

The Reliability of Cone Density Measurements in the Presence of Rods

Jessica I. W. Morgan^{1,2}, Grace K. Vergilio¹, Jessica Hsu³, Alfredo Dubra⁴, and Robert F. Cooper^{1,5}

¹ Scheie Eye Institute, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

² Center for Advanced Retinal and Ocular Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

³ Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

⁴ Department of Ophthalmology, Stanford University, Stanford, CA, USA

⁵ Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Correspondence: Jessica I. W. Morgan, 3400 Civic Center Blvd, Ophthalmology 3rd floor west, 3-112W, Philadelphia, PA 19104, USA. e-mail: jwmorgan@pennmedicine.upenn.edu

Received: 2 January 2018

Accepted: 14 May 2018

Published: 22 June 2018

Keywords: adaptive optics; photoreceptors; cone density; split detection

Citation: Morgan JIW, Vergilio GK, Hsu J, Dubra A, Cooper RF. The reliability of cone density measurements in the presence of rods. *Trans Vis Sci Tech.* 2018;7(3):21, <https://doi.org/10.1167/tvst.7.3.21>
Copyright 2018 The Authors

Purpose: Recent advances in adaptive optics scanning light ophthalmoscopy (AOSLO) have enabled visualization of cone inner segments through nonconfocal split-detection, in addition to rod and cone outer segments revealed by confocal reflectance. Here, we examined the interobserver reliability of cone density measurements in both AOSLO imaging modalities.

Methods: Five normal subjects (nine eyes) were imaged along the horizontal and vertical meridians using a custom AOSLO with confocal and nonconfocal split-detection modalities. The resulting images were montaged using a previously described semiautomatic algorithm. Regions of interest (ROIs) were selected from the confocal montage at 190 μm , and from split-detection and confocal montages at 900 and 1800 μm from the fovea. Four observers (three experts, one naïve) manually identified cone locations in each ROI, and these locations were used to calculate bound densities. Intraclass correlation coefficients and Dice's coefficients were calculated to assess interobserver agreement.

Results: Interobserver agreement was high in cone-only images (confocal 190 μm : 0.85; split-detection 900 μm : 0.91; split-detection 1800 μm : 0.89), moderate in confocal images at 900 μm (0.68), and poor in confocal images at 1800 μm (0.24). Excluding the naïve observer data substantially increased agreement within confocal images (190 μm : 0.99; 900 μm : 0.80; 1800 μm : 0.68).

Conclusions: Interobserver measurements of cone density are more reliable in rod-free retinal images. Moreover, when using manual cell identification, it is essential that observers are trained, particularly for confocal AOSLO images.

Translational Relevance: This study underscores the need for additional reliability studies in eyes containing pathology where identifying cones can be substantially more difficult.

Introduction

Since its invention, confocal adaptive optics scanning laser ophthalmoscopy (AOSLO)¹ has enabled the routine visualization of the cone photoreceptor mosaic geometry in normal and diseased retina, noninvasively.² However, it was several years before technical improvements to confocal AOSLO design allowed regular visualization of the rod

mosaic.³ Even more recently, another AOSLO imaging technique has enabled visualization of what is likely the cone inner segment mosaic. This technique, dubbed “nonconfocal split-detection,”⁴ involves collecting nonconfocal light backscattered by the retina that is equally split between two synchronized point detectors. While split-detection imaging has enhanced investigators' view of the cone mosaic, it is as yet unable to show cone inner segments near the fovea or rod inner segments in

most subjects, with the exception of subjects who have very short axial lengths or some subjects with pathological conditions.

With the ability to visualize the photoreceptor mosaic comes the ability to quantify mosaic parameters using numerous metrics.^{5,6} To date, the most routinely used metric for assessing photoreceptor images is cone density and the critical step for its direct measurement is the identification of all cones within a region of interest (ROI). Automated cone identification algorithms, in particular those using local maxima in intensity to identify cone locations,^{7,8} fail to identify only cones in images where rods and cones are interspersed because rods, like cones, appear as bright dots in confocal AOSLO images. Moreover, even manual identification of cones in images with both photoreceptor types became more difficult because graders now need to distinguish whether a photoreceptor in the image is a cone or rod and the multimodal wave-guided reflectance of cones outside of the parafoveal region can appear as a group of small bright dots. Thus, while the ability to image the rods represented a substantial improvement for noninvasive visualization of the retina, it came with the added complication that cone mosaic parameters became difficult to quantify in confocal images containing both rods and cones.

What follows is an investigation of the impact of rod photoreceptor presence on the interobserver reliability of normal manual cone density measurements. We assessed interobserver reliability of cone density measurements from AOSLO images where either only the cone mosaic is visualized (foveal confocal images and perifoveal split-detection images) or the rod and cone mosaics are interleaved (perifoveal confocal images). Understanding the reliability of reported cone density values is essential for characterizing retinal phenotypes and will be valuable as clinical trials testing experimental therapeutics consider incorporating cone density as an outcome measure.

Methods

This study was approved by the institutional review board at the University of Pennsylvania, and followed the tenets of the Declaration of Helsinki. Following explanation of the study, subjects gave informed consent and voluntarily participated.

Nine eyes of five subjects ages 24 to 35 with no known retinal pathology were included in this study. Axial lengths for each subject's eyes were obtained

using an IOL Master (Carl Zeiss Meditec, Dublin, CA). AOSLO image scale was determined by acquiring images of a Ronchi ruling positioned at the focal plane of a lens with a 19-mm focal length to determine the conversion between image pixels and degrees. We then used a proportional axial length method to approximate the retinal magnification factor (in microns/degree) to convert the angular scale to microns/pixel.⁹ Subjects' pupils were cyclopleged with phenylephrine hydrochloride (2.5%) and tropicamide (1%).

