

# Report From the NEI/FDA Endpoints Workshop on Age-Related Macular Degeneration and Inherited Retinal Diseases

Karl Csaky,<sup>1</sup> Frederick Ferris III,<sup>2</sup> Emily Y. Chew,<sup>2</sup> Prashant Nair,<sup>3</sup> Janet K. Cheetham,<sup>4</sup> and Jacque L. Duncan<sup>5</sup>

<sup>1</sup>Retina Foundation of the Southwest, Dallas, Texas, United States

<sup>2</sup>National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States

<sup>3</sup>Washington, District of Columbia, United States

<sup>4</sup>Foundation Fighting Blindness, Columbia, Maryland, United States

<sup>5</sup>Department of Ophthalmology, University of California, San Francisco, California, United States

Correspondence: Jacque L. Duncan, University of California, San Francisco, 10 Koret Way, K113, San Francisco, CA, USA; jacque.duncan@ucsf.edu.

Invited speakers and discussants are listed on page 3457.

Submitted: June 1, 2017

Accepted: June 2, 2017

Citation: Csaky K, Ferris F III, Chew EY, Nair P, Cheetham JK, Duncan JL. Report from the NEI/FDA Endpoints Workshop on age-related macular degeneration and inherited retinal diseases. *Invest Ophthalmol Vis Sci.* 2017;58:3456–3463. DOI:10.1167/iovs.17-22339

The Association for Research in Vision (ARVO), together with the Foundation Fighting Blindness (FFB), organized the National Eye Institute (NEI)–Food and Drug Administration (FDA) endpoints workshop on age-related macular degeneration (AMD) and inherited retinal diseases (IRD) on November 9, 2016. The public workshop convened basic researchers, clinicians, and regulatory authorities in an interactive forum to discuss the latest clinical findings on AMD and IRD with the goal of optimizing clinical trial design and developing trial endpoints.

Specifically, the aim of the workshop was to assemble stakeholders from academic and regulatory spheres to discuss clinical data on potential primary and secondary clinical trial endpoints; patient stratification for disease monitoring, interventions, and evaluation of treatment response; and identification of future research avenues. The workshop was jointly organized by Dr. Frederick Ferris III, MD, Clinical Director at the NEI; Dr. Emily Chew, MD, Deputy Clinical Director at the NEI; Dr. Karl Csaky, MD, PhD, Managing Director and Head of Molecular Ophthalmology Laboratory at the Retina Foundation of the Southwest; Dr. Jacque L. Duncan, MD, Professor of Clinical Ophthalmology at the University of California, San Francisco; and Janet K. Cheetham, PharmD, from the FFB. The workshop format was a series of presentations followed by panel discussions, in which FDA representatives offered regulatory perspectives on potential endpoints for clinical trial design. The following key topics were discussed.

## Structural Endpoints in Geographic Atrophy

Dr. Frank Holz provided data on the utility of fundus autofluorescence (FAF) in detecting loss of the retinal pigment epithelium (RPE). He presented additional studies demonstrating that RPE loss detected on FAF correlates with photoreceptor loss observed using spectral-domain optical coherence tomography (SD-OCT). Geographic atrophy (GA) lesions are often complex and multifocal, and Dr. Holz described semi-

automated image processing software to detect and quantify such lesions.<sup>1</sup>

Dr. Srinivas Sadda summarized the advantages of SD-OCT for assessing GA area, namely, precision, reproducibility, strong correlation with function, safety, and patient comfort. Additionally, frequent assessments are possible, potentially increasing measurement precision and reducing the duration of and sample sizes needed in clinical trials. Further, he presented data showing a significant correlation between SD-OCT and FAF for imaging GA, as well as OCT data that have been shown to correlate with functional deficits in patients with GA.

Additionally, Dr. Sadda discussed en face OCT, noting that by restricting the en face OCT images to include only the choroid, the contrast needed for delineating GA can be greatly increased. Delineating GA in this manner facilitates quantification through automated techniques, such as instrument software that can automatically generate a map of atrophy based on hypertransmission data. Any segmentation errors that result from automated segmentation can be corrected by manual inspection of en face images and OCT B-scans at reading centers. He noted that choroidal hypertransmission can be used not only to define GA boundaries but also to track GA progression over time in a reproducible and automated manner, even by graders with little experience; in contrast, most autofluorescence-based tools for GA segmentation still require some user input. However, he noted that challenges remain, including questions about the specificity of OCT based on choroidal hypertransmission. Reflecting these concerns, he summarized the proposed consensus definition of GA using OCT from retinal experts that includes a region of hypertransmission of at least 250  $\mu\text{m}$ ; a zone of disrupted or attenuated RPE (with associated basal laminar deposit) of at least 250  $\mu\text{m}$ ; signs of overlying photoreceptor degeneration with outer nuclear layer (ONL) thinning, loss of external limiting membrane, and loss of ellipsoid and interdigitation zones; and absence of features suggestive of an RPE tear. Dr. Sadda noted that the size threshold of 250  $\mu\text{m}$  was chosen on the basis of pilot studies



aimed at determining the minimum area required for repeatable measurements of GA by graders using OCT.

Dr. Emily Chew noted that color fundus photography (CFP) and FAF have been widely used to measure the growth of GA in patients' eyes over time in clinical studies. However, there is significant variability in measurements. Dr. Chew described the grading of GA on CFP images in the Age-Related Eye Disease Study (AREDS). She noted that at least two of the following criteria on CFP should be met for categorizing an image as GA: circular or oval shapes of lesions, sharp and well-defined borders, absence of RPE, and a diameter greater than 430  $\mu\text{m}$ . Through manual computer planimetry, the area of the lesion is measured, and the change in lesion area over time can be reliably determined. The criteria used for grading FAF images for GA include homogeneous, black lesions with well-defined borders, and a minimum diameter of 430  $\mu\text{m}$ . The presence of a halo surrounding the lesion is included in the AREDS2 grading.

