

Test–retest reliability of Cirrus HD-optical coherence tomography retinal layer thickness measurements in people with multiple sclerosis

Anna Bacchetti , Brenna McCormack , Ting-Yi Lin , Rozita Doosti, Gelareh Ahmadi, Omar Ezzedin, Nicole Pellegrini, Evan Johnson, Anna Kim , Gabriel Otero-Duran, Devon J Bonair, Elle Lawrence, Ernest Lievers , Simidele Davis, Sooyeon Park, Madeline Inserra, Ananya Gulati, Kathryn C Fitzgerald , Elias S Sotirchos , Peter A Calabresi  and Shiv Saidha 

Multiple Sclerosis Journal—
Experimental, Translational
and Clinical

April–June 2025, 1–14

DOI: 10.1177/
20552173251340957

© The Author(s), 2025.
Article reuse guidelines:
sagepub.com/journals-
permissions

Abstract

Background: Optical coherence tomography (OCT) allows evaluation of inter-eye differences (IEDs) in peri-papillary retinal nerve fiber layer (pRNFL) and macular ganglion cell-inner plexiform layer (GCIPL) thicknesses to identify unilateral optic nerve involvement (UONI), which is included in the 2024 revised McDonald diagnostic criteria for multiple sclerosis (MS).

Objectives: To evaluate the test–retest reliability of pRNFL and GCIPL thicknesses/IEDs in people with MS, other neurological disorders, and healthy controls using Cirrus HD-OCT.

Methods: 509 participants underwent Cirrus HD-OCT, acquiring two macular and optic disc scans per eye within each session. Scans meeting OSCAR-IB quality control criteria were included in final analyses (959 eyes), with no clinical/demographic exclusions (reflecting a real-world clinical setting). Reliability was assessed using coefficients of variation (COVs), intraclass correlation coefficients (ICCs), and Bland–Altman limits of agreement (LOA). IED consistency was evaluated using difference-in-differences (DiDs) of test–retest measurements.

Results: GCIPL demonstrated superior reliability (ICC: 0.998, COV: 0.40%, LOA: -1.29 to 1.35 μm) to pRNFL (ICC: 0.989, COV: 1.18%, LOA: -3.59 to 3.70 μm) thickness. Inter-eye absolute DiDs [pRNFL: 2.00 μm (standard deviation (SD) 1.73); GCIPL: 0.64 μm (SD 0.67)] were lower than IED thresholds proposed for identifying UONI.

Conclusions: The excellent reliability of GCIPL and pRNFL thicknesses/IEDs support OCT for identifying UONI to diagnose MS.

Keywords: Diagnosis, imaging, multiple sclerosis, OCT, reliability, test–retest

Date received: 20 February 2025; accepted 23 April 2025

Introduction

Multiple sclerosis (MS) is an inflammatory, demyelinating, and neurodegenerative disorder of the central nervous system.¹ Despite its varied clinical presentation, involvement of the visual system is almost universal in MS,^{2–4} with optic neuritis (ON) being the most frequent clinical manifestation of afferent visual pathway injury in people with MS (PwMS).^{5,6} The 2024 revised McDonald diagnostic criteria for MS will include optic nerve involvement as a fifth topographical region for

dissemination in space.⁷ This update reflects evidence that demyelinating optic nerve injury, detectable by optical coherence tomography (OCT), visual evoked potentials, and/or magnetic resonance imaging, is often present early in the MS disease course and improves the diagnostic performance of the 2017 revised McDonald criteria for MS diagnosis.^{8–10}

An important method for detecting unilateral optic nerve involvement (UONI) and establishing dissemination in

Correspondence to:
Shiv Saidha,
Department of Neurology,
Division of
Neuroimmunology &
Neuroinfectious diseases,
Johns Hopkins University
School of Medicine, 600 N
Wolfe St, Baltimore, MD,
21287, USA.
ssaidha2@jhmi.edu

Anna Bacchetti,
Department of Neurology,
Johns Hopkins University



School of Medicine,
Baltimore, Maryland, USA.

Brenna McCormack,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Ting-Yi Lin,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Rozita Doosti,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Gelareh Ahmadi,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Omar Ezzedin,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Nicole Pellegrini,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Evan Johnson,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Anna Kim,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Gabriel Otero-Duran,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Devon J Bonair,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Elle Lawrence,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Ernest Lievers,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Simidele Davis,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Sooyeon Park,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

space is OCT-derived inter-eye differences (IEDs) in peri-papillary retinal nerve fiber layer (pRNFL) and/or ganglion cell-inner plexiform layer (GCIPL) thicknesses.^{11–14} Regardless of using Cirrus HD-OCT or Spectralis, two of the most commonly utilized OCT platforms,¹⁵ optimal IEDs that have been identified for distinguishing MS eyes with, from without, a known history of prior ON (i.e. identifying UONI), are $\geq 6 \mu\text{m}$ for pRNFL and/or $\geq 4 \mu\text{m}$ for GCIPL thicknesses. These IEDs either approximate or exceed the 95th percentile for IEDs in healthy controls (HCs).^{13,16}

Cirrus HD-OCT is a high-resolution, commercially available spectral-domain OCT platform. Previous studies demonstrate high test–retest reliability and reproducibility of spectral-domain OCT measurements in HCs and individuals with ophthalmological diseases.^{17,18} However, data on OCT measurement reliability, particularly in a large cohort of PwMS, and across a wide range of measures, is limited.^{19,20} This is particularly important to elucidate, in order to ensure appropriate utilization and interpretation, as OCT begins to transition into mainstream clinical use for the diagnosis of MS.

This study aims to determine the test–retest reliability of pRNFL and GCIPL measurements in eyes of a large cohort of PwMS using Cirrus HD-OCT, including IEDs in these measurements, thereby validating the practicality and robustness of proposed pRNFL and/or GCIPL IEDs for detecting UONI and supporting dissemination in space, per the 2024 revised McDonald diagnostic criteria for MS.

Material and methods

Participants

PwMS and people with other neurological disorders (PwOND) were prospectively recruited from the Johns Hopkins MS Center between 2017 and 2024. HCs were also recruited from among Johns Hopkins staff. The diagnosis of MS was confirmed by the treating neurologist according to the 2017 revised McDonald criteria.²¹ According to the Lublin classification, participants with MS were categorized as having one of the following MS subtypes: relapsing-remitting MS (RRMS), secondary progressive MS (SPMS), or primary progressive MS (PPMS).²² Some participants attended multiple visits during the study period, with each visit considered as an independent test–retest assessment. Demographic and clinical data were collected, including disease duration, history of ON with date and side

of occurrence, as well as 100% and 2.5% contrast visual acuity.

