Artifact Rates for 2D Retinal Nerve Fiber Layer Thickness Versus 3D Neuroretinal Rim Thickness Using Spectral-Domain Optical Coherence Tomography

Elli A. Park, Edem Tsikata, Jenny Jyoung Lee, Eric Shieh, Boy Braaf, Benjamin J. Vakoc, Brett E. Bouma, Johannes F. de Boer, and Teresa C. Chen

1 Boston University School of Medicine, Boston, MA, USA
2 Department of Ophthalmology, Glaucoma Service, Massachusetts Eye and Ear Infirmary, Boston, MA, USA
3 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
4 Harvard Medical School, Boston, MA, USA
5 Jules Stein Eye Institute at University of California, Los Angeles, Los Angeles, CA, USA
6 Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, MA, USA
7 LaserLaB Amsterdam, Department of Physics and Astronomy, Vrije Universiteit, Amsterdam, The Netherlands
8 Department of Ophthalmology, Amsterdam University Medical Centers, Amsterdam, The Netherlands

Purpose: To compare the rates of clinically significant artifacts for two-dimensional peripapillary retinal nerve fiber layer (RNFL) thickness versus three-dimensional (3D) neuroretinal rim thickness using spectral-domain optical coherence tomography (SD-OCT).

Methods: Only one eye per patient was used for analysis of 120 glaucoma patients and 114 normal patients. For RNFL scans and optic nerve scans, 15 artifact types were calculated per B-scan and per eye. Neuroretinal rim tissue was quantified by the minimum distance band (MDB). Global MDB neuroretinal rim thicknesses were calculated before and after manual deletion of B-scans with artifacts and subsequent automated interpolation. A clinically significant artifact was defined as one requiring manual correction or repeat scanning.

Results: Among glaucomatous eyes, artifact rates per B-scan were significantly more common in RNFL scans (61.7%, 74 of 120) compared to B-scans in neuroretinal rim volume scans (20.9%, 1423 of 6820) (95% confidence interval [CI], 31.6–50.0; \( P < 0.0001 \)). For clinically significant artifact rates per eye, optic nerve scans had significantly fewer artifacts (15.8% of glaucomatous eyes, 13.2% of normal eyes) compared to RNFL scans (61.7% of glaucomatous eyes, 25.4% of normal eyes) (glaucoma group: 95% CI, 34.1–57.5, \( P < 0.0001 \); normal group: 95% CI, 1.3–23.3, \( P = 0.03 \)).

Conclusions: Compared to the most commonly used RNFL thickness scans, optic nerve volume scans less frequently require manual correction or repeat scanning to obtain accurate measurements.

Translational Relevance: This paper illustrates the potential for 3D OCT algorithms to improve in vivo imaging in glaucoma.

Introduction

Spectral-domain optical coherence tomography (SD-OCT) plays an essential role in the diagnosis and management of glaucoma by providing quantitative, structural measurements of the retinal nerve fiber layer (RNFL), the optic nerve, and the macula. Of these, the most commonly used glaucoma measurement is two-dimensional (2D) peripapillary RNFL thickness. Global RNFL thickness has strong diagnostic capability for glaucoma, with area under the receiver...
operating characteristic values ranging from 0.677 to 0.978,2–4 and it correlates significantly with functional testing.5–12 However, its value is limited by a high rate of artifacts, with 7.1% to 58.5% of RNFL scans affected.13–22 Studies show that these artifacts may cause clinically significant errors in measurement17,23,24 and lead to false glaucoma classifications in as many as 39.0% of scans.24–26

The minimum distance band (MDB) quantifies three-dimensional (3D) neuroretinal rim tissue from a high-density SD-OCT optic nerve volume scan protocol using custom-designed research software.27–29 It measures the closest distance from the end of the retinal pigment epithelium/Bruch’s membrane (RPE/BrM) complex to the internal limiting membrane (ILM) and is similar to the low-density commercially available BMO-MRW protocol (i.e., Bruch’s membrane opening minimum rim width).27–29 MDB thickness has demonstrated equal or better diagnostic capability for glaucoma compared to 2D RNFL thickness and 3D BMO-MRW, but the rate of clinically significant artifacts for 3D neuroretinal rim measurements is still unknown.27,28 One study found that an average of 11 out of 193 frames (or 5.7% of frames) examined in optic nerve volume scans had artifacts,29 and studies of the BMO-MRW have reported artifacts in 62.6% to 84.0% of scans.19,30 However, these papers did not analyze artifacts as a primary objective, and their different definitions, exclusion criteria, and methods of rate calculation make comparison to 2D artifacts challenging.

Therefore, we performed a cross-sectional, observational clinical study to compare the rate of artifacts in 2D RNFL thickness scans versus 3D optic nerve volume scans using MDB neuroretinal rim thickness. We calculated both (1) the artifact rate per B-scan and (2) the clinically significant artifact rate per eye. The former equals the percentage of B-scans having at least one artifact. The latter reflects the percentage of eyes whose scans would require manual correction or complete repetition in order to obtain accurate measurements and thus offers more practical information than a pure artifact rate. To our knowledge, only two studies have described the clinical significance of 3D artifacts in high-density volume scans, but these examined the peripapillary region and the macula, not the neuroretinal rim.21,31 In contrast, our study, to the best of our knowledge, is the first to compare the clinically significant artifact rates for 2D RNFL thickness versus 3D neuroretinal rim thickness in normal and glaucoma patients, using the Spectralis SD-OCT machine (Heidelberg Engineering, Heidelberg, Germany). We hypothesized that 3D MDB neuroretinal rim thickness has fewer clinically significant artifacts than traditional peripapillary RNFL thickness and therefore less frequently requires manual correction and repeat scanning to ensure accurate measurements.