Dr. Chew also described the use of slope analysis to monitor change in GA progression on CFP and FAF images from a range of studies. This approach is based on the reasoning that successful interventions would help change the slope of GA growth significantly, rendering such slope changes a potential outcome measure for monitoring GA progression. Finally, Dr. Chew described efforts to develop the AMD severity scale, similar to the Early Treatment Diabetic Retinopathy Study (ETDRS) scale, which is used for diabetic retinopathy. Such a scale might help evaluate systemic treatments for both eyes.<sup>2</sup>

Dr. Philip Rosenfeld discussed preliminary findings from a phase II clinical trial in which drusen volume was explored as an endpoint for preventing the progression of exudative AMD in eyes with drusen only. For AMD, early-stage approaches based on anatomic endpoints might include monitoring the progression of disease in drusen-only eyes to exudative AMD, progression of drusen to nascent GA and/or GA (the AREDS has shown that most GA arises from preexisting drusen), progression on the AREDS severity scale, change in drusen area and/or volume, and changes in the anatomy of retinal layers and choroid. Dr. Rosenfeld described an SD-OCT imaging approach using the Cirrus instrument that allows the quantitative assessment of drusen volume and area changes over time by constructing RPE elevation maps. In eyes with drusen only, he presented data that indicated that the volume of drusen gradually increases, leading to either exudative neovascular AMD or GA. However, in some rare cases, drusen volume gradually and spontaneously decreases, with no apparent sequelae.<sup>3</sup> Next, Dr. Rosenfeld reported the findings of a retrospective study of 89 patients with dry AMD in one eye and neovascular AMD in the fellow eye. That study revealed that eyes with a baseline drusen volume greater than or equal to 0.03  $\text{mm}^3$  were more likely to develop GA and neovascularization at 12 months and 24 months.<sup>4</sup> These findings lend support for considering a composite endpoint in clinical trials. Such an endpoint would test whether an intervention can prevent the formation of exudative neovascular AMD, prevent the conversion of drusen to nascent GA or GA, or arrest the growth of drusen in drusen-only eyes.

### Regulatory Perspectives for Ophthalmic Diagnostic Devices

Mr. Bradley Cunningham, Chief of the Diagnostic and Surgical Devices branch of the FDA's Center for Devices and Radiological Health (CDRH), stated that therapeutic devices, which are used to treat a condition or manage symptoms, are categorized into implants, prosthetics, and surgical devices. Diagnostic devices, which are intended to collect quantitative or qualitative information to help diagnose, monitor, or screen

### Invited Speakers and Discussants [Commercial Relationship Disclosure]

- Paul Sieving**, National Eye Institute, Bethesda, MD, USA [None]  
**Frederick Ferris III**, National Eye Institute, National Institutes of Health, Bethesda, MD, USA [None]  
**Emily Y. Chew**, National Eye Institute, Bethesda, MD, USA [None]  
**Karl Csaky**, Retina Foundation of the Southwest, Dallas, TX, USA [Acucela (C), AGTC (C), Allergan (F, C), Genentech (C, R), GSK (C), Heidelberg Engineering (C), Janssen Biotech (F)]  
**Jacque L. Duncan**, University of California, San Francisco, CA, USA [Neurotech USA, Inc. (F)]  
**Frank Holz**, University of Bonn, Bonn, Germany [Acucela (C), Allergan (F, C, R), Bayer (F, C, R), Genentech (F, C), GSK (F), Heidelberg Engineering (F, C, R), Merz (C), Novartis (F, C, R), Optos (F), Roche (C, R), Thea (C), Zeiss (F)]  
**Srinivas Sadda**, Doheny Eye Institute, Los Angeles, CA, USA [Carl Zeiss Meditec (F), Centervue (C), Optos (F, C, R), Topcon (R)]  
**Philip J. Rosenfeld**, Bascom Palmer Eye Institute, Palm Beach Gardens, FL, USA [Carl Zeiss Meditec (F, C)]  
**Bradley Cunningham**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Wiley Chambers**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Malvina Eydelman**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Joffre Angelo Green**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Bernard Lepri**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Michelle Tarver**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**Wilson Bryan**, U.S. Food and Drug Administration, Silver Spring, MD, USA [None]  
**David G. Birch**, Retina Foundation of the Southwest, Dallas, TX, USA [None]  
**Glenn Jaffe**, Duke University, Durham, NC, USA [Heidelberg Engineering (C)]  
**Hendrik Scholl**, Wilmer Eye Institute, Baltimore, MD, USA and University of Basel, Basel, Switzerland [Foundation Fighting Blindness Clinical Research Institute (FFB CRI) and a grant to FFB CRI by the U.S. Department of Defense USAMRMC TATRC, Fort Meade, Maryland (grant numbers W81-XWH-07-1-0720 and W81XWH-09-2-0189)]  
**Maureen Maguire**, University of Pennsylvania, Philadelphia, PA, USA [Genentech/Roche (C)]  
**Cynthia Owsley**, University of Alabama at Birmingham, Birmingham, AL, USA [Abbott Medical Optics (C), Centers for Disease Control and Prevention (F), EyeSight Foundation of Alabama (F), Genentech (C), National Eye Institute/NIH (F), National Institute on Aging/NIH (F), board member, Prevent Blindness (S), Research to Prevent Blindness (F)]  
**Catherine Cukras**, National Eye Institute, Bethesda, MD, USA [None]  
**Richard Weleber**, Oregon Health and Sciences University, Portland, OR, USA [4D Molecular Therapeutics (C), AGTC (F, R, S), Foundation Fighting Blindness (R, S), NightStarRx (C), Sanofi (F), Shire (C), (P)]  
**Albert Maguire**, University of Pennsylvania, Philadelphia, PA, USA [(P)]  
**Jose-Alain Sahel**, Vision Institute, Paris, France [Banque publique d'Investissement (F), Chronocam (I), Chronolife (I), ERC Synergy "HELMHOLTZ" (F), Foundation Fighting Blindness (F), Genesignal (C), GenSight Biologics (F, I, C), LabEx LIFESENSES (ANR-10-LABX-65) (F), Pixium Vision (I, C)]  
**Gislin Dagnelie**, Johns Hopkins University, Baltimore, MD, USA [Second Sight Medical Products (F, P)]  
**Jane Moseley**, European Medicines Agency, London, UK [None]