No exclusion criteria were applied to the demographic or clinical characteristics of the participants, in order to recruit and assess a study population that is reflective of a real-world clinical setting. Various retinal and/or ocular diseases have the potential to cause asymmetric thinning of the pRNFL and/or GCIPL, and should be considered when evaluating IEDs.²³ This in itself underpins the importance of thorough quality control (QC) of OCT scans. That being said, this study focused on assessing the reliability of OCT measures, rather than comprehensively evaluating the relationships of IEDs with medical history and related factors. While retinal and/or ocular diseases may influence IEDs, they are not expected to affect retinal measurement reliability, provided they do not affect the quality of the OCTs (i.e. if they do not interfere with layer segmentation).^{17,18}

Data collection was standardized to ensure consistency.

Optical coherence tomography

All participants underwent spectral-domain Cirrus HD-OCT (model 5000, software version 8.1; Carl Zeiss Meditec, Dublin, California). Optic disc scans were acquired using the Optic Disc Cube 200×200 protocol to evaluate pRNFL thicknesses, and macular scans using the Macular Cube 512×128 protocol to assess GCIPL thicknesses, both generated through the Cirrus HD-OCT incorporated automated segmentation.²⁰ We used the Cirrus HD-OCT's built-in segmentation algorithm for generalizability to real-world clinical practice, as it provides immediate results, whereas custom, in-house segmentation techniques, although potentially more reliable, may be time-consuming and impractical in routine clinical practice.²⁴

Experienced technicians obtained two sets of scans per eye under low-light conditions. After the initial acquisition of optic nerve and macular scans from the right and then the left eye, participants were instructed to sit back for a brief moment, before re-acquiring all scans from each eye once more. All OCT scans underwent QC assessments according to the OSCAR-IB criteria [(O) obvious problems, (S) poor signal strength, (C) centration of scan, (A) algorithm failure, (R) retinal pathology (I) illumination and (B) beam placement].²⁵ Only those eyes for which both scans met the OSCAR-IB criteria and had a signal strength of 7 or greater were included in the analysis, ensuring that only high-quality scans were considered.

Reporting of OCT methods and results adhered to the Advised Protocol for OCT Study Terminology and Elements (APOSTEL) recommendations.^{26,27}

Statistical analyses

Statistical analyses were conducted using STATA version 18 (StataCorp, College Station, TX). Statistical significance was defined as $p < 0.05$. Comparisons of demographic and clinical variables were performed using the chi-square test for categorical variables, ANOVA for normally distributed, and the Kruskal–Wallis test for non-normally distributed continuous variables. The mean absolute differences (MADs) and the mean absolute percentage differences (MAPDs), along with their standard deviations (SDs), were calculated for all groups to summarize the variability of pRNFL and GCIPL thicknesses.

The intra-visit reproducibility of pRNFL and GCIPL thicknesses was assessed using intra-visit intraclass correlation coefficients (ICCs) and coefficients of variation (COVs). The ICC was calculated using a two-way mixed-effects model with absolute agreement, with values greater than 0.9 indicating excellent reproducibility. The COV was defined as the SD divided by the average thickness of each set, expressed as a percentage. Bland–Altman analyses were performed to compare the differences in thickness measurements, illustrating the mean differences (agreement at the cohort level) and the limits of agreement with a 95% confidence interval (CI) [reflecting variability and agreement at the individual level].²⁸

The consistency of IEDs was evaluated by analyzing the difference-in-differences between eyes across separate tests, determined as the absolute value of the difference between IEDs from two repeated measurements.

Regression analyses were conducted to evaluate the impact of clinical factors (100% and 2.5% contrast visual acuity, time elapsed since ON, and disease duration in PwMS) and layer thicknesses on OCT reliability, measured by the MAPD. These correlations were adjusted for age and gender using a linear mixed-effects regression model.

Ethical approval and data availability

Johns Hopkins University Institutional Review Board approved the study and written informed consent was obtained from all participants. Data related to this study are available based upon reasonable request and subject to proper inter-institutional data-sharing agreements.

Results

Baseline characteristics of study participants

A total of 509 participants were enrolled, with 495 people (959 eyes) included in the test–retest analyses (Figure 1). The cohort consisted of 385 PwMS, 64 HCs, and 46 PwOND (Table 1). Among PwMS, there were 324 RRMS, 15 PPMS, and 46 SPMS. The median number of visits per patient was 1 (range: 1–4), and per eye was 1 (range: 1–4). The PwOND group (Table S1) included aquaporin-4 seropositive neuromyelitis optica spectrum disorder ($n = 13$), myelin oligodendrocyte glycoprotein antibody-associated disease ($n = 12$), prior idiopathic ON ($n = 10$), neurosarcoidosis ($n = 3$), acute idiopathic ON ($n = 2$), stiff person syndrome ($n = 2$), chronic relapsing inflammatory optic neuropathy ($n = 1$), cobalamin C disease ($n = 1$), epilepsy ($n = 1$), and transverse myelitis ($n = 1$).

The mean age was highest in PwMS (46.1 years; SD 11.9 years), lowest in HCs (37 years; SD 14.7 years), and intermediate in PwOND (44.3 years; SD 20.2 years). Among the PwMS subtypes, the higher mean age was observed in the PPMS (58 years; SD 5.2 years) and SPMS (55.6 years; SD 9.2 years), relative to the RRMS cohort. There was a higher proportion of females across all groups, with PwMS (79.5% females) exhibiting a significantly greater percentage of females than males ($p = 0.001$), as compared to the HCs (59.4% females). As expected, individuals with SPMS had a significantly longer disease duration (15.8 years, SD 8.3 years) relative to those with RRMS (9.7 years, SD 7.1 years) and PPMS (9.7 years, SD 6.9 years). The median time elapsed between ON and OCT acquisition was 119 months (IQR: 61–192 months) for PwMS and 32 months (IQR: 12–96 months) for PwOND.

Regarding 2.5% and 100% visual acuity, eyes of HCs and PwMS exhibited better scores than PwOND ($p < 0.001$ for all).

Comparison of OCT reliabilities

GCIPL thickness reliability

The mean GCIPL thickness was significantly lower in PwMS and PwOND, as compared to HCs ($p < 0.001$), with mean values of 73.46 μm (SD 9.85 μm) in PwMS, 71.94 μm (SD 12.08 μm) in PwOND, and 81.27 μm (SD 6.55 μm) in HCs (Table 2). The MADs between the two measurement sessions were minimal across all groups ($p = 0.81$), being 0.41 μm (SD 0.53 μm) in PwMS, 0.39 μm (SD 0.52 μm) in HCs, and 0.44 μm (SD 0.55 μm) in PwOND,

Madeline Inserra,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Ananya Gulati,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Kathryn C Fitzgerald,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Elias S Sotirchos,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Peter A Calabresi,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

Shiv Saidha,
Department of Neurology,
Johns Hopkins University
School of Medicine,
Baltimore, Maryland, USA.

