Accuracy and Reproducibility of Automated Drusen Segmentation in Eyes with Non-Neovascular Age-Related Macular Degeneration

Muneeswar Gupta Nittala, Humberto Ruiz-Garcia, and SriniVas R. Sadda

Purpose. To evaluate the accuracy and reproducibility of drusen quantification by an automated drusen segmentation algorithm in spectral domain optical coherence tomography (SD-OCT) images of eyes with non-neovascular age-related macular degeneration (AMD).

Methods. Drusen segmentation was performed using both a commercial automated algorithm (Cirrus OCT RPE analysis tool) and manual segmentation in 44 eyes of 30 subjects with dry AMD who underwent volume OCT scanning. The drusen (space between outer RPE layer and Bruch's membrane) was segmented automatically using an automated RPE tool and manually by 3D-OCTOR software. Drusen area and volume were calculated in all eyes. Age and visual acuity data were also collected. Reproducibility of manual and automated measurements was assessed by intraclass correlation (ICC).

Results. The mean age of subjects was 78.24 (±9.4; range, 56–97 years). The mean logMAR (logarithm of the minimum angle of resolution) visual acuity was 0.4 (Snellen equivalent, ~20/50) (standard deviation, 0.40; range, 0–1.3). The mean (standard deviation) drusen area was 5.05 (3.67) mm^2 with manual segmentation and 4.66 (3.51) mm^2 with the automated RPE tool; the absolute difference was 2.63 (2.5) mm^2. The mean drusen volume was 1.49 (0.42) mm^3 with manual segmentation and 1.42 (0.43) mm^3 with the automated RPE tool; the absolute difference was 1.42 (0.43) mm^3. The agreement between manual and automated measurements of drusen volume (highest ICC = 0.95) was better than the agreement for drusen area (ICC = 0.65).

Conclusions. The quantification of drusen area and volume using an automated RPE yielded better agreement for volume than for area when compared with human expert manual segmentation. Using this software, drusen volume measurements may be a useful tool for quantifying drusen burden in clinical trials and clinical practice. (*Invest Ophthalmol Vis Sci.* 2012;53:8319–8324) DOI:10.1167/iovs.12-10582

From the Doheny Eye Institute and the Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, California.

Supported by the Beckman Macular Research Center and a Physician Scientist Award from Research to Prevent Blindness. SriniVas R. Sadda shares in royalties from intellectual property licensed to Topcon Medical Systems by the Doheny Eye Institute. SriniVas R. Sadda also serves on the scientific advisory board for Heidelberg Engineering and receives research support from Carl Zeiss Meditec, Optos, and Optovue, Inc.

Submitted for publication July 12, 2012; revised October 10, 2012; accepted November 4, 2012.

Disclosure: **M.G. Nittala**, None; **H. Ruiz-Garcia**, None; **S.R. Sadda**, Carl Zeiss Meditec (F), Optos (F), Optovue, Inc. (F), Heidelberg Engineering (S), P

Corresponding author: SriniVas R. Sadda, Doheny Eye Institute, 1450 San Pablo Street, Los Angeles, CA 90033; SSadda@doheny.org.

Investigative Ophthalmology & Visual Science, December 2012, Vol. 53, No. 13

Age-related macular degeneration (AMD) is the most common cause of legal blindness among patients aged 50 years or older in the developed world.[1–3] Drusen, particularly large, soft drusen, are a distinguishing feature and a characteristic physical sign of non-neovascular AMD (NNVAMD).[1,4] Drusen are described as focal deposits of extracellular debris between the basal lamina of the retinal pigment epithelium (RPE) monolayer and the inner collagenous layer of the Bruch's membrane.[1,3,5] Disturbance of the outer neurosensory retinal layers[6] and the RPE overlying the drusen have been described previously, and imaging studies have reported a correlation between the presence of drusen and neurosensory retinal injury.[7] Because drusen play an important role in the progression of advanced atrophic AMD,[3,6,8,9] techniques and tools to facilitate the characterization[3] and detection[8] of drusen would appear to be of value. Initial studies[3,10,11] have demonstrated that automatic identification and quantification of drusen requires accurate segmentation of the RPE band and Bruch's membrane. Segmentation errors, however, may be a potentially significant problem, particularly as the disease advances with progressive disruption of the RPE band.

