same OCT device followed by automated segmentation (HRA 5.6.4.0) and data extraction. Subsequently, all scans were re-segmented with an updated software version (HRA 6.0.7.0). To assess level of agreement between the two algorithms, Bland-Altman Plots (BAP) were constructed. Paired samples T-test and linear regression analyses were used to investigate for differences in mean thickness and proportional bias respectively.

Results – Overall, the updated version showed an overestimation of 0.16 μm [95%CI 0.097 – 0.23, $p<0.001$] for the global pRNFL thickness compared to the earlier version. The largest differences were found for the nasal inferior (mean Δ 0.29 μm , $p<0.001$) and temporal inferior (mean Δ 0.43 μm , $p<0.001$) sectors. Inspection of the BAP revealed that the difference between the two versions could be up to 6 μm for the global mean. There was no proportional bias for the global mean ($\beta=0.003$, $p=0.245$) nor for any of the separate sectors.

Conclusion – The data show a significant difference in pRNFL thickness measurements between two subsequent versions of the same segmentation software. Although the mean difference was relatively small, the differences within the individual subject could be considerably higher than the known atrophy rate of 1 $\mu\text{m}/\text{year}$ in MS.

INTRODUCTION

Optical coherence tomography (OCT) is a non-invasive imaging technique that allows for detailed imaging and quantification of the thickness of individual retinal layers.(1, 2) Having gained wide acceptance in the ophthalmology practice for the evaluation of various primary ocular disorders, OCT is employed more and more in the assessment of neurodegenerative disorders as well, particularly in multiple sclerosis (MS).(3, 4) Cross-sectional studies in MS patients have demonstrated a link between atrophy of the peripapillary retinal nerve fiber layer (pRNFL), and various (para)clinical parameters of neurodegeneration including physical disability, cognitive impairment and brain atrophy.(5-7) Current studies focus on longitudinal atrophy measurements in the hope of being able to predict which MS patients will develop physical or cognitive disability.

The assessment of retinal atrophy has improved considerably with the introduction of spectral domain OCT (compared to the previously used time-domain OCT), allowing for reliable quantification of individual retinal layers.(8) In addition, automated retinal layer segmentation algorithms have not only replaced the time consuming and demanding task of manual segmentation but have also increased the accuracy of retinal layer thickness measurements. Subsequent updates in these algorithms aim at refining these aspects by decreasing the processing time and increasing accuracy.(9) **The Spectralis OCT by Heidelberg Engineering is frequently used for clinical and research purposes in a MS setting and gets updated regularly with newer versions of the segmentation algorithm.**

Nevertheless, in a longitudinal setting there is the potential of biased measurements if not all subsequent scans are segmented using the same software version. This problem can occur due to the difference in the manner of segmentation (for example how the algorithm handles

artefacts due to blood vessels). Therefore, the aim of this study was to investigate whether there is a systematic difference in pRNFL thickness measurement between subsequent updates of the pRNFL segmentation algorithm provided by the Heidelberg Spectralis OCT.

METHODS

Participants

For this cross-sectional study, scans were used from participants of the prospective Amsterdam MS Cohort. The inclusion criteria for this cohort have been described previously.⁽¹⁰⁾ At baseline, this cohort consisted of 230 patients and 63 healthy controls (HCs). Patients were retested after a period of two years. Both the baseline and the follow-up scans were used in this study. All examinations (clinical, OCT etc.) were performed on the same day.

This study was approved by the medical ethics committee and scientific research committee of the VU University Medical Center and was conducted in accordance with the declaration of Helsinki. Written informed consent was obtained from every participant.

OCT acquisition and segmentation

OCT imaging was performed in all subjects on a Spectralis spectral domain OCT (Heidelberg Engineering, Heidelberg, Germany) with dual beam simultaneous imaging and the eye tracking function enabled.⁽¹¹⁾ All scans, baseline as well as the follow-up scans, were acquired using the same acquisition software (version 1.7.1.0.). In order to measure pRNFL

thickness a 12° ring scan (1536 A-scans, no predetermined automatic real time) manually centred around the optic nerve head was performed. Follow-up scans were acquired using the automatic follow-up function in order to be sure the scans were made at the exact same location. The scans were segmented using software provided by the manufacturer, HRA / Spectralis Viewing Module 5.6.4.0 and the data was exported for statistical analysis. Subsequently, the scans were re-segmented using an updated version of the software, HRA / Spectralis Viewing Module 6.0.7.0, and again exported. After both segmentation procedures, quality control was performed according to the OSCAR-IB criteria(12) and scans were rejected if they failed one or more criteria. Special attention was paid to the algorithm failure criterion. Scans with obvious algorithm failures were rejected. In contrast, scans with small, debatable segmentation deviations (i.g. due to blood vessels), which would normally be accepted for any other study, were included. These scans were not corrected manually in order to be able to investigate the potential bias of the segmentation algorithms, instead of the bias caused by manual correction.

