

Validation of Optical Coherence Tomography Retinal Segmentation in Neurodegenerative Disease

Bryan M. Wong^{1,2}, Richard W. Cheng¹, Efrem D. Mandelcorn^{3,4}, Edward Margolin^{3,4}, Sherif El-Defrawy^{3,4}, Peng Yan^{3,4}, Anna T. Santiago⁵, Elena Leontieva¹, Wendy Lou⁶, ONDRI Investigators⁷, and Wendy Hatch^{3,4}, and Christopher Hudson^{1,3}

¹ University of Waterloo, School of Optometry and Vision Science, Waterloo, Ontario, Canada

² University of Toronto, Faculty of Medicine, Toronto, Ontario, Canada

³ University of Toronto, Department of Ophthalmology and Vision Sciences, Toronto, Ontario, Canada

⁴ Kensington Eye Institute, Toronto, Ontario, Canada

⁵ Baycrest, Rotman Research Institute, Toronto, Ontario, Canada

⁶ University of Toronto, Dalla Lana School of Public Health, Toronto, Ontario, Canada

⁷ See the Appendix for the names and affiliations of the ONDRI Investigators.

Correspondence: Christopher Hudson, University of Waterloo, School of Optometry and Vision Science, Optometry Building, Room 335, University of Waterloo, 200 Columbia St W, Waterloo, Ontario N2L 3G1, Canada. e-mail: chris.hudson@uwaterloo.ca

Wendy Hatch, University of Toronto, Ophthalmology and Vision Sciences, 340 College St, Suite 501, Toronto, Ontario M5T 3A9, Canada. e-mail: whatch@KensingtonHealth.org

Received: 12 February 2019

Accepted: 24 April 2019

Published: 11 September 2019

Keywords: optical coherence tomography; retinal thickness; retinal nerve fibre layer; automated segmentation; Parkinson's disease

Citation: Wong BM, Cheng RW, Mandelcorn ED, Margolin E, El-Defrawy S, Yan P, Santiago AT, Leontieva E, Lou W, ONDRI Investigators, Hatch W, Hudson C. Validation of optical coherence tomography retinal segmentation in neurodegenerative disease. *Trans Vis Sci Tech.* 2019;8(5):6, <https://doi.org/10.1167/tvst.8.5.6>
Copyright 2019 The Authors

Purpose: This study assessed agreement between an automated spectral-domain optical coherence tomography (SD-OCT) retinal segmentation software and manually corrected segmentation to validate its use in a prospective clinical study of neurodegenerative diseases (NDD).

Methods: The sample comprised 30 subjects with NDD, including vascular cognitive impairment, frontotemporal dementia, Parkinson's disease, and Alzheimer's disease. Macular SD-OCT scans were acquired and segmented using Heidelberg Spectralis. For the central foveal B scan of each eye, eight segmentation lines were examined to determine the proportion of each line that the software erroneously delineated. Errors in four lines were manually corrected in all B scans spanning a 6-mm circle centered on the foveola. Mean volume and thickness measurements for four retinal layers (total retina, retinal nerve fiber layer [RNFL], inner retinal layers, and outer retinal layers) were obtained before and after correction.

Results: The outer plexiform layer line had one of the lowest mean error ratios (2%), while RNFL had the highest (23%). Agreement between automated software and trained observer was excellent (ICC > 0.98) for retinal thickness and volume of all layers. Mean volume differences between software and observers for the four layers ranged from -0.003 to 0.006 mm³. Mean thickness differences ranged from -1.855 to 1.859 μ m.

Conclusions: Despite occasional small errors in software-generated retinal sublayer segmentation, agreement was excellent between software-derived and observer-corrected mean volume and thickness sublayer measurements.

Translational Relevance: Automated SD-OCT segmentation software generates valid measurements of retinal layer volume and thickness in NDD subjects, thereby avoiding the need to manually correct nonobvious delineation errors.

Introduction

Spectral-domain optical coherence tomography (SD-OCT), also known as Fourier-domain OCT, has revolutionized the clinical management of many retinal and ocular diseases, such as diabetic macular edema, age-related macular degeneration (AMD), and glaucoma.^{1–3} SD-OCT is capable of detecting previously irresolvable retinal structures in the living human eye and commercially available SD-OCTs are now able to detect change in thickness as small as 1 μm .⁴ This permits visualization of the retinal sublayers. Mean thickness and volume can be determined for predefined sectors of each retinal sublayer in the SD-OCT image. The objective evaluation of retinal and optic nerve morphology, the relatively high sensitivity to detect change in morphology compared with subjective clinical evaluation, the speed of acquisition, and the relatively low cost of running the technology has made SD-OCT a common investigative technique in eye care clinics and hospital ophthalmology units worldwide, especially for the evaluation of the optic nerve head in patients with glaucoma and of retinal thickness in patients with maculopathies.

SD-OCT may also have a role as a surrogate biomarker to assess neurodegenerative disease (NDD) because the death of cortical cells is suggested to trigger retinal ganglion cell death, or vice versa.⁵ A number of relatively small cross-sectional studies have found retinal thinning in Parkinson's Disease (PD) and Alzheimer's Disease (AD).^{6–9} The retinal nerve fiber layer (RNFL) in particular appears to be thinned (compared with controls or other comparison groups) and this has been thought to reflect retinal ganglion cell death secondary to the retrograde degeneration of the cortical neurons.^{8,10} In addition, the thickness of the neural retina has been shown to correlate with reduction in cortical gray matter volume in patients with early-onset AD.¹¹ Current clinical tests for the assessment of NDDs typically offer low specificity and sensitivity, and are often reliant upon subjective evaluation or clinical intuition,^{12,13} with the result that diagnosis can rely upon a positive response to Levodopa-therapy or upon a definitive decline in functional outcome measures over a short period of time. There is a dire need for the development of noninvasive objective tools to improve the clinical management of people with NDD. An absence of objective tests with high sensitivity and specificity to noninvasively categorize

and detect change has been identified as a barrier to the development of new treatments of NDD.¹⁴

