

Reproducibility of Diabetic Macular Edema Estimates From SD-OCT Is Affected by the Choice of Image Analysis Algorithm

Elliott H. Sohn,¹ John J. Chen,¹ Kyungmoo Lee,² Meindert Niemeijer,^{1,2} Milan Sonka,^{1,2} and Michael D. Abramoff¹⁻⁴

¹Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa

²Department of Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa

³Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa

⁴Department of Veterans Affairs, Iowa City, Iowa

Correspondence: Michael D. Abramoff, Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, 200 Hawkins Drive PFP, Iowa City, IA 52242; michael-abramoff@uiowa.edu

EHS and JJC are joint first authors.

Submitted: June 18, 2012

Accepted: May 1, 2013

Citation: Sohn EH, Chen JJ, Lee K, Niemeijer M, Sonka M, Abramoff MD. Reproducibility of diabetic macular edema estimates from SD-OCT is affected by the choice of image analysis algorithm. *Invest Ophthalmol Vis Sci*. 2013;54:4184-4188. DOI:10.1167/iovs.12-10420

PURPOSE. To evaluate the intersession repeatability of retinal thickness measurements in patients with diabetic macular edema (DME) using the Heidelberg Spectralis optical coherence tomography (OCT) algorithm and a publicly available, three-dimensional graph search-based multilayer OCT segmentation algorithm, the Iowa Reference Algorithm.

METHODS. Thirty eyes from 21 patients diagnosed with clinically significant DME were included and underwent consecutive, registered macula-centered spectral-domain optical coherence scans (Heidelberg Spectralis). The OCT scans were segmented into separate surfaces, and the average thickness between internal limiting membrane and outer retinal pigment epithelium complex surfaces was determined using the Iowa Reference Algorithm. Variability between paired scans was analyzed and compared with the retinal thickness obtained from the manufacturer-supplied Spectralis software.

RESULTS. The coefficient of repeatability (variation) for central macular thickness using the Iowa Reference Algorithm was 5.26 μm (0.62% [95% confidence interval (CI), 0.43-0.71]), while for the Spectralis algorithm this was 6.84 μm (0.81% [95% CI, 0.55-0.92]). When the central 3 mm was analyzed, the coefficient of repeatability (variation) was 2.46 μm (0.31% [95% CI, 0.23-0.38]) for the Iowa Reference Algorithm and 4.23 μm (0.53% [95% CI, 0.39-0.65]) for the Spectralis software.

CONCLUSIONS. The Iowa Reference Algorithm and the Spectralis software provide excellent reproducibility between serial scans in patients with clinically significant DME. The publicly available Iowa Reference Algorithm may have lower between-measurement variation than the manufacturer-supplied Spectralis software for the central 3 mm subfield. These findings have significant implications for the management of patients with DME.

Keywords: OCT, diabetic retinopathy, image analysis, macular edema, retina

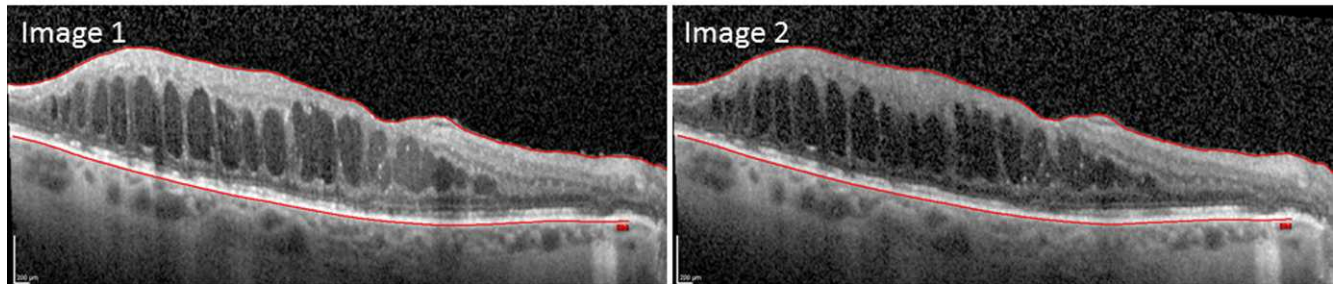
Optical coherence tomography (OCT) has become an essential tool in diagnosing and monitoring patients with diabetic macular edema (DME).^{1,2} Initially, time-domain OCT using Stratus OCT (Carl Zeiss Meditec, Inc., Dublin, CA) was used clinically to evaluate patients with DME.³ This has largely been supplanted by spectral-domain OCT technology (SD-OCT), which allows for faster scans and higher resolution. One of the most commonly used SD-OCT devices, Spectralis OCT (Heidelberg Spectralis; Heidelberg Engineering, Heidelberg, Germany), has a scan rate of 40,000 axial scans per second and an axial resolution of 5 μm .