Statistical analysis

All statistical analyses were performed using SPSS version 22.0. Normality was assessed visually by means of histograms. Differences in subject characteristics were tested using student T-test (parametric variables), Mann-Whitney-U test (nonparametric variables) and chi square test (categorical variables). Bland-Altman plots were constructed by calculating the mean difference (systematic bias) and 95% limits of agreement of the thickness measured by the two software versions.(13) The thickness measured by version 5.6.4.0 was always subtracted from the thickness measured by version 6.0.7.0, meaning that a positive difference indicates that version 6.0.7.0 measured a thicker pRNFL. Mean differences in retinal layer

thickness between the two versions were tested using the paired samples T-test. Proportional bias was tested by means of linear regression analyses. **Linear regression was also used to test the association between the absolute difference in pRNFL thickness between the two versions and the number of averaged B-scans (expressed by automated real time, ART).** Subgroup analyses were performed comparing the systematic bias in patients and HCs. **Lastly, longitudinal changes in pRNFL thickness were calculated for both software versions. Subsequently, the difference in the amount of progressive atrophy between the two versions was tested using paired samples T-test.** All analyses were performed for the mean thickness of the entire pRNFL (global mean) as well as the six individual sectors: temporal superior (TS), temporal (T), temporal inferior (TI), nasal inferior (NI), nasal (N) and nasal superior (NS). Statistical significance was set at $p < 0.05$.

RESULTS

For this cross-sectional study, 802 scans from 230 patients and 211 scans from 63 HCs were available for review. After quality control, 150 patient scans (18.7%) and 25 HC scans (11.8%) were rejected. This led to the inclusion of 652 scans (353 baseline and 299 follow-up) from 213 MS patients and 186 scans (113 baseline and 73 follow-up) from 61 HCs.

The baseline characteristics of these 213 patients and 61 HCs are shown in Table 1.

Compared to patients HCs were younger (mean difference 3.6 years, $p=0.001$). Patients had a mean disease duration of 20.2 years and showed a considerable level of disability which is reflected in the median EDSS score of 4.0. Most patients had a relapsing remitting disease course. As expected patients showed considerably more atrophy of the pRNFL compared to HCs (mean difference 11.20 μm , $p<0.001$).

Mean thickness by sector

Table 2 shows the average of the global mean and the six separate sectors for all scans as calculated by both software versions. There was a significant difference for the global mean, with version 6.0.7.0 showing a thicker pRNFL compared to version 5.6.4.0 (mean difference 0.16 μm , $p<0.001$). The largest difference in mean thickness between the two versions was found for the NI and TI sectors (NI 0.29 μm , $p<0.001$; TI 0.43 μm , $p<0.001$). The N sector showed a significant difference as well, with a mean difference of 0.14 μm , $p=0.015$. On the contrary, there was no significant difference in mean thickness between the two software versions for the T, TS or NS sector. **The median difference was 0.0 μm for the global mean as well as for all of the sectors.**

Bland-Altman plots and proportional bias

Figure 1 shows the Bland-Altman plot for the global mean of all scans combined. The dotted lines represent the bias (mean difference) and 95% levels of agreement. As illustrated, most scans showed no or only a small difference of 1 or 2 μm . Nevertheless, a considerable part of the scans showed larger differences with a maximum of up to 6 μm . There was no

proportional bias ($\beta = 0.003$, $p = 0.245$) for the global mean pRNFL thickness, meaning that the difference in thickness measured by the two segmentation software versions was independent of the thickness of the pRNFL.

As described earlier, only the NI, TI and N sectors showed a significant difference in thickness measured by the two software versions. The Bland-Altman plots for these three sectors are displayed in Figure 2. All three sectors, and especially the NI and TI sectors, showed considerable variability in the difference in thickness measured, with outliers of up to $-27 \mu\text{m}$ for the NI sector. This sector has a high density of retinal vessels. Again, there was no proportional bias for any of the sectors.