The Ontario Neurodegenerative Disease Research Initiative (ONDRI) is a province-wide research collaboration studying diseases that can result in dementia and how to improve the diagnosis and treatment of NDD, including subjects with AD and PD, with over 600 participants recruited across 13 clinical sites.¹⁵ In the ONDRI experimental design, thickness of RNFL and other retinal layers measured by SD-OCT is incorporated as an outcome measure. As well as analysis of retinal layers in the ONDRI disease cohorts, it is important to validate the agreement of the automated retinal sublayer segmentation software with manually corrected data in this cohort. Recent work by Ctori and Huntjens¹⁶ showed excellent repeatability and reproducibility with the SD-OCT sublayer segmentation software in young, healthy controls. Krebs et al.,¹⁷ however, found SD-OCT sublayer segmentation errors that were described as “clinically relevant” in approximately one-third of their AMD cohort. To our knowledge, no previous study has systematically evaluated the segmentation software in participants with NDD. Therefore, the main objective of this study was to validate the retinal sublayer SD-OCT segmentation software in patients with these diseases. By studying the effect that small segmentation errors have on retinal sublayer thickness and volume, a basis for the analysis requirements of SD-OCT images acquired as part of the ONDRI protocol can be established.

Materials and Methods

Data Collection and Image Selection

This cross-sectional, single-center study evaluated the SD-OCT images of 30 participants with one of the following NDD: vascular cognitive impairment (VCI; $n = 13$), frontotemporal dementia (FTD; $n = 6$), PD ($n = 6$), and AD/mild cognitive impairment (AD/MCI; $n = 5$). Participants were recruited from the Toronto Western Hospital¹⁵ between September 2015 and August 2016. Participants provided informed consent to participate in this study. The study protocol followed the tenets of the Declaration of Helsinki and was approved by the institutional review boards at the University of Western Ontario, University Health Network, University of Toronto, and the University of Waterloo.

General inclusion and exclusion criteria are outlined in the ONDRI protocol.¹⁵ Specific ocular

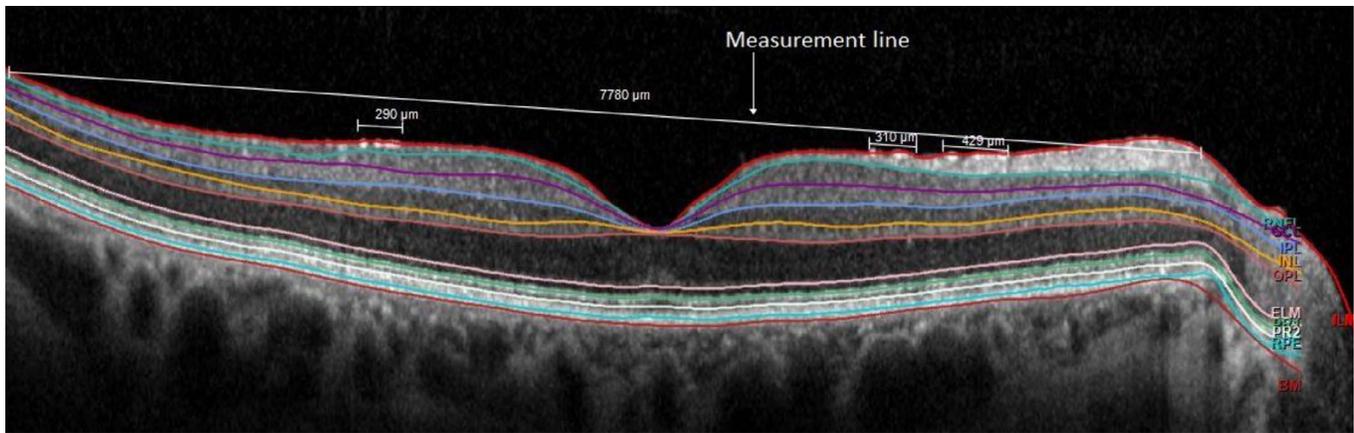


Figure 1. Line error analysis for the ILM segmentation line with small segmentation errors. The measured lengths of the three smaller line segments are summed, divided by the length of the long straight line, then multiplied by 100% to acquire the error ratio.

exclusion criteria were as follows: intraocular pressure (IOP) greater than 22 mm Hg in either eye, IOP difference greater than 5 mm Hg between eyes, optic nerve head cup-to-disc ratio (C/D) greater than or equal to 0.7, C/D asymmetry greater than 0.2, presence of a disc hemorrhage or neuroretinal rim notch in either eye, and wet AMD in either eye.

Participants underwent SD-OCT imaging using the Heidelberg Spectralis HRA + OCT, acquisition software version 6.0.13.0 (Heidelberg Engineering GmbH, Heidelberg, Germany). The preset Posterior Pole Scan Protocol was used, with scan fixation on the fovea. This protocol used a 30° horizontal × 25° vertical volume scan in high-speed mode, which included 768 A scans per B scan, and 61 B scans. For each eye, three images were acquired by a trained ophthalmic technician at the Kensington Eye Institute (Department of Ophthalmology and Vision Sciences, Toronto, ON).

Scans were automatically segmented using the Heidelberg Eye Explorer software (HEYEX version 6.3.4.0). One eye of each participant was randomly selected for analysis. One reference scan was chosen for analysis from the three that were acquired where the segmentation at initial evaluation was devoid of any obvious errors. Examples of obvious errors are those that were visible on quick inspection due to image acquisition errors (e.g., the inner limiting membrane [ILM] boundary following a hyperreflective vitreous base instead of the ILM) or pathology (e.g., the ILM boundary following an epiretinal membrane [ERM] instead of the ILM). The chosen scans also were required to have a quality score of at least 20 and automatic real time (ART) value of at

least 9. All measurements in this study were obtained using HEYEX.

Part 1: Frequency of Segmentation Line Error

On the central foveal scan for each eye, the length of eight boundary lines (ILM, retinal nerve fiber layer [RNFL], inner plexiform layer [IPL], inner nuclear layer [INL], outer plexiform layer [OPL], external limiting membrane [ELM], retinal pigment epithelium [RPE], and Bruch's membrane [BM]; Fig. 1) was measured by one of two trained observers (BW and RC) using a straight line (the measurement line) drawn from the nasal to temporal edges of each segmentation line. For all the segmentation lines in a given B scan, the end points of the measurement line were defined as the temporal and nasal locations at which all the boundary lines were correctly identified by the automated software. The proportion of the boundary line that deviated from what was deemed to be correct by a trained observer (Fig. 1) was also measured. The lengths of measured errors for each segmentation line were summed, then the sum was divided by the total length of the initially drawn line to acquire an “error ratio” (%) for the segmentation line of interest:

$$\left(\frac{\text{Sum of length of errors}}{\text{Total length of segmentation line}} \times 100 \right).$$

The mean, median, and range of error ratios were then calculated for each boundary line. Pairwise Wilcoxon rank sum tests with Holm-Bonferroni probability adjustment were performed for each group of lines to determine if there were differences in error ratios between NDD groups.