Repeatability of OCT measurements has an important impact on how data can be interpreted in both clinical practice and clinical trials. This has been recognized and explored in detail most extensively for eyes in normal subjects. Interestingly, Spectralis OCT repeatability was shown to be higher when compared to other OCT devices in measuring the retinal thickness in healthy subjects,^{4,5} possibly related to active eye tracking as well as the ability to register subsequent images, as is now standard on most devices. Only two studies, however,

have addressed the reproducibility of SD-OCT in patients with DME.^{6,7} Reproducibility in DME could potentially be lower than in normal eyes because macular edema may distort the retinal layers and segmentation by OCT, which has been a reported finding induced by subretinal fluid from neovascular age-related macular degeneration.^{8,9} Clinical OCT devices are equipped with manufacturer-specific segmentation algorithms that identify the retinal layers in the acquired OCT volumes. Thus, reproducibility potentially depends not only on the OCT imaging hardware, but also on the algorithm, both of which can be affected by macular edema.^{5,6,10}

The Iowa Reference Algorithm,¹¹⁻¹³ available in the public domain at <http://www.biomed-imaging.uiowa.edu/downloads>, has previously been developed and validated on OCT volumes from all clinically used OCT devices, including Spectralis. This allowed us to study the question whether the segmentation algorithm has a differential impact on reproducibility of DME quantification. The purpose of this pilot study was to compare the reproducibility of the manufacturer-supplied algorithm for Spectralis OCT volumes and the reproducibility of the Iowa

Heidelberg Spectralis software



Iowa Reference Algorithm

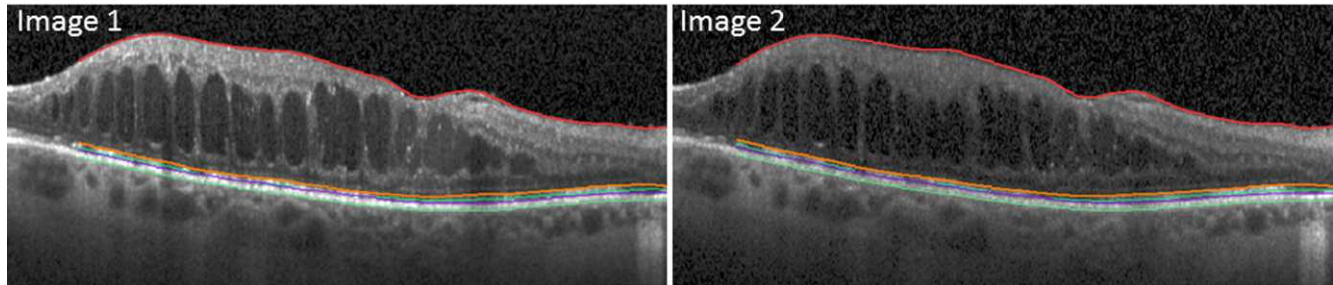


FIGURE 1. Heidelberg Spectralis software and the Iowa Reference Algorithm reliably identify the internal limiting membrane and the outer surface of the retinal pigment epithelium between serial scans (image 1 and image 2), which were used as the boundaries to calculate the retinal thickness in both algorithms. Also shown is the identification of the external limiting membrane (*orange*) and the inner/outer segment junction (*purple*) by the Iowa Reference Algorithm. The images between the two algorithms are identical cuts through the retina, but appear slightly different because the Spectralis images are exported from the Spectralis software, while the Iowa Algorithm images are the raw images with the boundaries annotated.

Reference Algorithm on the same Spectralis OCT volumes for patients with DME.

METHODS

The study protocol was approved by the Institutional Review Board for Human Subjects Research at the University of Iowa, and adhered to the tenets set forth in the Declaration of Helsinki.

Subjects

All subjects were recruited from a single academic clinic at the University of Iowa. Thirty eyes from 21 patients diagnosed with clinically significant DME as defined by the Early Treatment Diabetic Retinopathy Study (ETDRS) were included. Eyes were excluded from the study if there was significant media opacity resulting in a poor OCT signal.

