Reproducibility of Optical Coherence Tomography in Multiple Sclerosis

Deanna Cettomai, BS; Mathew Pulicken, MBBS, MHS; Eliza Gordon-Lipkin, BS; Amber Salter, MPH; Teresa C. Frohman, BA; Amy Conger, BA; Xiao Zhang, PhD; Gary Cutter, PhD; Laura J. Balcer, MD, MSCE; Elliot M. Frohman, MD, PhD; Peter A. Calabresi, MD

Background: Optical coherence tomography (OCT) is a promising new method of quantifying axon thickness in the retinal nerve fiber layer (RNFL) that has been used predominantly by ophthalmologists to monitor glaucoma. Optical coherence tomography is being considered as a potential outcome measure in multiple sclerosis (MS) clinical trials, but no data exist on the reproducibility of this technique in MS centers.

Objective: To determine the reproducibility of OCT measurement of mean RNFL thickness in the undilated eyes of healthy control subjects and patients with MS.

Design: Prospective analysis of 4 healthy controls to determine interrater, intrarater, and longitudinal reproducibility. Cross-sectional analysis of 3 cohorts of patients with MS (n=396) and healthy controls (n=153).

Setting: Multiple sclerosis clinics at 3 academic medical centers.

Patients or Other Participants: Healthy controls and patients with MS.

Main Outcome Measure: Thickness of RNFL.

Results: We found excellent agreement with respect to interrater (intraclass correlation [ICC], 0.89), intrarater (ICC, 0.98), and intervisit (ICC, 0.91) results. Mean RNFL thickness did not vary significantly among research centers for patients with MS (93, 92, and 90 µm) or among healthy controls (103, 105, and 104 µm) by site.

Conclusions: We demonstrate that mean RNFL thickness can be reproducibly measured by trained technicians in an MS center using the OCT-3 model. The RNFL measures from cohorts of age-matched controls and patients with MS from 3 different research centers were remarkably similar.

Arch Neurol. 2008;65(9):1218-1222

Optical coherence tomography (OCT) is a noninvasive high-resolution technique that uses near-infrared light to generate cross-sectional tomographic images of tissues, including the retinal nerve fiber layer (RNFL). Optical coherence tomography is used to monitor retinal ganglion cell axon loss in glaucoma, diabetic retinopathy, traumatic optic neuropathy, chiasmal lesions, and optic neuritis. Recently, OCT has been studied in patients with multiple sclerosis (MS) (hereinafter referred to as MS patients), of whom 80% experience visual impairment. Decreased RNFL thickness has been demonstrated in patients with a history of optic neuritis. Two studies showed that the eyes without optic neuritis among MS patients have decreased RNFL thickness compared with the eyes of control subjects, suggesting that retinal ganglion cell axonal loss occurs separately from acute optic neuritis in MS patients. In addition, RNFL thickness correlates with low-contrast visual acuity and contrast sensitivity. This suggests that OCT can be used to monitor axonal injury and visual dysfunction in MS and may be a useful outcome measure in clinical trials.

Whether reproducibility studies completed on earlier OCT models are applicable to the OCT-3 model (Carl Zeiss Meditec, Dublin, California) is unclear because foveal thickness measurements obtained using the prototype OCT scanner and OCT-3 are not directly comparable. The reproducibility of RNFL thickness has been examined using OCT-3 in healthy subjects and cohorts with glaucoma, ocular hypertension, macular edema, and diabetes mellitus but not in MS cohorts. These studies were performed by ophthalmologists on subjects with pharmacologically dilated pupils. All found good reproducibility of RNFL measurements. However, ocular symptoms in
MS, including nystagmus, can have an important effect on visual fixation—an essential component in obtaining high-quality OCT scans—and, therefore, present unique challenges in MS patients, with the potential to decrease OCT reproducibility in this cohort. Other possible impediments to the reliable use of OCT in an MS center include the facts that neurology patients do not routinely have their pupils dilated and that neurology office staff are not trained in the use of slitlamp examination. We hypothesized that, despite these issues, OCT could be performed reproducibly in the setting of an MS center.