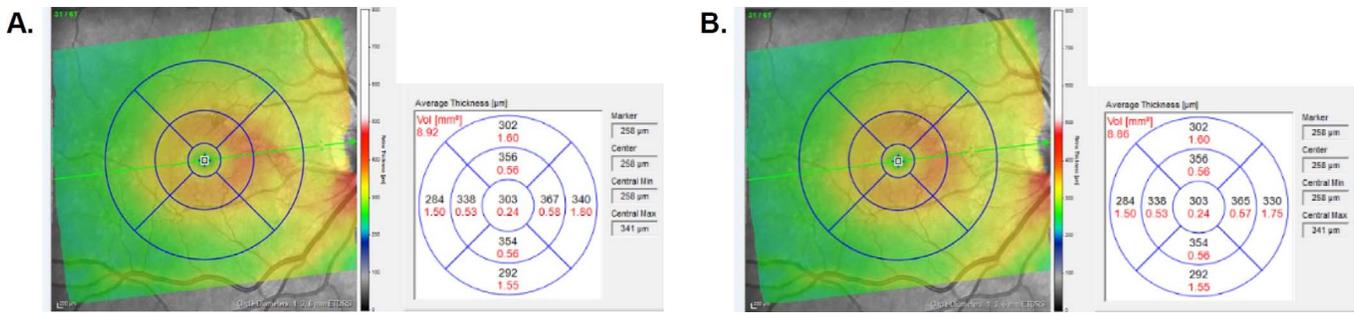


Figure 2. ETDRS grid with concentric circles of 1-, 3-, and 6-mm diameters overlaid on the en face image of a macula (right eye, OD) showing measurements (a) before and (b) after manual correction. The color-scaled images display a thickness map of the macula, while the adjacent grid contains the volume (red) and average thickness (black) of each macular sector. In this eye, the nasal inner and nasal outer sectors have reduced total retinal thickness and volume after manual correction of the segmentation lines.

Part 2: Agreement Between Software-Derived and Trained Observer-Derived Volume and Thickness Values

After automatic segmentation of the retinal layers by the Spectralis software, a grid with concentric circles of 1-, 3-, and 6-mm diameters was centered on the fovea (Fig. 2). Volume and average thickness measurements were obtained for the full retina plus each individual retinal layer inside each of the nine Early Treatment Diabetic Retinopathy Study (ETDRS) 1-, 3-, and 6-mm grid sectors. For each image, one of two trained observers (BW and RC) manually corrected erroneous portions of ILM, RNFL, OPL, and BM boundary lines in all cross-sectional B scans enclosed by the ETDRS grid, plus one scan immediately superior and inferior to the grid. The number of B scans manually corrected for each eye ranged from 55 to 57 depending on the dimensions of the eye.

After manual correction, volume and average thickness measurements were obtained for the total

retina (full thickness between ILM and BM segmentation lines) and each individual retinal layer inside the nine sectors of the grid. Volume and average thickness for both the software-generated and manually corrected scans were then calculated for the following layers in each sector as follows: (1) total retina, (2) RNFL, (3) inner retinal layers (IRL; sum of values of RNFL, ganglion cell layer [GCL], IPL, INL, and OPL), and (4) outer retinal layers (ORL; difference of total retina minus all inner retinal layers). Figure 3 illustrates how these layers are defined on a B scan.

Intraclass correlation coefficients (ICC)¹⁸ and Bland-Altman analyses were used to determine the differences in volume and average thickness of retinal layers from scans segmented by the automated software versus manual correction by the trained observers (R software R studio version 1.0.316; The R project for Statistical Computing, Vienna, Austria). Macula SD-OCT images comparing measurements before and after manual correction are shown in Figure 2.

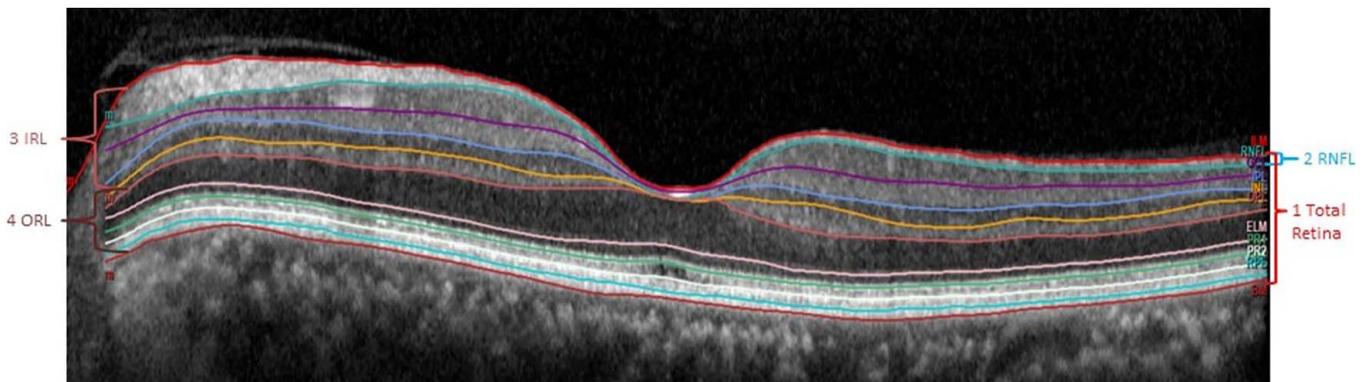


Figure 3. Foveal B scan showing the retinal layers of interest in this study. Total retina is measured from the ILM to the BM; IRL is a sum of the measurements of RNFL, GCL, IPL, INL, and OPL; ORL is the difference of the total retina minus the IRL.