Study Design and Imaging Protocol

All 30 eyes were imaged after mydriasis using the Spectralis HRA+OCT (Spectralis software version 5.3; Heidelberg Engineering). Each eye was scanned in two sessions, with a recovery interval of 10 minutes between the two sessions, during which the patient left the chair prior to reimaging. The scan protocol captured the central $20^\circ \times 15^\circ$ with a minimum of 19 sections and an automatic real-time averaging of nine frames. To minimize extrinsic factors, such as patient fixation and the operator's ability to consistently place the macular grid over the same points during each scan, the two OCT sessions were registered to one another using the Spectralis registration software. Volumes were stored in the manufacturer standard .vol format.

Image Analysis

The retinal thickness of the central 1 and 3 mm subfields as defined by the ETDRS¹⁴ was obtained from the manufacturer-supplied Spectralis software. The Spectralis software seems to determine average retinal thickness as the distance between the inner limiting membrane and the signal from the outer border of the RPE in two dimensions, though the algorithms have not been published. The average thickness of the central 1 and 3 mm subfields was determined using the Iowa Reference Algorithm by segmenting four surfaces in the same OCT volumes: the internal limiting membrane, external limiting membrane, inner/outer segment (IS/OS) junction, and the outer surface of the retinal pigment epithelium (RPE) complex (Fig. 1).^{13,15} Mean subfield retinal thickness was obtained by averaging the distance between internal limiting membrane and RPE complex for all A-scans in each central subfield. No correction of segmentation errors was performed.

Statistical Analysis

Mean central subfield thicknesses from the Iowa Reference Algorithm and Spectralis software were compared using the paired Student's *t*-test. The coefficient of repeatability was calculated as 1.96 times the standard deviation between subfield thicknesses, and the coefficient of variation was calculated as standard deviation divided by the mean. The 95% confidence intervals (95% CI) and comparisons of the coefficient of variation were analyzed using the log transformation method.¹⁶ Bland-Altman plots were calculated to further compare the two algorithms.¹⁷ To examine if increased macular thickness had an influence on repeatability of measurements, the coefficients of repeatability and variation were compared between eyes with a retinal thickness higher than 400 μm and eyes with a retinal thickness less than 400 μm . This threshold of 400 μm was chosen because it was

TABLE. Intersession Coefficient of Repeatability and Variation for the Iowa Reference Algorithm and the Heidelberg Spectralis Software

	Mean Macular Thickness, μm (95% CI)	Coefficient of Variation, % (95% CI)	Coefficient of Repeatability, μm
Iowa Reference Algorithm, central 1 mm	435.60 (186.60–684.59)	0.62 (0.43–0.71)	5.26
Spectralis software, central 1 mm	429.15 (180.31–677.99)	0.81 (0.55–0.92)	6.84
Iowa Reference Algorithm, central 3 mm	407.23 (265.90–548.56)	0.31 (0.23–0.38)	2.46
Spectralis software, central 3 mm	405.02 (274.12–535.91)	0.53 (0.39–0.65)	4.23
Iowa Reference Algorithm, central 1 mm > 400 μm	534.81 (336.65–732.98)	0.64 (0.43–0.71)	6.70
Iowa Reference Algorithm, central 1 mm < 400 μm	336.38 (246.25–426.50)	0.49 (0.47–0.79)	3.25
Spectralis software, central 1 mm > 400 μm	529.80 (339.98–719.62)	0.87 (0.66–1.11)	9.0
Spectralis software, central 1 mm < 400 μm	328.50 (234.45–422.55)	0.55 (0.40–0.67)	3.56

roughly the average of the central macular thickness in our patient population.

RESULTS

Thirty eyes of 21 patients with clinically significant DME (11 male and 10 females) with an average age of 59.9 ± 11.3 years were included in the study. The average central macular thickness (CMT), defined as the central 1 mm on the ETDRS grid, calculated by the Iowa Reference Algorithm, was $435.60 \mu\text{m}$ (95% CI, 186.60–684.59 μm); calculated by the Spectralis software it was $429.15 \mu\text{m}$ (95% CI, 180.31–677.99 μm), a difference that is not significant ($P = 0.85$). See the Table. Comparing the CMT between the two algorithms revealed a coefficient of repeatability of $25.02 \mu\text{m}$ and a coefficient of variation of 2.96% (95% CI, 2.29–3.86). The average central 3

mm thickness was $407.23 \mu\text{m}$ (95% CI, 265.90–548.56 μm) for the Iowa Reference Algorithm and $405.02 \mu\text{m}$ (95% CI, 274.12–535.91 μm) for the Heidelberg Spectralis algorithm, also a nonsignificant difference ($P = 0.902$). Comparing the central 3 mm thickness between the two algorithms gave a coefficient of repeatability of $15.73 \mu\text{m}$ and a coefficient of variation of 1.86% (95% CI, 1.24–2.08).

