Individual Prognostication of Disease Activity and Disability Worsening in Multiple Sclerosis With Retinal Layer Thickness z Scores

Ting-Yi Lin, MD, Seyedamirhosein Motamedi, PhD, Susanna Asseyer, MD, Claudia Chien, PhD, Shiv Saidha, MD, Peter A. Calabresi, MD, Kathryn C. Fitzgerald, ScD, Sara Samadzadeh, MD, Pablo Villoslada, MD, PhD, Sara Llufriu, MD, PhD, Ari J. Green, MD, Jana Lizrova Preiningerova, MD, PhD, Axel Petzold, MD, PhD, Letizia Leocani, MD, PhD, Elena Garcia-Martin, MD, PhD, Celia Oreja-Guevara, MD, PhD, Olivier Outteryck, MD, Patrick Vermersch, MD, PhD, Laura J. Balcer, MD, MSCE, Rachel Kenney, PhD, Philipp Albrecht, MD, Orhan Aktas, MD, Fiona Costello, MD, Jette Frederiksen, MD, Antonio Uccelli, MD, Maria Cellerino, MD, Elliot M. Frohman, MD, Teresa C. Frohman, MD, Judith Bellmann-Strobl, MD, Tanja Schmitz-Hübsch, MD, Klemens Ruprecht, MD, Alexander U. Brandt, MD, Hanna G. Zimmermann, PhD,* and Friedemann Paul, MD*

Correspondence
Dr. Paul
friedemann.paul@charite.de

Abstract

Background and Objectives

Retinal optical coherence tomography (OCT) provides promising prognostic imaging biomarkers for future disease activity in multiple sclerosis (MS). However, raw OCT-derived measures have multiple dependencies, supporting the need for establishing reference values adjusted for possible confounders. The purpose of this study was to investigate the capacity for age-adjusted z scores of OCT-derived measures to prognosticate future disease activity and disability worsening in people with MS (PwMS).

Methods

We established age-adjusted OCT reference data using generalized additive models for location, scale, and shape for peripapillary retinal nerve fiber layer (pRNFL) and ganglion cell-inner plexiform layer (GCIP) thicknesses, involving 910 and 423 healthy eyes, respectively. Next, we transformed the retinal layer thickness of PwMS from 3 published studies into age-adjusted z scores (pRNFL-z and GCIP-z) based on the reference data. Finally, we investigated the association of pRNFL-z or GCIP-z as predictors with future confirmed disability worsening (Expanded Disability Status Scale score increase) or disease activity (failing of the no evidence of disease activity [NEDA-3] criteria) as outcomes. Cox proportional hazards models or logistic regression analyses were applied according to the original studies. Optimal cutoffs were identified using the Akaike information criterion as well as location with the log-rank and likelihood-ratio tests.

---

*These authors contributed equally as senior authors.

From the Charité - Universitätsmedizin Berlin (T.-Y.L., S.M., S.A., C.C., S. Samadzadeh, J.B.-S., T.S.-H., A.U.B., H.G.Z., F.P.); Experimental and Clinical Research Center (S.M., S.A., C.C., S. Samadzadeh, J.B.-S., T.S.-H., A.U.B., H.G.Z., F.P.), a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC) (T.-Y.L., S.M., S.A., C.C., S. Samadzadeh, J.B.-S., T.S.-H., A.U.B., H.G.Z., F.P.); Neuroscience Clinical Research Center (S.M., S.A., C.C., J.B.-S., T.S.-H., H.G.Z., F.P.); Department of Psychiatry and Psychotherapy (C.C.), Charité - Universitätsmedizin Berlin, Germany; Department of Neurology (S. Saidha, P.A.C., K.C.F.); Department of Epidemiology (K.C.F.), Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; Department of Regional Health Research and Molecular Medicine (S. Samadzadeh), University of Southern Denmark, Odense; Department of Neurology (S. Samadzadeh), Slagelse Hospital, Denmark; Department of Neurology (P. Villoslada), Hospital Del Mar - Pompeu Fabra University; Neuroimmunology and Multiple Sclerosis Unit (S.L.), Hospital Clinic Barcelona and IDIBAPS, Barcelona, Spain; Department of Neurology (A.J.G.), University of California San Francisco; Department of Neurology (J.L.P.), Charles University in Prague, Czech Republic; Moorfield's Eye Hospital (A.P.), The National Hospital for Neurology and Neurosurgery, Queen Square Institute of Neurology, University College London, United Kingdom; Neuro-ophthalmology Expert Center (A.P.), Amsterdam UMC, Netherlands; Experimental Neurophysiology Unit (L.L.), Institute of Experimental Neurology (INSPE), San Raffaele Scientific Institute; Vita-Salute San Raffaele University (L.L.), Milan, Italy; Miguel Servet University Hospital (E.G.-M.), Zaragoza; Department of Neurology (C.O.-G.), Hospital Clínico Universitario San Carlos, Madrid, Spain; Department of Neurology (O.O., P. Vermersch); Department of Neuroradiology (O.O., P. Vermersch), Centre Hospitalier Universitaire de Lille, France; Departments of Neurology (L.J.B., R.K.), Population Health and Ophthalmology, NYU Grossman School of Medicine, NY; Department of Neurology (P.A., O.A.), Heinrich-Heine-University, Düsseldorf, Germany; Departments of Clinical Neurosciences and Surgery Cumming School of Medicine (F.C.), University of Calgary, Alberta, Canada; Clinic of Optic Neuritis and Clinic of Multiple Sclerosis (J.F.), Department of Neurology, Rigshospitalet - Glostrup, Denmark; Department of Neurosciences (A.U., M.C.), Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Italy; Laboratory of Neuroimmunology (E.M.F., T.C.F.), Professor Lawrence Steinman, Stanford University School of Medicine, Palo Alto, CA; Department of Neurology (K.R., F.P.), Charité - Universitätsmedizin Berlin; and Einstein Center Digital Future (H.G.Z.), Berlin, Germany.

Go to Neurology.org/NN for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the authors.

Glossary

aHR = adjusted hazard ratio; **AIC** = Akaike information criterion; **DMTs** = disease-modifying therapies; **GAMLSS** = General Additive Model for Location, Scale, and Shape; **GCIP** = ganglion cell-inner plexiform layer; **HC** = healthy control; **HD** = high definition; **IQR** = interquartile range; **MS** = multiple sclerosis; **NEDA-3** = no evidence of disease activity; **OCT** = optical coherence tomography; **pRNFL** = peripapillary retinal nerve fiber layer; **PwMS** = people with MS.

Results

In the first cohort (n = 863), 172 PwMS (24%) had disability worsening over a median observational period of 2.0 (interquartile range [IQR]:1.0–3.0) years. Low pRNFL-z (≤-2.04) were associated with an increased risk of disability worsening (adjusted hazard ratio (aHR) [95% CI] = 2.08 [1.47–2.95], $p = 3.82e^{-5}$). In the second cohort (n = 170), logistic regression analyses revealed that lower pRNFL-z showed a higher likelihood for disability accumulation at the two-year follow-up (reciprocal odds ratio [95% CI] = 1.51[1.06–2.15], $p = 0.03$). In the third cohort (n = 78), 46 PwMS (59%) did not maintain the NEDA-3 status over a median follow-up of 2.0 (IQR: 1.9–2.1) years. PwMS with low GCIP-z (≤−1.03) had a higher risk of showing disease activity (aHR [95% CI] = 2.14 [1.03–4.43], $p = 0.04$). Compared with raw values with arbitrary cutoffs, applying the z score approach with optimal cutoffs showed better performance in discrimination and calibration (higher Harrell's concordance index and lower integrated Brier score).

Discussion

In conclusion, our work demonstrated reference cohort–based z scores that account for age, a major driver for disease progression in MS, to be a promising approach for creating OCT-derived measures useable across devices and toward individualized prognostication.

Introduction

The clinical course of multiple sclerosis (MS) is highly heterogenous across individual patients.[1,2] The possibility of relapse-associated neuroaxonal damage and progression independent of relapse activity throughout all disease courses heightens the need for biomarkers that identify poor disease prognosis and may aid in optimal treatment.[3-5] Neuroimaging provides evidence of active inflammation or degeneration in the CNS in MS.[6] Numerous neuroimaging biomarkers have been established over recent decades, and studies have investigated their ability for reflecting disease activity in MS.[7]

Afferent visual pathway damage is one of the key manifestations of MS.[8] Retinal optical coherence tomography (OCT) allows for high-resolution quantification of the retinal structures, making it a good tool to measure retinal neuroaxonal integrity.[9,10] Two important OCT-derived measures, peripapillary retinal nerve fiber layer (pRNFL) thickness and combined macular ganglion cell and inner plexiform layer (GCIP) thickness, have been suggested to be useful imaging biomarkers for disease monitoring in people with MS (PwMS).[7,11-13] Along with others, we have previously shown that PwMS with lower pRNFL and GCIP thicknesses have a higher risk of developing future disability worsening and increased rate of clinical relapse.[14-19] These results support that OCT-derived measures can be used as imaging biomarkers for prognostication of MS disease activity.

One pivotal challenge for translating biomarkers from research to routine clinical practice is to implement them on an individual patient level.[20] Despite being a promising biomarker for future disease activity, the pathologic thresholds of OCT measures defined in previous studies were rather arbitrary.[14-19] A standardized definition of pathologic findings can improve the clinical applications of these imaging biomarkers.[21] Moreover, raw OCT-derived retinal layer thickness measures may vary according to age, sex, ethnicity, and certain comorbidities.[22-24] In addition, most current clinical studies were conducted with 2 commercially available OCT devices: Spectralis spectral-domain (SD) OCT and Cirrus high-definition (HD) OCT. Retinal layer thickness measures derived from these 2 devices are not interchangeable because of the differences in segmentation methods.[25,26] One attempt to tackle this obstacle in previous studies was to generate conversion equations for OCT measures derived from these 2 devices, or to use a consistent segmentation algorithm that can be applied across OCT platforms.[23,26] However, the need for further standardization of OCT-derived measures remains. The z score has the benefits of providing relative position information to simplify interpretation and has been applied in other candidate biomarkers in MS.[21] In this study, we proposed using z scores to create clinical applicable imaging biomarker candidates and applied them on data from 3 published studies[14,16,27] to test their clinical applicability.