The custom AOSLO used in this study has been previously described.^{4,10} A dental impression was used to align the subject to the AOSLO. An 848-nm superluminescent diode with a full-width at half-maximum (FWHM) bandwidth of 26 nm (Superlum, Cork, Ireland) was used for wavefront sensing, and a 97-actuator deformable mirror (Alpao SAS, France) provided the aberration correction. Confocal and nonconfocal split-detection images were acquired simultaneously at 16.7 frames per second over a 1° by 1° field of view using a superluminescent diode centered at 795 nm with FWHM of 15.3 nm (Superlum) and three photomultiplier tubes (PMT; Hamamatsu Corporation, Japan) configured as previously described.⁴

Subjects were instructed to fixate (using the imaged eye) as steadily as possible at a target while the AOSLO image sequences were acquired along all four retinal meridians out to approximately 1800 μm from the fovea. A custom strip-registration algorithm was used for intraframe strip based registration and dewarping of the AOSLO images.¹¹ Reference frames for registration were chosen manually from the confocal image sequence, and 50 frames of the confocal AOSLO images were registered and averaged. The same transformations that were applied to the confocal images were applied to the 50 simultaneously acquired nonconfocal split-detection images, as described in other simultaneous multimodal imaging paradigms.¹² Averaged confocal and nonconfocal split-detection images were then automatically montaged (Fig. 1) using a previously described algorithm.¹³ Square 80, 93, and 100 μm per side ROIs were extracted at 190, 900, and 1800 μm , respectively, from the fovea along all four meridians in the confocal montages and at 900 and 1800 μm in the nonconfocal split-detection montages. When necessary, ROIs were minimally displaced to avoid the shadows of retinal blood vessels. One subject's montage did not extend out to 1800 μm inferior, resulting in a total of 178 ROIs for the study.

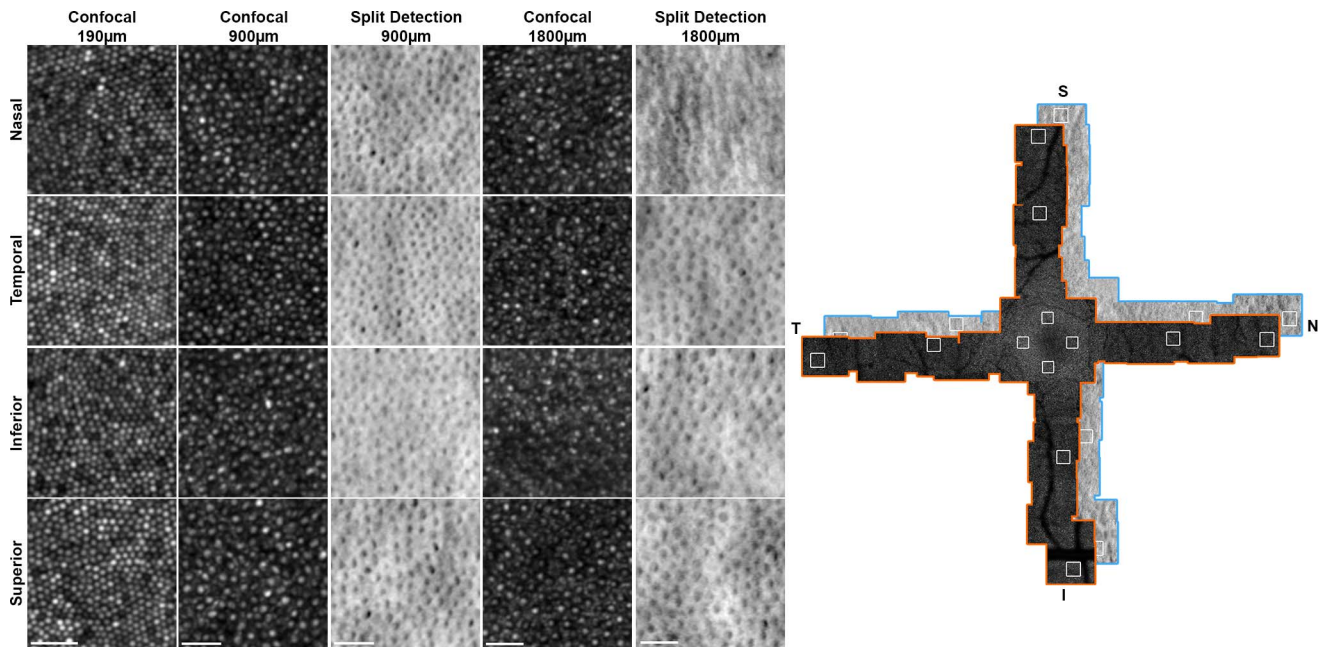


Figure 1. Example confocal and split-detection AOSLO images and montages showing the photoreceptor mosaic in the right eye of subject 11048. Twenty exemplar ROIs used for manual cone identification and cone density analysis are shown in the *left panels*. *White boxes* within the montage on the *right* outline the locations of the ROIs: 190 μm (confocal only), 900 μm and 1800 μm (both confocal and split-detection) along the superior, inferior, nasal, and temporal meridians. *Scale bars* for the ROI images: 25 μm .

Three expert graders and one naïve grader manually identified cone locations once per ROI. The naïve grader was instructed on the use of the custom software and was verbally instructed on how to make cone selections using a test set of images separate from the images included in this study. The order in which the 178 ROIs were presented was randomized, and graders were masked to the subject, eye, and meridian. Graders were able to adjust the image brightness and contrast in both linear and logarithmic displays to aid in determining the presence of cone photoreceptors. Graders were instructed to mark cell centers by manually clicking on each cell. These locations were used to determine the Voronoi mosaic and only cone locations whose Voronoi cells were fully contained within the ROI were used for the cone density calculation. Bound cone density was then calculated by dividing the number of bound Voronoi cells by their area.⁵

We undertook two different analyses to assess interobserver agreement for cone density measurements and cone identifications. First, interobserver agreement in cone density measurements was assessed at 190 for confocal images and at 190, 900, and 1800 μm for both confocal and nonconfocal split-detection images using intraclass correlation (ICC) coefficients with 95% confidence intervals (CIs). Agreement was

assessed between the three expert reviewers, as well as between the three expert and one naïve graders. Agreement between cone density measurements from confocal and nonconfocal split-detection images at identical retinal locations were also compared between observers. A paired *t*-test was used to assess whether cone density in split-detection images was significantly different from cone density in confocal images for each observer.