patients, may be hardware based, software based, combinations thereof, or mobile medical apps. Medical devices fall into three classes based on risk to patients (Class I, II, or III) and are associated with variable levels of FDA oversight. Class III devices (highest risk) receive the most FDA oversight.

Class I diagnostic devices include adaptometers (indicated for the measurement of the time needed for retinal adaptation and the minimum light threshold), perimeters (indicated for determining the extent of the peripheral visual field of a patient), and microperimeters (indicated for generating retinal sensitivity maps). Class II diagnostic devices include ophthalmic cameras (intended to take photographs of the eye and surrounding area), scanning laser ophthalmoscopes (SLO) (indicated for imaging various structures in the eyes, such as posterior segment and choroidal circulation), and OCT devices (indicated for viewing retinal layers as well as retinal and choroidal vasculature, for quantifying parameters such as retinal thickness and retinal nerve fiber layer, and as a diagnostic aid for retinal diseases such as AMD, macular edema, retinal detachment, diabetic retinopathy, and glaucoma). While these devices have limitations based on the cleared indications for use (IFU), they continue to be implemented in uses for which the devices were not cleared. For example, rod-cone break time and rod intercept time (RIT) are commonly used measures; however, there are no adaptometers currently cleared for automated measurement of these parameters. OCT devices are implicated for use in the measurement of drusen volume or atrophy, the ellipsoid zone (EZ), or the junctional zone; however, OCT devices are not cleared for these measurements.

To obtain clearance for qualitative indications, the agency recommends that images be taken from patients with various forms of the condition as well as disease-free individuals. The images should be compared with images obtained using legally marketed predicate devices from the same eye, ideally by masked graders following standardized criteria. To obtain clearance for quantitative indications, the agency recommends that the device not only capture images, but also demonstrate precision (repeatability and reproducibility) and agreement to a legally marketed predicate device. Lastly, Mr. Cunningham emphasized that FDA clearance does not imply that the agency has seen and evaluated evidence to support all potential device uses in clinical practice or studies. Hence, when diagnostic devices are incorporated into clinical trials of therapeutic medical products, device performance must be carefully considered, especially when the device is being considered for determination of patient enrollment criteria or assessment of trial endpoints.

### Highlights From Panel Q&A on Structural Endpoints in GA

- Dr. Wiley Chambers, MD, Deputy Director of the Division of Transplant and Ophthalmology Products of the Center for Drug Evaluation and Research (CDER), reminded the attendees that under the agency's Investigational New Drug (IND) pathway it is possible to use a product in a clinical trial as a means to seek eventual FDA approval even if the product is not labeled for that particular use. The agency strongly encourages validation studies that would gather support for labeling.
- On the question of whether GA onset might be acceptable as a potential trial endpoint, Dr. Chambers noted that any measurable change should be both statistically and clinically significant. Further, precisely defining the onset of GA would be crucial, given the variability seen in past studies. Surrogate endpoints,

which predict a future clinically significant outcome, can be used in certain cases as the basis of approval of a new product. Preventing photoreceptor loss, for example, would be considered a clinically meaningful endpoint, given the established link between photoreceptor loss and visual function. The threshold of such a therapeutic effect remains to be established, but if photoreceptor loss can be prevented at least to the extent of the fuzzy border, as seen on OCT, around the GA lesion, that might be considered a potential trial endpoint.

- Dr. Malvina Eydelman, Director of the agency's CDRH Division of Ophthalmic and Ear, Nose, and Throat Devices, emphasized that the agency often looks to the scientific community for consensus on such thresholds. Hence, it would be up to investigators to determine a consensus threshold through well-controlled, reproducible studies.
- On the question of using upstream measures such as drusen volume changes and step changes on the AMD severity scale, Dr. Chambers' response was that this would not be recommended at the present time. Notwithstanding their usefulness for research studies, drusen characteristics have not served as the basis of approval of any products. The agency does not recognize different categories of AMD on the basis of this scale. That said, he added, drusen characteristics might be useful in defining patient enrollment criteria in clinical trials and with further validation may serve as surrogate endpoints in trials. At the present time, he noted, drusen changes or step changes on the AMD scale would not be valid clinical trial endpoints for drug or biologic approval. In contrast, Dr. Eydelman added that the AMD severity scale could serve as a potential historical control in studies performed for device approval.
- On the question of using fellow eyes as comparators in clinical trials of GA progression, Dr. Chambers noted that the agency is not prepared to accept fellow eyes as substitutes for population controls in drug and biologic approvals at the present time. In contrast, Dr. Eydelman noted that for device approvals, fellow eyes may be acceptable as controls in some cases, depending on the device's impact on the fellow eye, among other factors.
- On using OCT as a potential trial endpoint for measuring drusen or GA, Dr. Eydelman noted that the more automated the measurements, the greater the likelihood of precision and accuracy. Further, she added that any change measured should be significantly greater than the device's margin of error, regardless of the model of device being used.
- Finally, Dr. Chambers emphasized that for drug and biologic approvals the agency shows a clear preference for functional over anatomic endpoints. Visual function includes elements such as visual field, contrast sensitivity, and other light sensitivity measures. That said, he added, given the variability in measuring visual function, the agency is willing to consider anatomic endpoints.