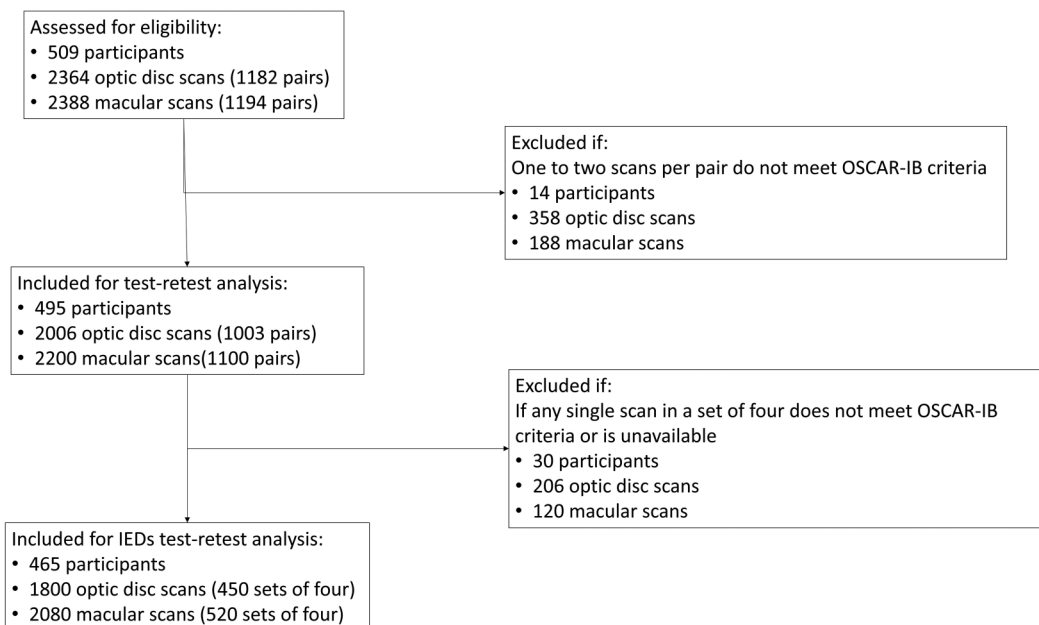


Figure 1. CONSORT flow chart illustrating participant inclusion and exclusion criteria. Scans were grouped as follows: test–retest pairs (two scans for one eye in a single session), IEDs test–retest sets of four (two scans for both eyes in a single session). For macular scans, 775 eyes had a single test–retest session, and 149 had multiple sessions (126 eyes with two sessions, 19 with three, and 4 with four sessions). For optic disc scans, 712 eyes had a single session, and 134 had multiple sessions (114 eyes with two sessions, 17 with three, and 3 with four sessions). For IEDs test–retest sessions, 368 participants had a single session for macular scans, and 69 had multiple sessions (58 participants with two sessions, 9 with three, and 2 with four sessions). For optic disc scans, 318 participants had a single session, and 61 had multiple sessions (52 participants with two sessions, 8 with three, and 1 with four sessions). GCIPL: Ganglion cell-inner plexiform layer; IEDs: inter-eye differences; pRNFL: peri-papillary retinal nerve fiber layer.

indicating excellent agreement at the cohort level. Similarly, the MADs were consistent across MS subtypes: 0.41 μm (SD 0.53 μm) in RRMS, 0.43 μm (SD 0.56 μm) in PPMS, and 0.44 μm (SD 0.57 μm) in SPMS; ($p = 0.94$), further supporting excellent measurement agreement.

The ICC values for GCIPL thicknesses demonstrated consistently high reliability across all groups, further indicating excellent agreement: 0.998 in PwMS, 0.995 in HCs, and 0.998 in PwOND. Similarly, among MS subtypes, the ICCs showed excellent consistency: 0.998 in RRMS, 0.995 in PPMS, and 0.997 in SPMS. COVs for GCIPL thicknesses were similarly low, strongly supporting measurement precision, being 0.41% in PwMS, 0.34% in HCs, and 0.45% in PwOND. According to MS subtypes, the COVs were also consistently low: 0.40% in RRMS, 0.41% in PPMS, and 0.46% in SPMS, further underscoring the excellent reliability of GCIPL measurements.

pRNFL thickness reliability

Similar to GCIPL thickness, the mean pRNFL thickness was significantly lower in PwMS and PwOND,

as compared to HCs ($p < 0.001$), being 84.83 μm (SD 12.62 μm) in PwMS, 83.37 μm (SD 15.80 μm) in PwOND, and 92.34 μm (SD 9.12 μm) in HCs (Table 3). The MADs in pRNFL thickness were consistently low across all groups ($p = 0.38$), reflecting excellent agreement between measurement sessions, although generally higher than those observed for GCIPL thicknesses: 1.40 μm (SD 1.20 μm) in PwMS, 1.56 μm (SD 1.36 μm) in HCs, and 1.23 μm (SD 1.03 μm) among PwOND. Among MS subtypes, the MADs trended towards being different ($p = 0.06$), being 1.36 μm (SD 1.18 μm) in RRMS, 1.81 μm (SD 1.47 μm) in PPMS, and 1.59 μm (SD 1.24 μm) in SPMS. Along these lines, the MAPDs were slightly higher in the PPMS group (2.20%, SD 1.81%), as compared to both the SPMS (1.99%, SD 1.64%) and RRMS (1.62%, SD 1.42%) cohorts ($p = 0.02$).

ICCs for pRNFL thicknesses demonstrated excellent reliability across all groups: 0.989 in PwMS, 0.975 in HCs, and 0.995 in PwOND, indicating robust agreement in pRNFL measurements. Among MS subtypes, ICCs showed slightly higher reliability in PPMS and SPMS, as compared to RRMS. The

Table 1. Baseline demographics and clinical characteristics.

	MS	RRMS	PPMS	SPMS	OND	HC	p-value (MS vs OND vs HC)	p-value (RRMS vs PPMS vs SPMS)
Subjects, n (eyes)	385 (746)	324 (629)	15 (28)	46 (89)	46 (86)	64 (127)	-	-
People visits, median (range)	1 (1-4)	1 (1-4)	1 (1-2)	1 (1-3)	1 (1-3)	1 (1-3)	-	-
Eye visits, median (range)	1 (1-4)	1 (1-4)	1 (1-2)	1 (1-3)	1 (1-3)	1 (1-3)	-	-
Optic neuritis history, eyes (%)	220 (29.5)	199 (31.6)	2 (7.1)	19 (21.3)	37 (43)	-	<0.001 ^b	0.004^b
Age, years, mean (SD)	46.07 (11.94)	44.16 (11.54)	57.98 (5.19)	55.62 (9.18)	44.25 (20.23)	36.96 (14.69)	<0.001 ^a	<0.001 ^a
Females, n (%)	306 (79.5)	266 (82.1)	9 (60)	31 (67.4)	30 (65.2)	38 (59.4)	0.001^b	0.01^b
Ethnicity, n (%)							<0.001 ^b	0.14 ^b
White American	302 (78.4)	252 (77.8)	15 (100)	35 (76.1)	24 (52.2)	41 (64.1)		
Black American	70 (18.2)	59 (18.2)	0 (0)	11 (23.9)	13 (28.3)	12 (18.8)		
Other	13 (3.4)	13 (4)	0 (0)	0 (0)	9 (19.6)	11 (17.2)		
Disease duration, years, mean (SD)	10.5 (7.5)	9.7 (7.1)	9.7 (6.9)	15.8 (8.3)	-	-	-	<0.001 ^c
Monocular 100% VA, mean (SD)	57.8 (9.8)	58.9 (8.3)	51.8 (16.7)	53 (12.9)	49.7 (14.4)	59.5 (7.8)	<0.001 ^c	0.005^c
Monocular 2.5% VA, mean (SD)	29.4 (11.9)	30.6 (11)	25.8 (12.3)	22.8 (14.6)	18.8 (15.3)	33.6 (9.8)	<0.001 ^c	0.005^c

^aOne-way ANOVA.^bChi-squared test.^cKruskal–Wallis test. Statistically significant differences are seen in bold.