Both the Iowa Reference Algorithm and Spectralis software consistently segmented the boundaries of the retina layers well as evidenced in Figure 1, providing excellent intersession repeatability (Fig. 2). The intersession coefficient of repeatability and variation for repeat scans of the CMT was $5.26 \mu\text{m}$ (0.62% [95% CI, 0.43–0.71]) for the Iowa Reference Algorithm and $6.84 \mu\text{m}$ (0.81% [95% CI, 0.55–0.92]) for the Heidelberg Spectralis software—slightly lower for the Iowa Reference Algorithm, but not significantly different as demonstrated by

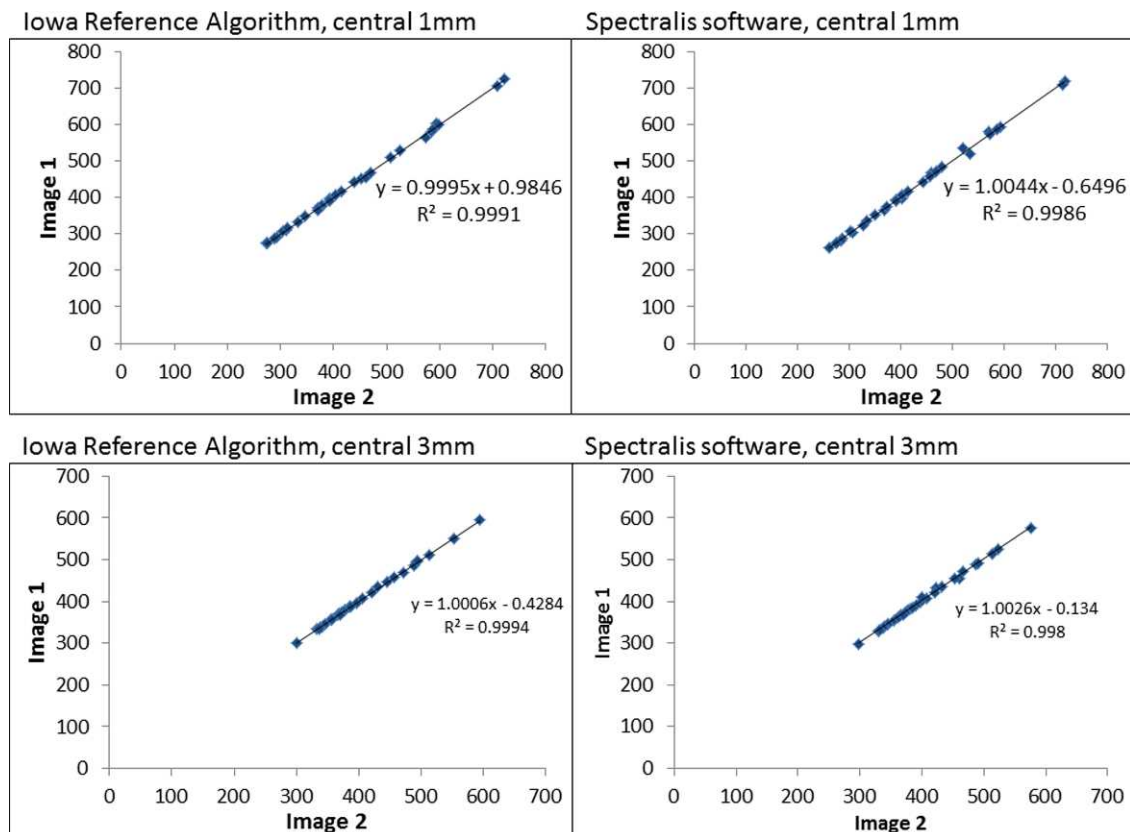


FIGURE 2. The retinal thickness of the central 1 and 3 mm of the two serial scans are plotted against each other to demonstrate excellent intersession repeatability for the Iowa Reference Algorithm and the Heidelberg Spectralis software.