### Structural Endpoints With Functional Associations in IRD

Dr. David Birch reported on studies in which OCT images of patients with retinitis pigmentosa (RP) show a transition zone of retinal tissue that lies between the severely affected and healthy regions and in which the EZ merges with the RPE. The findings of that study revealed that a decrease in EZ width is significant if greater than  $0.44^\circ$  ( $128 \mu\text{m}$ ). Analysis of the EZ

area changes from volume scans revealed that EZ area loss correlates with the visual field loss, and the EZ contour marks the edge of the usable visual field. Similarly, analysis revealed that the rate of visual field loss varies with location relative to EZ, with the highest rates of decline immediately inside and outside the EZ edge.<sup>5</sup> Taken together, these findings suggest that the transition zone contains locations that might be sensitive to treatment. Thus, EZ width or area measured using SD-OCT can provide a reliable potential outcome measure for patients with vision loss tied to photoreceptor degeneration.

Dr. Glenn Jaffe noted that disruption of the EZ, which encompasses the outer portion of the photoreceptor inner segments and is packed with mitochondria, is observed on SD-OCT in eyes with macular telangiectasia (MacTel) type 2. Hence, Dr. Jaffe's team aimed to quantify the loss of EZ by segmentation on SD-OCT, overlay the SD-OCT findings with microperimetry maps of retinal sensitivity, and determine the suitability of using EZ area loss as a potential clinical trial endpoint. While automated segmentation is still hampered by various artifacts, manual delineation of the EZ layer using en face OCT found a correlation with retinal sensitivity as measured by microperimetry.

Dr. Hendrik Scholl described the ProgStar study, a multicenter, international, natural history study of patients with Stargardt disease. Quantification of atrophic lesions detected on FAF, followed by grouping into two categories, namely, “definitely decreased autofluorescence (DDAF)” and “questionably decreased autofluorescence,” indicated that the rate of progression of DDAF depended on the baseline area of atrophy, the homogeneity of the background signals from the area surrounding the atrophic lesion, and the unifocal or multifocal nature of the lesions. The overall mean growth rate of DDAF in this cohort was approximately 0.67 mm<sup>2</sup>/year, and heterogenous background and multifocal lesions were associated with higher progression rates. Thus, Dr. Scholl noted, FAF tracking of lesion growth might serve as a promising endpoint in clinical trials of Stargardt disease aimed at slowing disease progression. Segmentation analysis of SD-OCT revealed that reduced FAF area was correlated with loss of RPE. The FAF DDAF area also correlated with the loss of the ONL, particularly after the outer nuclear complex (ONC) was lost. Thus, structural changes on FAF can be correlated with those seen on SD-OCT, with the inner and outer segments being the most affected layers, followed by the RPE and ONL. The loss of RPE area was tightly correlated with DDAF area loss. While inner segment/outer segment loss was also correlated with DDAF area loss, it appeared to precede DDAF area loss. Taken together, SD-OCT-derived measures could serve as potentially sensitive outcome measures of disease progression in interventional trials of Stargardt disease.

Microperimetry measurement of retinal sensitivity revealed an overall small mean loss after 1 year, albeit a statistically significant loss compared with baseline sensitivity. Correlation of microperimetry data with FAF and SD-OCT measures is currently ongoing and may help determine whether microperimetry is a suitable functional outcome measure for disease progression in Stargardt disease. Such structure-function correlations could help elucidate the functional consequences of the findings of retinal imaging and provide further support for their use as outcomes measures of therapeutic intervention.

Dr. Jacques L. Duncan discussed various aspects of adaptive optics SLO (AOSLO), including its ability to provide images with single-cell resolution. Because the technique allows the imaging of photoreceptor inner and outer segments, RPE, and retinal vasculature, it is particularly well suited for the early detection of retinal diseases marked by photoreceptor loss or changes to retinal vasculature and for tracking cone degeneration in clinical trials and patient care. To illustrate the point,

Dr. Duncan presented examples in which AOSLO helped detect cone loss in eyes with medication-induced photoreceptor damage that was otherwise undetectable on SD-OCT. Dr. Duncan presented examples demonstrating that AOSLO can measure cone spacing changes with significant repeatability as indicated by interobserver, intervisit (separated by no more than 2 months), and interocular agreement.<sup>6</sup> Further, AOSLO-derived cone spacing measurements at or near the center of the fovea were correlated in a nonlinear manner with visual function, and increasing cone spacing was correlated with decreasing visual acuity. Until approximately 50% of the cones were lost, visual acuity appeared to be relatively normal, decreasing below 20/25 after that threshold.<sup>7</sup> These findings suggest that combining AOSLO with other techniques such as SD-OCT, microperimetry, and visual acuity could increase the sensitivity of each technique for imaging disease progression and treatment response in patients with retinal degeneration.

### Regulatory Pathways for Devices in IRD

Dr. Joffre Angelo Green, Acting Chief of the Contact Lens and Retinal Devices Branch of CDRH Division of Ophthalmic and Ear, Nose, and Throat Devices, provided a general overview of possible regulatory pathways to market therapeutic devices used to treat IRD. He discussed three general regulatory pathways: premarket notification, premarket approval, and humanitarian device exemption. He also discussed the Early Feasibility Study (EFS), wherein small clinical trials are conducted in the United States during an early stage of development. The EFS program aims to provide patients in the United States access to technology that may benefit them and to encourage innovation within the United States. He explained that any type of device may be evaluated through an EFS, which may be conducted to study a novel device or a modified existing device, to address an unmet clinical need, to study devices for compassionate or emergency use, or to support new indications for a marketed device. He noted that the best time to approach the agency with an EFS proposal is after the device design, intended use, and the purpose of the EFS have been established, but before expensive and time-consuming nonclinical testing has commenced. The agency recommends early and regular consultation throughout the device development process.