HC: healthy controls; MS: multiple sclerosis; OND: other neurological disorders; PPMS: primary progressive multiple sclerosis; RRMS: relapsing-remitting multiple sclerosis; SD: standard deviation; SPMS: secondary progressive multiple sclerosis; VA: visual acuity.

Table 2. OCT GC IPL thicknesses and reliability measures.

	MS	RRMS	PPMS	SPMS	OND	HC	p-value (MS vs HC vs OND)	p-value (RRMS vs SPMS vs PPMS)
Total scan pairs, n	879	742	35	102	81	140	-	-
Thickness, mean μm (SD)	73.46 (9.85)	73.81 (9.98)	75.41 (7.08)	70.24 (9.08)	71.94 (12.08)	81.27 (6.55)	< 0.001 ^a	< 0.001 ^a
Absolute difference, mean μm (SD)	0.41 (0.53)	0.41 (0.53)	0.43 (0.56)	0.44 (0.57)	0.44 (0.55)	0.39 (0.52)	0.81 ^a	0.94 ^a
Absolute difference, mean % (SD)	0.58 (0.77)	0.57 (0.76)	0.57 (0.76)	0.65 (0.84)	0.63 (0.79)	0.48 (0.66)	0.18 ^a	0.56 ^a
COV %	0.41	0.40	0.41	0.46	0.45	0.34	-	-
ICC (95% CI)	0.998 (0.997–0.998)	0.998 (0.997–0.998)	0.995 (0.990–0.998)	0.997 (0.995–0.998)	0.998 (0.997–0.999)	0.995 (0.993–0.997)	-	-

^aKruskal–Wallis test. Statistically significant differences are seen in bold. CI: confidence interval; COV: coefficient of variation; HC: healthy controls; ICC: intraclass correlation coefficient; MS: multiple sclerosis; OND: other neurological diseases; PPMS: primary progressive multiple sclerosis; RRMS: relapsing-remitting multiple sclerosis; GC IPL: ganglion cell-inner plexiform layer.

Table 3. OCT pRNFL thicknesses and reliability measures.

	MS	RRMS	PPMS	SPMS	OND	HC	p-value (MS vs HC vs OND)	p-value (RRMS vs SPMS vs PPMS)
Total scan pairs, n	798	677	31	90	73	132	-	-
Thickness, mean μm (SD)	84.83 (12.62)	85.32 (12.76)	84.23 (9.64)	81.81 (11.94)	83.37 (15.80)	92.34 (9.12)	< 0.001 ^a	< 0.001 ^a
Absolute difference, mean μm (SD)	1.40 (1.20)	1.36 (1.18)	1.81 (1.47)	1.59 (1.24)	1.23 (1.03)	1.56 (1.36)	0.38 ^a	0.06 ^a
Absolute difference, mean % (SD)	1.69 (1.47)	1.62 (1.42)	2.20 (1.81)	1.99 (1.64)	1.51 (1.28)	1.68 (1.44)	0.67 ^a	0.02 ^a
COV %	1.19	1.15	1.56	1.41	1.07	1.19	-	-
ICC (95% CI)	0.989 (0.988–0.991)	0.990 (0.989–0.992)	0.972 (0.942–0.986)	0.986 (0.979–0.991)	0.995 (0.992–0.997)	0.975 (0.964–0.982)	-	-

^aKruskal–Wallis test. Statistically significant differences are seen in bold. CI: confidence interval; COV: coefficient of variation; HC: healthy controls; ICC: intraclass correlation coefficient; MS: multiple sclerosis; OND: other neurological disorders; PPMS: primary progressive multiple sclerosis; RRMS: relapsing–remitting multiple sclerosis; SD: standard deviation; SPMS: secondary progressive multiple sclerosis; pRNFL: peri-papillary retinal nerve fiber layer.

COVs for pRNFL thicknesses further reflected excellent precision: 1.19% in PwMS, 1.19% in HCs, and 1.07% in PwOND. Within MS subtypes, the COVs were similarly low, at 1.15% in RRMS, 1.56% in PPMS, and 1.41% in SPMS. In general, while COVs for pRNFL thicknesses were excellent across cohorts and MS subtypes, they were generally higher than those observed for GCIPL thickness COVs.

Bland–Altman analyses

Bland–Altman analyses revealed minimal test–retest mean differences for GCIPL (PwMS = 0.02 μm ; HCs = 0.06 μm ; PwOND = 0.07 μm) and pRNFL thicknesses (PwMS = 0.06 μm ; HCs = 0.08 μm ; PwOND = -0.03 μm), indicating excellent test–retest agreement for both measurements at the cohort level.

The Bland–Altman plots display the 95% CI limits of agreement, revealing narrow limits of agreement (Figures 2 and 3) further underscoring the reliability of GCIPL and pRNFL OCT measurements. The limits of agreement with 95% CI were narrow for GCIPL (PwMS: -1.30 μm to 1.34 μm ; HCs: -1.20 μm to 1.32 μm ; PwOND: -1.30 μm to 1.45 μm) and pRNFL thicknesses (PwMS: -3.56 μm to 3.67 μm ; HCs: -3.99 μm to 4.14 μm ; PwOND: -3.19 μm to 3.14 μm), demonstrating excellent test–retest agreement at the individual level, particularly for GCIPL thicknesses. This highlights the precision of OCT measurements and reinforces their potential utility in the detection of UONI.

Inter-eye difference in differences

A total of 520 sets of four macular scans and 450 sets of four optic disc scans met OSCAR-IB QC criteria, thereby allowing the calculation of GCIPL and pRNFL IEDs consistency respectively. The overall absolute difference-in-differences values were low at 2.00 μm (SD 1.72 μm) for pRNFL thicknesses and particularly low at 0.64 μm (SD 0.67 μm) for GCIPL thicknesses across all study cohorts.