**Materials and Methods**

**Study Population**

The study was approved by the Massachusetts Eye and Ear Institutional Review Board and conducted in accordance with the tenets of the Declaration of Helsinki. Subjects were a cross-sectional cohort of normal and open-angle glaucoma patients who were recruited between April 2009 and January 2016 for the prospective longitudinal SD-OCT in Glaucoma Study, which is comprised of 2000 normal, glaucoma suspect, and glaucoma patients. All subjects underwent a complete ophthalmic examination in the Glaucoma Service at Massachusetts Eye and Ear by one glaucoma specialist. All eligible participants consented to Spectralis SD-OCT (Spectralis software version 5.4.8.0; Heidelberg Engineering) scanning on the same day as their eye examination, which included history, visual acuity, refraction, Goldmann applanation tonometry, slit-lamp biomicroscopy, gonioscopy, dilated ophthalmoscopy, ultrasound pachymetry (PachPen; Accutome, Inc., Malvern, PA), stereo disc photography (VISUCAM Pro NM; Carl Zeiss Meditec, Inc., Dublin, CA), and Humphrey visual field testing (Swedish Interactive Threshold Algorithm 24-2 test; Carl Zeiss Meditec).

Subjects with and without glaucoma were included if they had a spherical equivalent between −5 and +5 diopters, best-corrected visual acuity of 20/40 or better, and reliable visual field testing defined as ≤33% fixation losses, ≤20% false positives, and ≤20% false negatives. Subjects were excluded if they had anterior segment dysgenesis, corneal scarring or opacities, visual field loss due to a non-glaucoma condition (e.g., diabetic retinopathy), or a dilated pupil diameter less than 2 mm. Glaucoma patients had characteristic optic nerve changes with corresponding visual field defects. Visual fields were abnormal if three or more contiguous locations in the pattern standard deviation plot were depressed at the $P < 0.05$ level or if at least two contiguous locations were depressed at the $P < 0.05$ level and one at the $P < 0.01$ level. Glaucoma severity was defined as mild (mean deviation [MD] ≥ −6 dB), moderate (−12 dB ≤ MD < −6 dB), or severe (MD < −12 dB) based on the Hodapp–Parrish–Anderson criteria.32
Normal subjects had no ocular disease except for mild cataracts, normal visual field test results defined by a pattern standard deviation and glaucoma hemifield test within normal limits, and no cup-to-disc asymmetry greater than 0.2 between eyes. When both eyes were eligible for the study, one eye was selected randomly for inclusion.

**OCT Imaging: High-Density Research Scan Protocol and Research Software**

All SD-OCT imaging was performed after pupillary dilation. Imaging was performed using the Spectralis OCT (version 5.4.8.0), which has an acquisition rate of 40,000 A-lines per second. Each subject had 2D RNFL thickness scans and one 3D volume scan of the optic nerve. For 2D scans, RNFL thickness was calculated along a 12° peripapillary circle scan, which is approximately 3.5 to 3.6 mm in diameter, depending on the axial length. Patients with OCT signal strength of less than 15 were excluded from the analysis. The Spectralis OCT software automatically segmented the anterior and posterior RNFL to calculate the average RNFL thickness for the overall globe (360°), for each 90° quadrant (superior, temporal, inferior, nasal), and for designated 45° octants (superior–temporal, superior–nasal, inferior–naso, and inferior–temporal).

Volumetric scans were obtained using the high-density optic nerve volume scan protocol, which consists of 193 B-scans arranged in a raster pattern over a 20° × 20° area (approximately 6 mm × 6 mm, depending on the patient’s refraction). Each volume scan was analyzed using custom-designed research software written in C++ with OpenCV, ITK, and VTK libraries. The software automatically segmented the ILM and the RPE/BM complex in all 193 B-scans and then reconstructed the neuroretinal rim in 3D space. The disc border was defined by termination of the RPE/BM at 100 equally spaced circumference points around the optic nerve. MDB thickness was then calculated by measuring the closest distance between the disc border and the ILM in 3D space.

**Definition and Evaluation of Artifacts**

A clinically significant artifact is defined as one that requires the technician either to repeat the scan or to manually correct segmentation in order to obtain an accurate measurement. For 2D RNFL scans, repeat scanning and/or manual correction of segmentation errors would generally be required for RNFL scans that were decentered, missing data, or incorrectly segmented. For 3D optic nerve volume scans, manual correction in the clinic would be required if two measurements (i.e., with manual correction and interpolation and without manual correction and interpolation) differed by more than expected test–retest variability.

All 2D peripapillary RNFL thickness scans and 3D volume scans were examined for artifacts by one observer. For the 193 B-scans that comprised a 3D volume scan, only the B-scans intersecting the optic nerve were analyzed for artifacts. Because the size of the optic nerve varies from person to person, the number of B-scans examined was not always the same for every eye. On average, 56.2 ± 6.6 B-scans were examined in each 3D volume scan. In total, 6820 B-scans and 6331 B-scans were examined in the 120 glaucoma patients and 114 normal patients, respectively.