### Highlights From Panel Q&A on IRD

- On the question of whether the agency would accept EZ area loss as a surrogate for visual field loss in trials of RP or MacTel, Dr. Chambers reiterated that while an anatomic endpoint might be considered clinically significant, the critical factors are the extent and location of the change. Further, he noted that there are certain areas, particularly at the edge of the EZ, sometimes referred to as the “fuzzy border,” where the correlation with microperimetry data is not perfect. But if the extent of change is outside the questionable area, then this parameter is likely to be acceptable as a surrogate endpoint.
- Dr. Chambers noted that in cases in which a drug or biologic is being tested in a clinical trial but requires an unapproved device for the diagnosis of the condition being treated, the agency typically considers such drug/biologic-device pairs as combination products, and the drug/biologic and device are reviewed at the same time.
- On the question of whether the agency would accept abbreviated microperimetry measurements for elderly patients on account of the time and potential discomfort involved, Dr. Chambers reminded the audience that in

some instances anatomic measures that have clear clinical significance and reproducibly overcome the noise associated with the signal can be used as a surrogate for functional measures to support an application. Ultimately, the agency considers a variety of factors, rarely a single measure, in determining whether the benefits of the proposed product outweigh its risks.

- On the question of including both eyes in treatment trials for certain conditions, Dr. Chambers noted that the agency might require treatment data from both eyes for a subset of patients, depending on the condition and a variety of associated factors, such as contralateral progression rates and treatment impacts, as well as on whether the product will be used on one or both eyes once it is approved.

### Functional Endpoints in GA and IRD

Dr. Maureen Maguire noted that visual acuity changes may have a role in the evaluation of treatments for GA and other forms of retinal degeneration. Visual acuity is a sensitive safety indicator for interventions in clinical trials, and the pattern of visual acuity loss can influence the choice of the summary outcome measure in clinical trials. Previous studies have shown that visual acuity, when measured with the ETDRS chart or the E-ETDRS system, has high test-retest reliability (for AMD patients, the difference between scores in a test-retest study was found to be less than 5 letters in 84% of measurements and less than 10 letters in 93% of measurements). Importantly, the variability of the measure increases with decreasing visual acuity. Dr. Maguire noted that mean changes in visual acuity can be used as a primary outcome measure for intervention trials of most retinal diseases, instead of comparing the percentage of success or failure between different treatment groups based on a specific number of lines on the scale.<sup>8</sup> Using mean changes requires smaller sample sizes in trials, increases precision, and helps avoid issues of misclassification around a threshold of step changes on the scale. That said, comparing the proportions with a specific amount of change in visual acuity may be useful for addressing the mixtures of distributions of visual acuity loss that can be observed in patients who develop foveal involvement and those who do not, as well as in patients who respond to a treatment and those who do not, as a function of their genotype. Finally, Dr. Maguire noted that the choice of threshold for visual acuity changes in trials, namely, mean changes, 1-line, 2-line, or 3-line loss, depends on the expected distributions of visual acuity among the patients in the trial and on the expected amount of change in visual acuity, with a greater number of lines required when vision is poor.

Dr. Cynthia Owsley discussed the use of driving simulators. Driving is considered an instrumental activity of daily living in the United States and is a predominantly visual task. There are a number of published methods to assess driving performance, and these methods relate to driver safety; driver behavior and vehicle kinematics; driver-reported outcomes such as attitudes, skills, and difficulties; and performance in a driving simulator. While driving simulators have several advantages, they also have shortcomings, including oversimplification of the visual complexity of real-world conditions, inadequate representation of variable ambient lighting and glare seen in the real world, potential participant complacency due to the simulated conditions, and a lack of confirmed correlation between simulator and real-world performance. Though study sponsors have used driving simulator performance data in support of ophthalmic products in applications submitted to the FDA, they have been used to evaluate the safety but not the effectiveness of a drug or device. The agency has issued draft

guidelines to pharmaceutical manufacturers on the use of driving simulators to evaluate the safety of some types of psychoactive drugs.

Dr. Catherine Cukras discussed details of a study on dark adaptation in patients with various stages of AMD using a prototype of the AdaptDx device, which delivers an 80% focal bleach to a circular test spot on the retina located at 5° on the inferior visual meridian.<sup>9</sup> The study revealed that dark adaptation time tended to increase with increasing AMD severity, with RPD eyes having the largest RIT. Further, the method was found to be reproducible when tested on a group of 87 participants with RIT of less than 40 minutes 1 week after initial testing, suggesting that RIT might be a reliable potential functional outcome measure.

Dr. Karl Csaky described the use of low-luminance visual acuity, measured in the presence and absence of neutral density filters to alter the amount of light in test conditions, as a functional outcome measure. Previous studies have found that loss of visual acuity under low luminance can predict visual acuity loss from GA over a 2-year period in AMD patients. Studies have also shown that low-luminance deficit (LLD), which is the difference in visual acuity under normal and low-luminance conditions, correlates with AMD progression, with a maximum deficit reported to be greater than 9.3 letters in one study.<sup>10</sup> Independent analysis by different groups suggests that LLD measurements are reproducible and precise, as indicated by test-retest variability. However, studies have also shown that LLD does not change significantly over time.