Second, we compared the sensitivity and precision of cone identifications between pairs of observers for each image. We found observer co-located cones using the following method: First, we determined the average and standard deviation of the nearest neighbor distance across all coordinates from each expert observer within each image. We then clustered cone locations across all observers within each image by grouping cone locations that were located less than the mean nearest neighbor distance plus two standard deviations of the nearest neighbor distance of a given cone. Only one cone selection per observer was allowed in each cluster. We then assessed the similarity between cone identifications from each of the other observers independently. To do this, we considered one expert observer to be “ground truth” and found the number of true positives (denoted N_{TP}), false positives (comparison observer identified a cone when the “ground

truth” expert observer did not, denoted N_{FP}), and the number of false negatives (comparison observer did not identify a cone when the “ground truth” expert observer did, denoted N_{FN}) for each of the other observers, including only cones with bound Voronoi areas. Thus, the number of cone identifications made from each observer can be expressed as:

$$N_{comparison\ observer} = N_{TP} + N_{FP}$$

$$N_{ground\ truth\ expert} = N_{TP} + N_{FN}$$

In order to compare data sets from different observers, we then calculated the true positive rate, the false discovery rate, and Dice’s coefficient for each image as:

$$true\ positive\ rate = N_{TP}/N_{ground\ truth\ expert}$$

$$false\ discovery\ rate = N_{FP}/N_{comparison\ observer}$$

$$Dice's\ coefficient = 2N_{TP}/(N_{ground\ truth\ expert} + N_{ground\ truth\ expert})$$

where Dice’s coefficient is used as a metric for describing the similarity between two data sets.^{14–16} We considered all permutations between graders where each expert grader was considered the ground truth and all other observers’ cone identifications were compared to that expert’s identifications. (For example, first considering expert observer 1 as ground truth and comparing observer 2 to 1, 3 to 1, and 4 to 1. Then, considering expert observer 2 as ground truth, and comparing 1 to 2, 3 to 2, and 4 to 2, etc.) We did not perform an analysis considering the naïve observer’s cone identifications as ground truth.

Results

Cone density measurements by the three expert observers showed a range of agreement (Fig. 2). Confocal images of the parafoveal cone mosaic (190 μ m ROIs) demonstrated the highest agreement between expert observers when compared to other retinal eccentricities (Table 1). Measurements of cone density from split-detection images showed higher agreement between expert observers than measurements of cone density from confocal images at the same locations.

Using the scale described by Cicchetti¹⁷ (>0.75, excellent; 0.6–0.74, good; 0.4–0.59 fair; <0.4 poor), expert interobserver agreement was excellent in cone-only images (confocal OS 190 μ m: 0.99, CI: 0.98–1.0; split-detection IS 900 μ m: 0.95, CI: 0.91–0.97; split-

detection IS 1800 μ m: 0.89, CI: 0.82–0.94). Interobserver agreement was also excellent in confocal OS images at 900 μ m and good in confocal OS images at 1800 μ m (900 μ m: 0.80, CI: 0.68–0.88; 1800 μ m: 0.68, CI: 0.52–0.82). Observers 1 and 3 exhibited a significant difference toward a higher density (640 and 920 cones/mm², $P = 0.005$ and $P < 0.001$, respectively) in confocal OS images compared to split-detection IS images at the same location. Conversely, observer 2 showed a significant difference toward a higher density (550 cones/mm², $P < 0.001$) in split-detection IS images compared to confocal images at the same location (Figs. 3, 4). This corresponds to 5.3 and 6.5 more bound manual cone identifications over an ROI in confocal images in comparison to split-detection images for observers 1 and 3 respectively, and 4.7 more bound manual cone identifications over an ROI in split-detection images in comparison to confocal images for observer 2.

Dice’s coefficients between the expert observers showed high similarity on average for all eccentricities for both confocal and split-detection modalities (Table 2). On average Dice’s coefficients for confocal images were highest at the 190- μ m eccentricity (0.971), and decreased with increasing eccentricity for all expert observers (900 μ m, 0.925; 1800 μ m, 0.906). The standard deviation for Dice’s coefficients over all images were also higher at perifoveal retinal locations in comparison to 190 μ m. Dice’s coefficients for 900 μ m split-detection images were lower than those at 1800 μ m (900 μ m, 0.883; 1800 μ m, 0.916) and were also lower than confocal images.

Including cone density measurements made by the naïve observer with those from the experts decreased the interobserver agreement at all eccentricities (Table 3). In this case, interobserver agreement was excellent in cone-only images (confocal 190 μ m: 0.85, CI: 0.76–0.91; split-detection 900 μ m: 0.91, CI: 0.85–0.95; split-detection 1800 μ m: 0.89, CI: 0.82–0.94), good in confocal images at 900 μ m (0.68, CI: 0.55–0.80), and poor in confocal images at 1800 μ m (0.24, CI: 0.09–0.44). The naïve observer showed a significant difference ($P < 0.0001$) toward identifying more cones in split-detection images than in confocal images and showed higher variability in agreement between cone density measurements made in paired confocal and split-detection images at the same retinal location in comparison to expert observers (Fig. 4, bottom right). In comparison to the expert observers, Dice’s coefficients were reduced for the naïve observer at all eccentricities for confocal images (190 μ m,