In this study, we aimed to (1) derive reference z scores and percentiles for both pRNFL and GCIP thicknesses from a normative cohort adjusted for age as a pilot approach, (2) define standardized pathologic thresholds for both retinal layer thicknesses independent of age, (3) investigate the

clinical applicability of pathologic pRNFL thickness age-adjusted z scores (pRNFL-z) and GCIP thickness age-adjusted z scores (GCIP-z) as risk factors of future disability worsening and disease activity in PwMS, (4) explore the cross-device usability of z scores, and (5) compare z scores with absolute values.

Methods

Study Population

Healthy Control Cohort

For the derivation of pRNFL reference percentiles and z score values, we included OCT scans of healthy controls (HCs) from 4 European MS centers (218 HC from Charité – Universitätsmedizin Berlin, Germany [ethics approval: EA1/175/15 and EA1/163/12][24]; 32 HCs from Heinrich-Heine-University, Düsseldorf, Germany; 121 HCs from Hospital Clinic Barcelona, Spain; and 181 HCs from First Faculty of Medicine and General University Hospital in Prague, Czech Republic). GCIP reference percentiles and z score values were only derived from the 218 HCs at Charité – Universitätsmedizin Berlin, Germany. Inclusion criteria were age from 18 to 70 years and no history of neurologic or ophthalmologic disorder. Hypertension or diabetes without retinal involvements were not excluded.

IMSVISUAL Cohort

The details of PwMS from the IMSVISUAL cohort have been previously reported.[14] PwMS were recruited from 15 centers across North America and Europe. Inclusion criteria were (1) baseline age 18 years and older and (2) a diagnosis of clinically isolated syndrome (CIS), relapsing-remitting MS, or progressive MS according to the 2010 McDonald criteria[28] (the cohort was composed before the introduction of 2017 McDonald criteria). Participants with a history of (1) bilateral optic neuritis (ON) or (2) non-MS neurologic or ophthalmologic disorders were excluded. The outcome measure was confirmed EDSS score increase (confirmed in a consecutive visit 3 to 6 months later), defined as a minimum increase of 1.0 or 0.5 in those with a baseline score of <5.5 or ≥5.5, respectively.[29]

Berlin CIS Cohort

The design of this single-center study, as previously described,[16] included participants with either CIS or early relapsing-remitting MS (PweMS) as per the revised 2017 McDonald criteria.[30] All participants were between 18 and 70 years of age at the time of study inclusion, and the exclusion criteria were the same as in the IMSVISUAL cohort. The primary study outcome was not fulfilling the no evidence of disease activity (NEDA-3) criteria.[31] Each individual evidence of disease activity, including new clinical relapse, new brain MRI activity, and confirmed EDSS score increase (defined as a minimum increase of 1.5, 1.0, or 0.5 in those with a baseline score of 0, 1.0–5.5, or ≥5.5, respectively, and confirmed in a consecutive annual follow-up visit), was also explored in the analyses.

Sys4MS Cohort

For further validation of the findings, we included prospectively collected data from the Sys4MS cohort, a European longitudinal multicenter MS cohort. We only included data from centers that used Spectralis OCT (Charité-Universitätsmedizin Berlin, Germany; Hospital Clinic of Barcelona, Spain; and Ospedale Policlinico San Martino, Italy).[27,32] Data were collected at baseline and at the median 2-year follow-up. All participants included in our analyses were between 18 and 70 years and had a diagnosis of CIS, relapsing-remitting MS, or progressive MS, as defined according to the revised 2017 McDonald criteria. The primary study outcome was disease worsening (not fulfilling NEDA-3) at the 2-year follow-up, and each component was also analyzed as a separate outcome measure.[27,32]

Standard Protocol Approvals, Registrations, and Patient Consents

The relevant institutional review boards and ethics committees at each institution approved all study protocols. All study participants provided written informed consent.

Optical Coherence Tomography

For PwMS in the IMSVISUAL cohort, retinal OCT images were obtained using either Spectralis SD-OCT (Heidelberg Engineering, Heidelberg, Germany) or Cirrus HD-OCT (Carl Zeiss, Dublin, CA) devices. All OCT images of HCs and PwMS in the Berlin CIS and Sys4MS cohort were only acquired using Spectralis SD-OCT devices. The pRNFL thicknesses of HCs and PwMS in all 3 MS cohorts were obtained using device-incorporated segmentation algorithms while the GCIP thicknesses of HCs and PwMS in the Berlin CIS cohort were measured using SAMIRIX, a semiautomated segmentation pipeline. GCIP thickness was not included in the following analyses for both IMSVISUAL and Sys4MS cohorts because GCIP data were not collected in the IMSVISUAL cohort and a different macular scan segmentation method (Orion software®, Voxeleron Inc., Berkeley) was used in the Sys4MS cohort. Details of the scanning protocol and segmentation methodology for PwMS have been, respectively, described in previous studies.[14,16,24,27] The scanning protocol for HC can be found in the eMethods.

We only included eyes without a history of ON based on clinical record. Participants with bilateral non-ON eyes had their pRNFL and GCIP values calculated as the mean of both eyes, whereas those with unilateral non-ON eyes had only the unaffected eyes included in the analyses. All OCT scans underwent quality control assessment according to the OSCAR-IB criteria,[33,34] and data were reported in accordance with the APOSTEL recommendations.[35]

Statistical Analysis

To model the distribution of pRNFL and GCIP thicknesses and their associations with age, a generalized additive model for location, scale, and shape (GAMLSS) using Box-Cox t distribution with cubic splines was used to establish smoothed

age-adjusted percentiles and z score curves of both pRNFL and GCIP thicknesses in HC eyes.[36] The Bayesian information criterion and worm plots were used to evaluate the goodness-of-fit and kurtosis, which helped to select the final models with an optimal outcome distribution. To determine the reliability and stability of the selected GAMLSS-fitted trajectories, we ran 200 bootstrapped iterations to obtain the 95% CIs.

OCT measures of PwMS from the MS cohorts were fitted into the reference curves and transformed into corresponding age-adjusted z scores. All PwMS were allocated into 6 categories based on their z score values (lower than −2, −2 to −1, −1 to 0, 0 to 1, 1 to 2, and greater than 2). Baseline clinical characteristics were the same as in the original studies and presented as mean ± SD, median (interquartile range), or number (%). Group comparisons were conducted with analysis of variance, the Kruskal-Wallis test, and χ^2 tests where appropriate.

The associations of lower pRNFL or GCIP thicknesses with future disability worsening or disease activity in PwMS were investigated with Cox proportional hazards models. For the IMSVISUAL cohort, we computed the risk of confirmed EDSS score increase with low pRNFL-z as risk factors. Models were adjusted for sex, disease duration, baseline EDSS, and use of disease-modifying therapies (DMTs). For the Berlin CIS cohort, low GCIP-z were investigated as the risk factor of subsequent disease activity, as evidenced by not fulfilling the NEDA-3 criteria. Models were adjusted for sex, disease duration, T2w lesion volume, and DMTs used.[37] The results were summarized as multivariable-adjusted hazard ratios (aHRs) and 95% CIs.

We first examined the association between OCT measures and each outcome on a continuous scale with restricted cubic spline curves based on Cox proportional hazards models.[38,39] The 95% CIs were computed along a continuous spectrum of pRNFL-z or GCIP-z. To balance for overfitting and best fit of the splines, we chose the number of knots based on the lowest value of the Akaike information criterion (AIC). Interactions between OCT measures and the aforementioned covariates were examined by including interaction terms of the 2 factors in the models. Furthermore, apart from analyzing within a continuous scale, we also explored optimal thresholds of pRNFL-z and GCIP-z for further group classifications. We applied the AIC method to facilitate the choice of optimal cutoff numbers (minimum AIC value) and then determined the values of the cutoff points with the most significant results in log-rank and likelihood-ratio tests.[40] Afterward, PwMS were separated into pathologic or normal groups based on the predefined optimal cutoffs to evaluate the risks of meeting each outcome. Finally, we used the pRNFL-z/GCIP-z categories where the most optimal cutoffs were located as the reference group and compared the risks in other z score categories for fully adjusted Cox proportional hazards models.

We were not able to generate device-dependent retinal layer thickness reference curves because the OCT images of HC were only acquired using the Spectralis SD-OCT device. Nevertheless, we investigated the interchangeability of z scores across different devices by integrating the retinal layer thicknesses of PwMS acquired from the Cirrus HD-OCT device into the reference curve derived from the Spectralis SD-OCT device, generating the corresponding z score and testing the model's performance.

The predictive performance among different Cox proportional hazards models was compared between raw OCT values and z scores, as well as different estimated thresholds using the Harrell's concordance index (C-index)[41] and integrated Brier score.[42] We ran 100 iterations of fivefold cross-validation to confirm significance. The best-performing model had the highest Harrell's C-index and the lowest integrated Brier score.

Considering the limited follow-up time point in the validation cohort from the Sys4MS study, multivariable logistic regression models were also used to quantify the risk of future disease activity. We analyzed the associations between future disease activity and pRNFL-z or GCIP-z, with sex, disease duration, baseline EDSS, and DMTs considered as covariates. The results were presented with odds ratios (ORs) and 95% CIs.

Statistical analyses were conducted in R (version 4.2.0) with packages survival, survminer, splines, smoothHR, ranger, ggplot2, ggfortify, rms, pec, and sjPlot. Statistical significance was defined as a double-tailed p value <0.05.