Twelve types of OCT artifacts were recorded for 2D peripapillary RNFL thickness scans: (1) anterior RNFL misidentification, (2) posterior RNFL misidentification, (3) incomplete segmentation, (4) out of register, (5) cut out or “cookie monster”, (6) cut edge, (7) missing part, (8) decenteration, (9) motion artifact, (10) mirroring, (11) peripapillary atrophy (PPA)-associated error, and (12) posterior vitreous detachment (PVD)/epithelial retinal membrane (ERM)-associated error. Three types of OCT artifacts were recorded for 3D optic nerve volume scans: (1) ILM misidentification, (2) RPE/BM misidentification, and (3) incomplete segmentation of the RPE/BM. Each B-scan could have multiple types of artifacts, only one type, or none. Anterior RNFL misidentification (Fig., part A) and posterior RNFL misidentification (Fig., part B) occurred with incorrect segmentation of the anterior and posterior boundaries of the RNFL, respectively. Incomplete segmentation was defined as failure to segment the RNFL across the entire length of the rectangular display box.

Out-of-register artifact occurred if the display box captured only part of the RNFL due to superior or inferior shift of the scan. Cut-out or “cookie monster” artifact (Fig., part C) occurred when part of the RNFL was missing or appeared to be eaten out. Cut-edge artifact represented abrupt lateral truncation of the RNFL on one or both sides of the scan. Missing-part artifact (Fig., part D) was observed when a portion of the RNFL was missing due to probable shadowing from an overlying opacity. Decentration artifact (Fig., part E) arose from imperfect alignment of the circular scan by the technician and was defined as a shift of the center of the scan from the center of the optic nerve by at least 10%. Motion artifact occurred if patient movement was so great such that parts of the RNFL moved.
Artifact Rates for 3D Versus 2D Scans

Figure. Examples of artifact types found in 2D and 3D scans. (A) Anterior RNFL misidentification (white arrow). (B) Posterior RNFL misidentification (blue line). (C) Cut-out or “cookie monster” artifact. (D) Missing-part artifact. (E) Decentration was defined as a misalignment of the center of the scan circle from the center of the optic nerve head by more than 10%. (F, G) PPA-associated error occurred when the scan circle overlapped an area of PPA, producing inaccurate segmentation in the corresponding B-scan. (H) PVD/ERM-associated error was counted when a PVD or ERM resulted in incorrect segmentation of the RNFL. (I) ILM or cup surface misidentification (green line). (J) RPE/BM misidentification (red line). (K) Incomplete segmentation of the RPE/BM complex (red line).

outside of the rectangular display. Mirror artifact was recorded if part of the retinal image appeared flipped over onto itself. PPA-associated artifact (Fig., part F) occurred when the scan circle overlapped with an area of PPA, which usually precluded accurate RNFL segmentation (Fig., part G). PVD/ERM-associated artifact (Fig., part H) was counted when a PVD or ERM was seen in the rectangular display and caused an RNFL segmentation error. For optic nerve scans, ILM misidentification (Fig., part I) occurred with incorrect segmentation of the ILM or cup surface. Likewise, RPE/BM misidentification (Fig., part J) occurred with incorrect determination of the RPE/BM complex. Incomplete segmentation of the RPE/BM (Fig., part K) was counted when the algorithm correctly delineated the RPE/BM but stopped short before reaching the true termination of the RPE/BM.

Interpolation and Correction of Artifacts in 3D Volume Scans

For each 3D volume scan, B-scans containing artifacts were manually deleted, and custom-designed software subsequently interpolated the missing frame.
using information from neighboring scans. MDB thickness measurements were calculated twice, once before and once after correction for artifacts. Three-dimensional volume scans were considered to have clinically significant artifacts if the difference in pre-interpolation and post-interpolation MDB thickness measurements was greater than the test–retest variability of 17.45 μm for glaucoma patients and 16.05 μm for normal patients, as determined in a study of MDB reproducibility (Kim J, Men C, Ratanawongphaibul K, et al., unpublished data, 2020). Theoretically, such scans would have required either repeat scanning and/or manual correction in the clinic with computerized interpolation in order to obtain accurate measurements for those eyes.

### Statistical Analysis

All statistical analyses were performed using R 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Demographic and ocular characteristics were compared with two-tailed Student’s *t*-tests for continuous variables, χ² tests for categorical variables, and two-sample proportion test with Yates’ continuity correction for equality of proportions. To calculate the frequency of artifacts for 2D RNFL scans, the number of 2D scans having at least one artifact was divided by the total number of 2D scans.

To calculate the frequency of artifacts per B-scan for 3D volume scans, for each individual 3D volume scan the number of B-scans with artifact was divided by the number of B-scans examined in the volume scan. The mean artifact rate was then determined for each study group. As a secondary analysis, the total artifact rate per B-scan for 3D volume scans was also calculated using a second method—dividing the total composite number of B-scans with any artifact by the total composite number of B-scans examined in the entire study group (6820 B-scans in the glaucoma group and 6331 B-scans in the normal group). The results of this second calculation were used to compare 3D versus 2D artifact rates per B-scan, as such a comparison requires a proportional test instead of a *t*-test.