Table 1. Mean and Median Error Ratios (no Units) for Segmentation Lines From Each Neurodegenerative Disease Group

Summary of Line Segmentation Data by Disease Group

Boundary	AD (<i>n</i> = 5)			FTD (<i>n</i> = 6)		
	Mean (SD)	Median	Range [Min, Max]	Mean (SD)	Median	Range [Min, Max]
ILM	0.13 (0.13)	0.11	[0.00, 0.34]	0.12 (0.06)	0.09	[0.07, 0.24]
RNFL	0.15 (0.11)	0.13	[0.04, 0.33]	0.23 (0.14)	0.25	[0.06, 0.38]
IPL	0.05 (0.05)	0.03	[0.00, 0.14]	0.04 (0.02)	0.04	[0.01, 0.07]
INL	0.10 (0.02)	0.11	[0.08, 0.13]	0.07 (0.04)	0.06	[0.02, 0.13]
OPL	0.02 (0.04)	0.00	[0.00, 0.09]	0.06 (0.05)	0.04	[0.02, 0.16]
ELM	0.10 (0.02)	0.11	[0.07, 0.12]	0.12 (0.07)	0.11	[0.04, 0.22]
RPE	0.13 (0.16)	0.04	[0.00, 0.35]	0.06 (0.11)	0.01	[0.00, 0.28]
BM	0.06 (0.02)	0.05	[0.04, 0.09]	0.07 (0.05)	0.09	[0.00, 0.12]

Table 1. Extended

Summary of Line Segmentation Data by Disease Group

Boundary	PD (<i>n</i> = 6)			VCI (<i>n</i> = 13)		
	Mean (SD)	Median	Range [Min, Max]	Mean (SD)	Median	Range [Min, Max]
ILM	0.20 (0.11)	0.18	[0.08, 0.39]	0.19 (0.12)	0.18	[0.04, 0.50]
RNFL	0.17 (0.18)	0.10	[0.07, 0.53]	0.15 (0.13)	0.10	[0.02, 0.47]
IPL	0.07 (0.04)	0.08	[0.00, 0.13]	0.09 (0.04)	0.09	[0.03, 0.15]
INL	0.08 (0.06)	0.06	[0.03, 0.20]	0.17 (0.10)	0.20	[0.03, 0.35]
OPL	0.04 (0.03)	0.03	[0.00, 0.08]	0.05 (0.04)	0.05	[0.00, 0.15]
ELM	0.09 (0.09)	0.05	[0.01, 0.24]	0.07 (0.05)	0.05	[0.02, 0.20]
RPE	0.22 (0.16)	0.21	[0.00, 0.49]	0.18 (0.14)	0.20	[0.04, 0.53]
BM	0.21 (0.25)	0.13	[0.00, 0.69]	0.10 (0.10)	0.07	[0.00, 0.29]

Results

Part 1: Frequency of Segmentation Line Error

Table 1 summarizes the error ratios of the eight segmentation lines in each disease group. The ILM segmentation line had one of the highest mean error ratios across all four NDD groups, at 13% for AD, 12% for FTD, 20% for PD, and 19% for VCI. RPE was also found to have relatively high mean error ratios compared with the other boundary lines, at 13% for AD, 22% for PD, and 18% for VCI. In the AD and FTD groups, the highest ratio was found with the RNFL line, at 15% and 23%, respectively.

Compared with other segmentation lines, OPL was found to have the lowest mean ratio for AD (2%), PD (4%), and VCI groups (5%), and second lowest for FTD (6%). IPL had low ratios for FTD (4%), AD

(5%), PD (7%), and VCI (9%). BM had low error ratios for most NDD groups, with 6% for AD, 7% for FTD, 10% for VCI, although it had one of the highest ratios in the PD group (21%).

Pairwise Wilcoxon rank sum tests with Holm-Bonferroni probability adjustment comparing boundary error rates across the four disease groups showed that any observable boundary error rate variability was not statistically significant.

Part 2: Agreement Between Software-Derived and Trained Observer-Derived Volume and Thickness Values

Based on ICC analyses, there was excellent agreement between trained observer-derived and software-derived total retinal volume (0.999) and thickness (0.996), RNFL volume (0.998) and thickness (0.978), IRL volume (0.999) and thickness