Data Availability

All data used for analysis are presented in the tables and figures in this article. Anonymized data from the IMSVISUAL, Berlin CIS, and Sys4MS cohort are available on reasonable request from any qualified investigator and after obtaining ethics approval. The HC pRNFL/GCIP data and source code used for age-adjusted pRNFL/GCIP reference database are provided in a public repository (github.com/neurodial/TL-HC-OCT-Zscore.git), which can be used on reasonable request with permission from the corresponding author.

Results

Healthy Controls

We assessed 910 eyes of 552 HCs from 4 European centers (mean [SD] age: 38.8 [13.3] years; 338 women [61.2%], 214 men [38.8%]) to derive the pRNFL thickness reference percentiles and z scores (mean [SD] pRNFL: 99.4 [8.8] μm). The characteristics of these 4 HC cohorts are presented in eTable 1. Because of variations in macular scan segmentation algorithms across different centers, reference curves for GCIP thickness were derived only from 423 eyes of 218 HCs at Charité – Universitätsmedizin Berlin (mean [SD] GCIP: 70.8

Figure 1 pRNFL and GCIP Age-Adjusted *z* Score Reference Curves
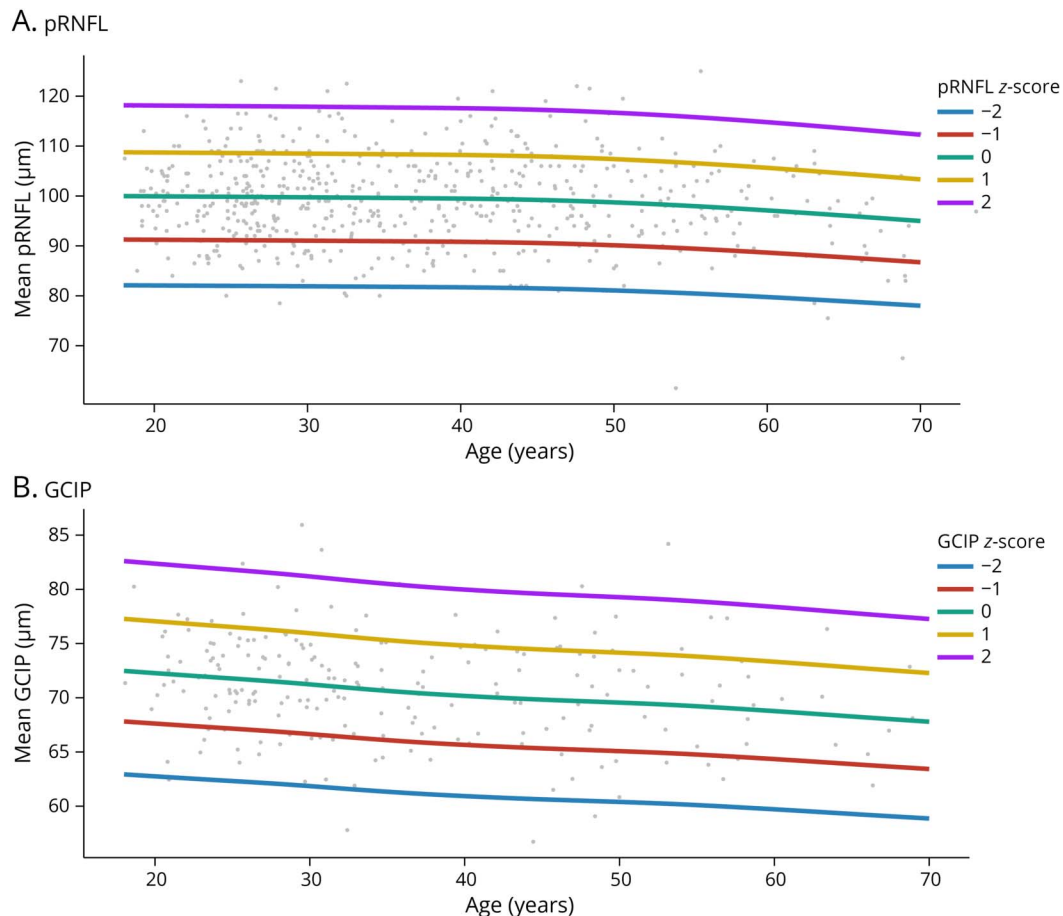
GCIP = combined macular ganglion cell and inner plexiform layer thickness; pRNFL = peripapillary retinal nerve fiber layer thickness.

[4.9] μm). All OCT scans from HCs were acquired with the Spectralis SD-OCT device. The distribution of pRNFL and GCIP thicknesses and their associations with age were modeled by means of GAMLSS, and the age-adjusted *z* score reference curves are presented in Figure 1. The estimated values of the 1st, 5th, 20th, 50th, 80th, 95th, and 99th pRNFL and GCIP percentiles, along with the 95% bootstrapped CIs are summarized in Table 1 and Table 2. Our results demonstrated a clear pattern, showing an inverse correlation between age and both pRNFL and GCIP thicknesses. An inflection point was observed at around 45 years of age in the pRNFL thickness reference curves, and a steeper decrease was noticed thereafter.

IMSVISUAL Cohort: pRNFL Thinning as Risk Factors of Future Disability Accumulation

Of the 863 PwMS (mean [SD] age: 40.9 [11.4] years; 570 women [66.0%], 293 men [34.0%]) included from the IMSVISUAL cohort, 722 (mean [SD] age: 40.7 [11.6] years; 483 women [66.9%], 239 men [33.1%]) had retinal OCT images acquired with the Spectralis SD-OCT device. The OCT measures of eyes from those 722 PwMS were then fitted to the reference curves generated from HC eyes and

transformed into age-adjusted *z* scores accordingly. The baseline clinical and demographic characteristics of the PwMS by pRNFL-z categories are listed in Table 3. There were no significant differences in sex among different pRNFL-z groups. Yet, PwMS within the lower pRNFL-z groups had longer disease duration, higher EDSS scores, and more progressive MS at baseline.

In the IMSVISUAL cohort, 172 (23.8%) of the 722 PwMS had a confirmed increase of EDSS scores (35 had a baseline EDSS score ≥5.5; 114 had a baseline EDSS score between 1.0 and 5.5; 23 had a baseline EDSS score = 0) during a median (IQR) follow-up of 2.0 (1.0–3.0) years. The association between pRNFL-z on a continuous scale and the risk of future disability increase was evident, with a higher risk observed in those with lower pRNFL-z (Figure 2A). The optimal cutoff value of pRNFL-z, as evidenced by maximally selected rank statistics, for disability increase risk stratification was −2.04 (Figure 2B). Participants with pRNFL-z of −2.04 or lower had a multivariable-adjusted hazard ratio of 2.08 (95% CI 1.47–2.95; $p = 3.82e^{-5}$) for disability increase, compared with those with higher pRNFL-z (>−2.04). The model was validated with 100-times repeated fivefold cross-validation to

Table 1 Estimated pRNFL Percentiles With Bootstrap Confidence Intervals Calculated From HC Eyes

Age	1st Percentile	5th Percentile	20th Percentile	50th Percentile	80th Percentile	95th Percentile	99th Percentile
20	78.9 (76.5–81.3)	85.4 (83.4–86.8)	92.6 (91.1–93.9)	99.9 (98.6–101.1)	107.3 (105.9–108.6)	114.6 (113.2–116.5)	121.4 (119.1–123.9)
25	78.9 (76.5–81.2)	85.3 (83.6–86.5)	92.5 (91.3–93.5)	99.8 (98.8–100.7)	107.2 (106.1–108.3)	114.6 (113.4–115.9)	121.3 (119.3–123.3)
30	78.8 (76.4–80.9)	85.3 (83.7–86.3)	92.4 (91.4–93.2)	99.7 (98.9–100.4)	107.1 (106.2–107.9)	114.5 (113.3–115.7)	121.2 (119.2–123.0)
35	78.8 (76.4–80.8)	85.2 (83.9–86.2)	92.3 (91.5–93.1)	99.6 (98.9–100.5)	107.0 (106.0–108.1)	114.4 (113.2–115.6)	121.1 (119.2–122.9)
40	78.7 (76.3–80.7)	85.1 (83.8–86.0)	92.2 (91.4–93.1)	99.5 (98.8–100.4)	106.9 (105.8–108.1)	114.3 (112.9–115.7)	121.0 (119.0–123.1)
45	78.5 (76.2–80.4)	84.9 (83.8–85.8)	92.0 (91.1–92.8)	99.3 (98.5–100.3)	106.7 (105.3–108.0)	114.0 (112.3–115.7)	120.7 (118.6–123.1)
50	78.1 (75.8–80.0)	84.5 (83.3–85.5)	91.6 (90.6–92.5)	98.8 (97.8–100.0)	106.2 (104.7–107.7)	113.5 (111.7–115.3)	120.1 (117.8–122.7)
55	77.6 (75.1–79.6)	83.9 (82.5–85.1)	90.9 (89.9–92.1)	98.1 (96.8–99.5)	105.4 (103.7–107.2)	112.7 (110.9–114.7)	119.3 (116.6–122.0)
60	76.9 (74.1–79.2)	83.1 (81.6–84.7)	90.1 (88.7–91.4)	97.2 (95.6–98.8)	104.4 (102.4–106.4)	111.6 (109.6–113.6)	118.2 (115.2–120.9)
65	76.0 (73.2–78.8)	82.2 (80.4–84.3)	89.1 (87.3–90.8)	96.2 (94.1–98.0)	103.3 (100.8–105.5)	110.4 (107.9–112.9)	116.9 (113.5–120.2)
70	75.2 (72.1–78.5)	81.3 (79.0–83.9)	88.1 (85.7–90.4)	95.1 (92.3–97.4)	102.2 (99.1–104.8)	109.2 (106.0–112.1)	115.6 (111.5–119.5)

Abbreviations: HC = healthy control; pRNFL = peripapillary retinal nerve fiber layer thickness.

correct for overoptimism (C-index = 0.588). Taking the lowest pRNFL-z category (≤−2), where the optimal cutoff is located, as the reference group, a fully adjusted model showed a lower hazard ratio in all other categories, compared with the reference group (Figure 2C).