Spectral domain optical coherence tomography (SD-OCT) has now become a widely used technology in ophthalmology.[12,13] The high resolution, sensitivity, and speed of SD-OCT have allowed the outer retinal and subretinal substructures to be visualized with unprecedented detail. This improvement in outer retinal imaging quality provides an opportunity to segment and quantify these substructures, including pathological features of diseases, such as drusen. Indeed, several studies have suggested that SD-OCT may be a superior modality for drusen detection and characterization compared with color photography.[3,8,10,12,13]

More recently, SD-OCT instruments have incorporated automated software algorithms to quantify drusen volumes and areas. These algorithms rely on the detection of the RPE surface and an estimated fit of the original or expected RPE baseline (assuming no elevation). Although the reproducibility of these algorithms for quantifying drusen has been evaluated previously, their accuracy for measuring drusen burden has not been well characterized.

In this report, we compare OCT drusen area and volume determined by automated algorithms against those obtained by human expert OCT graders using a standardized grading protocol.

Methods

Forty-four eyes of 30 consecutive subjects with early or intermediate NNVAMD who underwent SD-OCT (Cirrus HD-OCT; Carl Zeiss Meditec, Inc., Dublin, CA) imaging at the Doheny Eye Institute were included in this retrospective study. Approval for the collection and analysis of image data was obtained from the institutional review

board of the University of Southern California. All subjects were treated in accordance with the Declaration of Helsinki. Early and intermediate NNVAMD were defined, in accordance with Age Related Eye Diseases Study (AREDS) classification, as AREDS category 2 (numerous small drusen or a few medium drusen) and AREDS category 3 (numerous medium or at least one large drusen), respectively. Eyes with significant media opacity resulting in poor OCT image quality, geographic atrophy visible on the clinical examination, evidence of choroidal neovascularization (CNV), or other retinal or optic nerve diseases were excluded. Eyes with evidence of retinal thinning or possible nongeographic or geographic atrophy evident on OCT were not specifically excluded. All subjects included also underwent a complete ophthalmic examination, as well as color fundus photography and SD-OCT imaging.

All SD-OCT imaging was performed on a single instrument (Cirrus OCT; Carl Zeiss Meditec, Inc.) using a macular cube protocol of 512 by 128 (128 B-scans and 512 A-scans per B-scan). Automated quantification of drusen area and volume within the 6-mm cube were performed using the Cirrus 6.0 software [510(k) cleared by U.S. Food and Drug Administration (U.S. FDA)] (Carl Zeiss Meditec, Inc.) RPE analysis segmentation algorithms. For manual grading, raw OCT data was exported and then imported into previously described and validated SD-OCT reading center grading software known as 3D-OCTOR.[14,15] The 3D-OCTOR software (Doheny Eye Institute, Los Angeles, CA) has been optimized to facilitate manual segmentation of various user-defined retinal boundaries and calculate thickness and volume of segmented structures.

Grading Procedure

As described previously,[16] the Doheny Image Reading Center (DIRC) OCT protocol specifies that drusen and other RPE elevations be defined as the space between the outer border of the highly reflective RPE band (the outermost of the bright outer bands) and the inner border of the choroid (Fig. 1). The inner border of the choroid is located by the graders by recognizing the thin hyperreflective band believed to represent Bruch's membrane. Drusen borders were drawn on all OCT B-scans of the volume cube and drusen areas, and volumes were computed by the OCTOR software. All OCT manual segmentation was performed by trained, certified DIRC OCT senior graders. Although intergrader reproducibility for identifying these boundaries has been previously established,[16] all OCT B-scans were regraded by a second grader to define manual segmentation reproducibility for this dataset. All segmentation was performed in double-masked fashion, with graders masked as to each other's segmentation as well as the results of the automated segmentation.