We also examined whether the **systemic systematic** bias in pRNFL thickness measurement was different in patients compared to HCs. The patient scans showed a mean difference of $0.15 \mu\text{m}$ between the two versions and the HC scans showed $0.19 \mu\text{m}$. This difference of $0.039 \mu\text{m}$ was statistically not significant ($p=0.634$).

To get an indication whether scan 'quality' (is this right?) influenced the segmentation results we investigated the association between ART and the difference in pRNFL thickness between the two software versions. The median ART was 31 with a range of 1 - 100. There was no significant association for the global mean ($\beta = -0.001$, $p=0.451$). The N and T sectors did however show a small, yet significant, inverse association (N $\beta = -0,003$, $p=0.047$; T $\beta = -0.002$, $p=0.008$). The other four sectors did not show any significant associations.

The effect on longitudinal atrophy measurements

39 HCs: show the same amount of longitudinal atrophy: op een hoop gooien of ???

A total of 178 MS patients completed the two year follow-up. Using the initial software version (5.6.4.0) for both the baseline and follow-up visits, these patients showed a decline in pRNFL thickness of $-0.95 \mu\text{m}$ (95%CI $-0.67 - -1.22$, $p < 0.001$). When using the later software version (6.0.7.0) the amount of atrophy was slightly less, $-0.71 \mu\text{m}$ (95%CI $-0.46 - -0.96$, $p < 0.001$). This difference of $0.18 \mu\text{m}$ was statistically significant ($p = 0.006$).

The amount of atrophy in the most extreme case was also calculated. In this scenario, where the baseline scans would be segmented with version 5.6.4.0 and the follow-up scans with version 6.0.7.0 the amount of atrophy the patients showed was $-0.63 \mu\text{m}$ (95%CI -0.37 to -0.89 , $p < 0.001$) This was significantly lower than the amount of atrophy showed when all scans were segmented with version 5.6.4.0 (mean difference $0.32 \mu\text{m}$, $p < 0.001$) but not when they were all segmented with version 6.0.7.0 (mean difference $0.08 \mu\text{m}$, $p = 0.136$).

DISCUSSION

This study shows that there is a significant, systematic difference in the pRNFL thickness measurement between two subsequent software updates of the retinal layer segmentation algorithm provided by the Heidelberg Spectralis OCT. Although the mean difference was relatively small (yet significant), the difference between the two algorithms in individual scans could be as high as $6 \mu\text{m}$ for the global mean pRNFL thickness. This last finding is an important one if we plan on using OCT as a surrogate marker for neurodegeneration in individual patients. The expected rate of pRNFL atrophy in MS patients is about $1-2 \mu\text{m}/\text{year}$ (5, 14, 15) and is even lower in patients with a longer disease duration.(16) A large part of our scans showed a degree of difference in pRNFL thickness measurement between the two versions that could mask or overestimate the true pRNFL atrophy, which is presumed to be

caused by retrograde trans-synaptic degeneration.(17) In addition, the study shows that using different software versions does in fact influence longitudinal atrophy measurements. To our knowledge, this is the first study that has investigated this methodological issue. The findings are relevant and should be considered by regulatory authority for approval of OCT segmentation software. Producers will need to be aware of the need to potentially update their segmented normative databases.

Current Spectralis OCT machines provide pRNFL thickness measurements directly after scanning. This process is done automatically and according to the last installed segmentation algorithm. However, even when the follow-up scanning function is used, the previous (baseline) scan retains its previous segmentation. So, if the OCT device gets updated with a new version of the segmentation algorithm in between patient assessments, a measurement error can be introduced easily. It is therefore important to make sure that the baseline as well as each subsequent follow-up scan is segmented by the same software version. In the case of the Heidelberg Spectralis OCT this can be done quite easily and quickly by opening the previous scans and intentionally segmenting them again. In our experience, it is, in most scans, visible that the line between the RNFL and ganglion cell layer replaces when the scan is re-segmented. In contrast, the line indicating the inner limiting membrane remains very stable. For illustration, Figure 3 shows three examples of (obvious) differences in segmentation between the two software versions. All three scans were of high quality and the segmentation would be accepted in all cases. Nevertheless, there was a difference in pRNFL thickness between the two software versions.

There are other factors which need to be considered because they can also influence pRNFL thickness measurements, for instance the interobserver and intraobserver variability. However, it has been shown that both the inter- as well as intraobserver variability are generally high when only those scans that have passed the OSCAR-IB criteria are selected, as

is the case in our study.(18) One specific confounder that is especially relevant in longitudinal studies is off-axis beam placement, which can introduce a large error.(19) To our knowledge, the potential error of off-center beam placement is not dependent on the software version. Nevertheless, while exact beam placement at subsequent scans can be quite tricky, making sure every scan is segmented with the same software version is rather simple.