To calculate the rate of clinically significant artifacts per eye for 3D volume scans, the number of eyes for which the pre- and post-interpolation MDB thickness measurements differed by more than test–retest variability was determined. This number was then divided by the total number of eyes. For 2D scans, the rate of clinically significant artifacts per eye was equal to the rate per B-scan, as a clinically significant artifact is one that would require correction or a repeat scan. Results were stated as means ± standard deviation unless otherwise specified. Results were statistically significant for *P* < 0.05.

### Results

#### Patient Characteristics

This study included a total of 234 participants. The normal and glaucoma groups were similar in eye laterality, race, and spherical equivalent (Table 1). There were more female subjects in the normal group than in the glaucoma group (72 out of 114 normal subjects vs. 52 out of 120 glaucoma subjects; 95% confidence interval [CI], 1.3–3.8; *P* = 0.004) (Table 1). Glaucoma patients were significantly older than normal patients (67.8 ± 12.3 years vs. 54.4 ± 16.2 years; 95% CI, 9.8–17.2 years; *P* < 0.0001) and had lower mean deviation (–11.9 ± 7.5 dB vs. –1.8 ± 2.3 dB; 95% CI, –11.5 to –8.7 dB; *P* < 0.0001) and higher pattern standard deviation (8.3 ± 3.3 dB vs. 1.6 ± 0.4 dB; 95% CI, 6.1–7.3 dB; *P* < 0.0001) on visual field testing (Table 1).

#### Frequency of 2D Artifacts for Normal Versus Glaucoma Patients

Artifacts in 2D peripapillary RNFL thickness scans were more common among glaucoma subjects than normal subjects (61.7% vs. 25.4%; 95% CI, 23.6–48.9; *P* < 0.0001) (Table 2). Posterior RNFL misidentification was the most common type of artifact observed in glaucomatous eyes and was the primary source of difference between the two groups (41.7% for glaucoma subjects vs. 7.0% for normal subjects; 95% CI, 23.8–45.5; *P* < 0.0001) (Table 2). Five artifacts described in the literature were not seen in our study population: incomplete segmentation, motion artifact, out of register, cut edge, and mirroring (Table 2).

#### Frequency of 3D Artifacts for Normal Versus Glaucoma Patients

All 3D volume scans contained at least one B-scan with an artifact. Glaucoma and normal subjects had a similar rate of artifacts in 3D optic nerve volume scans (20.9% vs. 19.5%; 95% CI, –1.7 to 4.6; *P* = 0.38) (Table 3). However, ILM misidentification occurred more commonly in glaucomatous eyes (12.5% vs. 8.4%; 95% CI, 1.2–7.1; *P* = 0.006), as well as RPE/BM misidentification (4.2% vs. 2.9%; 95% CI, 0.2–2.6; *P* = 0.03), and incomplete segmentation of the RPE/BM occurred more frequently in normal eyes (5.2% vs. 8.7%; 95% CI, –5.5 to –1.6; *P* = 0.0005) (Table 3).
Table 1. Demographic Characteristics of the Normal and Glaucoma Study Populations

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Glaucoma</th>
<th>Normal</th>
<th>Estimate (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes, n</td>
<td>120</td>
<td>114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OD/OS</td>
<td>62/58</td>
<td>57/57</td>
<td>1.1 (0.6, 1.8)</td>
<td>0.90</td>
</tr>
<tr>
<td>Male/female</td>
<td>68/52</td>
<td>42/72</td>
<td>2.2 (1.3, 3.8)</td>
<td>0.004a</td>
</tr>
<tr>
<td>Age (y), mean ± SD (range)</td>
<td>67.8 ± 12.3 (22.2 – 92.1)</td>
<td>54.4 ± 16.2 (15.8 – 88.8)</td>
<td>13.5 (9.8, 17.2)</td>
<td>&lt;0.0001a</td>
</tr>
<tr>
<td>Race, n</td>
<td></td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>73</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>32</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>7</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>8</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refractive errorb</td>
<td>−0.6 ± 1.8</td>
<td>−0.5 ± 1.8</td>
<td>−0.2 (−0.6, 0.3)</td>
<td>0.50</td>
</tr>
<tr>
<td>Visual field (dB), mean ± SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean deviation</td>
<td>−11.9 ± 7.5</td>
<td>−1.8 ± 2.3</td>
<td>−10.1 (−11.5, −8.7)</td>
<td>&lt;0.0001a</td>
</tr>
<tr>
<td>Pattern standard deviation</td>
<td>8.3 ± 3.3</td>
<td>1.6 ± 0.4</td>
<td>6.7 (6.1, 7.3)</td>
<td>&lt;0.0001a</td>
</tr>
</tbody>
</table>

*a Results are statistically significant (P < 0.05).
*b Three patients in the glaucoma group and one in the normal group were excluded from the calculation of mean spherical equivalent because their refractive error was either not available (n = 3) or not relevant due to the eye having no useful vision (n = 1).