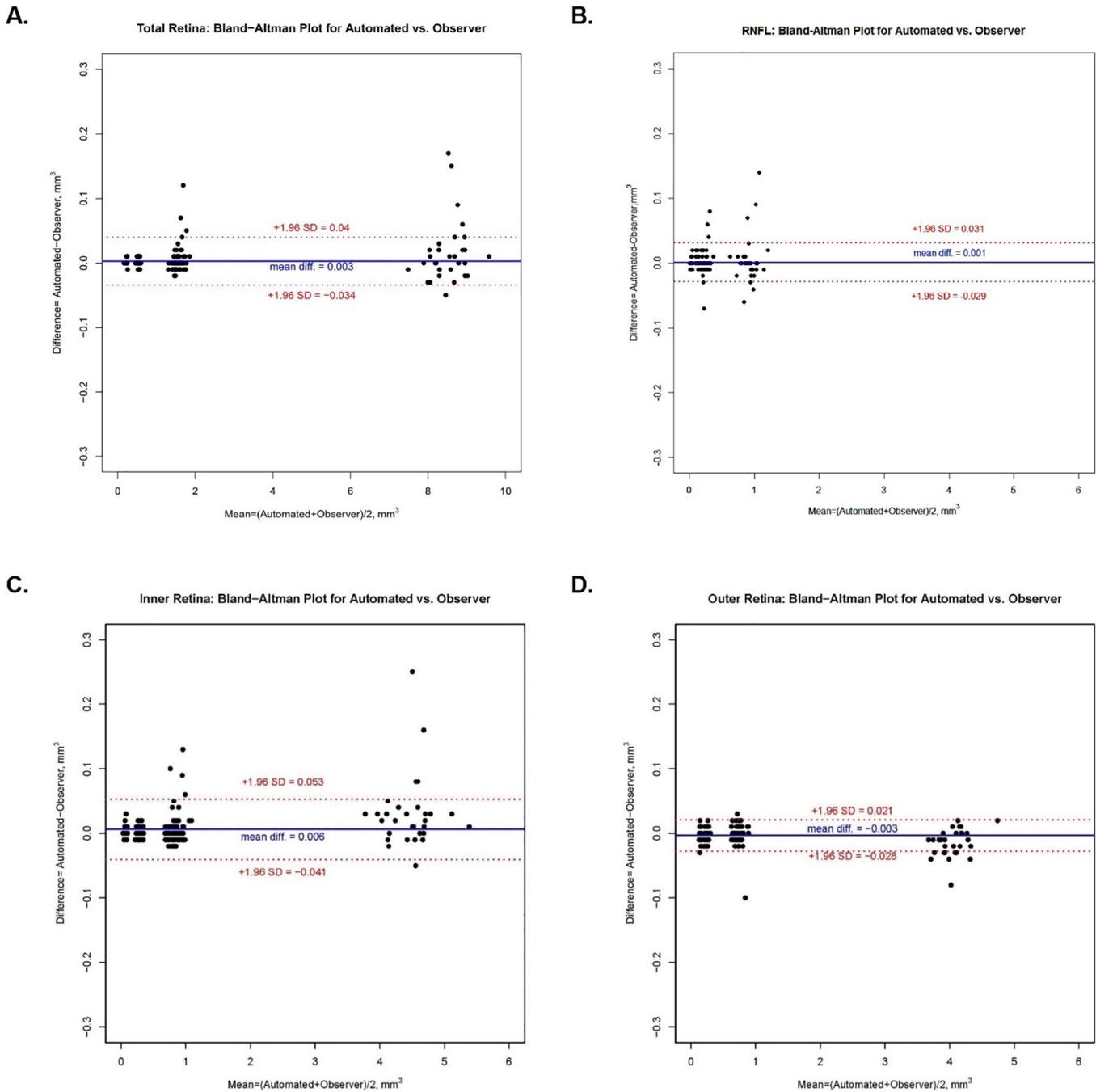


Figure 4. Bland-Altman plots illustrating the difference (automated – observer) in volume versus the mean ($[(\text{automated} + \text{observer}) / 2]$) volume for automated delineation and manual correction of retinal segmentation lines for (a) total retina, (b) RNFL, (c) IRL, and (d) ORL.

(0.991), and ORL volume (0.999) and thickness (0.979) (Table 2).

Bland-Altman Plots (Figs. 4, 5) illustrate the mean differences (software generated – manual generated) in volume of 0.003, 0.001, 0.006, and -0.003 mm^3 , respectively, for total retina, RNFL, IRL, and ORL. Respective mean differences in thickness between

software and observers for the four above groups of layers were 0.367, 0.492, 1.855, and $-1.488 \mu\text{m}$. Table 3 illustrates the 95% limits of agreement for the nine sectors of the ETDRS grid for each group of retinal layers.

Analyzing volume by individual macular sectors, the nasal outer and superior outer sectors showed

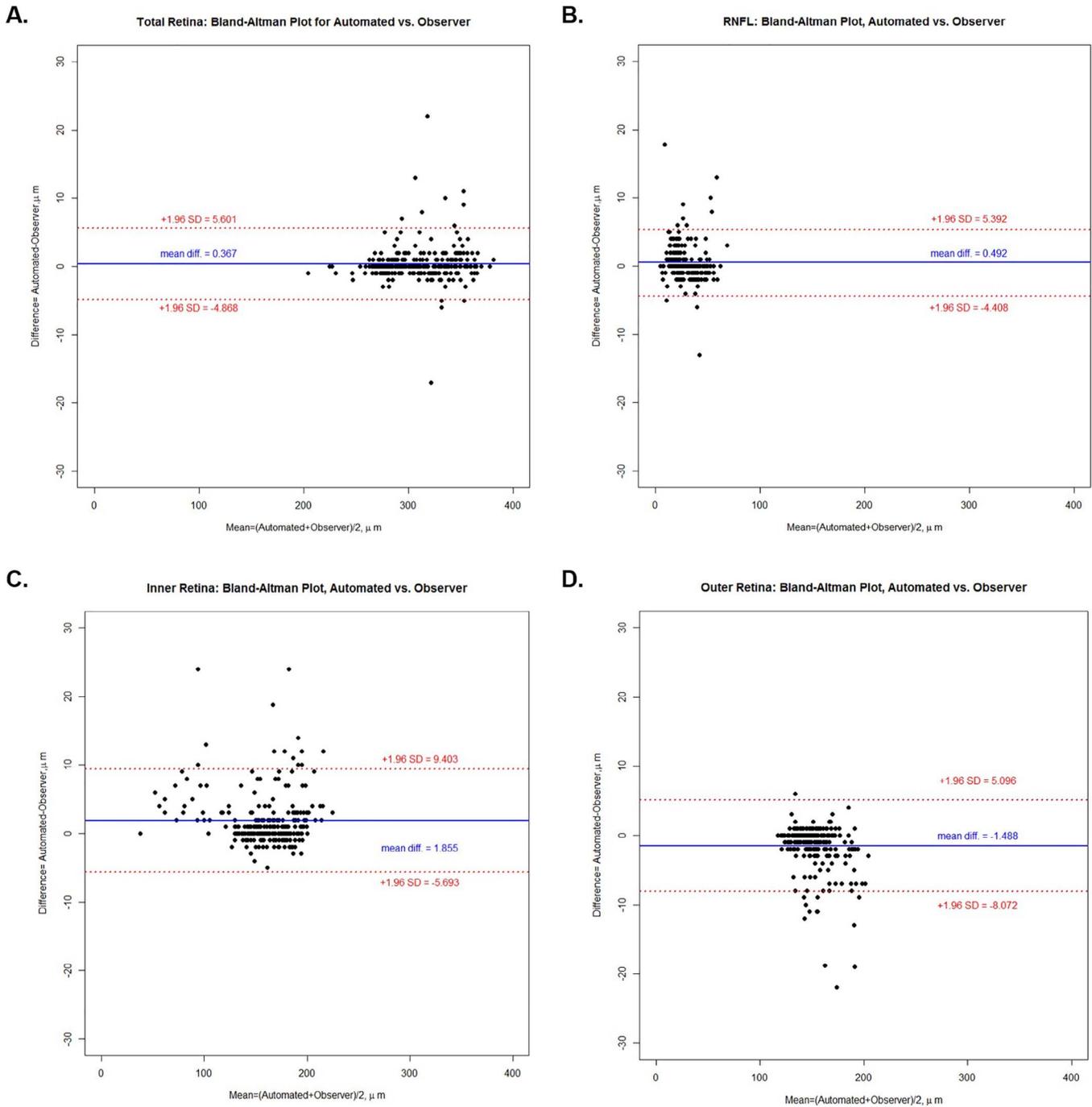


Figure 5. Bland-Altman plots illustrating the difference (automated – observer) in thickness versus the mean $([automated + observer] / 2)$ thickness for automated delineation and manual correction of retinal segmentation lines for (a) total retina, (b) RNFL, (c) IRL, and (d) ORL.