One of the strengths of this study is the statistical robustness due to the large sample size. We chose to analyse scans from both patient as well as HCs. One could argue that patient scans are generally of poorer quality (due to problems with fixation for example) and are thus more prone to (small) segmentation errors leading to a difference in thickness measurements. However, we could not find a significant difference in the ~~systemic~~ systematic bias between patient scans and HC scans. Although the ART did not affect the the difference in global pRNFL thickness, there were small but significant associations for the N and T sectors, which may indicate that higher ART is associated with a smaller difference between the software versions. However, as all scans were of sufficient quality, we cannot say with certainty that the differences in segmentation occur more easily with lower quality scans. Another strength of this study is that we performed OSCAR-IB quality control. We only included those scans that would pass the quality control if they were to be used in any other study or clinical practice.

A limitation of this study is the fact that we did not investigate whether this problem also occurs with OCT machines and segmentation software by other manufacturers. ~~For example, a number of studies in MS have been performed with the Cirrus OCT by Carl Zeiss Meditec (N=15) and 3D-OCT-1000 / 2000 by Topcon Corporation (N=5).~~ A recent meta-analysis by Petzold et al. showed that of all the studies investigating pRNFL atrophy in MS patients with and without optic neuritis, seventeen have been performed with the Spectralis OCT, fourteen with the Cirrus OCT by Carl Zeiss Meditec and four with the 3D-OCT-1000 / 2000 by

Topcon Corporation.(20) A literature review on longitudinal SD-OCT studies revealed that six have been performed with the Spectralis OCT(16, 21-25) and five with the Cirrus OCT(15, 26-29). It is plausible that this problem might occur with these OCT devices as well, but this still remains to be investigated and these groups are well placed to investigate this point.

In conclusion, this study has demonstrated that updates in pRNFL segmentation software lead to a significant difference in pRNFL thickness measurement, especially when analysing individual patient scans. Therefore, we recommend that all scans are segmented with the same software version when assessing pRNFL atrophy in a longitudinal setting.

ACKNOWLEDGEMENTS

The VUmc MS Center Amsterdam received research support for OCT projects from Teva Pharmaceuticals Netherlands B.V., Amsterdam, The Netherlands.

DECLARATION OF CONFLICT OF INTEREST

DC nothing to disclose.

AP is supported by the University College London Comprehensive Bio-medical Research Centre, the Moorfields Biomedical Research Centre, the Dutch MS Research Foundation, ERN-EYE and reports grants from the Dutch MS Society, grants from VKC foundation, during the conduct of the study; and reports personal fees and grants from Novartis, outside

the submitted work; and is a member of the steering committee of the OCTiMS study by Novartis.

BU has received personal compensation for consulting from Biogen Idec, Genzyme, Merck Serono, Novartis, Roche and TEVA.

LB reports institutional support from TEVA.