To investigate whether the z scores could be interchangeable between different devices, we included 148 PwMS with retinal OCT images acquired from the Cirrus HD-OCT device and fitted their pRNFL thickness into the reference z score curves generated from Spectralis SD-OCT. The optimal pRNFL-z cutoff using the Cirrus HD-OCT device was −0.8. PwMS with lower pRNFL-z (≤−0.8) also had a higher hazard ratio of 1.76 (95% CI 1.02–3.07; $p = 0.044$) for disability increase, compared with those with higher pRNFL-z.

A multivariable logistic regression model with pRNFL-z as a dependent variable also showed that PwMS with lower pRNFL-z had a higher probability of future EDSS worsening (reciprocal OR [95% CI] = 1.17, 95% CI 1.01–1.34; $p = 0.032$) after the 2-year follow-up (eFigure 1). Compared with the continuous analysis, the use of pRNFL-z for dichotomization

Table 2 Estimated GCIP Percentiles With Bootstrap Confidence Intervals Calculated From HC Eyes

Age	1st Percentile	5th Percentile	20th Percentile	50th Percentile	80th Percentile	95th Percentile	99th Percentile
20	61.1 (59.2–64.1)	64.5 (63.0–66.8)	68.4 (67.0–70.2)	72.3 (71.1–74.2)	76.3 (75.0–78.2)	80.4 (78.5–82.5)	84.3 (81.6–87.6)
25	60.6 (58.8–63.2)	64.1 (62.7–65.6)	67.9 (66.8–68.9)	71.8 (70.8–72.6)	75.7 (74.7–76.7)	79.8 (78.2–81.3)	83.7 (81.1–86.2)
30	60.2 (58.3–62.7)	63.6 (62.3–65.1)	67.4 (66.3–68.4)	71.2 (70.3–72.1)	75.2 (74.0–76.2)	79.2 (77.5–80.8)	83.1 (80.3–85.3)
35	59.7 (57.7–61.8)	63.1 (61.7–64.4)	66.8 (65.7–67.8)	70.6 (69.5–71.5)	74.5 (73.3–75.4)	78.6 (76.8–79.8)	82.4 (79.6–84.5)
40	59.3 (57.3–61.4)	62.7 (61.4–64.1)	66.4 (65.2–67.5)	70.2 (69.0–71.3)	74.1 (72.9–75.2)	78.1 (76.7–79.3)	81.9 (79.6–83.8)
45	59.0 (56.8–61.1)	62.3 (60.8–64.0)	66.0 (64.5–67.5)	69.8 (68.4–71.3)	73.7 (72.2–75.1)	77.7 (76.1–79.2)	81.5 (79.1–83.6)
50	58.8 (56.7–61.0)	62.1 (60.7–63.9)	65.8 (64.2–67.4)	69.6 (68.0–71.2)	73.4 (71.8–75.1)	77.4 (75.6–79.0)	81.1 (78.4–83.4)
55	58.5 (56.4–60.8)	61.8 (60.4–63.6)	65.5 (64.1–67.2)	69.2 (67.7–70.9)	73.1 (71.3–74.8)	77.0 (75.0–78.9)	80.7 (78.0–83.2)
60	58.1 (55.8–60.5)	61.4 (59.7–63.3)	65.0 (63.5–66.6)	68.8 (67.0–70.5)	72.6 (70.7–74.5)	76.5 (74.4–78.4)	80.2 (77.6–82.8)
65	57.7 (54.8–60.4)	61.0 (59.0–63.2)	64.6 (62.5–66.5)	68.3 (66.1–70.3)	72.1 (69.6–74.4)	76.0 (73.2–78.2)	79.6 (76.4–82.2)
70	57.3 (53.8–60.6)	60.5 (57.1–63.7)	64.1 (60.8–67.1)	67.8 (64.3–70.9)	71.5 (67.9–74.9)	75.4 (71.4–79.0)	79.1 (74.4–83.3)

Abbreviations: GCIP = combined macular ganglion cell and inner plexiform layer thickness; HC = healthy control.

Table 3 Baseline Characteristics of PwMS From the IMSVISUAL Study by pRNFL Age-Adjusted *z* Score Categories

Variable	pRNFL age-adjusted *z* score					
	≤−2	−2 to −1	−1 to 0	0 to 1	1 to 2	>2
Total, no. (%)	130 (18.0%)	191 (26.5%)	201 (27.8%)	132 (18.3%)	61 (8.4%)	7 (1.0%)
Age, y, mean (SD)	43.3 (11.8)	42.0 (10.9)	39.8 (11.6)	39.2 (11.9)	37.5 (11.5)	40.0 (16.1)
Sex, no. (%)						
Female	83 (63.8%)	126 (66.0%)	131 (65.2%)	88 (66.7%)	50 (82.0%)	5 (71.4%)
Male	47 (36.2%)	65 (34.0%)	70 (34.8%)	44 (33.3%)	11 (18.0%)	2 (28.6%)
Disease course, no. (%)						
CIS	5 (3.8%)	12 (6.3%)	22 (10.9%)	19 (14.4%)	11 (18.0%)	1 (14.3%)
RRMS	82 (63.1%)	145 (75.9%)	156 (77.6%)	97 (73.5%)	45 (73.8%)	5 (71.4%)
PMS	43 (33.1%)	34 (17.8%)	23 (11.4%)	16 (12.1%)	5 (8.2%)	1 (14.3%)
Disease duration, y, median (IQR)	11.7 (6.6–19.3)	7.7 (3.1–15.9)	5.6 (2.2–11.3)	4.1 (1.2–9.4)	4.0 (2.0–7.0)	5.0 (1.7–11.5)
pRNFL, μm, mean (SD)	72.6 (7.1)	86.1 (2.7)	94.7 (2.7)	102.3 (2.5)	111.8 (2.8)	119.6 (2.9)
EDSS score, median (IQR)	3.5 (1.5–5.0)	2.0 (1.5–3.5)	2.0 (1.5–3.0)	1.5 (1.0–3.0)	1.5 (1.0–2.5)	1.5 (1.0–1.5)

Abbreviations: CIS = clinically isolated syndrome; EDSS = Expanded Disability Status Scale; pRNFL = peripapillary retinal nerve fiber layer thickness; PMS = progressive multiple sclerosis; PwMS = people with multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis.

led to a substantially higher rate of EDSS score increase, by incremental increases of cutoff levels (eTable 2).

Sys4MS Cohort: pRNFL Thinning as Risk Factors of Future Disease Activity

The associations of disability accumulation with low pRNFL-z were also supported by results in the Sys4MS validation cohort (eTable 3), either as a continuous (reciprocal OR [95% CI] = 1.51, 95% CI 1.06–2.15; p = 0.024, eFigure 2) or categorical variable (eTable 4). PwMS with lower pRNFL-z also tend to have a higher risk of disease worsening (p = 0.055), albeit no association was found with new clinical relapses (p = 0.524) and developing new T2 lesions (p = 0.436) (eTable 5).

Berlin CIS Cohort: GCIP Thinning as Risk Factors of Future Disease Activity

The clinical characteristics of the 78 PwMS in the Berlin CIS cohort were identical as previously reported,[16] 62 (79.5%) of which had a diagnosis of RRMS and the remaining 16 (20.5%) were considered CIS. At baseline, the mean (SD) age was 33.7 (7.4) years, the median (IQR) time since disease onset was 12.1 (11.8–12.7) months, and the mean (SD) GCIP thickness was 69.7 (7.1) μm. Most participants were women (N = 50, 64.1%), and the median (IQR) follow-up duration was 23.9 (23.3–24.7) months. The baseline clinical and demographic characteristics of the PweMS by GCIP-z categories are listed in Table 4. There were no significant differences in clinical characteristics among different GCIP-z groups.

During follow-up, 46 (59.0%) of the 78 PwMS did not maintain the NEDA-3 status (new clinical relapse: N = 23

(29.5%), new brain MRI activity: N = 38 (48.7%), confirmed EDSS score increase: N = 9 (11.5%)). Using adjusted hazard ratio curves, PweMS with lower GCIP-z had an increased risk of subsequent disease activity during follow-up (Figure 3A). The optimal cutoff values of GCIP-z for stratifying PweMS with a higher risk of future disease activity in the Berlin CIS cohort were −1.03 (Figure 3B). PweMS with lower GCIP-z had a higher risk of not maintaining the NEDA-3 status during follow-up (aHR [95% CI] = 2.14 [1.03–4.43], p = 0.042).

For group comparison, given the numbers of participants within the highest (>2) and lowest (≤−2) z score categories were limited, we merged those 2 groups into the adjacent ones. Afterward, we considered the GCIP-z category where the optimal cutoff was located (≤−1) as the reference group, and a fully adjusted model revealed a significantly lower hazard ratio in those within the highest GCIP-z category compared with the reference group (Figure 3C).