3D MDB Thickness Calculations With and Without Correction for Artifacts

Table 4 shows mean global MDB thickness calculated before and after manual correction with subsequent computerized interpolation for artifacts. As expected, glaucoma subjects had lower mean global MDB thickness compared to normal subjects, both before interpolation (169.3 ± 50.6 μm vs. 303.8 ± 44.5 μm; 95% CI, −146.8 to −122.3; P < 0.0001) and after interpolation (176.1 ± 50.7 μm vs. 309.4 ± 43.0 μm; 95% CI, −145.4 to −121.2; P < 0.0001) (Table 4). Correction for artifacts did not significantly change the mean global MDB thickness for either the glaucoma subjects (95% CI, −19.7 to 6.0; P = 0.30) or the normal subjects (95% CI, −17.0 to 5.8; P = 0.33) (Table 4), with mean absolute percentage differences of 4.9% and 2.3%, respectively (95% CI, 1.2–3.9; P = 0.0003) (Table 4).

Frequency of 2D Versus 3D Artifacts

Three-dimensional volume scans were less likely to have clinically significant artifacts compared to 2D RNFL thickness scans for both the glaucoma group (15.8% of eyes vs. 61.7% of eyes; 95% CI, 34.1–57.5; P < 0.0001) and the normal group (13.2% of eyes vs. 25.4% of eyes; 95% CI, 1.3–23.3; P = 0.03) (Table 5). The rate of clinically significant artifacts per eye in 3D volume scans was similar between the glaucoma and normal groups (95% CI, −7.2 to 12.5; P = 0.69) (Table 5). Three-dimensional volume scans also had a lower rate of artifacts per B-scan compared to 2D RNFL scans for glaucoma patients (20.9% of 6820 B-scans vs. 61.7% of 120 B-scans; 95% CI, 31.6–50.0; P < 0.0001; table not shown), but not for normal patients (95% CI, −2.5 to 14.5; P = 0.14; table not shown).

Differences in Artifact Rates Among Patients with Mild, Moderate, and Severe Glaucoma

For 2D scans, patients with severe glaucoma had a significantly higher artifact rate than those with moderate glaucoma (75.5% vs. 48.7%; P = 0.01) (Supplementary Fig.); however, there was no significant difference between mild and moderate glaucoma (55.9% vs. 48.7%; P = 0.54) or mild and severe glaucoma (55.9% vs. 75.5%; P = 0.06). Artifact rates in 3D scans were the
Table 2. Percentage of 2D RNFL Scans with Artifacts

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Percentage of B-Scans (%)</th>
<th>Estimate (%) (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Glaucoma (n = 120)</td>
<td>Normal (n = 114)</td>
<td></td>
</tr>
<tr>
<td>Posterior RNFL misidentification</td>
<td>41.7</td>
<td>7.0</td>
<td>34.7 (23.8, 45.5)</td>
</tr>
<tr>
<td>Decentration</td>
<td>20.0</td>
<td>19.3</td>
<td>0.7 (–10.2, 11.6)</td>
</tr>
<tr>
<td>Anterior RNFL misidentification</td>
<td>5.0</td>
<td>0.9</td>
<td>4.1 (–1.0, 9.2)</td>
</tr>
<tr>
<td>PVD/ERM-associated error</td>
<td>3.3</td>
<td>0</td>
<td>3.3 (–0.7, 7.4)</td>
</tr>
<tr>
<td>Cut out or “cookie monster”</td>
<td>1.7</td>
<td>2.6</td>
<td>–1.0 (–5.5, 3.6)</td>
</tr>
<tr>
<td>PPA-associated error</td>
<td>1.7</td>
<td>0</td>
<td>1.7 (–1.5, 4.8)</td>
</tr>
<tr>
<td>Missing part</td>
<td>0.8</td>
<td>0</td>
<td>0.8 (–16.3, 3.3)</td>
</tr>
<tr>
<td>Incomplete segmentation</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Motion artifact</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Out of register</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Cut edge</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Mirroring</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>Any artifact (total percentage, %)</td>
<td>61.7</td>
<td>25.4</td>
<td>36.2 (23.6, 48.9)</td>
</tr>
</tbody>
</table>

*Number of 2D RNFL scans with the specified artifact type divided by 120 for the glaucoma group and 114 for the normal group.
*Results are statistically significant (P < 0.05).
*Total percentage represents the total percentage of 2D RNFL scans with at least one artifact type. Scans with multiple artifacts were only counted once.

Table 3. Percentage of B-Scans with Artifacts in 3D Volume Scans

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Percentage of B-Scans (%)</th>
<th>Estimate (%) (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Glaucoma (n = 120)</td>
<td>Normal (n = 114)</td>
<td></td>
</tr>
<tr>
<td>ILM misidentification</td>
<td>12.5</td>
<td>8.4</td>
<td>4.2 (1.2, 7.1)</td>
</tr>
<tr>
<td>RPE/BM misidentification</td>
<td>4.2</td>
<td>2.9</td>
<td>1.4 (0.2, 2.6)</td>
</tr>
<tr>
<td>Incomplete segmentation of the RPE/BM</td>
<td>5.2</td>
<td>8.7</td>
<td>–3.6 (–5.5, –1.6)</td>
</tr>
<tr>
<td>Any artifact (total percentage, %)</td>
<td>20.9</td>
<td>19.5</td>
<td>1.4 (–1.7, 4.6)</td>
</tr>
</tbody>
</table>

*For 3D scans, the percentage of B-scans was calculated as the mean artifact rate; that is, for each of the 234 3D scans, the number of B-scans with the artifact type was divided by the number of B-scans analyzed. The mean rate was then determined for each study group.
*Results are statistically significant (P < 0.05).
*Total percentage represents the mean percentage of B-scans with any type of artifact. B-scans with multiple artifacts were only counted once.

same for patients with mild (22.4%), moderate (23.1%), or severe (18.2%) glaucoma (P > 0.05 for all comparisons) (Supplementary Fig.).