REFERENCES

1. Jaffe GJ, Caprioli J. Optical coherence tomography to detect and manage retinal disease and glaucoma. *Am J Ophthalmol*. 2004;137(1):156-69.
2. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. *Science*. 1991;254(5035):1178-81.
3. Balcer LJ. Clinical trials to clinical use: using vision as a model for multiple sclerosis and beyond. *J Neuroophthalmol*. 2014;34 Suppl:S18-23.
4. Frohman EM, Fujimoto JG, Frohman TC, Calabresi PA, Cutter G, Balcer LJ. Optical coherence tomography: a window into the mechanisms of multiple sclerosis. *Nat Clin Pract Neurol*. 2008;4(12):664-75.
5. Sepulcre J, Murie-Fernandez M, Salinas-Alaman A, Garcia-Layana A, Bejarano B, Villoslada P. Diagnostic accuracy of retinal abnormalities in predicting disease activity in MS. *Neurology*. 2007;68(18):1488-94.
6. Coric D, Balk LJ, Verrijp M, Eijlers A, Schoonheim MM, Killestein J, et al. Cognitive impairment in patients with multiple sclerosis is associated with atrophy of the inner retinal layers. *Mult Scler*. 2017;1352458517694090.
7. Saidha S, Sotirchos ES, Oh J, Syc SB, Seigo MA, Shiee N, et al. Relationships between retinal axonal and neuronal measures and global central nervous system pathology in multiple sclerosis. *JAMA Neurol*. 2013;70(1):34-43.
8. Galetta KM, Balcer LJ. Measures of visual pathway structure and function in MS: Clinical usefulness and role for MS trials. *Mult Scler Relat Disord*. 2013;2(3):172-82.
9. Kafieh R, Rabbani H, Kermani S. A review of algorithms for segmentation of optical coherence tomography from retina. *J Med Signals Sens*. 2013;3(1):45-60.
10. Balk LJ, Twisk JW, Steenwijk MD, Daams M, Tewarie P, Killestein J, et al. A dam for retrograde axonal degeneration in multiple sclerosis? *J Neurol Neurosurg Psychiatry*. 2014;85(7):782-9.
11. Balk LJ, Sonder JM, Strijbis EM, Twisk JW, Killestein J, Uitdehaag BM, et al. The physiological variation of the retinal nerve fiber layer thickness and macular volume in humans as assessed by spectral domain-optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53(3):1251-7.
12. Cruz-Herranz A, Balk LJ, Oberwahrenbrock T, Saidha S, Martinez-Lapiscina EH, Lagreze WA, et al. The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. *Neurology*. 2016;86(24):2303-9.
13. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141-51.
14. Talman LS, Bisker ER, Sackel DJ, Long DA, Jr., Galetta KM, Ratchford JN, et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. *Ann Neurol*. 2010;67(6):749-60.
15. Garcia-Martin E, Ara JR, Martin J, Almarcegui C, Dolz I, Vilades E, et al. Retinal and Optic Nerve Degeneration in Patients with Multiple Sclerosis Followed up for 5 Years. *Ophthalmology*. 2017;124(5):688-96.
16. Balk LJ, Cruz-Herranz A, Albrecht P, Arnow S, Gelfand JM, Tewarie P, et al. Timing of retinal neuronal and axonal loss in MS: a longitudinal OCT study. *J Neurol*. 2016.
17. Gabilondo I, Martinez-Lapiscina EH, Martinez-Heras E, Fraga-Pumar E, Llufrui S, Ortiz S, et al. Trans-synaptic axonal degeneration in the visual pathway in multiple sclerosis. *Ann Neurol*. 2014;75(1):98-107.
18. Balk LJ, Petzold A. Influence of the eye-tracking-based follow-up function in retinal nerve fiber layer thickness using fourier-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2013;54(4):3045.
19. Balk LJ, de Vries-Knoppert WA, Petzold A. A simple sign for recognizing off-axis OCT measurement beam placement in the context of multicentre studies. *PLoS One*. 2012;7(11):e48222.

20. Petzold A, Balcer LJ, Calabresi PA, Costello F, Frohman TC, Frohman EM, et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol.* 2017;16(10):797-812.
21. Graham EC, You Y, Yiannikas C, Garrick R, Parratt J, Barnett MH, et al. Progressive Loss of Retinal Ganglion Cells and Axons in Nonoptic Neuritis Eyes in Multiple Sclerosis: A Longitudinal Optical Coherence Tomography Study. *Invest Ophthalmol Vis Sci.* 2016;57(4):2311-7.
22. Serbecic N, Aboul-Enein F, Beutelspacher SC, Vass C, Kristoferitsch W, Lassmann H, et al. High resolution spectral domain optical coherence tomography (SD-OCT) in multiple sclerosis: the first follow up study over two years. *PLoS One.* 2011;6(5):e19843.
23. Pardini M, Botzkowski D, Muller S, Vehoff J, Kuhle J, Ruberte E, et al. The association between retinal nerve fibre layer thickness and N-acetyl aspartate levels in multiple sclerosis brain normal-appearing white matter: a longitudinal study using magnetic resonance spectroscopy and optical coherence tomography. *European journal of neurology.* 2016;23(12):1769-74.
24. Pisa M, Guerrieri S, Di Maggio G, Medagliani S, Moiola L, Martinelli V, et al. No evidence of disease activity is associated with reduced rate of axonal retinal atrophy in MS. *Neurology.* 2017;89(24):2469-75.
25. Klistorner A, Graham EC, Yiannikas C, Barnett M, Parratt J, Garrick R, et al. Progression of retinal ganglion cell loss in multiple sclerosis is associated with new lesions in the optic radiations. *European journal of neurology.* 2017;24(11):1392-8.
26. Wings KM, Murchison CF, Bourdette DN, Spain RI. Longitudinal optical coherence tomography study of optic atrophy in secondary progressive multiple sclerosis: Results from a clinical trial cohort. *Mult Scler.* 2017:1352458517739136.
27. Abalo-Lojo JM, Treus A, Arias M, Gomez-Ulla F, Gonzalez F. Longitudinal study of retinal nerve fiber layer thickness changes in a multiple sclerosis patients cohort: A long term 5 year follow-up. *Mult Scler Relat Disord.* 2017;19:124-8.
28. Saidha S, Al-Louzi O, Ratchford JN, Bhargava P, Oh J, Newsome SD, et al. Optical coherence tomography reflects brain atrophy in multiple sclerosis: A four-year study. *Ann Neurol.* 2015;78(5):801-13.
29. Narayanan D, Cheng H, Bonem KN, Saenz R, Tang RA, Frishman LJ. Tracking changes over time in retinal nerve fiber layer and ganglion cell-inner plexiform layer thickness in multiple sclerosis. *Mult Scler.* 2014;20(10):1331-41.