For GCIPL measurements, difference-in-differences values were similar across groups ($p=0.94$): 0.64 μm (SD 0.69 μm) in PwMS, 0.63 μm (SD 0.57 μm) in HCs, and 0.65 μm (SD 0.63 μm) in PwOND. GCIPL difference-in-differences values were also similar between MS subtypes ($p=0.5$): 0.63 μm (SD 0.68 μm) in RRMS, 0.61 μm (SD 0.71 μm) in SPMS, and 0.88 μm (SD 0.81 μm) in PPMS (Table 4).

For pRNFL, difference-in-differences values trended towards being different across cohorts ($p=0.05$), being highest in HCs: 1.95 μm (SD 1.70 μm) in PwMS, 2.48 μm (SD 1.85 μm) in HCs, and 1.71 μm (SD 1.53 μm) in PwOND. pRNFL difference-in-differences values also trended towards being different between MS subtypes ($p=0.05$), being 1.86 μm (SD 1.62 μm) in RRMS, 2.36 μm (SD 1.83 μm) in SPMS, and 2.77 μm (SD 2.80 μm) in PPMS.

Influence of absolute retinal layer thickness measurements and clinical factors on pRNFL and GCIPL reliability

Age and sex did not show associations with the GCIPL/pRNFL MAPDs. Minimal negative associations ($\beta = -0.01$, $p \leq 0.001$) were observed between MAPDs in GCIPL and GCIPL thicknesses, pRNFL MAPDs and pRNFL thicknesses, and GCIPL MAPDs and 2.5/100% visual acuity. The MAPDs in pRNFL thickness demonstrated a very weak positive association with disease duration in PwMS ($\beta=0.02$, $p=0.005$). For GCIPL MAPDs, no significant association was observed with disease duration. Minimal positive associations were found between MAPDs in pRNFL thicknesses and time since ON, when adjusting for age and sex ($\beta=0.003$, $p=0.04$). However, given the expected negative correlation due to ON-related swelling, we suspected disease duration in PwMS influenced the result. After adjusting for disease duration, the association was no longer significant.

Discussion

The findings of this study demonstrate the high reliability of OCT measurements for assessing GCIPL and pRNFL thicknesses in PwMS. The low MADs in both GCIPL and pRNFL thicknesses, coupled with excellent ICC values, underscore the consistency of these measurements across different cohorts, especially in MS, across MS subtypes, and in particular highlight the reliability of GCIPL thicknesses.

For GCIPL thicknesses, the MADs were minimal, remaining below 0.45 μm across all groups, with the ICCs above 0.994, indicating excellent reliability. The COVs further support the precision of GCIPL measurements, remaining below 0.46% across all groups, including MS subtypes.

When comparing our results to those of previous studies, it is notable that prior studies primarily focused on HCs and people with glaucoma, rather than PwMS. For instance, studies assessing GCIPL thickness reliability using Cirrus HD-OCT reported

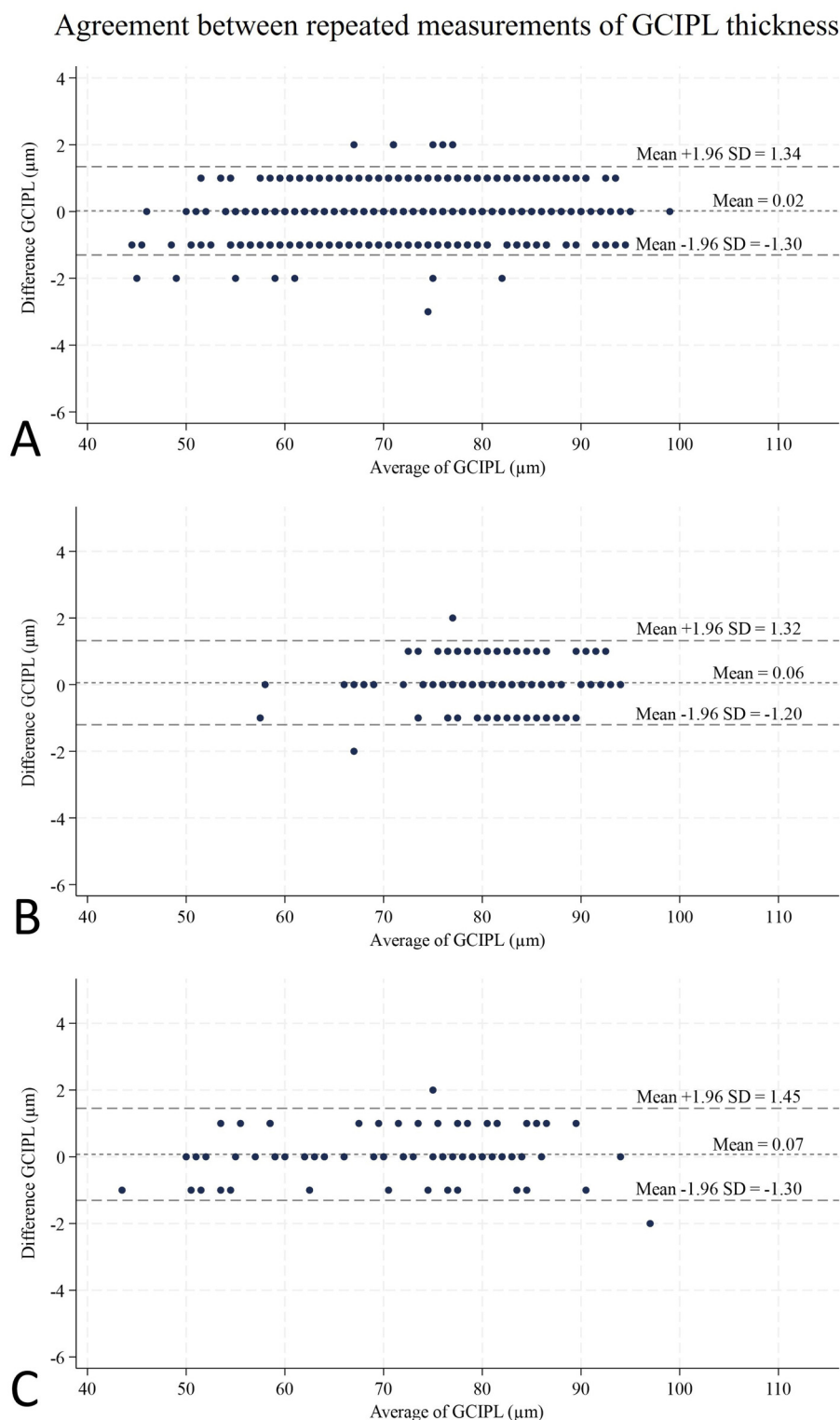


Figure 2. Bland–Altman plots of average GCIPL thicknesses. On the x-axis, the plot displays the mean thickness (μm) of the first and second readings, while the y-axis shows the difference in thickness (μm) between these two readings. The short-dashed line indicates the mean difference, representing the mean bias, and the long-dashed lines represent the limits of agreement (± 1.96 SD), reflecting the 95% confidence interval. Data are presented for (A) PwMS (879 scan pairs), (B) HCs (140 scan pairs), and (C) PwOND (81 scan pairs). GCIPL: Ganglion cell-inner plexiform layer; SD: standard deviation; HC: healthy control.