METHODS

SUBJECTS

Interrater, intrarater, and intervisit reproducibility studies were performed at the Johns Hopkins MS Center. We examined both eyes of 4 healthy subjects recruited from the staff of the neurology department. The cross-center comparison was performed using cross-sectional data obtained from age- and sex-matched MS patients (n=396) and healthy controls (n=153) at the MS centers of The Johns Hopkins University (JHU), the University of Pennsylvania (U Penn), and The University of Texas Southwestern Medical Center (UTSW). We included subjects with no history of intraocular surgery, glaucoma, retinal disease, diabetes, or hypertension and who completed informed consent. All MS disease subtypes were included. Data from the initial scans of all patients and controls at each center were included in the cross-center comparison.

OPTICAL COHERENCE TOMOGRAPHY

The RNFL measurements were obtained using the OCT-3 fast RNFL thickness protocol, which performs 3 consecutive 3.4-mm-diameter circular scans centered on the optic nerve head. In addition, OCT software (OCT 4.0, version A2; Carl Zeiss Meditec) generated a mean RNFL thickness measurement for 360° around the optic disc, 4 retinal quadrants, and 12 clock hour segments (30° for each hour position).

STUDY PROTOCOL

The RNFL scans were obtained on both eyes of 4 healthy subjects by 3 technicians during 5 consecutive weekly visits. Interrater and intervisit reproducibility were obtained from these data. At visit 3, 1 investigator performed 3 consecutive RNFL scans on each eye of each subject to determine intrarater reproducibility. All scans were performed without pupil dilation. In the cross-center comparison, 1 OCT scan was obtained of each eye of the MS patients and healthy controls.

STATISTICAL ANALYSIS

We used the intraclass correlation coefficient (ICC) as a summary measure for interrater, intrarater, and intervisit agreement. The ICC represents the proportion of variance in data explained by between-subject differences; the higher the ICC (maximum value, 1.0), the better the agreement between measures of the same patient. An ICC of less than 0.40 indicates poor reproducibility; of 0.40 to 0.75, fair to good reproducibility; and of greater than 0.75, excellent reproducibility.

In this study design, there was complex nesting, which requires large sample sizes for simultaneous estimates of desired measures; thus, we used random-effects general linear models (Proc MIXED in SAS; SAS Institute Inc, Cary, North Carolina) to compare ICCs between groups of patients, observers, and longitudinally, treating certain factors as fixed and others as random, depending on the ICC being estimated. Variance homogeneity and ICC homogeneity tests were used to validate assumptions made for estimates. We used only 1 eye in each analysis to enable consistency with the literature but present values for each eye. The high correlations between eyes and the relative consistency of the results demonstrate that little additional information was available by incorporating both eyes in the same analysis.

RESULTS

SUBJECT CHARACTERISTICS

We studied 8 eyes of 4 healthy subjects for the interrater, intervisit, and intrarater portions of this study. There were 2 men and 2 women, and the mean (SD) age was 23 (3) years (range, 20-27 years). In the cross-center comparison, MS patients and controls across 3 centers did not differ significantly in demographic characteristics (Table 1).

RNFL THICKNESS

The ICCs were first calculated by combining data from each eye of each subject and indicated excellent agreement (Table 2). Quadrant ICCs ranged from 0.66 to 0.98 and were slightly lower than mean RNFL ICCs, which ranged...
We found that all RNFL measurements showed excellent ICCs when examined for intrarater, interrater, and intervisit reproducibility. Intrarater reproducibility was stronger than intervisit reproducibility, indicating that reproducibility within a given eye on a given day is greater than reproducibility within a given eye on different days.\textsuperscript{21,30} Quadrant thicknesses were more variable than were mean RNFL thickness. The lower ICC for quadrants suggests that quadrant analyses, although potentially more sensitive to subtle changes, will decrease power in clinical trials owing to poorer reproducibility. However, mean RNFL thicknesses are sensitive to abnormalities in MS patients and highly reproducible, making them appropriate to use for comparisons. The major limitation of this portion of our investigation was the small sample size studied.