Table 1: Characteristics of the study cohort.

	Patients	Healthy controls	p-value
	N = 213	N = 61	
Age (years), mean (\pm SD)	53.8 (9.8)	50.2 (7.0)	0.001
Sex (female : male)	147 : 66	41 : 20	0.789
Disease duration (years), mean (\pm SD)	20.2 (6.9)	N/a	
EDSS, median [range]	4.0 [1.0 – 8.0]	N/a	
Type of MS			
RR	130 (61.0%)	N/a	
SP	56 (26.3%)		
PP	27 (12.7%)		
pRNFL thickness, mean (\pm SD)*	83.36 (7.65)	94.56 (14.14)	< 0.001

* measured with version 6.0.7.0

SD = Standard Deviation; EDSS = Expanded Disability Status Scale; MS = Multiple Sclerosis; RR = Relapsing Remitting; SP = Secondary Progressive; PP = Primary Progressive; pRNFL = peripapillary Retinal Nerve Fiber Layer

Table 2: Mean pRNFL thickness and proportional bias for the global mean and each of the six separate sectors as measured by the two software versions. Proportional bias was tested using linear regression analyses, resulting in regression coefficient β and corresponding p-value.

Sector	pRNFL thickness version 6.0.7.0 $\mu\text{m} (\pm \text{SD})$	pRNFL thickness version 5.6.4.0 $\mu\text{m} (\pm \text{SD})$	<i>p</i>-value	Proportional bias (β)	<i>p</i>-value
Global Mean	85.22 (13.87)	85.06 (13.83)	<0.001	0.003	0.245
Nasal Superior	92.55 (20.05)	92.47 (20.11)	0.114	-0.003	0.247
Nasal	63.68 (11.45)	63.55 (11.36)	0.015	0.006	0.120
Nasal Inferior	96.90 (23.15)	96.62 (23.24)	<0.001	-0.004	0.241
Temporal Inferior	124.37 (23.12)	123.94 (22.98)	<0.001	0.006	0.157
Temporal	59.60 (14.99)	59.62 (15.02)	0.402	-0.001	0.437
Temporal Superior	121.31 (21.50)	121.16 (21.32)	0.162	0.009	0.081

pRNFL = peripapillary Retinal Nerve Fiber Layer; SD = Standard Deviation

Figure 1: Bland-Altman Plot for the global mean pRNFL thickness. The difference in pRNFL thickness between the two segmentation software versions (y-axis) is plotted against the average of the two measurements (x-axis). The blue line represent the regression line to test for proportional bias. The middle broken line represents the mean difference (systematic bias) in pRNFL thickness for all scans combined. The top and bottom broken lines represent the 95% limits of agreement.

pRNFL = peripapillary Retinal Nerve Fiber Layer

Figure 2: Bland-Altman plots of the three sectors that showed a significant difference in pRNFL thickness between the two segmentation software versions. The difference in pRNFL thickness between the two segmentation software versions (y-axis) is plotted against the average of the two measurements (x-axis). The middle broken line represents the mean difference (systematic bias) in pRNFL thickness for all scans combined. The top and bottom broken lines represent the 95% limits of agreement. Notice that the scale on the y-axis has been enlarged in comparison to Figure 1.

Figure 3: Three examples of differences in segmentation between the two software versions. The top image is segmented with version 5.6.4.0 and the bottom image with version 6.0.7.0. The mean thickness of the entire pRNFL (global mean) is given in white.