the largest range of limits of agreement for the total retina, RNFL, and IRL layers. For the ORL, the largest range of limits of agreement were found in the superior outer and inferior outer sectors. With respect to average thickness, the nasal outer sector, followed by central macula, showed the largest

range of limits of agreement for total retina and IRL. For RNFL, the nasal outer sector also had the highest range, followed by the temporal inner sector. For ORL, the highest ranges were found in the central macula, followed by the temporal inner sector.

Table 2. ICC Values Illustrating Agreement Between Automated Software Versus Manual Correction

Layer	Volume	Average Thickness
Total retina	0.999	0.996
RNFL	0.998	0.978
IRL	0.999	0.991
ORL	0.999	0.979

Discussion

For SD-OCT to be clinically useful for the assessment of retinal morphology in NDD, it is essential to ensure that segmentation software is valid when compared with expert human assessment. In this study, we assessed both the segmentation agreement of the automated Spectralis software for eight boundary lines and the agreement between trained observer-derived and automated software-derived volume and thickness of retinal layers in the macula. In part 1 of the study, the highest mean error ratios were observed with the ILM, RNFL, and RPE segmentation lines, which means that the automated software delineated those boundary lines with errors more frequently than other lines. In part 2 of the study, we found excellent agreement between software generated and manually corrected retinal thickness and volume outcomes. Although the error ratios for some lines were as high as 0.23 (RNFL/FTD), excellent agreement between software and observer indicates that the small-scale segmentation errors do not have a large effect on the final measurements of the layers. Interestingly, the RPE line (22%) and the BM line (21%) were relatively high in the PD group.

Part 1: Frequency of Segmentation Line Error

Although we excluded obvious pathology, including large ERMs, from our sample, it is important to inspect images and the automated segmentation in retinas with pathology. ERMs can cause the automated software to misinterpret the ERM for the ILM. High error rates in the RPE line may be due to the poor definition and subtle contrast of the RPE against the adjacent photoreceptor outer segment layer, as suggested by Liu et al.¹⁹ Additionally, Lang et al.²⁰ suggest that the outer segment–RPE boundary is more difficult to visualize away from the fovea, as the photoreceptors transition from mostly cones around the fovea to mostly rods at the outer macula.

The BM, IPL, and OPL segmentation lines had the lowest error ratios. A likely reason for this is because the high contrast between the layers on either side of the line make it easier for the software algorithm to detect. However, Staurengi et al.²¹ suggested that the ONL may have been mislabeled by some OCT systems because its inner portion actually consists of Henle's fiber layer (HFL). According to a study by Lujan et al.,²² the reflectivity of HFL actually varies depending on the eccentricity of the OCT beam entering the pupil; the HFL appears thicker on the side of the fovea opposite to the direction that the beam is decentered. Consequently, the position of the OPL line and thickness of the measured ONL layer can vary depending on the position of OCT beam entry.

The GCL segmentation line was not included in our error analysis because of the difficulty in discriminating the contrast between its two surrounding layers (GCL and IPL) on the scan. Lang et al.²⁰ also report that this boundary tends to be

Table 3. Limits of Agreement for Volume and Average Thickness in Each Sector of the 6-mm Macular Grid Between Scans That Had Automated Delineated Lines Versus Manually Corrected Lines

Layer	Full Volume	Central Macula	Superior Inner	Temporal Inner	Inferior Inner	Nasal Inner	Superior Outer	Temporal Outer	Inferior Outer	Nasal Outer
Volume, mm ³										
Total retina	±0.095	±0.007	±0.010	±0.005	±0.006	±0.010	±0.019	±0.010	±0.015	±0.055
RNFL	±0.074	±0.004	±0.008	±0.007	±0.010	±0.012	±0.019	±0.012	±0.016	±0.048
IRL	±0.110	±0.018	±0.014	±0.015	±0.017	±0.016	±0.043	±0.013	±0.023	±0.062
ORL	±0.041	±0.020	±0.013	±0.017	±0.017	±0.017	±0.039	±0.017	±0.018	±0.017
Average thickness, μm										
Total retina		±6.790	±3.903	±2.747	±3.729	±5.882	±3.843	±1.851	±2.816	±10.107
RNFL		±4.348	±3.877	±6.761	±2.741	±4.358	±3.564	±2.367	±2.912	±8.384
IRL		±9.156	±6.893	±7.030	±8.128	±7.630	±4.420	±2.042	±3.556	±10.406
ORL		±10.402	±5.947	±7.505	±7.227	±6.509	±1.707	±1.622	±2.720	±2.298

indistinguishable in OCT images. Because the transition point between the GCL and IPL is extremely difficult to discern with current OCT technology, an alternative option for segmentation software could be to combine the two layers and label the resultant layer as the GCL–IPL complex instead, to prevent erroneous measurements for the individual layers.

Part 2: Agreement Between Software-Derived and Trained Observer-Derived Volume and Thickness Values

Our finding of excellent agreement between software and trained observer in NDD is in agreement with studies by Loh et al.²³ and Polo et al.,²⁴ which found excellent repeatability and validity of RNFL and total retinal thickness measurements using SD-OCT systems in a NDD population. To the best of our knowledge, our study is the first to analyze the HEYEX software for macular sublayer segmentation in NDD. Cetinkaya et al.²⁵ found excellent agreement with repeated measurements using the Spectralis HEYEX software on healthy participants for all individual retinal sublayer thickness values. Heussen et al.²⁶ compared automated software with manual correction in healthy participants, and found that manual correction of inner and outer retinal boundary errors yields mean differences of less than 6 μm , which is similar to the axial resolution of SD-OCT devices. This provides further support for the excellent agreement between software and manual correction that we found in our study.