Statistical Analysis

The main outcome measures for analysis were age, logMAR (logarithm of the minimum angle of resolution) visual acuity, drusen area, and drusen volume. The correlation of drusen measurements between eyes was assessed using bivariate Pearson correlations. The generalized estimating equation (GEE) method was used to adjust for correlations between eyes. Intergrader reproducibility of the manually derived drusen area and volume measurements was assessed with the intraclass correlation coefficient (ICC). A higher ICC indicates better reproducibility of the parameter. In addition, ICCs were also used to compare the human manually calculated values (for each grader separately). The ICC was calculated for reproducibility of manual segmentation and manual segmentation versus the automated RPE analysis tool. Bland-Altman plots were also generated to show the agreement between graders on drusen analysis. A P value of ≤ 0.05 was considered clinically significant. All statistical analysis was performed using SPSS18 Statistical Software (SPSS, Inc., Chicago, IL).

Results

Automated and manual drusen measurements were performed in 44 eyes of 30 subjects (data for all subjects are shown in Supplementary Table S1; see Supplementary Material and Supplementary Table S1, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.12-10582/-/DCSupplemental). The mean age of the subjects was 78.2 years (±9.4; range, 56–97), and 19 were female. The mean logMAR visual acuity in this cohort was 0.4 (Snellen equivalent, ~20/50) (standard deviation, 0.40; range, 0–1.3). Four of 44 eyes were noted to have OCT evidence of atrophy, manifested as thinning of the outer retinal layers with increased light transmission into the choroid. There was no significant correlation between the right and left eyes for manual and automated drusen measurements. The mean drusen volume within the macular cube scan as determined by the Cirrus 6.0 software (Carl Zeiss Meditec, Inc.) was 1.42 ± 0.43 mm^3, and the mean automated drusen area measurement was 4.66 ± 3.51 mm^2 (Table 1).

The mean manual drusen volume measurement (averaged between the graders) was 1.53 ± 0.41 mm^3. The mean manual area measurement (averaged between the graders) was 5.29 ± 3.69 mm^2. The mean absolute difference between graders was 0.12 ± 0.08 mm^3 for volume and 0.68 ± 0.56 mm^2 for area. The mean percentages of difference between graders were 7.82% and 16.64% for drusen volume and drusen area, respectively. It should be noted, however, that all cases with >10% difference in drusen volume between graders had small amounts of total drusen (volume of <2.47 mm^3 compared with a mean area of 5.3 mm^2 for the entire cohort). Similarly, all cases with >10% difference in drusen area had a small total drusen area (area of <2.6 mm^2 compared with a mean area of 5.3 mm^2 for the entire cohort). The agreement between graders was also studied using intraclass correlation coefficients and Bland-Altman plots (Table 2, Fig. 2). ICC values of over 0.98 were observed for both area and volume.

Drusen area and volume values were compared between manual and automated segmentation approaches. For this comparison, the values from the two graders were averaged for each case to produce a consensus manual result for each case. The mean absolute differences between these methods were 0.2 ± 0.1 mm^3 for volume and 2.8 ± 2.5 mm^2 for area. The mean percentages of difference were 14.3% and 136.1% for drusen volume and area, respectively. Similar to the intergrader comparison, the greater percent differences were observed in the cases with a smaller drusen burden. Regression analysis shows that a smaller drusen volume (<1.5 mm^3) predicts higher percent error (>50) for drusen area (odds ratio, 9.50; 95% confidence interval [CI], 2.32–38.87; $P = 0.002$). ICC showed excellent agreement (ICC = 0.95) between the manual and automated methods for drusen volume but only moderate agreement for area (ICC = 0.65). The levels of agreement are also illustrated in the Bland-Altman plot (Fig. 2). Cases showing a large (>3 mm^2) absolute difference between manual and automated measurements of drusen area were manually inspected. In all cases, the discrepancy was due to subtle differences in the placement of the RPE fit, or Bruch's membrane line, between the two methods (Fig. 3).