Our data (Tables 2 and 3) are comparable to previously reported ICCs. One study measuring intervisit reproducibility\textsuperscript{30} reported a mean RNFL thickness ICC of 0.83 and ICCs for quadrant thicknesses ranging from 0.62 to 0.81, whereas another group examining intrarater reproducibility\textsuperscript{33} found a mean RNFL thickness ICC of 0.95 with quadrant thickness ICCs varying from 0.79 to 0.97. Finally, a recent study of patients with glaucoma\textsuperscript{34} found an intrarater mean RNFL thickness ICC of 0.98 and an intervisit mean RNFL thickness ICC of 0.96.
The most reproducible RNFL measurement in our study was mean thickness, which correlates with previously published results.\textsuperscript{30,32} Averaging the mean RNFL thicknesses from several consecutive scans increased the ICC in one report. However, results from averaged and nonaveraged data were similar, indicating that a single scan can provide reliable results.\textsuperscript{30} Several groups previously reported quadrant thickness reproducibility. Four studies found that the nasal quadrant was the most variable,\textsuperscript{25,32-34} and another found the greatest variability in the superior quadrant, followed by the nasal quadrant.\textsuperscript{31} Our results indicate that the nasal quadrant thickness was the most reproducible, whereas superior quadrant thickness was the least reproducible. However, the mean superior quadrant thickness of our healthy subjects was much greater than the mean nasal quadrant thickness (superior, 139 µm, and nasal, 85 µm). A previous study\textsuperscript{35} found that macular sector thickness variation increased with increasing macular thickness. This seems to indicate that a similar phenomenon may be seen with RNFL quadrant thicknesses, thus potentially explaining the decreased reproducibility of the superior quadrant in our study.

In this study, we did not dilate pupils and found no effect on the quality of the data, which is in keeping with a previous report.\textsuperscript{30} Another study investigating whether the technicians’ experience affected reproducibility found that inexperienced technicians could generate useful measurements.\textsuperscript{25} In our study, 2 technicians had 8 months of OCT-3 experience, whereas 1 technician had 1 month of experience (D.C., M.P., and E.G.-L.). We found that all 3 technicians generated reproducible data.

Other groups have shown higher variation in patients with glaucoma and diabetes compared with healthy subjects.\textsuperscript{31,30} To our knowledge, the reproducibility of OCT in MS patients has not been reported previously. Our cross-sectional study was limited because different patients were studied at each site. Although not ideal, the only feasible way to compare large, geographically distant cohorts was to use age-matched subjects with similar demographic characteristics as found in our cohorts. Despite the potential for ocular abnormalities of MS to interfere with obtaining high-quality OCT scans, our cross-sectional data obtained from 3 research centers examining separate MS cohorts were virtually identical. This suggests that RNFL measurements are reproducible within diverse MS patient groups, which is encouraging for the potential use of OCT-3 as an outcome measure in clinical trials. Use of a single model of the same machine also offers advantages over magnetic resonance images obtained using many different models and machine types.

Validation of OCT as an imaging biomarker in MS is important because several aspects of the information it generates are unique. Imaging the RNFL allows direct measurement of the unmyelinated axons of the central nervous system.\textsuperscript{37} The capacity to image central nervous system axons quickly and noninvasively, to minimize expense, and to correlate structural abnormalities with visual dys-
function add to the appeal of OCT as an imaging biomarker and outcome measure in clinical trials.  

We have demonstrated that OCT RNFL thicknesses obtained in an MS clinic show excellent interrater, intrarater, and intervisit reproducibility in healthy controls. In addition, RNFL measurements from MS and control cohorts from 3 different academic MS centers were remarkably similar. This makes OCT an attractive potential outcome measure for clinical trials of axonal-protective therapeutics and as a potential marker for disease progression in MS patients.

Accepted for Publication: March 11, 2008.

Correspondence: Peter A. Calabresi, MD, Pathology Bldg, Ste 627, The Johns Hopkins Hospital, 600 N Wolfe St, Baltimore, MD 21287 (calabresi@jhmi.edu).


Financial Disclosure: None reported.

Funding/Support: This study was supported by grant TR3760-A3 from the National Multiple Sclerosis Society (Dr Calabresi), The Nancy Davis Foundation (Dr Calabresi), and the American Academy of Neurology Med-