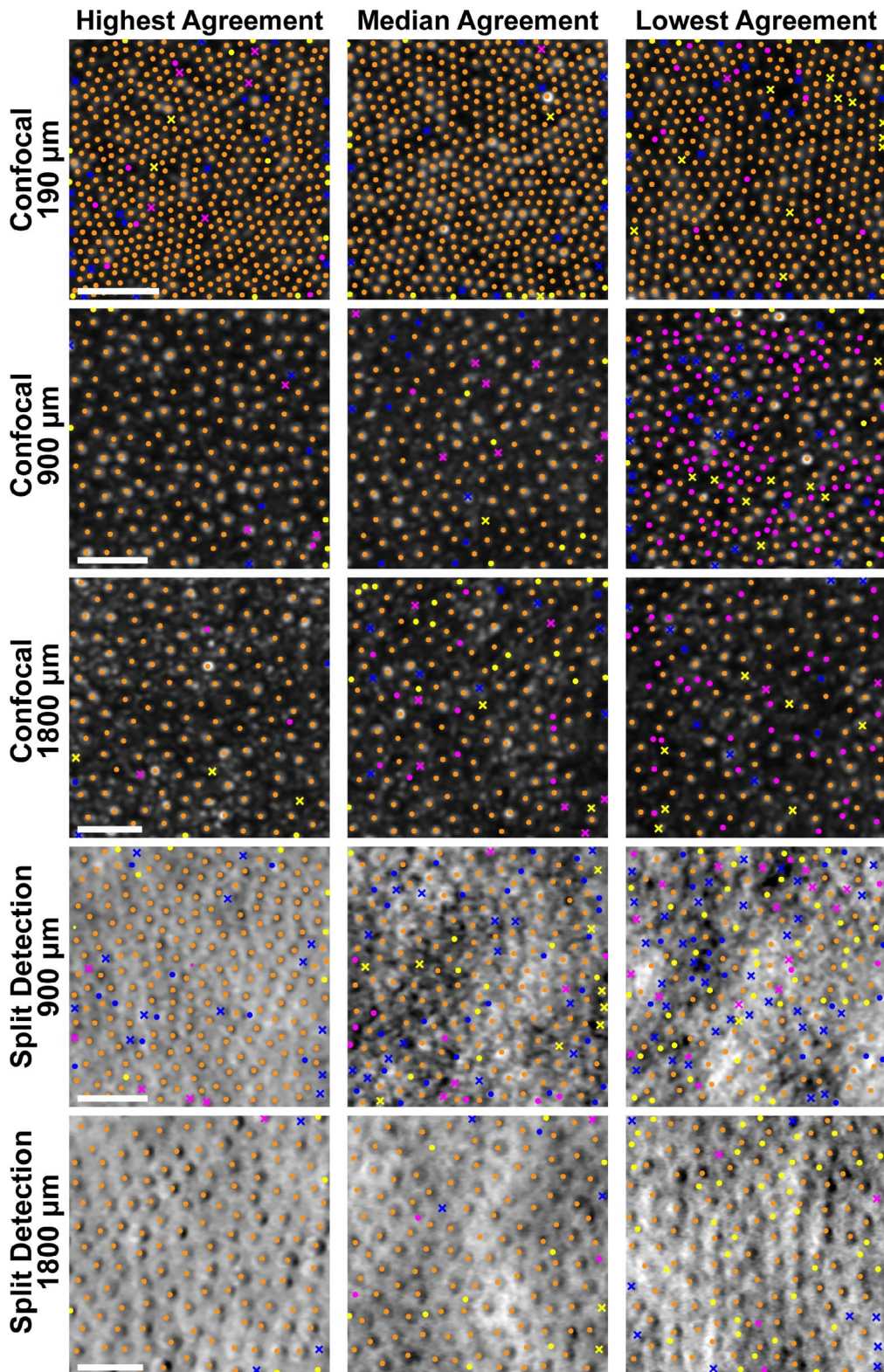


Figure 2. The range in agreement between cone density measurements (highest, median, and lowest) made by the three expert observers for all five ROI types: confocal 190 μm , confocal 900 μm , confocal 1800 μm , split-detection 900 μm , and split-detection 1800 μm . Manually identified cones identified by all three expert observers are marked in *orange*. Locations identified as a cone by only one of the three expert observers are denoted by a *colored dot*. Observers 1, 2, and 3 are *yellow*, *blue*, and *magenta*, respectively. Locations identified as cones by two of the three expert observers are denoted by an *X*, with a color corresponding to the observer who *did not*

identify that location as a cone. Confocal images at 190 μm showed the highest agreement for cone density measurements between expert observers. Cone density measurements made using split-detection images showed a higher interobserver agreement than confocal images at the same retinal eccentricity. The lowest agreement confocal image at 900 μm corresponds to the outlier in Figure 4C, where it is likely that observer 3 misidentified rods as cones in this confocal image. *Scale bars: 25 μm .*

0.941; 900 μm , 0.859; 1800 μm , 0.872) but not for split detection (900 μm , 0.808; 1800 μm , 0.910; Table 2).

Discussion

Our high interobserver agreement for confocal images at 190 μm is consistent with previous reports in normal cone-only parafoveal images.^{18,19} Our results in confocal images show that while interobserver reliability is considered good or excellent at all retinal locations studied, there is a measurable decrease with eccentricity (Table 1). We attribute part, but not all, of this decrease to the presence of rods in perifoveal images because interobserver reliability for split-detection images (presumed to show cones not rods) was higher than interobserver reliability for paired confocal images. However, the decrease in interobserver reliability with eccentricity is not entirely based on infiltration of rods to the images because interobserver reliability for split-detection images was lower than for cone density measurements in 190 μm confocal images and also decreased with eccentricity, though to a lesser extent than in confocal images.

The true positive rate, false discovery rate, and Dice's coefficient also showed differences with retinal eccentricity and imaging modality. As would be expected based on the ICC agreement between cone densities, Dice's coefficient and the true positive rate decreased with increasing eccentricity while the false discovery rate increased with eccentricity for confocal images. Surprisingly, Dice's coefficient was higher for 900 μm confocal images than for 900 μm split-detection images for comparisons involving expert observer 2, despite the fact that cone density measurements showed higher interobserver agreement in split-detection images compared to confocal images. This result could occur if cell selections made by expert observer 2 were shifted relative to the cell selections made by expert observers 1 and 3, resulting in pairs of "false positives" and "false negatives" that would both fall outside of the clustering area used for determining locations considered to have one-to-one pairing. Such pairs of "false positives" and "false negatives" would cause Dice's coefficient to be

reduced while not affecting cone density measurements.