Discussion

Our study has demonstrated that 3D optic nerve volume scans are less likely to have clinically significant artifacts compared to 2D peripapillary RNFL thickness scans for both glaucomatous eyes (15.8% vs. 61.7%; 95% CI, 34.1–57.5; P < 0.0001) and normal eyes (13.2% vs. 25.4%; 95% CI, 1.3–23.3; P = 0.03) (Table 5). They also had fewer artifacts per B-scan for glaucomatous eyes (20.9% of 6820 B-scans vs. 61.7% of 120 B-scans; 95% CI, 31.6–50.0; P < 0.0001; table not shown). Stated differently, for patients with glaucoma, only 15.8% of 3D optic nerve volume scans had artifacts that necessitated manual correction or repeat scanning in the clinic, as opposed to 61.7% of 2D RNFL thickness scans (Table 5). Only one other study has reported a rate of clinically significant artifacts in high-density volume scans, and this study determined a clinically significant artifact rate of 7.5% for 3D RNFL volume
Table 4. Global MDB Thickness Derived from 3D Volume Scans, Pre- and Post-Interpolation

<table>
<thead>
<tr>
<th>Global MDB Thickness</th>
<th>Glaucoma ((n = 120))</th>
<th>Normal ((n = 114))</th>
<th>Estimate (95% CI)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-interpolation (µm), mean ± SD</td>
<td>169.3 ± 50.6</td>
<td>303.8 ± 44.5</td>
<td>-134.5 (-146.8, -122.3)</td>
<td>&lt;0.0001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Post-interpolation (µm), mean ± SD</td>
<td>176.1 ± 50.7</td>
<td>309.4 ± 43.0</td>
<td>-133.3 (-145.4, -121.2)</td>
<td>&lt;0.0001&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Estimate (95% CI), mean ± SD</td>
<td>-6.8 (-19.7, 6.0)</td>
<td>-5.6 (-17.0, 5.8)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(P)</td>
<td>0.30</td>
<td>0.33</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean absolute percentage difference (pre- vs. post-interpolation) (%), mean ± SD</td>
<td>4.9 ± 6.7</td>
<td>2.3 ± 3.6</td>
<td>2.6 (1.2, 3.9)</td>
<td>0.0003&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Results are statistically significant \((P < 0.05)\).

Table 5. Percentage of Clinically Significant Artifacts in 2D and 3D Scans

<table>
<thead>
<tr>
<th>Clinically Significant Artifacts Per Eye (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>3D Scans</th>
<th>2D Scans</th>
<th>Estimate (%) (95% CI)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glaucoma ((n = 120))</td>
<td>15.8</td>
<td>61.7</td>
<td>45.8 (34.1, 57.5)</td>
<td>&lt;0.0001&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Normal ((n = 114))</td>
<td>13.2</td>
<td>25.4</td>
<td>12.3 (1.3, 23.3)</td>
<td>0.03&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Estimate (95% CI)</td>
<td>2.7 (-7.2, 12.5)</td>
<td>36.2 (23.6, 48.9)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(P)</td>
<td>0.69</td>
<td>&lt;0.0001&lt;sup&gt;b&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<sup>a</sup>For 3D scans, the percentage of clinically significant artifacts was calculated as the percentage of eyes for which the removal of B-scans with artifacts and subsequent automated interpolation resulted in an absolute change in global MDB thickness greater than test–retest variability of 17.45 µm for the glaucoma group and 16.05 µm for the normal group. For 2D scans, the percentage of clinically significant artifacts was equal to the artifact rate per B-scan, because a clinically significant artifact was defined as one that would require manual correction or a repeat scan to ensure accurate measurements.

<sup>b</sup>Results are statistically significant \((P < 0.05)\).

To our knowledge, our study is the first to evaluate the rate of clinically significant artifacts for 3D optic nerve scans, and our study also revealed that 3D volume scans require fewer repeat scans and less manual correction in the clinic than the most commonly used 2D RNFL thickness scan.

Three-dimensional scans may be less susceptible to artifacts than 2D scans because they carry a higher density of information (e.g., 193 B-scans in a raster pattern vs. a single circular B-scan).<sup>29,31</sup> If one B-scan contains an artifact, neighboring B-scans can provide sufficient information to correct the affected frame or B-scan. In contrast, the accuracy of 2D peripapillary RNFL thickness relies on a single frame, so there are no neighboring scans to correct for any bad data. Not only are 3D optic nerve volume scans less sensitive to artifacts compared to 2D RNFL thickness scans, but, among glaucomatous eyes, they also have fewer artifacts per B-scan from the outset (20.9% of 6820 B-scans vs. 61.7% of 120 B-scans; 95% CI, 31.6–50.0; \(P < 0.0001\); table not shown). One explanation for the lower rate of artifacts in 3D optic nerve versus 2D RNFL scans for glaucoma patients is the higher reflectivity of the RPE/BM complex in comparison to the posterior RNFL border.<sup>33–36</sup> Loss of RNFL reflectivity in the setting of glaucomatous nerve fiber layer thinning interferes with accurate segmentation in 2D RNFL thickness scans, but there is no evidence of such reflectivity loss associated with the RPE/BM.<sup>33,35</sup> In fact, we found that 41.7% of 2D scans had posterior RNFL misidentification but only 4.2% of B-scans in 3D volume scans had RPE/BM misidentification in glaucomatous eyes (Tables 2 and 3). Decentration offers another explanation for differences in artifact rates between 2D and 3D scans, because 2D RNFL scans in this study required manual centration by the technician, whereas 3D optic nerve volume scans have automatic centration by the computer algorithm which effectively eliminates this type of error. We identified decentration in 20.0% and 19.3% of 2D RNFL scans among glaucomatous and normal eyes, respectively (Table 2); these percentages fall within the range of the 10.4% and 27.8% reported in the literature.<sup>13,21</sup>