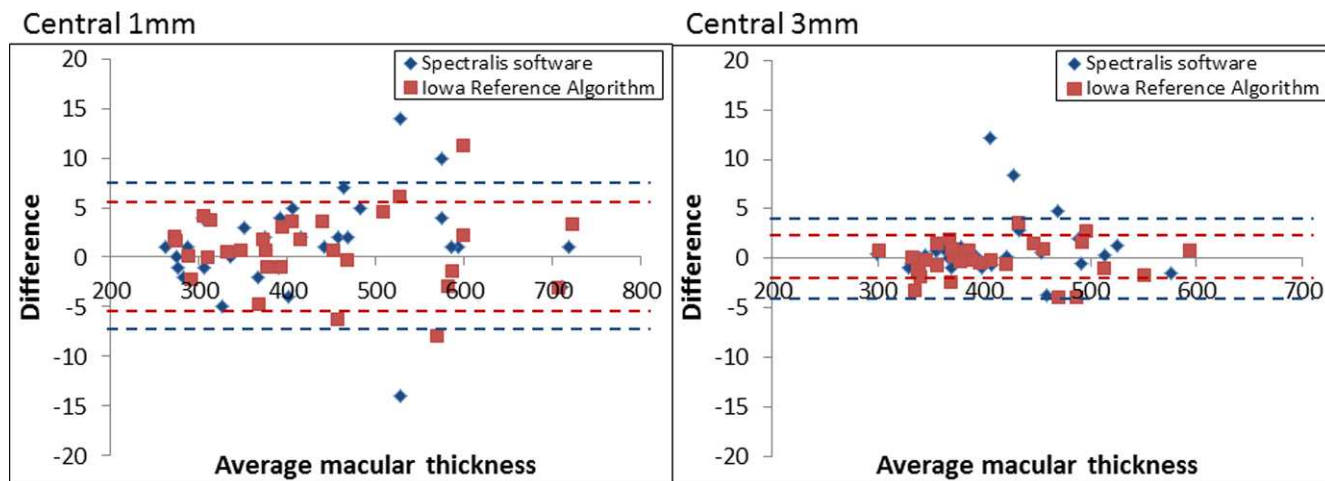


FIGURE 3. Bland-Altman plot demonstrating the variability between serial scans for the Iowa Reference Algorithm and the Heidelberg Spectralis software for the central 1 and 3 mm. The 95% CI are shown for both the Iowa Reference Algorithm (red dotted line) and the Heidelberg Spectralis software (blue dotted line).

the overlapping 95% confidence intervals (Table). When the central 3 mm was analyzed, the Iowa Reference Algorithm showed a significantly lower coefficient of repeatability and variation of 2.46 μm (0.31% [95% CI, 0.23–0.38]) compared with 4.23 μm (0.53% [95% CI, 0.39–0.65]) for the Heidelberg Spectralis software (Table). Bland-Altman plots were calculated (Fig. 3).

When the central 1 mm was compared, the patients with the highest intersession variability typically had large amounts of macular edema that were centered adjacent to the fovea, which placed the center of the OCT analysis on the edge of the edema. Therefore a small change in locating the center of the OCT analysis introduced differences in retinal thickness measurements. This was less pronounced when the central 3 mm was identified, because it encompassed larger areas of the retina. For the central 3 mm analysis, the higher variability of the Spectralis software was mostly due to two patients. Further analysis of these patients revealed that one had an error in registration within the Spectralis software, but not with the Iowa algorithm. The other patient had vitreomacular traction, which introduced segmentation error in the Spectralis software that was not encountered with the Iowa Reference Algorithm.

There was a trend for larger degrees of macular edema to have an increase in the variability between serial scans, although this was not found to be statistically significant. In patients with CMT greater than 400 μm , the coefficient of repeatability and variation was 9 μm (0.87%) compared with a coefficient of repeatability and variation of 3.56 μm (0.55%) for patients with CMT less than 400 μm when analyzed by the Heidelberg Spectralis software (Table). This trend was less pronounced when analyzed by the Iowa Reference Algorithm; here the coefficient of repeatability and variation was 6.7 (0.64%) and 3.25 μm (0.49%) for greater than 400 μm and less than 400 μm , respectively (Table).