Assessing each component of the NEDA-3 criteria, on a continuous scale, the associations between GCIP-z and risk of reaching each outcome all showed similar patterns of failing the NEDA-3 status. The optimal GCIP-z thresholds for new clinical relapse, new brain MRI activity, and confirmed EDSS score increase were −1.3, −0.17, and 0.56, respectively. For clinical relapse, PweMS with lower GCIP-z showed an increased risk of having new clinical relapse (aHR [95% CI] = 2.81 [1.19–6.66], p = 0.019). For brain MRI activity, higher rates of developing new lesions, albeit borderline significant, were found in those with lower GCIP-z (HR [95% CI] = 1.55

Figure 2 Association Between Risk of Disability Accumulation and pRNFL *z* Scores in PwMS
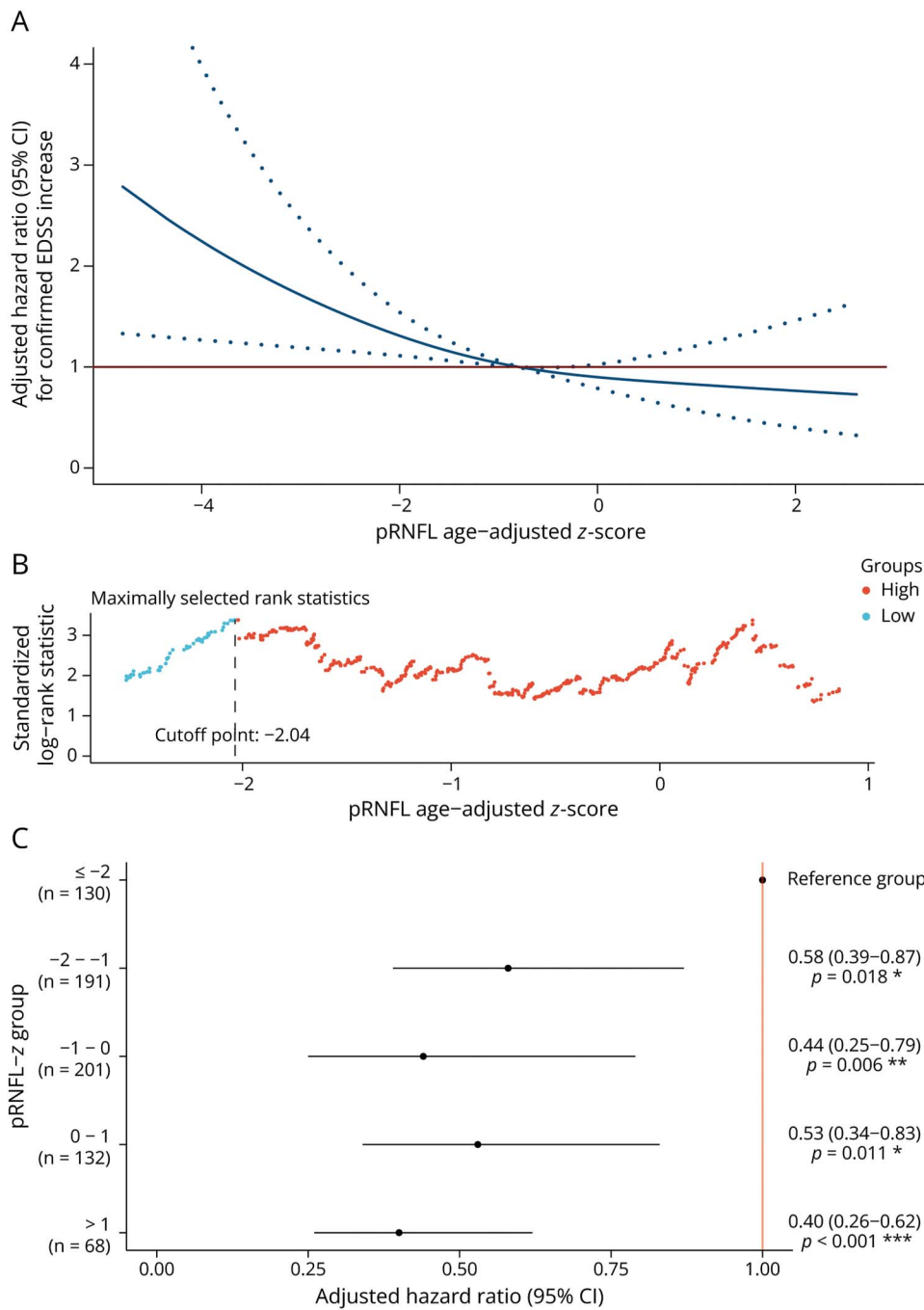
Multivariable analyses adjusted for sex, disease duration, and use of disease-modifying therapies at baseline. (A) Multivariable-adjusted hazard ratio for confirmed EDSS score increase to pRNFL age-adjusted *z* score on a continuous scale. The solid blue line indicates the multivariable-adjusted hazard ratio, with dashed blue lines showing the 95% CIs derived from restricted cubic spline regressions with 3 knots. The solid red line is the reference line for no association at a hazard ratio of 1.0. (B) Optimal pRNFL age-adjusted *z* score cutoff point for determination of PwMS with a higher risk of disability accumulation. (C) Multivariable-adjusted hazard ratio for confirmed EDSS score increase according to categories of pRNFL age-adjusted *z* score. EDSS = Expanded Disability Status Scale; pRNFL = peri-papillary retinal nerve fiber layer thickness; PwMS = people with multiple sclerosis.

[0.92–3.03], $p = 0.087$). Finally, the optimal cutoff for EDSS score increase did not have the power to identify the association between the 2 parameters (aHR [95% CI] = 3.08 [0.77–12.39], $p = 0.112$).

To decrease the complexity of defining multiple cutoffs for different disease activity outcome measures, we applied the threshold for not fulfilling the NEDA-3 status (−1.03) to each component. The results showed a similar pattern that lower GCIP-z were still associated with a higher risk of failing each

NEDA-3 outcome. However, only new clinical relapses showed significant results (aHR [95% CI] = 2.37 [1.06–5.92], $p = 0.023$).

Comparison Between Raw OCT Measures and *z* Score Values

To obtain accurate estimates of predictive performance, we compared the absolute values and *z* scores of retinal layer thickness between different estimated thresholds (tertile and optimal) with 100-times repeated fivefold cross-validation. The optimal cutoff points using *z* scores

Table 4 Baseline Characteristics of PweMS From the Berlin CIS Study by GCIP Age-Adjusted *z* Score Categories

Variable	GCIP age-adjusted *z* score					
	≤−2	−2 to −1	−1 to 0	0 to 1	1 to 2	>2
Total, no. (%)	5 (6.4%)	12 (15.4%)	27 (34.6%)	18 (23.1%)	11 (14.1%)	5 (6.4%)
Age, y, mean (SD)	35.8 (3.9)	31.5 (7.9)	33.4 (9.0)	35.4 (6.9)	33.1 (5.6)	33.8 (5.5)
Sex, no. (%)						
Female	3 (60%)	7 (58.3%)	19 (70.4%)	12 (66.7%)	7 (63.6%)	2 (40%)
Male	2 (40%)	5 (41.7%)	8 (29.6%)	6 (33.3%)	4 (36.4%)	3 (60%)
Disease course, no. (%)						
CIS	2 (40%)	2 (16.7%)	6 (22.2%)	3 (16.7%)	1 (9.1%)	2 (40%)
RRMS	3 (60%)	10 (83.3%)	21 (77.8%)	15 (83.3%)	10 (90.9%)	3 (60%)
Disease duration, y, median (IQR)	1.0 (1.0–1.1)	1.0 (0.9–1.0)	1.0 (1.0–1.1)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.1)
GCIP, μm, mean (SD)	59.2 (1.6)	64.9 (1.5)	68.8 (1.4)	73.0 (1.5)	77.9 (1.4)	83.1 (1.8)
Total T2w lesion count, no., median (IQR)	13 (11–53)	8 (6–12)	24 (7–40)	14 (5–20)	6 (5–11)	18 (1–29)
Total T2w lesion volume, cm^3, median (IQR)	1.5 (0.7–2.7)	0.6 (0.5–1.1)	2.3 (0.9–5.4)	1.1 (0.4–2.1)	0.5 (0.3–1.0)	2.1 (0.1–2.3)
EDSS score, median (IQR)	1.5 (1.0–1.5)	1.0 (0–2.0)	1.5 (1.0–2.0)	2 (1.0.–2.0)	1.5 (0.5–1.5)	1.5 (1.0–2.0)

Abbreviations: CIS = clinically isolated syndrome; EDSS = Expanded Disability Status Scale; GCIP = combined macular ganglion cell and inner plexiform layer thickness; PweMS = people with early multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis.

showed better performance in discrimination power and calibration with the Harrell's C-index and integrated Brier scores (Table 5). In conclusion, *z* scores of both retinal layers using optimal cutoffs provided well-performed predictions.

Discussion

To mitigate age-related effects on the interpretation of OCT measures, we established age-adjusted *z* score reference curves for both pRNFL and GCIP thicknesses and tested their clinical utility for clinical course prognostication in 3 MS cohorts of adult patients. Our findings suggest that PwMS with lower pRNFL-z and GCIP-z are at greater risk of future disability worsening, as well as clinical and radiologic disease activity. In addition, we found that models using *z* scores as predictors of MS disease course outperformed those of absolute thickness values and that age-adjusted *z* scores may have utility across OCT devices. Our collective study findings demonstrate that pRNFL-z and GCIP-z can serve as clinically feasible prognostic biomarkers for identifying PwMS at higher risk of having a worse prognosis.

Prognostic biomarkers that identify poor clinical outcomes for initiation of higher efficacy immunotherapies are of importance in slowing down irreversible neurologic deficits for PwMS. With increasing knowledge of retinal imaging in MS, there has been growing research and clinical interest in investigating the potential of retinal OCT measures as

biomarkers for predicting disease course, as well as disease monitoring. We have previously demonstrated the potential of both low pRNFL and GCIP thicknesses as risk factors of disability accumulation and disease activity.[14-16] In this study, both pRNFL and GCIP thicknesses were included to validate the results under different methodological approaches. Previous IMSVISUAL study[14] used pRNFL thickness tertiles as arbitrary cutoffs to separate 879 PwMS into 3 groups, with the cutoffs for pRNFL measured by Spectralis SD-OCT at 87 and 97 μm and those for Cirrus HD-OCT at 87 and 98 μm. The results showed that PwMS with baseline pRNFL thicknesses in the lowest tertile group had a 2-fold risk of disability worsening after a median follow-up of 2.0 years (range 0.5–5 years) compared with those within the intermediate and highest tertile groups. The risk tended to be higher in those with longer follow-up durations, with hazard ratios of 1.36 and 3.81 at 1–3 and 3–5 years, respectively.