To the best of our knowledge, ours is one of only two studies that have evaluated artifact rates related to the newer high-density MDB neuroreti-
Artifact Rates for 3D Versus 2D Scans

TVST | September 2020 | Vol. 9 | No. 10 | Article 10 | 9

ternal rim parameter, which represents the full potential of SD-OCT optic nerve imaging for glaucoma. We observed artifacts in 20.9% and 19.5% of B-scans in 3D optic nerve volume scans in glaucomatous and normal eyes, respectively (Table 3). In contrast, Tsikata and associates 29 reported that an average of 11 out of 193 B-scans per volume scan, or 5.7% of B-scans, had artifacts; however, they analyzed artifacts as a secondary objective and may not have captured all the types of errors that we identified. Very few studies describe artifacts related to other neuroretinal rim parameters, such as BMO-MRW and rim thickness, but their varying artifact definitions and methods of calculation make comparison to our results challenging. For example, some studies used the Zeiss Cirrus HD-OCT or included errors in the color thickness map and deviation map as part of their definition of artifact.37,38 In contrast, our study used the Heidelberg Spectralis OCT and recorded errors in both B-scan and volume scan measurements. In addition, many papers have reported an artifact rate on a per-exam basis (i.e., percentage of volume scans having at least one B-scan with artifact), with rates ranging from 62.6% to 84.0% of exams for BMO-MRW and 12.1% to 92.9% of exams for 2D rim parameters.19,30,37–40 These prior studies of BMO-MRW and 2D rim parameters did not calculate clinically significant artifact rates, so we could not compare our artifact rate results with these studies. When we defined artifact rate on a per-exam basis, we found that 100% of volume scans had at least one B-scan with an artifact. However, despite 100% of volume scans having at least one B-scan with an artifact, we determined that these artifacts were clinically significant (i.e., requiring manual correction or scan repetition) in only 15.8% and 13.2% of glaucomatous and normal eyes, respectively (Table 5). Future investigations should use consistent artifact definitions and rate calculations to enable meaningful comparisons among studies.

For 2D RNFL thickness scans, we determined that artifacts were more common among glaucoma patients than normal patients (61.7% vs. 25.4%; 95% CI, 23.6–48.9; \( P < 0.0001 \)) (Table 2). Our findings agree with others who have shown that artifacts related to RNFL thickness measurements are more common in eyes with glaucoma than in eyes without.13,17,20,21 This may be explained by glaucomatous thinning and loss of reflectivity of the RNFL, which create a challenge for automated segmentation.33–36 Theoretically, such changes would primarily impact segmentation of the posterior RNFL, as it becomes less easily distinguished from the underlying retinal layers. Indeed, posterior RNFL misidentification was the primary source of difference between the two groups in our study (41.7% for the glaucoma group vs. 7.0% for the normal group; 95% CI, 23.8–45.5; \( P < 0.0001 \)) (Table 2). Notably, our total artifact rate for glaucoma patients (61.7%) is outside the 7.1% to 58.5% range reported in the literature.13–22 This may be due to our broad inclusion of several different types of artifacts, as opposed to only segmentation errors, which were the focus of other papers.15,17,19,20

We discovered significant differences in the types of 3D artifacts found between glaucoma and normal patients. Specifically, ILM and RPE/BM misidentification were more common among glaucomatous eyes, whereas incomplete segmentation of the RPE/BM was more common among normal eyes (Table 3). The increased rates of cup surface (ILM) and disc border (RPE/BM) misidentification in glaucoma patients may be related to a combination of decreased RNFL reflectivity and increased PPA, respectively. In contrast, normal eyes have a thicker layer of tissue overlying the RPE/BM; the thicker layer may attenuate the signal from the RPE/BM, making it more difficult to fully segment in normal eyes. Ultimately, however, these differences appear to balance each other, as the overall artifact rates were similar between groups (20.9% for glaucoma patients vs. 19.5% for normal patients; 95% CI, –1.7 to 4.6; \( P = 0.38 \)) (Table 3). Three-dimensional MDB neuroretinal rim thickness may therefore provide an advantage over 2D RNFL thickness by offering a parameter that is equally reliable between eyes with and without glaucoma.