DISCUSSION

The results of this pilot study show that the Spectralis manufacturer-supplied algorithm and Iowa Reference Algorithm segmentation algorithms may have a differential impact on the reproducibility of DME quantification by SD-OCT. There is no significant difference between the retinal thicknesses measured in these subjects with DME by the two algorithms

overall, so they are essentially measuring the same entity. However, in this sample of subjects with DME, the reproducibility of the manufacturer-supplied algorithm for Spectralis and the reproducibility of the Iowa Reference Algorithm on the same OCT volumes were significantly different for the 3 mm but not the 1 mm central subfield (CMT). Both provide good reproducibility for the CMT when compared to prior studies of OCT reproducibility on DME using other commonly employed OCT machines, such as time-domain Stratus and spectral-domain Cirrus.^{3,7} The coefficient of variation of the CMT was lower for both the Iowa algorithm (0.62%) and the Spectralis algorithm (0.81%) than the coefficient of variation previously found by Forooghian et al. for both Cirrus OCT (2.42%) and Stratus OCT (2.63%).⁷ Wolf-Schnurrbusch et al. showed that Spectralis OCT had a lower coefficient of variation when compared to other OCT devices in normal eyes.⁵

The coefficient of variation of the manufacturer-supplied Spectralis software for CMT is 6.84 μm (5.26 μm for the Iowa algorithm). This is the most useful number clinically from our study because it provides the threshold for which a change in CMT is statistically significant in a patient with DME, where changes below this number are likely to be lost in variability between measurements. This number can be used to evaluate disease progression or detect true change in response to therapeutic interventions in both clinical practice and clinical trials. Interestingly, there was a trend toward worse repeatability for patients with increased DME as measured by CMT; the coefficient of repeatability was 9 μm for patients with a CMT greater than 400 μm , compared with a coefficient of repeatability of 3.56 μm for patients with CMT less than 400 μm using the Spectralis software. In measuring the CMT, the largest variability in intersession repeat scans was caused by the center of the OCT analysis falling on the edge of macular edema. In these cases, a small change in where the center of the OCT analysis is performed creates a large difference in the retinal thickness measured. This likely explains the trend for an increase in variability in patients with larger degrees of DME. Therefore, one can take the amount of DME into account when interpreting the repeatability of subsequent scans with use of the Spectralis software.

The Iowa Reference Algorithm may have better reproducibility for analyzing Spectralis OCT volumes than the manufacturer's supplied algorithm. Specifically, the coefficient of repeatability and variation is slightly lower at 5.26 μm

(0.62%) for the 1 mm central subfield, and $2.46 \mu\text{m}$ (0.31%) for the 3 mm central subfield, significantly lower than with the manufacturer's supplied Spectralis software. The most likely explanation for this higher robustness is that the Iowa Reference Algorithm uses all three-dimensional information when identifying and segmenting the retinal layers. Thus we conclude that incorporation of three-dimensional data provides useful information that is most likely lost in the currently available segmentation algorithms, which we assume are all two-dimensional although the manufacturer-supplied algorithms have not been made public. However, the caveat to our analysis is the inclusion of two patients that largely contributed to the increased variability for the central 3 mm analysis in the Spectralis software. In one patient there was an error in registration performed by the Spectralis software. The other patient had vitreomacular traction that introduced segmentation errors within the Spectralis software, but the OCT volume was correctly segmented with the Iowa Reference Algorithm. Because such patients are seen among those with typical DME, they were included in our study. Interestingly, these two patients did not have much of an effect on the central 1 mm analysis, where the patients with the highest variability had the peak of the macular edema located just adjacent to the fovea as described above.

It is remarkable that such excellent repeatability can be achieved given that the axial resolution of SD-OCT is on the order of 4 to $6 \mu\text{m}$.⁵ The Iowa Reference Algorithm can analyze the OCT volumes from all major SD-OCT devices. Potentially the measured differences that have been found between retinal layer thickness in different SD-OCT devices are related to the manufacturer-specific algorithms used in these devices.^{5-7,10,18} Because our results show that the layer segmentation algorithm affects the between-measurement variability, a publicly available published algorithm such as the Iowa Reference Algorithm has the potential to eliminate cross-device variability.