In addition to disability accumulation, NEDA-3 is another important measure of treatment response in clinical trials.[31] Loss of NEDA-3 status is associated with increased evidence of brain[43] and retinal[44] neurodegeneration. To further investigate the potential of OCT measures for prognostication, both Zimmermann et al.[15] and Lin et al.[16] expanded the outcome parameters by including NEDA-3 as a measure of disease activity. The former study included 97 PweMS from 2 centers with a median follow-up duration of 2.0 years (IQR 1.8–2.5 years) and used OCT measures acquired only from the Spectralis SD-OCT device. The tertile dividers for pRNFL

Figure 3 Association Between Risk of Future Disease Activity and GCIP *z* Scores in PweMS
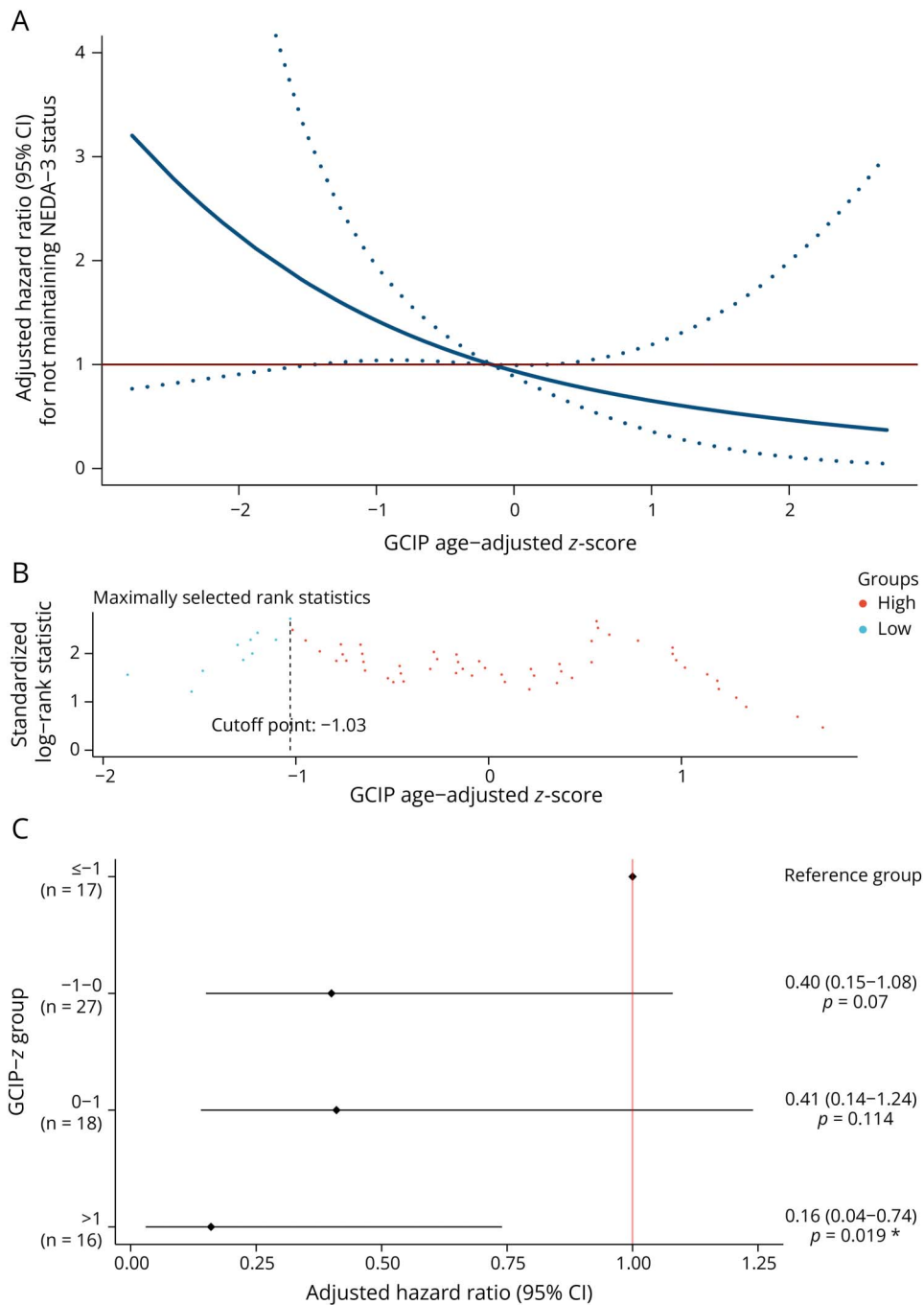
Multivariable analyses adjusted for sex, disease duration, and use of disease-modifying therapies at baseline. (A) Multivariable-adjusted hazard ratio for new disease activity to GCIP age-adjusted *z* score on a continuous scale. The solid blue line indicates the multivariable-adjusted hazard ratio, with dashed blue lines showing the 95% CIs derived from restricted cubic spline regressions with 3 knots. The solid red line is the reference line for no association at a hazard ratio of 1.0. (B) Optimal GCIP age-adjusted *z* score cutoff point for determination of PwMS with higher risk of having new disease activity. (C) Multivariable-adjusted hazard ratio for new disease activity according to categories of GCIP age-adjusted *z* score. GCIP = combined macular ganglion cell and inner plexiform layer thickness; PweMS = people with early multiple sclerosis.

thickness were 96 and 104 μm and for GCIP thickness were 69.3 and 74.2 μm. PweMS in the lowest pRNFL thickness tertile group had twice the risk of not maintaining the NEDA-3 status compared with those in the highest tertile group. Of interest, GCIP thickness at baseline seemed to have superior prognostic utility over pRNFL thickness, with more than a threefold increase in the risk of not meeting the NEDA-3 criteria in the lowest GCIP thickness tertile group compared with the highest. Similar results were found using median splits (pRNFL: 100 μm; GCIP: 70.4 μm) instead of tertiles as

cutoffs. Of the 3 components of the NEDA-3 criteria, the aforementioned results were mainly driven by new clinical relapses.

The expansion in the number of available DMTs over the past decade, as well as anticipated future increase in new DMTs with novel mechanisms of action, highlights the need to standardize biomarker measures that account for potential confounders because it is a crucial step toward precision medicine.[45] Furthermore, considering the highly variable

Table 5 Performance Comparison of Raw OCT Measures, z Scores, and Different Threshold Definitions

Method	pRNFL for EDSS score increase		GCIP for not fulfilling NEDA-3 criteria	
	C-index	Integrated Brier score	C-index	Integrated Brier score
Optimal z score cutoff	0.588 ± 0.155	0.133 ± 0.005	0.580 ± 0.161	0.164 ± 0.008
Optimal raw value cutoff	0.576 ± 0.151	0.134 ± 0.004	0.561 ± 0.149	0.166 ± 0.009
Tertile z score cutoff	0.578 ± 0.149	0.137 ± 0.005	0.563 ± 0.152	0.167 ± 0.006
Tertile raw value cutoff	0.576 ± 0.158	0.138 ± 0.006	0.559 ± 0.169	0.169 ± 0.009

Abbreviations: C-index = concordance index; EDSS = Expanded Disability Status Scale; GCIP = combined macular ganglion cell and inner plexiform layer thickness; NEDA-3 = no evidence of disease activity; OCT = optical coherence tomography; pRNFL = peripapillary retinal nerve fiber layer thickness.

disease course of MS, a composition of biomarkers reflecting different pathophysiologic features could provide a valuable tool for disease monitoring. Our previous work has shown an additive and potentially synergistic effect by combining OCT-derived measures with serum neurofilament light chain levels.[16] Serum neurofilament light chain is a biomarker that predominantly reflects acute axonal injury[46] while retinal atrophy reflects chronic neuroaxonal degeneration.[47] The combination of the 2 biomarkers can add extra values to each other. However, a critical concern for a multimodal approach to integrating biomarkers is heterogeneity. Therefore, standardization of different biomarker representations with normative data is indispensable to combine biomarkers and provide further individualized clinical insights.

The z score approach for OCT measures allows for straightforward calibration of possible influences, including physiologic parameters relevant to retinal atrophy as well as different OCT devices. In addition, for clinical interpretability, the z score is more advantageous compared with raw values. Retinal atrophy may simply seem as a normal characteristic of aging, supporting the need for optimizing the results observed in studies.[48,49] However, the best-fitting models for associations between retinal thinning and age have yet to be confirmed in large international cohort studies. While some studies showed a linear association between the 2 parameters, others demonstrate a nonlinear correlation with a faster thinning rate occurring after the age of 40.[23,50] In our study, we applied GAMLSS to select the optimal outcome distribution and found that GCIP thinning remained relatively stable as age increased while pRNFL thickness showed an accelerated thinning rate after around the age of 45. Age-adjusted reference tables for both pRNFL and GCIP thickness were generated accordingly, allowing clinicians to interpret retinal layer thickness values without interference from the factor of age.

For OCT devices, one straightforward solution to address discrepancies between devices is to define different thresholds for each platform, as previously described.[14] Alternatively, conversion equations have been developed to relate OCT measures from different devices: pRNFL (Cirrus HD-OCT value = −5.0 + 1.0 * Spectralis SD-OCT value) and GCIP (Cirrus HD-OCT value = −4.5 + 0.9 * Spectralis SD-OCT value) based on data from 546 HCs in 11 centers across North America and Europe.[23] Nevertheless, both of the proposed approaches may increase the burden for clinicians to interpret the results before applying the methods to routine clinical use. In our study, we investigated the applicability of simply applying OCT measures obtained from Cirrus HD-OCT on reference z scores generated by Spectralis SD-OCT devices. Although the optimal pRNFL-z for identifying future disability accumulation between the 2 platforms were different (−2.04 vs −0.8), our results showed that the reference z scores generated from Spectralis SD-OCT remain applicable to measures derived from Cirrus HD-OCT devices. Nevertheless, because we only have a relatively small sample size of OCT images acquired from the Cirrus HD-OCT platform, the z score threshold we calculated for the Cirrus HD-OCT devices may not be optimal. Therefore, further investigation into device-specific reference z scores is warranted.