The size of an ROI is known to influence cone density.^{8,20} The ROI sizes selected for this study were chosen based on the method described by Cooper et al.⁵ and increased with eccentricity such that ROIs at 900 and 1800 μm would contain close to 100 cones, while the parafoveal ROI, which is known to exhibit tight triangular packing, would contain a large enough continuous area of the mosaic to capture subtle variations in cone packing. On average over all observers, images at 190 μm had 409 bound cones identified within the ROI in comparison to 139 bound cones at 900 μm and 91 bound cones at 1800 μm . Thus, a single cone misidentification at the perifoveal eccentricities represents a greater percentage of total cones selected and thus, could have a larger impact on interobserver agreement and Dice's coefficient than at the parafoveal eccentricity.

Our results showed reduced reliability between observers using both ICC and Dice's coefficient when cone identifications from the naïve observer were considered. The naïve observer, new to the cone identification process, had only spent a couple of months working with adaptive optics retinal images. In contrast, the expert observers all had several years of experience in the field, and specifically, identifying cones in confocal images. In addition, all three expert reviewers had been exposed to nonconfocal split-detection imaging of cone inner segments for at least a year. However, only one of the expert reviewers had previous experience identifying cones in this imaging modality. Thus, the graders overall were less exposed to split-detection than confocal images. Given that the interobserver agreement for density was higher in

Table 1. Agreement in Cone Density Measurements Made by Three Expert Observers

Eccentricity (μm)	Image Type	ICC	
		Coefficient	CI Range
190	Confocal	0.99	0.98–1.0
900	Confocal	0.80	0.68–0.88
	Split-detection	0.95	0.91–0.97
1800	Confocal	0.68	0.52–0.81
	Split-detection	0.89	0.82–0.94

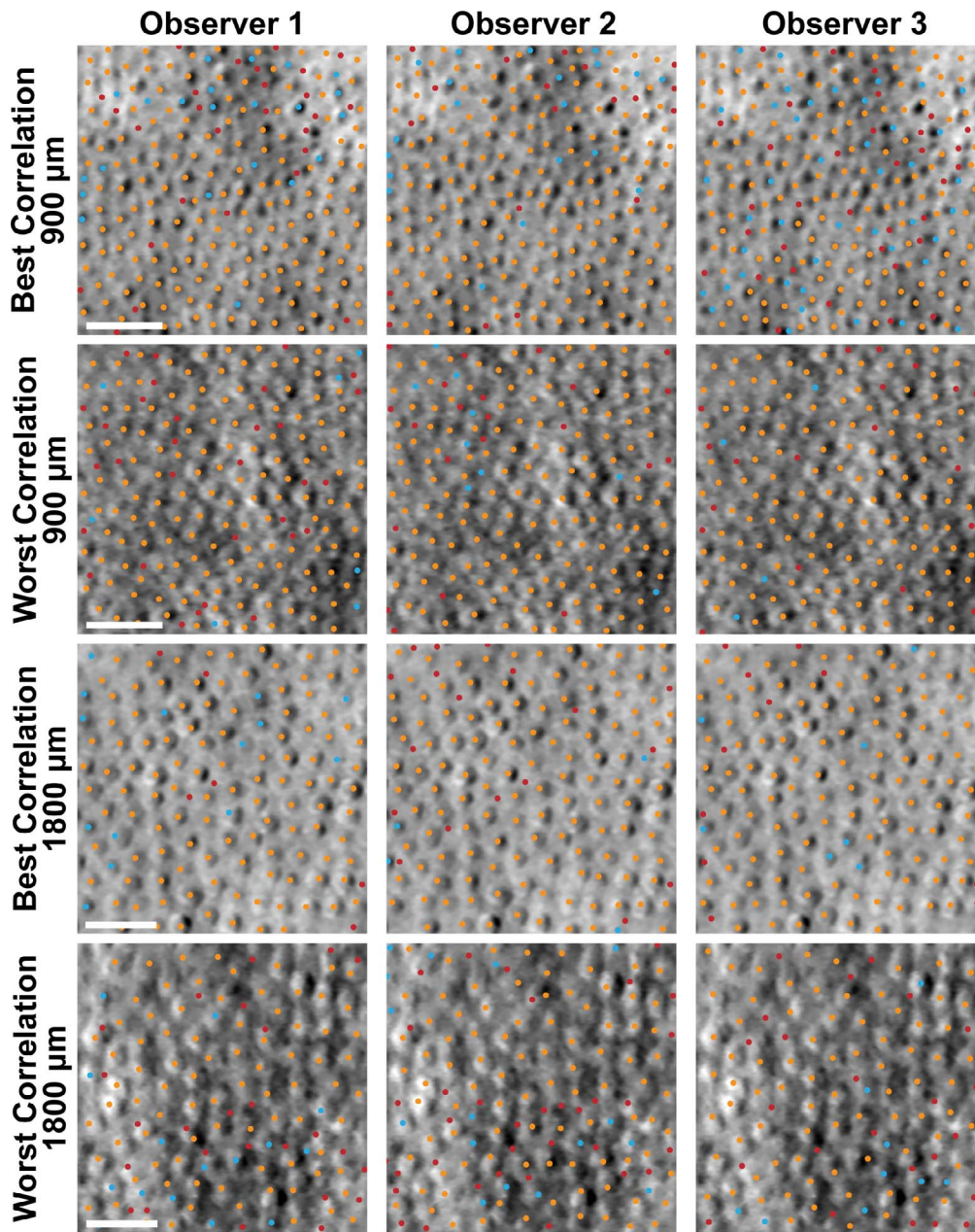


Figure 3. Cone identifications from paired confocal and split-detection images showing the highest and lowest correlation on average between cone density measurements overlaid on split-detection images for each of the three expert observers. *Orange dots* show locations where an expert observer identified a cone in both the split-detection and confocal image. *Red and blue dots* show locations where a cone was identified in either the confocal or split-detection image only (*red*, confocal; *blue*, split-detection). *Scale bars:* 25 μm .