A limitation of this pilot study is the small sample size. It is possible that a larger sample size would be able to detect a significant difference in reliability between the Iowa Reference Algorithm and the Spectralis software when comparing the CMT. A larger sample size may also demonstrate a significant difference in reliability when comparing larger and lesser degrees of macular edema. In addition, because of the small sample size, inclusion of the two aforementioned patients with the error in registration and vitreomacular traction had a large effect on repeatability. Larger numbers would aid in teasing out the reliability for measurements of the CMT and central 3 mm. As mentioned, we have not compared the variability of the manufacturer-supplied and the Iowa Reference Algorithm across multiple SD-OCT devices; we plan to do such a study.

In summary, both the Heidelberg Spectralis software and Iowa Reference Algorithm provide excellent reproducibility between serial scans in patients with clinically significant DME. The publicly available Iowa Reference Algorithm may have lower between-measurement variability than the manufacturer-supplied algorithm for the central 3 mm subfield. Lowering between-measurement variability is crucial for optimal management of patients with DME.

Acknowledgments

Supported in part by National Institutes of Health Grants R01 EY018853, R01 EY019112, and R01 EB004640.

Disclosure: **E.H. Sohn**, None; **J.J. Chen**, None; **K. Lee**, None; **M. Niemeijer**, None; **M. Sonka**, P; **M.D. Abramoff**, P

References

1. Hee MR, Puliafito CA, Wong C, et al. Quantitative assessment of macular edema with optical coherence tomography. *Arch Ophthalmol*. 1995;113:1019-1029.
2. Hee MR, Puliafito CA, Duker JS, et al. Topography of diabetic macular edema with optical coherence tomography. *Ophthalmology*. 1998;105:360-370.
3. Krzystolik MG, Strauber SF, Aiello LP, et al. Reproducibility of macular thickness and volume using Zeiss optical coherence tomography in patients with diabetic macular edema. *Ophthalmology*. 2007;114:1520-1525.
4. Menke MN, Knecht P, Sturm V, Dabov S, Funk J. Reproducibility of nerve fiber layer thickness measurements using 3D fourier-domain OCT. *Invest Ophthalmol Vis Sci*. 2008;49:5386-5391.
5. Wolf-Schnurrbusch UE, Ceklic L, Brinkmann CK, et al. Macular thickness measurements in healthy eyes using six different optical coherence tomography instruments. *Invest Ophthalmol Vis Sci*. 2009;50:3432-3437.
6. Lammer J, Scholda C, Prunte C, Benesch T, Schmidt-Erfurth U, Bolz M. Retinal thickness and volume measurements in diabetic macular edema: a comparison of four optical coherence tomography systems. *Retina*. 2011;31:48-55.
7. Forooghian F, Cukras C, Meyerle CB, Chew EY, Wong WT. Evaluation of time domain and spectral domain optical coherence tomography in the measurement of diabetic macular edema. *Invest Ophthalmol Vis Sci*. 2008;49:4290-4296.
8. Patel PJ, Chen FK, Ikeji F, et al. Repeatability of stratus optical coherence tomography measures in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2008;49:1084-1088.
9. Parravano M, Oddone F, Boccassini B, et al. Reproducibility of macular thickness measurements using Cirrus SD-OCT in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2010;51:4788-4791.
10. Kiernan DF, Hariprasad SM, Chin EK, Kiernan CL, Rago J, Mieler WF. Prospective comparison of cirrus and stratus optical coherence tomography for quantifying retinal thickness. *Am J Ophthalmol*. 2009;147:267-275, e2.
11. Lee K, Niemeijer M, Garvin MK, Kwon YH, Sonka M, Abramoff MD. Segmentation of the optic disc in 3-D OCT scans of the optic nerve head. *IEEE Trans Med Imaging*. 2010;29:159-168.
12. Garvin MK, Abramoff MD, Kardon R, Russell SR, Wu X, Sonka M. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search. *IEEE Trans Med Imaging*. 2008;27:1495-1505.
13. Garvin MK, Abramoff MD, Wu X, Russell SR, Burns TL, Sonka M. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans Med Imaging*. 2009;28:1436-1447.
14. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch Ophthalmol*. 1985;103:1796-1806.
15. Yin Y, Zhang X, Williams R, Wu X, Anderson DD, Sonka M. LOGISMOS-layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. *IEEE Trans Med Imaging*. 2010;29:2023-2037.
16. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ*. 1996;313:106.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307-310.
18. Legarreta JE, Gregori G, Punjabi OS, Knighton RW, Lalwani GA, Puliafito CA. Macular thickness measurements in normal eyes using spectral domain optical coherence tomography. *Ophthalmic Surg Lasers Imaging*. 2008;39:S43-S49.