In the abovementioned studies investigating pRNFL and GCIP thicknesses for prognostication of future disease activity, the pathologic definitions of both OCT measures were rather subjective,[51] raising the levels of uncertainty when interpreting the results. Therefore, we find it necessary to demonstrate the association between retinal layer thickness and risk of subsequent disability accumulation or disease activity on a continuous scale. Our results demonstrated a nonlinear, yet unidirectional, inverse association between the 2 variables. In the logistic regression analyses, our results also confirm that lower pRNFL-z are associated with future disability accumulation. Lower GCIP-z, on the contrary, still displayed a stronger link with new clinical relapses as in Cox proportional hazards models. These results enhanced the credibility and filled the knowledge gaps of previous studies. Nevertheless, clear definitions of pathologic findings are still required before accepting these biomarkers as clinically feasible. We found that the optimal cutoff for pRNFL-z as a risk factor of future disability accumulation is approximately one SD away from the optimal cutoff for GCIP-z as a risk factor of

future disease activity. The differences in cutoff values could be explained by the fact that (1) there were only 5 PweMS with GCIP-z lower than −2 in the Berlin CIS cohort, limiting the statistical power for further discrimination, and (2) compared with the IMSVISUAL cohort, PwMS in the Berlin CIS cohort were in a relatively early stage of the disease (median disease duration 1 year vs 6.5 years). Nevertheless, despite being able to define an optimal threshold for both retinal OCT measures, it remains unclear whether there should be a gray zone in measures around pathologic thresholds. In addition, clinicians should compromise on the sensitivity and specificity by using cutoff values that are more aligned with their goals (e.g., including more PwMS for screening or including less PwMS for more aggressive treatment options). Each of these issues require prospective investigations in larger clinical trials for resolution.

Our study contains several limitations. First, the HCs included in the generation of the reference database were relatively homogenous. Although our HCs were recruited at multiple centers, the majority of participants were of Caucasian ethnicity and solely from Europe. A previous study has shown that healthy African Americans seemed to have a thicker pRNFL than Caucasian Americans while there was no difference in GCIP thickness.[52] This highlights the need to establish separate reference curves for different races/ethnicities. The number of African Americans in our HC group was less than 5%, limiting the possibility to apply GAMLSS without overfitting and provide a meaningful reference database that accounts for ethnic effects. In addition, African American PwMS tend to have faster pRNFL and GCIP thinning rates in comparison with Caucasian American PwMS.[52] The ethnic homogeneity of our HC and MS cohorts could limit the generalizability of our results. As a pilot study, we were only able to account for one major influence on OCT measures. During our validation process, we also added biological sex as the second adjustable factor in the GAMLSS. However, the goodness-of-fit and normality in the female model were inferior to those in the male model. Therefore, we are confident that our GAMLSS is robust when solely adjusted for age. For comorbidities, because different comorbidities have different levels of impact on OCT measures, we simply excluded people with any medical history of ophthalmologic or neurologic disorders to maintain the normality of our models. Thus, in cases where people have ophthalmologic or neurologic comorbidities, the reference curves may differ. Comprehensive reference z score curves that consider other physiologic parameters or different OCT platforms are yet to be developed. Nevertheless, although direct interchangeability of z scores between devices is lacking, our utilization of the Spectralis SD-OCT reference curve on the Cirrus HD-OCT device still yielded significant outcomes. Furthermore, the IMSVISUAL cohort was composed before the introduction of the 2017 McDonald criteria, and the data might be considered outdated. To address this concern, we have included an additional multicenter MS cohort to emphasize the relevance of our findings. Finally, our ability

to build GCIP age-adjusted reference values and assess the utility of GCIP-z for predicting future disease progression within the IMSVISUAL cohort was hindered by variations in macular scan segmentation algorithms used by different centers. A consistent and representative segmentation method for macular scans should be established across centers to increase the possibility of multicenter standardization.

To conclude, our results support the potential of both pRNFL-z and GCIP-z as promising biomarkers with prognostic value for future disease progression in MS. Generating retinal layer thickness z scores that control for age and other confounding factors may be a robust approach for individual prognostication. Larger normative cohorts reflecting the general population should be constructed to establish influence-free clinical references for the interpretation of OCT results. This can enhance the clinical utility of the OCT biomarkers and add value to future clinical trial designs.

Disclosure

T.-Y. Lin has received compensation from ADA Health, unrelated to the presented work; S. Asseyer has received conference grant from Celgene and speaking honoraria from Bayer Healthcare, Roche, and Alexion; C. Chien has received research support from Novartis and Alexion and writing honoraria from the British Society for Immunology, as well as serves as a member of the Standing Committee on Science for the Canadian Institutes of Health Research (CIHR); S. Saidha has received consulting fees from Medical Logix for the development of CME programs in neurology and has served on scientific advisory boards for Biogen, Novartis, Genentech Corporation, TG therapeutics, Clene Pharmaceuticals & ReWind therapeutics. He has performed consulting for Novartis, Genentech Corporation, JuneBrain LLC, Innocare Pharma, Kiniksa pharmaceuticals and Lapix therapeutics. 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; P.A. Calabresi has received consulting fees from Lilly, Idorsia, and Novartis; and is PI on grants to Johns Hopkins from Genentech and the Myelin repair Foundation; P. Villoslada has received an honorarium from Heidelberg Engineering in 2014, has

Publication History

Appendix Authors

Name	Location	Contribution
Ting-Yi Lin, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Seyedamirhosein Motamedi, PhD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Susanna Asseyer, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Claudia Chien, PhD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center; Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Shiv Saidha, MD	Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Peter A. Calabresi, MD	Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Kathryn C. Fitzgerald, ScD	Department of Neurology, Johns Hopkins School of Medicine; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Sara Samadzadeh, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; Department of Regional Health Research and Molecular Medicine, University of Southern Denmark, Odense; Department of Neurology, Slagelse Hospital, Denmark	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Pablo Villoslada, MD, PhD	Department of Neurology, Hospital Del Mar - Pompeu Fabra University, Barcelona, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Sara Llufriu, MD, PhD	Neuroimmunology and Multiple Sclerosis Unit, Hospital Clinic Barcelona and IDIBAPS, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Ari J. Green, MD	Department of Neurology, University of California San Francisco	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Jana Lizrova Preiningerova, MD, PhD	Department of Neurology, Charles University in Prague, Czech Republic	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Axel Petzold, MD, PhD	Moorfield's Eye Hospital, The National Hospital for Neurology and Neurosurgery, Queen Square Institute of Neurology, University College London, United Kingdom; Neuro-ophthalmology Expert Center, Amsterdam UMC, Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Letizia Leocani, MD, PhD	Experimental Neurophysiology Unit, Institute of Experimental Neurology (INSPE), San Raffaele Scientific Institute; Vita-Salute San Raffaele University, Milan, Italy	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Elena Garcia-Martin, MD, PhD	Miguel Servet University Hospital, Zaragoza, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Celia Oreja-Guevara, MD, PhD	Department of Neurology, Hospital Clínico Universitario San Carlos, Madrid, Spain	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Olivier Outteryck, MD	Department of Neurology; Department of Neuroradiology, Centre Hospitalier Universitaire de Lille, France	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Patrick Vermersch, MD, PhD	Department of Neurology; Department of Neuroradiology, Centre Hospitalier Universitaire de Lille, France	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Laura J. Balcer, MD, MSCE	Departments of Neurology, Population Health and Ophthalmology, NYU Grossman School of Medicine, NY	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Rachel Kenney, PhD	Departments of Neurology, Population Health and Ophthalmology, NYU Grossman School of Medicine, NY	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Philipp Albrecht, MD	Department of Neurology, Heinrich-Heine-University, Düsseldorf, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Orhan Aktas, MD	Department of Neurology, Heinrich-Heine-University, Düsseldorf, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Fiona Costello, MD	Departments of Clinical Neurosciences and Surgery Cumming School of Medicine, University of Calgary, Alberta, Canada	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Jette Frederiksen, MD	Clinic of Optic Neuritis and Clinic of Multiple Sclerosis, Department of Neurology, Rigshospitalet - Glostrup, Denmark	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Antonio Uccelli, MD	Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Italy	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data

Appendix *(continued)*

Name	Location	Contribution
Maria Cellerino, MD	Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Italy	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Elliot M. Frohman, MD	Laboratory of Neuroimmunology, Professor Lawrence Steinman, Stanford University School of Medicine, Palo Alto, CA	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Teresa C. Frohman, MD	Laboratory of Neuroimmunology, Professor Lawrence Steinman, Stanford University School of Medicine, Palo Alto, CA	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Judith Bellmann-Strobl, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Tanja Schmitz-Hübsch, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Klemens Ruprecht, MD	Department of Neurology, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Alexander U. Brandt, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design

Appendix *(continued)*

Name	Location	Contribution
Hanna G. Zimmermann, PhD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin; Einstein Center Digital Future, Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Friedemann Paul, MD	Charité - Universitätsmedizin Berlin; Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité - Universitätsmedizin Berlin; Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC); Neuroscience Clinical Research Center, Charité - Universitätsmedizin Berlin; Department of Neurology, Charité - Universitätsmedizin Berlin, Germany	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data