Another visual function test is central mesopic sensitivity (CMS). The test measures retinal sensitivity to a white light stimulus and has been shown to be reproducible, as indicated by test-retest variability, and loss of CMS appears to be highly correlated with AMD progression.<sup>11</sup> However, similar to LLD, whether CMS varies reliably over time remains unclear. Finally, scotopic microperimetry (SM) is a functional measure of the sensitivity of rod photoreceptors. Preliminary findings suggest high variability in SM-measured rod sensitivity from point to point across the visual field in patients with AMD, underscoring the importance of precision in using SM as an outcome measure of visual function in trials. Preliminary findings also suggest that SM can track changes in rod sensitivity over time. However, the currently high test-retest variability of SM seen in studies means that testing protocols might need to be further honed before its suitability as a functional outcome measure can be established. Finally, Csaky discussed ways in which microperimetry data can be numerically represented as a functional outcome measure. One approach is to use the mean difference in sensitivity across the entire visual field, and another is to examine the change in five or more points on microperimetry. Both of these approaches have precedent in the FDA guidelines issued for glaucoma visual field sensitivity measurements.

Dr. Richard Weleber discussed visual field modeling and analysis (VFMA), a software program that provides a standardized mechanism to monitor the degree and extent of visual function in patients. The method allows the creation of hill-of-vision sensitivity surfaces, which are plots of retinal sensitivity used in visual field testing, as well as interactive three-dimensional (3D) representations of the hill of vision. Thus, the method enables volumetric analysis of the visual field.<sup>12</sup>

Whereas conventional perimetry produces mean sensitivity or deficiency (measured in decibels) as endpoints that summarily represent the average global retinal sensitivity, VFMA generates a volumetric endpoint (measured in decibel-stereadian units) that represents a global or local measure of the amount of sensitivity. Dr. Weleber noted that providing a total amount rather than an average value results in a more meaningful indicator of the gain or loss of visual function.

Further, unlike measures of conventional mean sensitivity, VFMA has been shown to be consistent across protocols, regardless of whether the central visual field or entire visual field is being evaluated. The topographic display and analysis afforded by VFMA is suitable for a range of perimetry methods, including full-field static perimetry, standard automated perimetry, modified two-color scotopic perimetry, and microperimetry. Taken together, VFMA might provide a standardized, unified framework for analyzing visual field sensitivity data and compelling 3D representations of visual function for research studies and clinical practice.

Dr. Jose-Alain Sahel discussed patient-centric testing to assess patient functioning. Naturalistic testing approximates the conditions of daily living and has numerous advantages over laboratory and simulator-based testing. For example, stimuli in naturalistic tests are often rich and complex and include unlimited fields of view; tasks are often learned well before the experiment and embedded within natural behavior; and the behaviors entail unrestricted head and body movements and complex sequential actions.<sup>13</sup> Dr. Sahel described a naturalistic setup called the Streetlab, which is an artificial virtual reality street developed at the Institut de la Vision in Paris. In addition, Dr. Sahel presented a study designed to monitor a locomotor task in twilight and photopic conditions in patients with RP. Finally, Dr. Sahel noted that while virtual reality tests may not be feasible at every test site, they can help verify the relevance of measurements obtained using standardized methods in clinical trials. Demonstrating a meaningful decline in visual function can also help identify target populations for clinical trials and lend support to eventual approval by regulatory agencies.

Dr. Albert Maguire described the Multi-luminance Mobility Test (MLMT), a relatively novel method that provides a measure of vision-related activities of daily living, namely ambulatory vision. While performance on MLMT does not correlate with performance on visual acuity tests, MLMT specifically measures the speed and accuracy with which a subject can navigate a mobility course under different lighting conditions, which range from a moonless summer night (1 lux) to a well-lit office environment (400 lux). Each light level is assigned a score code, from which an MLMT change score can be calculated over time as a measure of improvement in mobility. Based on input from the FDA to improve the rigor of MLMT, light levels in the testing room were standardized and continuously monitored using calibrated light meters, the 12 courses (each with identical course components, but different test patterns) used in the test were presented in randomized fashion to all subjects, subjects were acclimated to testing through a prior training session, talking was prohibited in the test room once the test began, and graders scored all performances (presented in randomized sequence) and monitored examiners' adherence to protocols. Dr. Maguire reported the results of an observational study designed to test the construct and content validity of MLMT (McCague S, et al. *IOVS* 2015;56:ARVO E-Abstract 4774). The MLMT scoring system was found to be reproducible across multiple testing components and final pass/fail outcomes. Finally, evaluation of the correlation between MLMT and other light sensitivity and visual field-related endpoints in a Phase 3 treatment trial of IRD revealed that the ceiling-adjusted MLMT score significantly correlated with both full-field light sensitivity and macular threshold measured through Humphrey visual fields. Taken together, the findings suggest that MLMT is a reliable test that can provide a clinically meaningful endpoint of functional vision.

Dr. Gislin Dagnelie outlined two types of patient-reported outcomes (PRO) tools used in clinical trials of retinal disease. These tools may be classified as "quality-of-life instruments" and "visual functioning questionnaires." The former group of instruments monitors changes in patients' well-being and may

be related to visual outcomes, and addresses patients' independence and physical and emotional states. The latter group of instruments helps rate the difficulties in performing vision-related activities of daily living based on their relevance to individual patients. Many of the PROs in current use are a combination of the two types of tools. Dr. Dagnelie discussed a relatively modern mixed PRO, namely, the IVI questionnaire.<sup>14</sup> The content of this questionnaire has been validated by expert panel and focus groups, and it includes three domains: mobility and independence, emotional well-being, and reading and accessing information. The shortcomings of the questionnaire include the fact that emotional well-being measured using the scale is not equivalent to visual function, despite good correlation, and that the respondent and item ranges on the logarithmic scale as originally designed were found to be discordant. These shortcomings were later addressed by scaling emotional well-being separately and expanding the item ranges on the scale. In addition, Dr. Dagnelie described a relatively modern Visual Function Questionnaire (VFQ) instrument, namely, the Veterans Administration Low Vision VFQ.<sup>15</sup> Dr. Dagnelie concluded that PROs can help determine patients' emotional well-being at baseline and follow-up, help assess the effects of disease progression and treatment on patients' well-being, help capture visual function as it applies to activities of daily living, and provide a potential patient-centered outcome measure.

### Regulatory Perspectives on Task Performance as a Measure of Function

Dr. Bernard Lepri, Clinical Review Scientist at the Center for Devices and Radiological Health (CDRH) Division of Ophthalmic and Ear, Nose, and Throat Devices, discussed the FDA regulatory perspective on the measurement of patients' visual task performance. The agency published a guidance document on retinal prostheses through the combined efforts of the Division of Ophthalmic and Ear, Nose, and Throat Devices and the Office of Science and Engineering Laboratories.<sup>16</sup> This guidance allows considerable latitude as the technology undergoes development over time. The guidance recommends that manufacturers collect extensive clinical data on patient enrollment, device performance and effectiveness, safety, and PROs, as well as nonclinical data, such as device durability, biocompatibility, hermeticity, sterility, and animal testing results. Study endpoints on safety and effectiveness may be structural or functional. The effectiveness-related functional endpoints include visual acuity, including low-luminance visual acuity, contrast sensitivity, visual fields, dark adaptation, mobility, orientation, and navigation tests, and activities of daily living. Effectiveness distinguishes visual function from functional vision. Functional vision is a measure of an individual's ability to perform activities of daily living, and measures real-world visual ability or disability as sustainable performance in real-world settings. Whereas visual function tests measure a threshold performance under controlled conditions in which often only a single parameter is varied, functional vision tests are often carried out in simulated environments. Simulated environments cannot perfectly replicate the real world. The agency recommends that functional vision tests be carried out in actual living and working environments encountered by patients, wherever possible. Dr. Lepri noted that measures of visual task performance must be directed toward difficulties encountered in daily life. Performance on these tasks may be primary or secondary endpoints, and must be evaluated in three domains, namely, orientation and mobility (evaluated by a trained independent professional), daily living in patients' home environments (evaluated by a trained, independent low-vision professional),

and reading and occupational needs (evaluated by a trained, independent, low-vision reading instructor or occupational therapist). Finally, Dr. Lepri noted that the agency's guidance on retinal prostheses for IRD does not address the issue of driving. However, past studies of intraocular lens implantation have evaluated the postoperative performance of subjects in simulated driving tests after lens implantation.

### Patient Voices in Medical Product Evaluation

Dr. Michelle Tarver, Medical Officer and Epidemiologist at the FDA CDRH, informed the group that increasing the use and transparency of patient input as evidence in decision making was a strategic priority for CDRH in 2016 to 2017. She noted that patients' perspectives are crucial along the entire development and evaluation process of medical devices, from discovery and ideation to postmarket monitoring. To this end, the NIH and FDA provided consensus definitions for the types of clinical outcome assessment<sup>17</sup>: any assessment that may be influenced by human choices, judgment, or motivation, and may support either direct or indirect evidence of treatment benefit. Such assessments may fall into one of four categories. Clinician-reported outcomes are based on reports from trained health care professionals following the observation of a patient's health condition; observer-reported outcomes are based on observations of a nonclinical individual other than the patient, such as a parent or spouse who is well positioned to regularly observe and report observable symptoms without including medical judgment or interpretation (one example is reports of the number of times a child rubs his or her eyes); performance outcomes are measurements based on a task performed by a patient under the instructions of a health care professional and are considered to have better face validity and reliability than self-reports, such as visual acuity testing; PROs, which are symptoms or other health conditions directly reported by patients without interpretation from anyone else and are often incorporated as endpoints in clinical studies. In keeping with the evolving nature of health conditions and device development, Dr. Tarver noted that the development of PROs as well as other clinical outcome assessments is an iterative process.

Dr. Tarver noted that when using clinical outcome assessments in trials, the target patient population should be defined, the concept of interest and context of use should be defined, the method of measuring the concept that will best define treatment benefit or safety risk should be identified, and well-defined assessment tools that can reliably measure each concept in the proposed context should be selected. As further supportive information, device premarket evaluation takes into account patient preference information. This may include qualitative or quantitative assessment of the relative desirability of different alternatives among outcomes or other intervention-related choices. Such information often indicates which endpoints, attributes, and factors are valued by patients and influence their perspectives on device risks and benefits. Patient preference studies also provide an estimate of the tradeoffs that patients are willing to make or demonstrate. Such information can guide endpoint choice and requisite effect size in clinical studies as well as subgroup considerations, labeling expansion, and indication changes for medical devices.

### Highlights From Panel Q&A on Functional Endpoints in GA and IRD

- Dr. Jane Moseley, Senior Scientific Officer at the European Medicines Agency (EMA), offered the EMA's perspectives on endpoints for clinical trials of pharmaceuticals for

retinal diseases. She noted that the EMA is largely aligned with the FDA in terms of regulatory principles for retinal disease trial endpoints and considers validity, reliability, sensitivity to change, and clinical significance in approving trial endpoints. The European regulators recognize the shortfalls of best-corrected visual acuity (BCVA) as a functional endpoint measure for past and ongoing drug trials for retinal diseases and hence the need for novel functional endpoints, such as those related to functional vision and PROs. She urged trial sponsors to consult with the EMA regularly and early in the endpoint development process. For GA, area of GA could be acceptable as a primary efficacy variable in principle, but the European regulators would like this use to be supported by a positive effect on function. The challenge is to show that the apparent difference in anatomic progression translates into a functional benefit that can be weighed against safety issues.

- On the question of whether 5-letter or 10-letter changes in BCVA on the ETDRS would be preferable as potential trial endpoints, Dr. Chambers noted that the agency does not set out to determine the minimum threshold of change in visual acuity while considering drugs for approval. The agency's goal is to determine a clinically significant change that can be weighed against the safety risks associated with the proposed method or product. As a rule of thumb, the smaller the change, the lower the likelihood that it would outweigh risks. Hence, he noted, it is unlikely that the agency would change its current expectation for BCVA to any value less than 15 letters on the ETDRS for individual patients (mean changes less than 15 letters are a different matter, however, and must be weighed against the risks). Dr. Eydelman noted that her center at the agency has used 10-letter changes or fewer as the basis of approval of several devices for specific conditions. From the European regulatory perspective, Dr. Moseley concurred that a treatment effect based on mean change in BCVA would be acceptable, as would a responder definition based on a  $\geq 15$ -letter change relative to baseline. There may be situations in which a  $\geq 10$ -letter change in BCVA (as a responder definition) would be acceptable as a primary endpoint.
- On the question of whether the agency would consider a change in the hill of vision, for example, or a regional functional change, if the EZ area cannot be measured, Dr. Chambers responded that that agency would consider hill-of-vision measurements and added that the challenge would be to determine the extent of volume change that is clinically significant. The panelists added that the agency would likely require evidence supporting clinically meaningful volume changes. According to Dr. Moseley, visual field sensitivity could be acceptable as a primary endpoint for European regulators, but the proposed methodology and analysis must be fully clarified in the protocol, and the expected clinical relevance justified in the target population.
- On the question of whether the agency would consider contrast sensitivity as a potential trial endpoint, Dr. Eydelman noted that its use has been long debated in the scientific community, and the agency has recognized standards addressing recommendations for contrast sensitivity measurements, including definitions of clinically meaningful change. Dr. Chambers noted that for drug approvals the agency would be willing to consider changes in contrast sensitivity, but the changes have to be in multiple cycles or frequencies. Further, he added that from the agency's standpoint, measuring visual acuity in either high or low contrast can be clinically meaning-

ful. From the European standpoint, contrast sensitivity has never been proposed as a primary endpoint in any submitted development plan, but advice could be sought from European regulators in this regard.

- Regarding the agency's perspectives on specificity of PROs for consideration as potential outcome measures, Dr. Tarver noted that PROs do not need to be disease-specific but should measure the concepts that they are purported to measure in the intended population.
- On the question of whether patients should have a right to try new interventions when faced with no approved alternatives, Dr. Chambers noted that the agency does have a mechanism that allows physicians or sponsors to submit an IND application to trial interventions on individual patients, and such permissions are issued on a case-by-case basis. Dr. Tarver encouraged sponsors and patient advocacy groups to conduct patient preference studies, which can inform the agency's evaluation of risks and benefits of new interventions.

## References

- Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semi-automated image processing method for identification and quantification of geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2011;52:7640-7646.
- Vitale S, Clemons TE, Agron E, et al.; for the Age-Related Eye Disease Study 2 Research Group. Evaluating the validity of the age-related eye disease study grading scale for age-related macular degeneration: AREDS2 report 10. *JAMA Ophthalmol.* 2016;134:1041-1047.
- Yehoshua Z, Wang F, Rosenfeld PJ, Penha FM, Feuer WJ, Gregori G. Natural history of drusen morphology in age-related macular degeneration using spectral domain optical coherence tomography. *Ophthalmology.* 2011;118:2434-2441.
- Abdelfattah NS, Zhang H, Boyer DS, et al. Drusen volume as a predictor of disease progression in patients with late age-related macular degeneration in the fellow eye. *Invest Ophthalmol Vis Sci.* 2016;57:1839-1846.
- Birch DG, Locke KG, Felius J, et al. Rates of decline in regions of the visual field defined by frequency-domain optical coherence tomography in patients with RPGR-mediated X-linked retinitis pigmentosa. *Ophthalmology.* 2015;122:833-839.
- Zayit-Soudry S, Sippl-Swezey N, Porco TC, et al. Repeatability of cone spacing measures in eyes with inherited retinal degenerations. *Invest Ophthalmol Vis Sci.* 2015;56:6179-6189.
- Ratnam K, Carroll J, Porco TC, Duncan JL, Roorda A. Relationship between foveal cone structure and clinical measures of visual function in patients with inherited retinal degenerations. *Invest Ophthalmol Vis Sci.* 2013;28:5836-5847.
- Beck RW, Maguire MG, Bressler NM, Glassman AR, Lindblad AS, Ferris FL. Visual acuity as an outcome measure in clinical trials of retinal diseases. *Ophthalmology.* 2007;114:1804-1809.
- Jackson GR, Edwards JG. A short duration dark adaptation protocol for assessment of age-related maculopathy. *J Ocul Biol Dis Infor.* 2008;1:7-11.
- Wu Z, Ayton LN, Guymer RH, Luu CD. Low-luminance visual acuity and microperimetry in age-related macular degeneration. *Ophthalmology.* 2014;121:1612-1619.
- Hanout M, Horan N, Do DV. Introduction to microperimetry and its use in analysis of geographic atrophy in age-related macular degeneration. *Curr Opin Ophthalmol.* 2015;26:149-156.
- Weleber RG, Smith TB, Peters D, et al. VFMA: topographic analysis of sensitivity data from full-field static perimetry. *Trans Vis Sci Tech.* 2015;4(2):14.
- Lappi O. Eye tracking in the wild: the good, the bad, and the ugly. *J Eye Mov Res.* 2015;8;1:1-21.
- Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci.* 2006;47:4732-4741.
- Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci.* 2004;45:3919-3928.
- U.S. Food and Drug Administration. Guidance for Industry and FDA Staff - Investigational Device Exemption (IDE) Guidance for Retinal Prostheses. Available at: <http://www.fda.gov/RegulatoryInformation/Guidances/ucm341954.htm>. Accessed December 27, 2016.
- U.S. Food and Drug Administration. Clinical Outcome Assessment Qualification Program. Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm>. Accessed December 27, 2016.