Our data suggest that artifact rates increase with increasing glaucoma severity in 2D RNFL thickness scans, but not in 3D optic nerve scans. Specifically, we found that patients with severe glaucoma had a significantly higher rate of 2D RNFL thickness artifacts than patients with moderate glaucoma (75.5% vs. 48.7%; \( P = 0.01 \)) (Supplementary Fig.), but artifact rates in 3D optic nerve scans were the same regardless of glaucoma severity. This finding is consistent with previous studies that found that more advanced glaucoma stage was associated with a greater frequency and severity of artifacts in peripapillary RNFL thickness scans.13,23,24 The reason for this likely relates, again, to the loss of the reflectivity of the RNFL layer with worsening glaucoma.35,36,41 Loss of reflectivity hinders the ability of the machine to identify the posterior RNFL border in peripapillary RNFL thickness scans, which predisposes to algorithm failures.13,21,33 In contrast, there are no data to suggest that the RPE/BM loses reflectivity in the setting of glaucoma; therefore, 3D optic nerve volume scans, which rely on identification of the RPE/BM, may be less susceptible to artifacts related to such loss.
Finally, to the best of our knowledge, ours is the first study to evaluate the effects of manual correction and subsequent automated interpolation for high-density 3D optic nerve volume scans in order to calculate a 3D neuroretinal rim parameter, which could represent the full future potential of clinical SD-OCT optic nerve imaging. When looking at a large series of patients, we discovered that the group mean pre-interpolation values were similar to group mean post-interpolation values for MDB neuroretinal rim thickness (Table 4). This is consistent with low-density BMO-MRW optic nerve studies featuring manual correction of BMO points, a procedure that has shown minimal impact on the group mean BMO-MRW measurement.19,30,42

Studies of high-density peripapillary RNFL volume scans and high-density macular volume scans have likewise found that average group RNFL volume and average group macular volume values were similar before and after correction of artifacts and subsequent automated interpolation.21,31 On the other hand, before and after correction of artifacts and subsequent automated interpolation for high-density scans and high-density macular volume scans have likewise found that average group RNFL volume and average group macular volume values were similar before and after correction of artifacts and subsequent automated interpolation.21,31

The need to manually fix errors or repeat scans in the clinic requires extra time and attention, and parameters derived from 3D volume scans appear to reduce this need.

Our study has potential limitations. Of note, the glaucoma subjects were significantly older than the normal subjects by 13.4 years (95% CI, 9.8–17.2; P < 0.0001) and had fewer females (52 out of 120 glaucoma subjects vs. 72 out of 114 normal subjects; 95% CI, 1.3–3.8; P = 0.004) (Table 1). However, studies thus far have shown that gender is not associated with artifact frequency,13,44 and there is conflicting evidence for whether age is a significant factor13,20,34,44,45. Second, our patient population did not include subjects with pre-perimetric glaucoma, for whom SD-OCT measurements play an important role in management. Nevertheless, we believe that these patients are likely captured within the range of optic nerve and peripapillary RNFL characteristics defined by glaucoma and normal subjects.

Our study illustrates that 3D optic nerve volume scans carry fewer clinically significant artifacts compared to 2D RNFL thickness scans (15.8% of eyes vs. 61.7% of eyes; 95% CI, 34.1–57.5%; P < 0.0001) (Table 5) and have fewer artifacts per B-scan (20.9% of 6820 B-scans vs. 61.7% of 120 B-scans; 95% CI, 31.6–50.0; P < 0.0001; table not shown) for glaucoma patients. We further determined that the mean MDB thickness measurement generally does not significantly change after interpolation (Table 4). Previous studies have demonstrated that 3D MDB neuroretinal rim thickness has similar or better diagnostic capability for glaucoma compared to 2D RNFL thickness.27,28

Taken together with our study, these findings suggest that 3D MDB neuroretinal rim thickness offers an advantage to traditional 2D RNFL thickness not only in improved diagnostic capability but also in the reduced need for repeat scanning or manual correction. Although there is a commercially available low-density protocol (BMO-MRW) that quantifies neuroretinal rim tissue, this paper may encourage SD-OCT companies to incorporate high-density protocols with associated software analysis so that clinical use of SD-OCT can reach its full potential for glaucoma care.

Acknowledgments

Supported by the National Institutes of Health (Award #UL1 RR025758, TCC); Massachusetts Lions Eye Research Fund (TCC); American Glaucoma Society Mid-Career Award (TCC); Fidelity Charitable Fund (Harvard University, TCC); Department of Defense Small Business Innovation Research (Topic #DHP15-016, TCC); and by a grant from the Center for Biomedical OCT Research and Translation, awarded by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (P41EB015903, BB, BJV, BEB). The funding organizations had no role in the design or conduct of this research.

Disclosure: E.A. Park, None; E. Tsikata, None; J.J. Lee, None; E. Shieh, None; B. Braaf, Heidelberg Engineering (E); B.J. Yakoc, Terumo Corporation (P), Ninepoint Medical (P); B.E. Bouma, NIDEK, Inc. (P), Terumo Corporation (P, F), Ninepoint Medical (P), Heidelberg Engineering (P); J.F. de Boer, NIDEK, Inc. (P), Terumo Corporation (P), Ninepoint Medical (P), Heidelberg Engineering (P, F), T.C. Chen, None

References

2. Chen TC, Hoguet A, Junk AK, et al. Spectral-domain OCT: helping the clinician diagnose glau-
Artifact Rates for 3D Versus 2D Scans