split-detection images than the paired confocal images (both including and excluding the results of the naïve observer), this may show that observers require less training to agree on cone densities in the split-detection imaging modality as compared to the confocal imaging modality. Alternatively, the higher agreement could arise in part from the lack of visible rods in the split-detection images, since the confocal

images at the 190 μm ROI also showed better interobserver agreement (even when including the naïve observer) than the confocal images at greater eccentricities. Regardless, these results highlight the need of training for performing manual AOSLO image analysis, especially as AO imaging becomes more accessible and reading centers are developed for interpreting AO data sets.

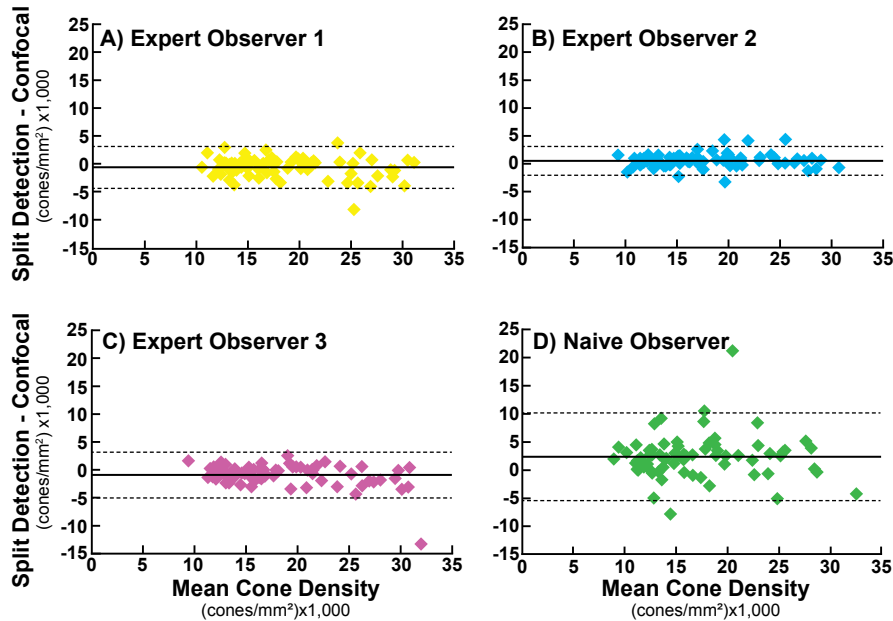


Figure 4. Bland-Altman plots for each observer show the difference between split-detection and confocal cone densities versus the mean of the paired densities for paired ROIs at 900 and 1800 μm . Observers 1 (A) and 3 (C) measured a higher cone density in confocal images (640 and 920 cones/ mm^2 , $P = 0.005$ and $P < 0.001$, respectively) while observer 2 (B) measured a higher cone density in split detection images (550 cones/ mm^2 , $P < 0.001$). Post-hoc evaluation of the outlier points in observers 1 and 3 were caused by erroneous cone identifications in confocal images (selection of rods) more often than erroneous cone identification in split-detection images. Observer 4 (D, naïve observer) measured a higher cone density in split detection images (2300 cones/ mm^2 , $P < 0.0001$) and showed more variability in agreement between confocal and split-detection measures of cone density at the same retinal locations.

Several automated algorithms for identifying normal cones in parafoveal (rod-free) confocal images^{7,8,21–23} and split-detection images^{14,23–25} have shown promise. However, none of these algorithms have demonstrated the ability to reliably identify cones separate from rods. As a result, confocal images at eccentricities beyond the parafovea required manual cone identifications for density analysis. Our goal for the present study was to understand how eccentricity and imaging modality, not cone identification algorithms, effect cone density reliability. For this reason, we chose to grade all images manually, rather than use semiautomated algorithms in some conditions and fully manual grading in others. Until reliable algorithms are developed for all retinal eccentricities in both normal and diseased images, manual or semiautomatic grading of cone locations will remain necessary, and limitations inherent to these methods, such as the reliability between graders, will remain present in cone mosaic measurements.

Both expert observers 1 and 3 each had one outlier image when comparing cone density measurements made from confocal versus split-detection images (Figs. 4A, 4C). Upon re-examination, it was apparent

that both outlier data points were caused by marking more rods as cones in the confocal image rather than missing cones that should have been marked in the split-detection image (Fig. 2, confocal 900 μm , lowest agreement). These types of errors again point to a yet unmet need for fully automated cone identification algorithms that utilize both confocal and split-detection image features for identifying cone locations. Until such time, one could consider allowing graders to identify cones in confocal and split-detection images of the same retinal location simultaneously or to review cone identifications prior to unmasking results to attempt to alleviate errors caused by manual grading.