References

1. Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. *N Engl J Med.* 2018; 378(2):169-180. doi:10.1056/NEJMra1401483
2. Pitt D, Lo CH, Gauthier SA, et al. Toward precision phenotyping of multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2022;9(6):e200025. doi:10.1212/NXI.0000000000200025
3. Leocani L, Rocca M, Comi G. MRI and neurophysiological measures to predict course, disability and treatment response in multiple sclerosis. *Curr Opin Neurol.* 2016;29(3):243-253. doi:10.1097/WCO.0000000000000333
4. Kappos L, Wolinsky JS, Giovannoni G, et al. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurol.* 2020;77(9):1132-1140. doi:10.1001/jamaneurol.2020.1568
5. Lublin FD, Häring DA, Ganjgahi H, et al. How patients with multiple sclerosis acquire disability. *Brain.* 2022;145(9):3147-3161. doi:10.1093/brain/awac016
6. Kuchling J, Paul F. Visualizing the central nervous system: imaging tools for multiple sclerosis and neuromyelitis optica spectrum disorders. *Front Neurol.* 2020;11:450. doi:10.3389/fneur.2020.00450
7. Graves JS, Oertel FC, Van der Walt A, et al. Leveraging visual outcome measures to advance therapy development in neuroimmunologic disorders. *Neurol Neuroimmunol Neuroinflamm.* 2022;9(2):e1126. doi:10.1212/NXI.0000000000001126
8. Balcer LJ, Miller DH, Reingold SC, Cohen JA. Vision and vision-related outcome measures in multiple sclerosis. *Brain.* 2015;138(Pt 1):11-27. doi:10.1093/brain/awu335
9. Oberwahrenbrock T, Weinhold M, Mikolajczak J, et al. Reliability of intra-retinal layer thickness estimates. *PLoS One.* 2015;10(9):e0137316. doi:10.1371/journal.pone.0137316
10. Petzold A, Wattjes MP, Costello F, et al. The investigation of acute optic neuritis: a review and proposed protocol. *Nat Rev Neurol.* 2014;10(8):447-458. doi:10.1038/nrneurol.2014.108
11. Petzold A, Balcer L, Calabresi PA, et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol.* 2017;16(10):797-812. doi:10.1016/S1474-4422(17)30278-8
12. Oberwahrenbrock T, Traber GL, Lukas S, et al. Multicenter reliability of semi-automatic retinal layer segmentation using OCT. *Neurol Neuroimmunol Neuroinflamm.* 2018;5(3):e449. doi:10.1212/NXI.0000000000000449
13. Britze J, Pihl-Jensen G, Frederiksen JL. Retinal ganglion cell analysis in multiple sclerosis and optic neuritis: a systematic review and meta-analysis. *J Neurol.* 2017; 264(9):1837-1853. doi:10.1007/s00415-017-8531-y
14. Martinez-Lapiscina EH, Arnow S, Wilson JA, et al. Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study. *Lancet Neurol.* 2016;15(6):574-584. doi:10.1016/S1474-4422(16)00068-5
15. Zimmermann HG, Knier B, Oberwahrenbrock T, et al. Association of retinal ganglion cell layer thickness with future disease activity in patients with clinically isolated syndrome. *JAMA Neurol.* 2018;75(9):1071-1079. doi:10.1001/jamaneurol.2018.1011
16. Lin T-Y, Vitkova V, Asseyer S, et al. Increased serum neurofilament light and thin ganglion cell-inner plexiform layer are additive risk factors for disease activity in early multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm.* 2021;8(5):e1051. doi:10.1212/NXI.0000000000001051
17. Cilingir V, Batur M. First measured retinal nerve fiber layer thickness in RRMS can be used as a biomarker for the course of the disease: threshold value discussions. *J Neurol.* 2021;268(8):2858-2865. doi:10.1007/s00415-021-10469-x
18. Wauschkuhn J, Solorza Buenrostro G, Aly L, et al. Retinal ganglion cell loss is associated with future disability worsening in early relapsing–remitting multiple sclerosis. *Eur J Neurol.* 2023;30(4):982-990. doi:10.1111/ene.15681
19. Lambe J, Fitzgerald KC, Murphy OC, et al. Association of spectral-domain OCT with long-term disability worsening in multiple sclerosis. *Neurology.* 2021;96(16): e2058-e2069. doi:10.1212/WNL.0000000000011788
20. Gafson A, Craner MJ, Matthews PM. Personalised medicine for multiple sclerosis care. *Mult Scler.* 2017;23(3):362-369. doi:10.1177/1352458516672017
21. Benkert P, Meier S, Schaedelin S, et al. Serum neurofilament light chain for individual prognostication of disease activity in people with multiple sclerosis: a retrospective modelling and validation study. *Lancet Neurol.* 2022;21(3):246-257. doi:10.1016/S1474-4422(22)00009-6
22. Koh VT, Tham YC, Cheung CY, et al. Determinants of ganglion cell-inner plexiform layer thickness measured by high-definition optical coherence tomography. *Investig Ophthalmol Vis Sci.* 2012;53(9):5853-5859. doi:10.1167/iovs.12-10414
23. Kenney R, Liu M, Hasanaj L, et al. Normative data and conversion equation for spectral-domain optical coherence tomography in an international healthy control cohort. *J Neuroophthalmol.* 2022;42(4):442-453. doi:10.1097/WNO.0000000000001717
24. Motamedi S, Gawlik K, Ayadi N, et al. Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline. *Front Neurol.* 2019;10:1117. doi:10.3389/fneur.2019.01117
25. Warner CV, Syc SB, Stankiewicz AM, et al. The impact of utilizing different optical coherence tomography devices for clinical purposes and in multiple sclerosis trials. *PLoS One.* 2011;6(8):e22947. doi:10.1371/journal.pone.0022947
26. Pierro L, Gagliardi M, Iuliano L, Ambrosi A, Bandello F. Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. *Investig Ophthalmol Vis Sci.* 2012;53(9):5912-5920. doi:10.1167/iovs.11-8644
27. Brune S, Høgestøl EA, de Rodez Benavent SA, et al. Serum neurofilament light chain concentration predicts disease worsening in multiple sclerosis. *Mult Scler.* 2022; 28(12):1859-1870. doi:10.1177/13524585221097296
28. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Ann Neurol.* 2011;69(2):292-302. doi:10.1002/ana.22366
29. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology.* 2014;83(3):278-286. doi:10.1212/WNL.0000000000000560
30. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018;17(2):162-173. doi:10.1016/S1474-4422(17)30470-2
31. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort. *JAMA Neurol.* 2015;72(2):152-158. doi:10.1001/jamaneurol.2014.3537
32. Andorra M, Freire A, Zubizarreta I, et al. Predicting disease severity in multiple sclerosis using multimodal data and machine learning. *J Neurol.* 2024;271(3): 1133-1149. doi:10.1007/s00415-023-12132-z
33. Tewarie P, Balk L, Costello F, et al. The OSCAR-IB consensus criteria for retinal OCT quality assessment. *PLoS One.* 2012;7(4):e34823. doi:10.1371/journal.pone.0034823
34. Schippling S, Balk LJ, Costello F, et al. Quality control for retinal OCT in multiple sclerosis: validation of the OSCAR-IB criteria. *Mult Scler.* 2015;21(2):163-170. doi:10.1177/1352458514538110
35. Aytulun A, Cruz-Herranz A, Aktas O, et al. APOSTEL 2.0 recommendations for reporting quantitative optical coherence tomography studies. *Neurology.* 2021;97(2): 68-79. doi:10.1212/WNL.0000000000012125
36. Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med.* 2004;23(19): 3053-3076. doi:10.1002/sim.1861
37. Tintore M, Rovira À, Río J, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain.* 2015;138(Pt 7):1863-1874. doi:10.1093/brain/awv105
38. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci.* 1996; 11(2):89-121. doi:10.1214/ss/1038425655

39. Molinari N, Daurès JP, Durand JF. Regression splines for threshold selection in survival data analysis. *Stat Med.* 2001;20(2):237-247. doi:10.1002/1097-0258(20010130)20:2<237::aid-sim654>3.0.co;2-i

40. Chang C, Hsieh MK, Chang WY, Chiang AJ, Chen J. Determining the optimal number and location of cutoff points with application to data of cervical cancer. *PLoS One.* 2017;12(4):e0176231. doi:10.1371/journal.pone.0176231

41. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* 2011;30(10):1105-1117. doi:10.1002/sim.4154

42. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med.* 1999;18(17-18):2529-2545. doi:10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5

43. Damasceno A, Damasceno BP, Cendes F. No evidence of disease activity in multiple sclerosis: Implications on cognition and brain atrophy. *Mult Scler.* 2016;22(1):64-72. doi:10.1177/1352458515604383

44. Pisa M, Guerrieri S, Di Maggio G, et al. No evidence of disease activity is associated with reduced rate of axonal retinal atrophy in MS. *Neurology.* 2017;89(24):2469-2475. doi:10.1212/WNL.0000000000004736

45. Chitnis T, Prat A. A roadmap to precision medicine for multiple sclerosis. *Mult Scler.* 2020;26(5):522-532. doi:10.1177/1352458519881558

46. Siller N, Kuhle J, Muthuraman M, et al. Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. *Mult Scler.* 2019;25(5):678-686. doi:10.1177/1352458518765666

47. Balk LJ, Cruz-Herranz A, Albrecht P, et al. Timing of retinal neuronal and axonal loss in MS: a longitudinal OCT study. *J Neurol.* 2016;263(7):1323-1331. doi:10.1007/s00415-016-8127-y

48. Budenz DL, Anderson DR, Varma R, et al. Determinants of normal retinal nerve fiber layer thickness measured by stratus OCT. *Ophthalmology.* 2007;114(6):1046-1052. doi:10.1016/j.ophtha.2006.08.046

49. Poon LYC, Antar H, Tsikata E, et al. Effects of age, race, and ethnicity on the optic nerve and peripapillary region using spectral-domain OCT 3d volume scans. *Transl Vis Sci Technol.* 2018;7(6):12. doi:10.1167/tvst.7.6.12

50. Talman LS, Bisker ER, Sackel DJ, et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. *Ann Neurol.* 2010;67(6):749-760. doi:10.1002/ana.22005

51. Iacobucci D, Posavac SS, Kardes FR, Schneider MJ, Popovich DL. Toward a more nuanced understanding of the statistical properties of a median split. *J Consum Psychol.* 2015;25(4):652-665. doi:10.1016/j.jcps.2014.12.002

52. Kimbrough DJ, Sotirchos ES, Wilson JA, et al. Retinal damage and vision loss in African American multiple sclerosis patients. *Ann Neurol.* 2015;77(2):228-236. doi:10.1002/ana.24308