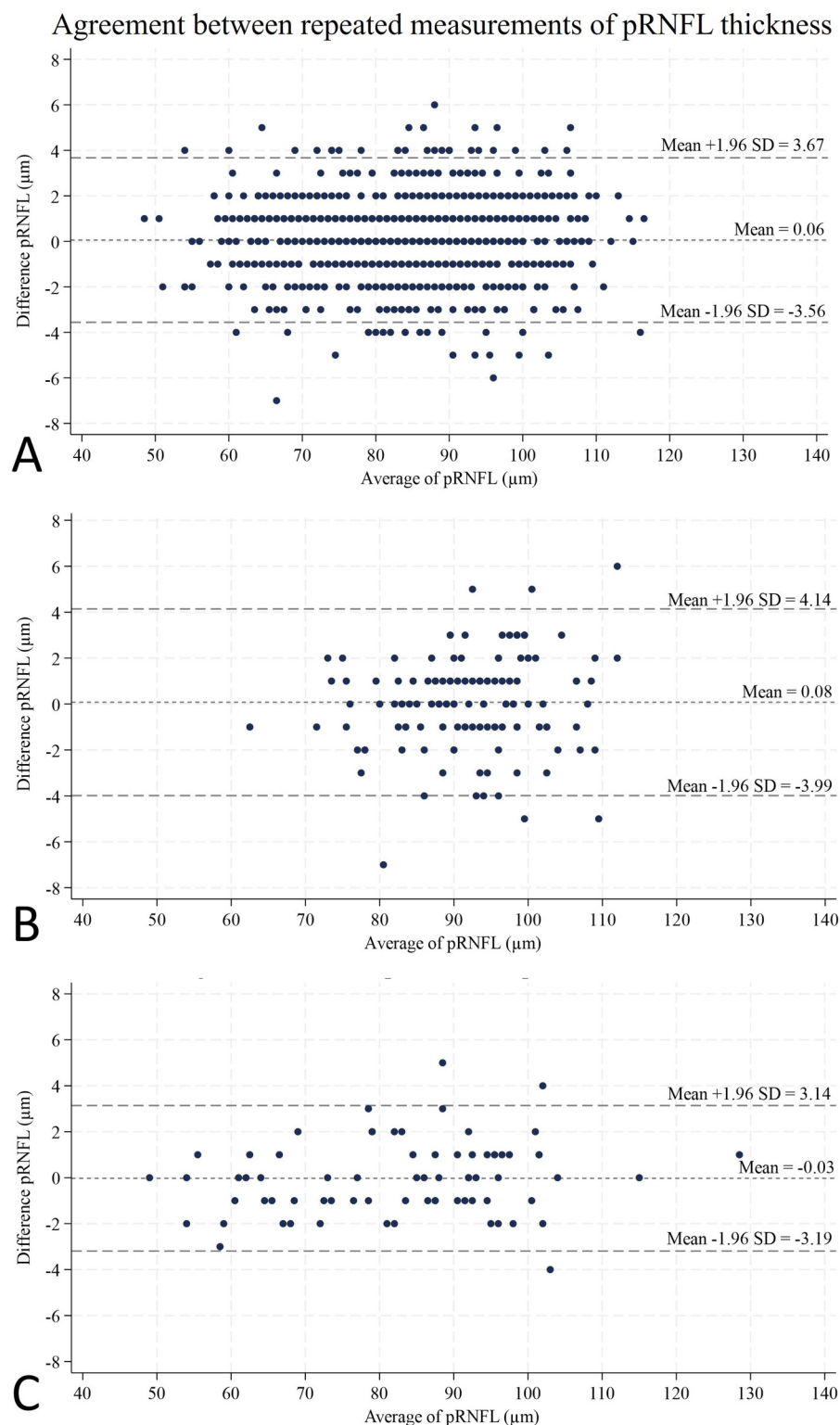


Figure 3. Bland–Altman plots of average pRNFL thicknesses. On the x-axis, the plot displays the mean thickness (μm) of the first and second readings, while the y-axis shows the difference in thickness (μm) between these two readings. The short-dashed line indicates the mean difference, representing the mean bias, and the long-dashed lines represent the limits of agreement (± 1.96 SD), reflecting the 95% confidence interval. Data are presented for (A) PwMS (798 scan pairs), (B) HCs (132 scan pairs), and (C) PwOND (73 scan pairs). pRNFL: peri-papillary retinal nerve fiber layer; SD: standard deviation; HC: healthy control.

Table 4. GCIPL and pRNFL absolute inter-eye difference in differences (DiDs) across all subgroups.

	MS (n = 361)	RRMS (n = 303)	PPMS (n = 14)	SPMS (n = 44)	OND (n = 41)	HC (n = 63)	p-value (MS vs HC vs OND)	p-value (RRMS vs SPMS vs PPMS)
Total GCIPL/pRNFL sets, n	415/359	353/307	16/13	46/39	37/31	68/60	-	-
GCIPL DiDs, mean μ m (SD)	0.64 (0.69)	0.63 (0.68)	0.88 (0.81)	0.61 (0.71)	0.65 (0.63)	0.63 (0.57)	0.94 ^a	0.50 ^a
pRNFL DiDs, mean μ m (SD)	1.95 (1.70)	1.86 (1.62)	2.77 (2.80)	2.36 (1.83)	1.71 (1.53)	2.48 (1.85)	0.05 ^a	0.18 ^a

^aKruskal–Wallis test. DiDs: difference in differences; GCIPL: ganglion cell-inner plexiform layer; HC: healthy controls; MS: multiple sclerosis; n: sample size; OND: other neurological disorders; PPMS: primary progressive multiple sclerosis; pRNFL: peri-papillary retinal fiber layer; RRMS: relapsing-remitting multiple sclerosis; SD: standard deviation; SPMS: secondary progressive multiple sclerosis.

COVs ranging from 0.67% to 0.82% and ICCs from 0.987 to 0.996.^{17,29} Despite the differences in patient populations, results of the current study are remarkably consistent with those from other diseases. Collectively, these studies confirm the high reliability and precision of pRNFL, and in particular GCIPL measurements.

IEDs of $\geq 4 \mu\text{m}$ for GCIPL thicknesses are likely to correctly identify UONI, provided there is no better explanation. In the current study, the limits of agreement with a 95% CI ranged from $-1.29 \mu\text{m}$ to $1.35 \mu\text{m}$ for GCIPL thicknesses, indicating excellent agreement of repeated measures of GCIPL thicknesses at the individual level, and lower than the optimal threshold in GCIPL IEDs of $4 \mu\text{m}$ or more for detecting UONI. Moreover, the IED-in-differences for GCIPL thicknesses in PwMS were $0.64 \mu\text{m}$ (SD $0.69 \mu\text{m}$), further supporting the reliability of these measurements in assessing subtle IEDs, which emphasizes the precision of OCT to identify UONI, for the purpose of demonstrating dissemination in space to support the diagnosis of MS.

Similarly, pRNFL measurements demonstrated high reliability, with MADs below $1.82 \mu\text{m}$ across all groups. The basis for the higher MAPD observed in PPMS, as compared to RRMS and SPMS, is unclear, although could be related to underpowering in the PPMS cohort, as well as cohort differences, since lower pRNFL thicknesses may be associated with slightly higher pRNFL test–retest variability. For pRNFL measurements, the ICCs exceeded 0.971, while the COVs were also consistently low, remaining below 1.56% in all groups, supporting overall excellent test–retest reliability of pRNFL thicknesses. Similar to the results of GCIPL reproducibility and reliability, our findings regarding excellent pRNFL reliability are also consistent with those from other patient populations. For example, test–retest intra-visit variability studies using Cirrus HD-OCT in HCs and people with glaucoma have showed COVs between 1.2% and 1.9%, and ICC values ranging from 0.975 to 0.994.^{17,30,31}

IEDs of $\geq 6 \mu\text{m}$ for pRNFL thicknesses are likely to correctly identify UONI, provided there is no better explanation. In our study, the limits of agreement with a 95% CI ranged from $-3.59 \mu\text{m}$ to $3.70 \mu\text{m}$ for repeated measures of pRNFL thickness, and the difference-in-differences values for pRNFL IEDs in PwMS were $1.95 \mu\text{m}$ (SD $1.70 \mu\text{m}$). Collectively, these findings highlight the consistency of OCT-derived pRNFL measurements, and similar to

GCIPL thicknesses/IEDs, the reliability and practicality of pRNFL and/or GCIPL IEDs for identifying UONI, to support dissemination in space and the diagnosis of MS.

A particular strength of our study relates to the assessment of a very large number of OCT scans, comprising over 1000 macular and optic disc test–retest pairs. Although primarily focused on reliability of OCT in PwMS, the inclusion of HCs and PwOND enhances the generalizability of our findings. This study offers valuable insights, supporting the robustness of OCT measurements in PwMS, demonstrating similar reliability metrics to earlier research focused on HCs and people with glaucoma. Some studies have examined the impact of certain OCT artifacts on measurement reliability,^{32,33} and although there is a notable absence of research specifically addressing scans that do not fulfill OSCAR-IB criteria, high-quality OCT scans are likely to be a fundamental component for ensuring the reproducibility and reliability of OCT measurements. Accordingly, only OCT scans fulfilling OSCAR-IB criteria were used in the current study, thereby maintaining rigorous QC. Our study does have a number of limitations warranting discussion. While this study did include HCs and PwOND, the bulk of power was in PwMS, and therefore findings of our reliability and reproducibility analyses of pRNFL and GCIPL measures should be very cautiously extended to non-MS populations. Moreover, this study only assessed test–retest reliability of pRNFL and GCIPL thicknesses using Cirrus HD-OCT, a commonly used OCT platform. The current analyses warrant similar investigation in MS populations using other OCT platforms, such as the Spectralis OCT device, which together with the Cirrus HD-OCT, account for the majority of OCT platforms used in PwMS.¹⁵ As mentioned, scans underwent rigorous QC based on OSCAR-IB criteria, meaning our results are applicable only when the same standards are followed. Proper training in OCT QC and IED assessment is essential for reliable MS diagnosis in clinical practice. Variations in training and QC could limit the generalizability of our findings. Additionally, the time since most recent ON was relatively long in the studied cohort, with only five eyes in the MS group and four in the OND group having an ON episode within 6 months prior to undergoing OCTs. Since post-ON swelling can persist for up to 6 months,³⁴ and may interfere with OCT reliability, further studies on the test–retest reliability of OCT in acute ON are needed. We did not assess the relationships between disease-modifying therapies and OCT reproducibility and

reliability, since in the absence of possible ocular side effects such as macular edema (which would have resulted in the scans being excluded from this study due to violating OSCAR-IB QC criteria),³⁵ they would not be expected to directly impact OCT measurement reliability. While ocular and/or retinal diseases may influence IEDs, they are not expected to affect retinal measurement reliability, including IED reliability, if they do not interfere with retinal layer segmentation.^{17,18}

In conclusion, this study underscores the robust reliability and precision of Cirrus HD-OCT-derived GCPL and pRNFL measurements, establishing their utility as valuable biomarkers for both clinical and research applications. These results support the reliability and practical utility therein of OCT IEDs proposed in the 2024 revised McDonald diagnostic criteria for identifying UONI to support dissemination in space for the diagnosis of MS.

Declaration of conflicting interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: PAC has received consulting fees from Lilly, Novartis, and Biogen and is Principal Investigator on a grant to JHU from Genentech.

SS has received consulting fees from Medical Logix for the development of CME programs in neurology and has served on scientific advisory boards for Biogen, Sanofi, EMD Serono, Novartis, Genentech Corporation, Amgen, Horizon therapeutics, Clene Pharmaceuticals & ReWind therapeutics. He has performed consulting for Novartis, Genentech Corporation, JuneBrain LLC, Innocare Pharma, Kiniksa pharmaceuticals, Lapix therapeutics, and Setpoint Medical. He is the PI of investigator-initiated studies funded by Genentech Corporation, Biogen, and Novartis. He previously received support from the Race to Erase MS foundation. He has received equity compensation for consulting from JuneBrain LLC and Lapix therapeutics. He was also the site investigator of trials sponsored by MedDay Pharmaceuticals, Clene Pharmaceuticals, and is the site investigator of trials sponsored by Novartis, as well as Lapix therapeutics.

ESS has consulted for Alexion, Amgen, TG Therapeutics, and Roche/Genentech, is site principal investigator for studies funded by Alexion, Roche/Genentech, UCB, and Ad Scientiam, and is principal investigator for an investigator-initiated study funded by Astoria Biologica.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National

Institute of Health (grant number R01NS082347), and the National Multiple Sclerosis Society under (grant numbers RG-1606-08768 and RG-1907-34405).

Data availability statement

The data that support the findings of this study are available on request from the corresponding author and through a data use agreement with Johns Hopkins University.

ORCID iDs

Anna Bacchetti  <https://orcid.org/0009-0000-9434-8531>
Brenna McCormack  <https://orcid.org/0009-0002-6678-7021>

Ting-Yi Lin  <https://orcid.org/0000-0001-9605-8095>
Anna Kim  <https://orcid.org/0009-0002-1188-5345>
Ernest Lievers  <https://orcid.org/0009-0006-5274-277X>
Kathryn C Fitzgerald  <https://orcid.org/0000-0003-3137-0322>
Elias S Sotirchos  <https://orcid.org/0000-0002-8812-1637>
Peter A Calabresi  <https://orcid.org/0000-0002-7776-6472>
Shiv Saidha  <https://orcid.org/0000-0001-6387-0714>

Supplemental material

Supplemental material for this article is available online.

References

1. Reich DS, Lucchinetti CF and Calabresi PA. Multiple sclerosis. *N Engl J Med* 2018; 378: 169–180.
2. Graham SL and Klistorner A. Afferent visual pathways in multiple sclerosis: a review. *Clin Exp Ophthalmol* 2017; 45: 62–72. 20160428.
3. Ikuta F and Zimmerman HM. Distribution of plaques in seventy autopsy cases of multiple sclerosis in the United States. *Neurology* 1976; 26: 26–28.
4. Toussaint D, Perier O, Verstaappen A, et al. Clinicopathological study of the visual pathways, eyes, and cerebral hemispheres in 32 cases of disseminated sclerosis. *J Clin Neuroophthalmol* 1983; 3: 211–220.
5. Costello F. Vision disturbances in multiple sclerosis. *Semin Neurol* 2016; 36: 185–195. 20160426.
6. Balcer LJ. Clinical practice. Optic neuritis. *N Engl J Med* 2006; 354: 1273–1280.
7. Montalban X. *Revisions of the McDonald criteria*. Copenhagen, Denmark: ECTRIMS, 2024.
8. Vidal-Jordana A, Rovira A, Calderon W, et al. Adding the optic nerve in multiple sclerosis diagnostic criteria: a longitudinal, prospective, multicenter study. *Neurology* 2024; 102: e200805. 20231213.
9. Vidal-Jordana A, Rovira A, Arrambide G, et al. Optic nerve topography in multiple sclerosis diagnosis: the utility of visual evoked potentials. *Neurology* 2021; 96: e482–e490. 20201216.
10. Bsteh G, Hegen H, Altmann P, et al. Diagnostic performance of adding the optic nerve region assessed

- by optical coherence tomography to the diagnostic criteria for multiple sclerosis. *Neurology* 2023; 101: e784–e793. 20230703.
11. Coric D, Balk LJ, Uitdehaag BMJ, et al. Diagnostic accuracy of optical coherence tomography inter-eye percentage difference for optic neuritis in multiple sclerosis. *Eur J Neurol* 2017; 24: 1479–1484. 20171009.
12. Patil SA, Joseph B, Tagliani P, et al. Longitudinal stability of inter-eye differences in optical coherence tomography measures for identifying unilateral optic nerve lesions in multiple sclerosis. *J Neurol Sci* 2023; 449: 120669. 20230503.
13. Nolan-Kenney RC, Liu M, Akhand O, et al. Optimal intereye difference thresholds by optical coherence tomography in multiple sclerosis: an international study. *Ann Neurol* 2019; 85: 618–629. 20190410.
14. Bsteh G, Hegen H, Altmann P, et al. Validation of inter-eye difference thresholds in optical coherence tomography for identification of optic neuritis in multiple sclerosis. *Mult Scler Relat Disord* 2020; 45: 102403. 20200717.
15. Balcer LJ, Balk LJ, Brandt AU, et al. The international multiple sclerosis visual system consortium: advancing visual system research in multiple sclerosis. *J Neuroophthalmol* 2018; 38: 494–501.
16. Xu SC, Kardon RH, Leavitt JA, et al. Optical coherence tomography is highly sensitive in detecting prior optic neuritis. *Neurology* 2019; 92: e527–e535. 20190123.
17. Wadhvani M, Bali SJ, Satyapal R, et al. Test-retest variability of retinal nerve fiber layer thickness and macular ganglion cell-inner plexiform layer thickness measurements using spectral-domain optical coherence tomography. *J Glaucoma* 2015; 24: e109–e115.
18. Parravano M, Oddone F, Boccassini B, et al. Reproducibility of macular thickness measurements using Cirrus SD-OCT in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci* 2010; 51: 4788–4791. 20100430.
19. Garcia-Martin E, Pueyo V, Pinilla I, et al. Fourier-domain OCT in multiple sclerosis patients: reproducibility and ability to detect retinal nerve fiber layer atrophy. *Invest Ophthalmol Vis Sci* 2011; 52: 4124–4131. 20110613.
20. Syc SB, Warner CV, Hiremath GS, et al. Reproducibility of high-resolution optical coherence tomography in multiple sclerosis. *Mult Scler* 2010; 16: 829–839. 20100608.
21. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 2018; 17: 162–173. 20171221.
22. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 2014; 83: 278–286. 20140528.
23. Petzold A, Chua SYL, Khawaja AP, et al. Retinal asymmetry in multiple sclerosis. *Brain* 2021; 144: 224–235.
24. Lang A, Carass A, Al-Louzi O, et al. Combined registration and motion correction of longitudinal retinal OCT data. *Proc SPIE Int Soc Opt Eng* 2016; 9784: 20160321.
25. Tewarie P, Balk L, Costello F, et al. The OSCAR-IB consensus criteria for retinal OCT quality assessment. *PLoS One* 2012; 7: e34823. 20120419.
26. Aytulun A, Cruz-Herranz A, Aktas O, et al. APOSTEL 2.0 Recommendations for reporting quantitative optical coherence tomography studies. *Neurology* 2021; 97: 68–79. 20210428.
27. Cruz-Herranz A, Balk LJ, Oberwahrenbrock T, et al. The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. *Neurology* 2016; 86: 2303–2309. 20160525.
28. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
29. Francoz M, Fenolland JR, Giraud JM, et al. Reproducibility of macular ganglion cell-inner plexiform layer thickness measurement with cirrus HD-OCT in normal, hypertensive and glaucomatous eyes. *Br J Ophthalmol* 2014; 98: 322–328. 20131204.
30. Mwanza JC, Chang RT, Budenz DL, et al. Reproducibility of peripapillary retinal nerve fiber layer thickness and optic nerve head parameters measured with cirrus HD-OCT in glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2010; 51: 5724–5730. 20100623.
31. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology* 2009; 116: 1257–1263. 20090522.
32. Yohannan J, Cheng M, Da J, et al. Evidence-based criteria for determining peripapillary OCT reliability. *Ophthalmology* 2020; 127: 167–176. 20190829.
33. Awadalla MS, Fitzgerald J, Andrew NH, et al. Prevalence and type of artefact with spectral domain optical coherence tomography macular ganglion cell imaging in glaucoma surveillance. *PLoS One* 2018; 13: e0206684. 20181205.
34. Costello F, Pan YI, Yeh EA, et al. The temporal evolution of structural and functional measures after acute optic neuritis. *J Neurol Neurosurg Psychiatry* 2015; 86: 1369–1373. 20150210.
35. Munoz-Ortiz J, Reyes-Guanes J, Zapata-Bravo E, et al. Ocular adverse events from pharmacological treatment in patients with multiple sclerosis—A systematic review of the literature. *Syst Rev* 2021; 10: 280. 20211028.