It is important to note that these data are likely a best-case: images from patients with retinal disease are typically of lower quality than the high-quality images from young normal controls that were included in our study. Therefore, image interpretation can be substantially more difficult in images of pathology compared with normal anatomy, and reports of intergrader agreement are lower in images of pathology compared with normal.^{26,27} Interestingly, Tanna et al.²⁶ show higher reliability for cone

Table 2. Similarity in Cone Identifications Made by Three Expert Observers and 1 Naïve Observer

Observer	Modality	Eccentricity (μm)	True Positive Rate Mean (SD)	False Discovery Rate Mean (SD)	Dice's Coefficient Mean (SD)
Expert 1 as "ground truth"					
Expert 2	Confocal	190	0.955 (0.015)	0.017 (0.016)	0.968 (0.010)
		900	0.870 (0.069)	0.014 (0.012)	0.922 (0.042)
		1800	0.860 (0.084)	0.030 (0.029)	0.909 (0.045)
	Split-detection	900	0.841 (0.062)	0.126 (0.082)	0.856 (0.063)
		1800	0.866 (0.090)	0.052 (0.069)	0.903 (0.074)
		1800	0.866 (0.090)	0.052 (0.069)	0.903 (0.074)
Expert 3	Confocal	190	0.959 (0.016)	0.021 (0.015)	0.969 (0.011)
		900	0.911 (0.055)	0.055 (0.058)	0.925 (0.041)
		1800	0.881 (0.065)	0.074 (0.061)	0.900 (0.039)
	Split-detection	900	0.922 (0.048)	0.053 (0.036)	0.933 (0.034)
		1800	0.904 (0.058)	0.037 (0.028)	0.932 (0.038)
		1800	0.904 (0.058)	0.037 (0.028)	0.932 (0.038)
Naïve	Confocal	190	0.890 (0.070)	0.009 (0.006)	0.936 (0.041)
		900	0.768 (0.125)	0.033 (0.041)	0.848 (0.084)
		1800	0.874 (0.078)	0.086 (0.060)	0.892 (0.062)
	Split-detection	900	0.730 (0.125)	0.081 (0.112)	0.801 (0.081)
		1800	0.879 (0.080)	0.047 (0.047)	0.913 (0.057)
		1800	0.879 (0.080)	0.047 (0.047)	0.913 (0.057)
Expert 2 as "ground truth"					
Expert 1	Confocal	190	0.983 (0.016)	0.045 (0.015)	0.968 (0.010)
		900	0.986 (0.012)	0.130 (0.069)	0.922 (0.042)
		1800	0.970 (0.029)	0.140 (0.084)	0.909 (0.045)
	Split-detection	900	0.874 (0.082)	0.159 (0.062)	0.856 (0.063)
		1800	0.948 (0.069)	0.134 (0.090)	0.903 (0.074)
		1800	0.948 (0.069)	0.134 (0.090)	0.903 (0.074)
Expert 3	Confocal	190	0.980 (0.015)	0.027 (0.012)	0.977 (0.010)
		900	0.973 (0.024)	0.109 (0.070)	0.928 (0.042)
		1800	0.946 (0.038)	0.121 (0.069)	0.909 (0.042)
	Split-detection	900	0.866 (0.080)	0.143 (0.056)	0.860 (0.063)
		1800	0.928 (0.075)	0.097 (0.080)	0.914 (0.072)
		1800	0.928 (0.075)	0.097 (0.080)	0.914 (0.072)
Naïve	Confocal	190	0.911 (0.079)	0.014 (0.008)	0.945 (0.045)
		900	0.832 (0.103)	0.071 (0.055)	0.873 (0.057)
		1800	0.833 (0.095)	0.164 (0.072)	0.833 (0.080)
	Split-detection	900	0.779 (0.102)	0.120 (0.133)	0.815 (0.077)
		1800	0.902 (0.101)	0.106 (0.100)	0.897 (0.097)
		1800	0.902 (0.101)	0.106 (0.100)	0.897 (0.097)
Expert 3 as "ground truth"					
Expert 1	Confocal	190	0.979 (0.015)	0.041 (0.016)	0.969 (0.011)
		900	0.945 (0.058)	0.089 (0.055)	0.925 (0.041)
		1800	0.926 (0.061)	0.119 (0.065)	0.900 (0.039)
	Split-detection	900	0.947 (0.036)	0.078 (0.048)	0.933 (0.034)
		1800	0.963 (0.028)	0.096 (0.058)	0.932 (0.038)
		1800	0.963 (0.028)	0.096 (0.058)	0.932 (0.038)
Expert 2	Confocal	190	0.973 (0.012)	0.020 (0.015)	0.977 (0.009)
		900	0.891 (0.070)	0.027 (0.024)	0.928 (0.042)
		1800	0.879 (0.069)	0.054 (0.038)	0.909 (0.042)
	Split-detection	900	0.857 (0.056)	0.134 (0.080)	0.860 (0.063)
		1800	0.903 (0.080)	0.072 (0.075)	0.914 (0.072)
		1800	0.903 (0.080)	0.072 (0.075)	0.914 (0.072)

Table 2. Continued

Observer	Modality	Eccentricity (μm)	True Positive Rate Mean (SD)	False Discovery Rate Mean (SD)	Dice's Coefficient Mean (SD)
Naïve	Confocal	190	0.904 (0.076)	0.014 (0.009)	0.941 (0.043)
		900	0.784 (0.117)	0.045 (0.042)	0.855 (0.072)
		1800	0.886 (0.075)	0.100 (0.066)	0.892 (0.064)
	Split-detection	900	0.749 (0.111)	0.095 (0.126)	0.807 (0.082)
		1800	0.913 (0.070)	0.070 (0.062)	0.921 (0.058)

density measurements in split-detection images versus paired confocal images for patients with Stargardt disease and retinitis pigmentosa GTPase regulator associated retinopathy, consistent with our results that cone identification is less difficult in split-detection images. More generally, it is important that investigators understand how a given disease affects the reliability of their cone density measurements, especially if those measurements are to be used for assessment of disease progression or as an outcome measure to determine the effect of interventions on cone morphology. This could be accomplished by establishing reliability of cone metrics for each disease, which are then applied to future studies of that disease. Alternatively, studies using cone density or other mosaic metrics could document reliability of grading within the study.

In summary, we show it is important to consider both AOSLO image modality and retinal eccentricity when evaluating the reliability of cone metrics. Studies, such as the present one that strive to understand the confidence with which cone density measurements are reported, will be required before cone density measurements can be used as a reliable biomarker for disease phenotype and progression monitoring or as an outcome measure to assess disease intervention.