This study found that the mean difference in total retinal volume between scans segmented by software and trained observer was 0.016 mm^3 , and the difference in total retinal volume in the central macular sector was 0.0007 mm^3 . Although no literature to date has discussed the amount of change required to be clinically significant in NDD populations, a study by Tah et al.²⁷ on 73 eyes with AMD reports that a change in volume of greater than 0.050 mm^3 or thickness of greater than 64 μm in the central 1-mm sector is needed to distinguish clinical change from measurement variability.²⁷ The differences in agreement found in this study are less than the values proposed in the study by Tah et al.,²⁷ and therefore are unlikely to be clinically significant.

Similar to volume differences, the mean differences between software- and observer-generated thickness values were small, with the largest difference being 1.855 μm for the IRL. The mean difference in total

retinal thickness in all sectors between software- and observer-corrected data was 0.367 μm . These overall low variabilities are unlikely to be clinically significant given that a normal foveal central subfield thickness is approximately 237 μm and that it is difficult to achieve a precision level better than 5 μm when manually delineating a boundary line using a computer mouse.²⁶

The highest variability between software and observer was found with the nasal outer sector for total retina, RNFL, and IRL, likely due to higher variability of thickening of the RNFL as more axons congregate to form the optic nerve. Although the small ERMs caused some automated segmentation errors, our analysis shows that they had little effect on volume or thickness measurements.

The analysis of total retina, RNFL and IRL all showed a positive mean difference (Figs. 4a–c, 5a–c) with a tendency for some points to be above the +1.96 SD confidence limits line, while ORL volume and ORL thickness (Figs. 4d, 5d, respectively) showed a negative mean difference (automated – observer) with a tendency for some individual points to be distributed below the –1.96 SD confidence limits line. This demonstrates a tendency for automated analysis to be greater than observer analysis for total retina, RNFL, and IRL, but lower for ORL. The observation might suggest that there is some bias either by the software or by the human observers, in the delineation of the segmentation of the retinal layers; however, the very high ICC values show that the effect was very small.

This study has some limitations to consider. The method of measuring error ratios of segmentation lines only took into account what proportion of the line was erroneous, but did not evaluate the magnitude or direction of the discrepancy in boundary identification. As a result, a line that has errors of equal magnitude but in opposite directions may not show a significant difference in volume or thickness before versus after correction, even if there actually was a difference. However, the Bland-Altman analysis takes equal and opposite differences into account. Nevertheless, the confidence intervals were narrow. A second limitation was that we only included images without obvious delineation errors from the automated software resulting from retinal pathologies or acquisition errors. However, our conclusions that the automated software is in excellent agreement with trained observers remains valid on condition that images with obvious errors are manually corrected. A third limitation is that although participants with wet AMD were excluded in this study, some participants had small drusen that could have disrupted the segmentation for the RPE

line. Despite these small drusen, the agreement found between software versus observer-derived volume and thickness measurements was still excellent. Finally, the methodology dictated that the automated segmentation analysis was always conducted first and then the manual correction was undertaken from that starting point. Although it might be interesting to examine expert-defined retinal segmentation performance without the initial advantage of starting with the automated segmentation, this never was the aim of the study.

Our findings indicate that in those SD-OCT images without obvious delineation errors, the SD-OCT software can validly measure retinal thickness and volume. Because each make of SD-OCT instrument has different properties, such as software segmentation algorithm, axial resolution, and signal-to-noise ratio, the results from this study apply specifically to the Heidelberg Spectralis SD-OCT but also have a level of general relevance. Future studies should assess the validity of the segmentation software for each individual retinal layer in order to investigate subtler potential neurodegenerative changes.

Future longitudinal analyses of retinal layer volume and thickness in the NDD cohorts, including those in the ONDRI study, can be performed efficiently using automated segmentation software, with the knowledge that in the absence of obvious delineation errors, manual correction is not required to yield valid measurements.

Acknowledgments

We thank Ann Lvin, Lori Henderson, Kari Stuart, and Vera Stiuso at the Kensington Eye Institute for technical assistance.

Supported by grants from the ONDRI through the Ontario Brain Institute, an independent nonprofit corporation, funded partially by the Ontario government, and from the Toronto Western Hospital Practice Plan (Dr. Robert Devenyi), and the Canadian Optometry Education Trust Fund (COETF).

This work was presented, in part, as a poster at the Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO) in Baltimore, MD on May 8, 2017.

Disclosure: **B.M. Wong**, None; **R.W. Cheng**, None; **E.D. Mandelcorn**, Novartis, Bayer, Optos, Bausch + Lomb (R); **E. Margolin**, None; **S. El-Defrawy**, None; **P. Yan**, None; **A.T. Santiago**, None; **E. Leontieva**,

None; **W. Lou**, None; **W. Hatch**, None; **C. Hudson**, None

References

1. Kim BY, Smith SD, Kaiser PK. Optical coherence tomographic patterns of diabetic macular edema. *Am J Ophthalmol*. 2006;142:405–412.
2. Sleiman K, Veerappan M, Winter KP, et al. Optical coherence tomography predictors of risk for progression to non-neovascular atrophic age-related macular degeneration. *Ophthalmology*. 2017;124:1764–1777.
3. Lisboa R, Leite MT, Zangwill LM, Tafreshi A, Weinreb RN, Medeiros FA. Diagnosing preperimetric glaucoma with spectral domain optical coherence tomography. *Ophthalmology*. 2012;119:2261–2269.
4. Wolf-Schnurrbusch UE, Ceklic L, Brinkmann CK, et al. Macular thickness measurements in healthy eyes using six different optical coherence tomography instruments. *Invest Ophthalmol Vis Sci*. 2009;50:3432–3437.
5. Ascaso FJ, Cruz N, Modrego PJ, et al. Retinal alterations in mild cognitive impairment and Alzheimer's disease: an optical coherence tomography study. *J Neurol*. 2014;261:1522–1530.
6. Hajee ME, March WF, Lazzaro DR, et al. Inner retinal layer thinning in Parkinson disease. *Arch Ophthalmol*. 2009;127:737–741.
7. Chorostecki J, Seraji-Bozorgzad N, Shah A, et al. Characterization of retinal architecture in Parkinson's disease. *J Neurol Sci*. 2015;355:44–48.
8. Kirbas S, Turkyilmaz K, Tufekci A, Durmus M. Retinal nerve fiber layer thickness in Parkinson disease. *J Neuroophthalmol*. 2013;33:62–65.
9. Gao L, Liu Y, Li X, Bai Q, Liu P. Abnormal retinal nerve fiber layer thickness and macula lutea in patients with mild cognitive impairment and Alzheimer's disease. *Arch Gerontol Geriatr*. 2015;60:162–167.
10. Cunha JP, Proenca R, Dias-Santos A, et al. OCT in Alzheimer's disease: thinning of the RNFL and superior hemiretina. *Graefes Arch Clin Exp Ophthalmol*. 2017;255:1827–1835.
11. den Haan J, Janssen SF, van de Kreeke JA, Scheltens P, Verbraak FD, Bouwman FH. Retinal thickness correlates with parietal cortical atrophy in early-onset Alzheimer's disease and controls. *Alzheimers Dement (Amst)*. 2018;10:49–55.