Discussion

In this study, we observed better agreement between manual and automated drusen volume measurements than for area measurements. Intergrader agreement, in contrast, was high for both area and volume. Although the mean percent differences for some comparisons appeared high, this was largely due to cases with small total amounts of drusen, where

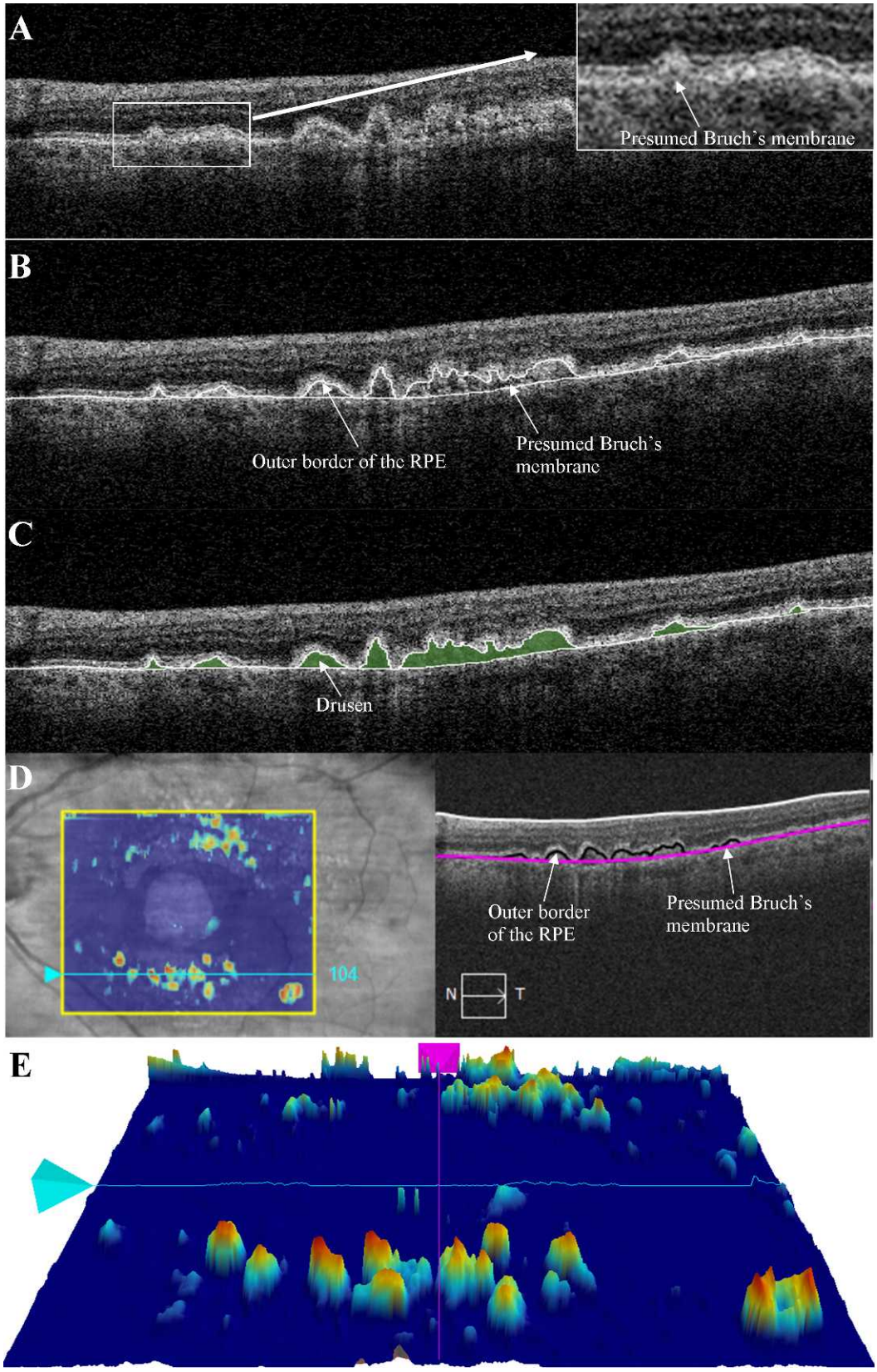

Figure 1. Optical coherence tomography (OCT) B-scan manual drusen segmentation using computer-assisted grading software (3D-OCTOR) (**A**–**C**) and using an automated RPE analysis tool (**D**, **E**). (**A**) Raw OCT scan image. (**B**) Manually graded boundaries. (**C**) Calculated space (drusen). (**D**) Infrared fundus image with drusen map (*left side of image*) and automated-software-graded boundaries (*right side of image*). (**E**) Drusen volume map generated by the automated RPE tool.

Table 1. Drusen Measurements (Area and Volume) and Absolute and Percent Errors between Manual and Automated Segmentation

	Drusen Area, mm²		Drusen Volume, mm³	
	Mean (±SD)	Range	Mean (±SD)	Range
Grader 1*	5.05 (±3.67)	0.1 to 13.6	1.49 (±0.42)	0.9 to 2.48
Grader 2*	5.53 (±3.75)	0.1 to 13.8	1.57 (±0.41)	0.96 to 2.6
Absolute difference between graders	0.68 (±0.56)	0 to 2.3	0.12 (±0.08)	0 to 0.31
% Error between graders	16.64 (±13.57)	0 to 54.55	7.82 (±5.67)	0 to 26.67
Automated tool*	4.66 (±3.51)	0.1 to 13.5	1.42 (±0.43)	0.83 to 2.31
Absolute difference between methods†	2.8 (±2.5)	0.2 to 12.4	0.2 (±0.1)	0 to 0.5
% Error between methods‡	68.3 (±48.5)	4.0 to 196.8	14.5 (±10.1)	1.0 to 39.5

* Original values.
† Difference between manual and automated segmentation.
‡ Percent error between manual and automated segmentation.

Table 2. ICCs for Drusen Quantification by Manual* and Algorithm Segmentation

	ICC	95% CI
Drusen area		
Between graders	0.99	0.98 to 0.99
Between manual and algorithm	0.64	0.34 to 0.80
Drusen volume		
Between graders	0.98	0.96 to 0.99
Between manual and algorithm	0.94	0.89 to 0.97

* Manual results represent the average from two graders.

minute discrepancies in quantification could yield large percentages of difference. Inspection of cases with apparent discordance in absolute drusen area measurements between manual and automated methods revealed that the source of the discrepancy in nearly all cases was the positioning of the outer of the two lines (the RPE fit, or Bruch's membrane line). This line appeared to ride above or below the actual (as determined by a human expert) Bruch's membrane position in some locations along the B-scans. This is perhaps not surprising, as this outer boundary is in fact a "fit," or approximation, of the outer wall curvature. Thus, it would not be expected to precisely follow small-amplitude oscillations in the ocular

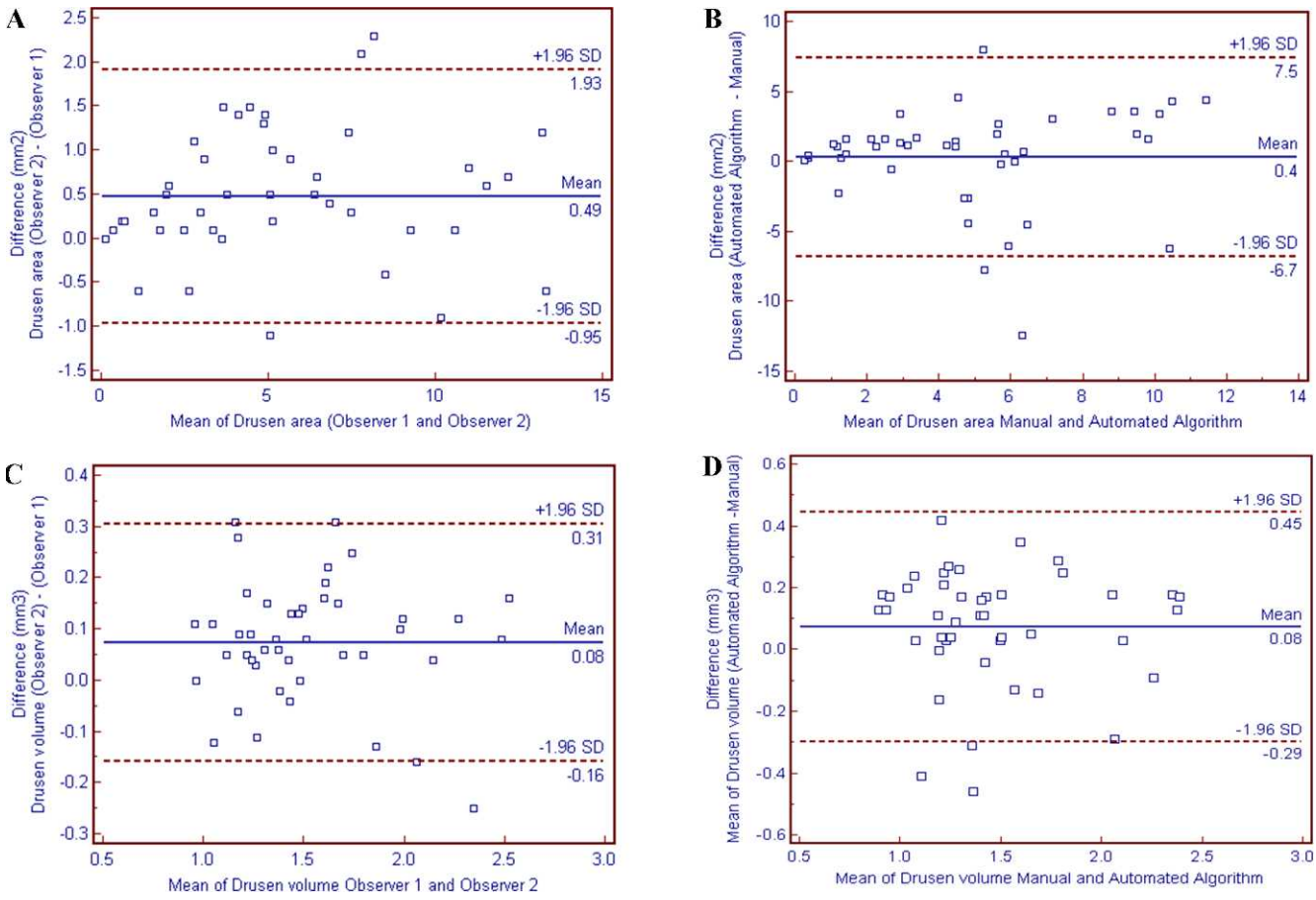

Figure 2. Bland-Altman plots demonstrate the intergrader reproducibility for area (mm²) and volume (mm³) of drusen, and *dotted line* indicates the 95% CI limits (upper and lower). (**A**) Drusen area between graders. (**B**) Drusen area between manual grading and automated algorithm. (**C**) Drusen volume between graders. (**D**) Drusen volume between manual grading and automated algorithm.

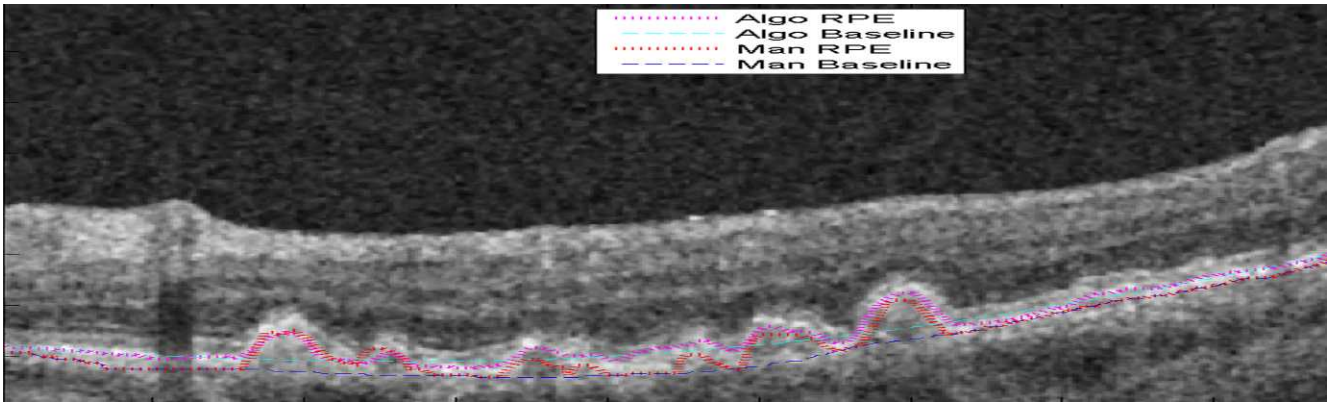

Figure 3. The differences between the placement of the RPE and Bruch's membrane line between the two (algorithm and manual) methods. Note that the algorithm-determined Bruch's membrane (or RPE fit) line (*dashed green*) is placed higher than the manually segmented line (*dashed blue*). Algo, algorithm; Baseline, location of Bruch's membrane; Man, manual.

surface that are typical of real eyes of patients. It is interesting but perhaps not surprising that this discrepancy had a significant impact on the area measurements but not on the volume determinations. Since subtle shifts in the outer boundary position could truncate or extend the base (i.e., diameter) of a detected drusenoid elevation, these shifts could result in significant changes in area, since area would change by the square. Truncation or inclusion of the peripheral edges of the drusen, however, would have a substantially smaller effect on drusen volume, since the thickness of the drusen would be relatively small in these areas (especially compared to the thickness at the center or apex of the druse) and would not contribute much to the overall volume of the individual drusen or the total drusen burden in the eye. This observation suggests that drusen volume, rather than drusen area, may be the preferred metric for use in OCT-based drusen quantification analyses.

The absolute accuracy of the Cirrus automated drusen segmentation algorithm did not appear to be influenced by the level of drusen burden, as evidenced by the lack of any notable trend in the Bland-Altman Plots.

OCT-based quantification of drusen burden may be a useful tool for study of AMD, particularly with multiple agents currently under study for treatment of non-neovascular AMD. Traditionally, drusen quantification has been performed by planar color imaging methods. However, reliable quantification by these techniques may be challenging, particularly when the borders of the drusen are indistinct. As an example, drusen area and drusen size are known to be important indicators of AMD progression.[17–19] In "AREDS Report No. 18," however, the investigators suggested that drusen size rather than drusen area should be used to assess AMD severity because of the difficulty in measuring drusen area.[18] Compared with planar imaging, SD-OCT provides excellent visualization of the morphologic structure for drusen and, in particular, their axial extent.[13] The axial information facilitates the identification of drusen borders and could potentially yield more reliable measurements. Further comparative studies are required to compare drusen detection by color photography versus that by OCT to better understand the relationship between drusen burden quantified by the two techniques. However, the existence of commercial U.S. FDA-cleared automated drusen quantification tools in OCT systems makes this approach attractive and potentially broadly applicable in clinical practice, should drusen quantification prove to be of clinical value. Previous studies[20–22] have demonstrated that OCT-based automated drusen quantification is reproducible. Chiu et al.[22] showed good reproducibility of drusen volume measurements using custom automatic drusen segmentation software. In this paper, we also demonstrated that it is accurate (at least for volume) against a reading center standard.

Our study is not without limitations. First, the overall sample (44 eyes) is still relatively small. On other hand, detailed manual segmentation of hundreds of B-scans from dense OCT volume cubes is an extremely laborious task, which limits the ability to generate very large datasets. Second, we did not have color photographs for this particular cohort to compare OCT drusen areas with color photograph-derived measurements. While this would have been an interesting analysis, this has been evaluated in other studies (Philip JR, et al. *IOVS* 2011;52:ARVO E-Abstract 139) and was not the main focus of our project. Third, we only evaluated drusen area and volume. Other potential OCT-based metrics (which may become areas of future study), such as circularity/topology, were not studied.

On the other hand, our study also has several strengths, including the use of certified, experienced reading center OCT graders and the demonstration of a high degree of reproducibility between the graders.

In summary, we observed that drusen area and volume could be computed reproducibly by manual segmentation of dense OCT volume datasets. In addition, we observed that automated measurements of drusen volume by commercial (Cirrus OCT; Carl Zeiss Meditec, Inc.) OCT algorithms demonstrated better agreement with manually derived values, with less-good agreement observed for area. Our findings suggest that automated OCT-derived drusen volume measurements may be a reliable and accurate tool for quantifying drusen burden in clinical trials and clinical practice.

References

1. Schmitz-Valckenberg S, Steinberg JS, Fleckenstein M, Visvalingam S, Brinkmann CK, Holz FG. Combined confocal scanning laser ophthalmoscopy and spectral-domain optical coherence tomography imaging of reticular drusen associated with age-related macular degeneration. *Ophthalmology*. 2010;117:1169–1176.
2. Abdelsalam A, Del Priore L, Zarbin MA. Drusen in age-related macular degeneration: pathogenesis, natural course, and laser photocoagulation-induced regression. *Surv Ophthalmol*. 1999;44:1–29.
3. Schlanitz FG, Ahlers C, Sacu S, et al. Performance of drusen detection by spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2010;51:6715–6721.

4. Khanifar AA, Koreishi AF, Izatt JA, Toth CA. Drusen ultrastructure imaging with spectral domain optical coherence tomography in age-related macular degeneration. *Ophthalmology*. 2008;115:1883–1890.
5. Spaide RF, Curcio CA. Drusen characterization with multimodal imaging. *Retina*. 2010;30:1441–1454.
6. Wolf-Schnurrbusch UE, Enzmann V, Brinkmann CK, Wolf S. Morphologic changes in patients with geographic atrophy assessed with a novel spectral OCT-SLO combination. *Invest Ophthalmol Vis Sci*. 2008;49:3095–3099.
7. Schuman SG, Koreishi AF, Farsiu S, Jung SH, Izatt JA, Toth CA. Photoreceptor layer thinning over drusen in eyes with age-related macular degeneration imaged in vivo with spectral-domain optical coherence tomography. *Ophthalmology*. 2009; 116:488–496.
8. Smith RT, Chan JK, Nagasaki T, et al. Automated detection of macular drusen using geometric background leveling and threshold selection. *Arch Ophthalmol*. 2005;123:200–206.
9. Paulus de Jong TVM. Age-related macular degeneration. *N Engl J Med*. 2006;355:1474–1485.
10. Jain N, Farsiu S, Khanifar AA, et al. Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs. *Invest Ophthalmol Vis Sci*. 2010; 51:4875–4883.
11. Smith RT, Sohrab MA, Pumariega N, et al. Dynamic soft drusen remodelling in age-related macular degeneration. *Br J Ophthalmol*. 2010;94:1618–1623.
12. Freeman SR, Kozak I, Cheng L, et al. Optical coherence tomography-raster scanning and manual segmentation in determining drusen volume in age-related macular degeneration. *Retina*. 2010;30:431–435.
13. Yi K, Mujat M, Park BH, et al. Spectral domain optical coherence tomography for quantitative evaluation of drusen and associated structural changes in non-neovascular age-related macular degeneration. *Br J Ophthalmol*. 2009;93:176–181.
14. Sadda SR, Keane PA, Ouyang Y, Updike JF, Walsh AC. Impact of scanning density on measurements from spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2010;51:1071–1078.
15. Sadda SR, Joeres S, Wu Z, et al. Error correction and quantitative subanalysis of optical coherence tomography data using computer-assisted grading. *Invest Ophthalmol Vis Sci*. 2007;48:839–848.
16. Joeres S, Tsong JW, Updike PG, et al. Reproducibility of quantitative optical coherence tomography subanalysis in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2007;48:4300–4307.
17. Davis MD, Gangnon RE, Lee LY, et al. The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS report no. 17. *Arch Ophthalmol*. 2005;123: 1484–1498.
18. Ferris FL, Davis MD, Clemons TE, et al. A simplified severity scale for age-related macular degeneration: AREDS report no.18. *Arch Ophthalmol*. 2005;123:1570–1574.
19. Klein R, Klein BE, Kndtson MD, Meuer SM, Swift M, Gangnon RE. Fifteen-year cumulative incidence of age-related macular degeneration: the Beaver Dam Eye Study. *Ophthalmology*. 2007;114:253–262.
20. Gregori G, Wang F, Rosenfeld PJ, et al. Spectral domain optical coherence tomography imaging of drusen in nonexudative age-related macular degeneration. *Ophthalmology*. 2011;118: 1373–1379.
21. Freeman SR, Kozak I, Cheng L, et al. Optical coherence tomography-raster scanning and manual segmentation in determining drusen volume in age-related macular degeneration. *Retina*. 2010;30:431–435.
22. Chiu SJ, Izatt JA, O'Connell RV, Winter KP, Toth CA, Farsiu S. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Invest Ophthalmol Vis Sci*. 2012;53:53–61.