Table 3. Agreement in Cone Density Measurements Made by Three Expert Observers and 1 Naïve Observer

Eccentricity (μm)	Image Type	ICC	
		Coefficient	CI Range
190	Confocal	0.85	0.76–0.91
	Split-detection	0.91	0.85–0.95
900	Confocal	0.68	0.55–0.80
	Split-detection	0.91	0.85–0.95
1800	Confocal	0.24	0.09–0.44
	Split-detection	0.89	0.82–0.94

Acknowledgments

We thank Joseph Carroll for providing cone identification software and for helpful discussions, and Gui-shuang Ying for statistical advice.

Supported by NIH R01EY028601; NIH U01EY025477; NIH P30 EY001583; Foundation Fighting Blindness; Research to Prevent Blindness Stein Innovation Award; the F. M. Kirby Foundation; and the Paul and Evanina Mackall Foundation Trust.

Disclosure: **J.I.W. Morgan**, P (US Patent 8226236), F (AGTC); **G.K. Vergilio**, None; **J. Hsu**, None; **A. Dubra**, P (US Patent 8226236), C (Boston Micro-machines Corporation and Meira Gtx); **R.F. Cooper**, None

References

- Roorda A, Romero-Borja F, Donnelly WJ III, Queener H, Hebert TJ, Campbell MCW. Adaptive optics scanning laser ophthalmoscopy. *Opt Express*. 2002;10:405–412.
- Morgan JI. The fundus photo has met its match: optical coherence tomography and adaptive optics ophthalmoscopy are here to stay. *Ophthalmic Physiol Opt*. 2016;36:218–239.
- Dubra A, Sulai Y, Norris JL, et al. Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope. *Biomed Opt Express*. 2011;2:1864–1876.
- Scoles D, Sulai YN, Langlo CS, et al. In vivo imaging of human cone photoreceptor inner segments. *Invest Ophthalmol Vis Sci*. 2014;55:4244–4251.

5. Cooper RF, Wilk MA, Tarima S, Carroll J. Evaluating descriptive metrics of the human cone mosaic. *Invest Ophthalmol Vis Sci.* 2016;57:2992–3001.
6. Litts KM, Cooper RF, Duncan JL, Carroll J. Photoreceptor-based biomarkers in AOSLO retinal imaging. *Invest Ophthalmol Vis Sci.* 2017;58: BIO255–BIO267.
7. Li KY, Roorda A. Automated identification of cone photoreceptors in adaptive optics retinal images. *J Opt Soc Am A.* 2007;24:1358–1363.
8. Garrioch R, Langlo C, Dubis AM, Cooper RF, Dubra A, Carroll J. Repeatability of in vivo parafoveal cone density and spacing measurements. *Optom Vis Sci.* 2012;89:632–643.
9. Bennett AG, Rudnicka AR, Edgar DF. Improvements on Littmann's method of determining the size of retinal features by fundus photography. *Graefes Arch Clin Exp Ophthalmol.* 1994;32:361–367.
10. Dubra A, Sulai Y. Reflective afocal broadband adaptive optics scanning ophthalmoscope. *Biomed Opt Express.* 2011;2:1757–1768.
11. Dubra A, Harvey Z. Registration of 2D images from fast scanning ophthalmic instruments. *Lecture Notes Computer Sci.* 2010;6204:60–71.
12. Morgan JIW, Dubra A, Wolfe R, Merigan WH, Williams DR. In vivo autofluorescence imaging of the human and macaque retinal pigment epithelial cell mosaic. *Invest Ophthalmol Vis Sci.* 2009;50:1350–1359.
13. Chen M, Cooper RF, Han GK, Gee J, Brainard DH, Morgan JI. Multi-modal automatic mounting of adaptive optics retinal images. *Biomed Opt Express.* 2016;7:4899–4918.
14. Cunefare D, Cooper RF, Higgins B, et al. Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images. *Biomed Opt Express.* 2016;7:2036–2050.
15. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26: 297–302.
16. Srensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* 1948;5:1–34.
17. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessments.* 1994;6:284–290.
18. Liu BS, Tarima S, Visotcky A, et al. The reliability of parafoveal cone density measurements. *Br J Ophthalmol.* 2014;98:1126–1131.
19. Bidaut Garnier M, Flores M, Debellemanniere G, et al. Reliability of cone counts using an adaptive optics retinal camera. *Clin Experiment Ophthalmol.* 2014;42:833–840.
20. Lombardo M, Serrao S, Ducoli P, Lombardo G. Influence of sampling window size and orientation on parafoveal cone packing density. *Biomed Opt Express.* 2013;4:1318–1331.
21. Chiu SJ, Lokhnygina Y, Dubis AM, et al. Automatic cone photoreceptor segmentation using graph theory and dynamic programming. *Biomed Opt Express.* 2013;4:924–937.
22. Bukowska DM, Chew AL, Huynh E, et al. Semi-automated identification of cones in the human retina using circle Hough transform. *Biomed Opt Express.* 2015;6:4676–4693.
23. Cunefare D, Fang L, Cooper RF, Dubra A, Carroll J, Farsiu S. Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Sci Rep.* 2017;7:6620.
24. Liu J, Jung H, Dubra A, Tam J. Automated photoreceptor cell identification on nonconfocal adaptive optics images using multiscale circular voting. *Invest Ophthalmol Vis Sci.* 2017;58:4477–4489.
25. Bergeles C, Dubis AM, Davidson B, et al. Unsupervised identification of cone photoreceptors in non-confocal adaptive optics scanning light ophthalmoscope images. *Biomed Opt Express.* 2017;8:3081–3094.
26. Tanna P, Kasilian M, Strauss R, et al. Reliability and repeatability of cone density measurements in patients with Stargardt disease and RPGR-associated retinopathy. *Invest Ophthalmol Vis Sci.* 2017;58:3608–3615.
27. Abozaid MA, Langlo CS, Dubis AM, Michaelides M, Tarima S, Carroll J. Reliability and repeatability of cone density measurements in patients with congenital achromatopsia. *Adv Exp Med Biol.* 2016;854:277–283.