12. Pillai JA, Bermel R, Bonner-Jackson A, et al. Retinal nerve fiber layer thinning in Alzheimer's disease: a case-control study in comparison to normal aging, Parkinson's disease, and non-Alzheimer's dementia. *Am J Alzheimers Dis Other Demen.* 2016;31:430–436.
13. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol.* 2012;71:266–273.
14. Subramaniam NS, Bawden CS, Waldvogel H, Faull RML, Howarth GS, Snell RG. Emergence of breath testing as a new non-invasive diagnostic modality for neurodegenerative diseases. *Brain Res.* 2018;1691:75–86.
15. Farhan SM, Bartha R, Black SE, et al. The Ontario Neurodegenerative Disease Research Initiative (ONDRI). *Can J Neurol Sci.* 2017;44:196–202.
16. Ctori I, Huntjens B. Repeatability of foveal measurements using Spectralis optical coherence tomography segmentation software. *PLoS One.* 2015;10:e0129005.
17. Krebs I, Smretschnig E, Moussa S, Brannath W, Womastek I, Binder S. Quality and reproducibility of retinal thickness measurements in two spectral-domain optical coherence tomography machines. *Invest Ophthalmol Vis Sci.* 2011;52:6925–6933.
18. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
19. Liu X, Shen M, Huang S, Leng L, Zhu D, Repeatability Lu F. and reproducibility of eight macular intra-retinal layer thicknesses determined by an automated segmentation algorithm using two SD-OCT Instruments. *PLoS One.* 2014;9:e87996.
20. Lang A, Carass A, Hauser M, et al. Retinal layer segmentation of macular OCT images using boundary classification. *Biomed Opt Express.* 2013;4:1133–1152.
21. Staurengi G, Sadda S, Chakravarthy U, Spaide RF; for the International Nomenclature for Optical Coherence Tomography (IN·OCT) Panel. Proposed lexicon for anatomic landmarks in normal posterior segment spectral-domain optical coherence tomography: the IN·OCT consensus. *Ophthalmology.* 2014;121:1572–1578.
22. Lujan BJ, Roorda A, Knighton RW, Carroll J. Revealing Henle's fiber layer using spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2011;52:1486–1492.
23. Loh EH, Ong YT, Venketasubramanian N, et al. Repeatability and reproducibility of retinal neuronal and axonal measures on spectral-domain optical coherence tomography in patients with cognitive impairment. *Front Neurol.* 2017;8:359.
24. Polo V, Garcia-Martin E, Bambo MP, et al. Reliability and validity of Cirrus and Spectralis optical coherence tomography for detecting retinal atrophy in Alzheimer's disease. *Eye (Lond).* 2014;28:680–690.
25. Cetinkaya E, Duman R, Duman R, Sabaner MC. Repeatability and reproducibility of automatic segmentation of retinal layers in healthy subjects using Spectralis optical coherence tomography. *Arg Bras Oftalmol.* 2017;80:378–381.
26. Heussen FM, Ouyang Y, McDonnell EC, et al. Comparison of manually corrected retinal thickness measurements from multiple spectral-domain optical coherence tomography instruments. *Br J Ophthalmol.* 2012;96:380–385.
27. Tah V, Keane PA, Esposti SD, et al. Repeatability of retinal thickness and volume metrics in neovascular age-related macular degeneration using the Topcon 3DOCT-1000. *Indian J Ophthalmol.* 2014;62:941–948.

Appendix: ONDRI Investigators

Robert Bartha, Robarts Research Institute, Western University, London, Ontario, Canada
 Sandra E. Black, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
 Michael Borrie, St. Joseph's Health Care London, London, Ontario, Canada
 Dale Corbett, University of Ottawa, Ottawa, Ontario, Canada
 Elizabeth Finger, St. Joseph's Health Care London, London, Ontario, Canada
 Morris Freedman, Baycrest Hospital, Toronto, Ontario, Canada
 Barry Greenberg, Johns Hopkins University, Baltimore, Maryland, USA
 David A. Grimes, Ottawa Hospital, Ottawa, Ontario, Canada
 Robert A. Hegele, Western University, London, Ontario, Canada
 Christopher Hudson, School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario, Canada
 Anthony E. Lang, Toronto Western Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada

Mario Masellis, Department of Medicine (Neurology), Sunnybrook HSC, University of Toronto, Toronto, Ontario, Canada

William E. McIlroy, Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada

Paula M. McLaughlin, Western University, London, Ontario, Canada

Manuel Montero-Odasso, St. Joseph's Health Care London, London, Ontario, Canada

David G. Munoz, St. Michael's Hospital, Toronto, Ontario, Canada

Douglas P. Munoz, Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada

J. B. Orange, School of Communication Sciences & Disorders, Western University, London, Ontario, Canada

Michael J. Strong, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada

Stephen C. Strother, Baycrest Hospital, Toronto, Ontario, Canada

Richard H. Swartz, Department of Medicine (Neurology), Sunnybrook HSC, University of Toronto, Toronto, Ontario, Canada

Sean Symons, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

Maria Carmela Tartaglia, Toronto Western Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada

Angela Troyer, Baycrest Hospital, Toronto, Ontario, Canada

Lorne Zinman, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada