Glaucoma Research Community and FDA Look to the Future, II: NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium: Measures of Structural Change and Visual Function

Robert N. Weinreb[1] *and Paul L. Kaufman*[2]

The 2010 NEI-FDA Glaucoma Clinical Design and Endpoints Symposium was a follow-up to a similar March 2008 public meeting where the glaucoma research community, National Eye Institute (NEI) of the National Institutes of Health (NIH), and Food and Drug Administration Center for Drug Development and Research (FDA CDER) convened to discuss adopting new endpoint measures for assessing glaucoma therapies in clinical trials to "facilitate bringing safe and efficacious pharmacotherapies to the U.S. market." A report published in this journal described the discussion and set the stage for the September 24, 2010, follow-up meeting reported here.[1]

This 2010 meeting on glaucoma endpoints was organized by the Association for Research in Vision and Ophthalmology (ARVO) and co-chaired by Robert N. Weinreb, MD, FARVO, of the University of California, San Diego, and Paul L. Kaufman, MD, FARVO, of the University of Wisconsin. The symposium planning committee included Frederick L. Ferris III, MD, of the National Eye Institute of the NIH, and Wiley A. Chambers, MD, Malvina B. Eydelman, MD, and Robert L. Kramm, MD, MSE, of the FDA. Attendees were mainly researchers, clinicians, policymakers, and representatives from industry and vision associations.

Glaucoma is a serious problem in the United States, diagnosed in more than 2.3 million Americans, mostly 40 years of age and older.[2] It accounts for 9% to 12% of all cases of blindness.[3] Approximately an equal number of people have glaucoma but do not realize it. Glaucoma affects more than 60 million people worldwide; more than 10% are consequently legally blind in both eyes.[4]

This report describes the most recent meeting comparing several imaging technologies (optical coherence tomography [OCT], stereoscopic optic disc photography, scanning laser polarimetry [SLP], and confocal scanning laser ophthalmoscopy [CSLO]) for assessing structural changes related to glaucoma progression and treatment. The conference objectives were to advance the understanding of these technologies and instruments, including their relationship to visual function and their role in diagnosing and treating different stages of glaucoma.

The meeting was the fourth in a series in which the vision community and FDA convened to discuss the FDA requirements for adding new endpoints to the evaluation of ophthalmic treatments and products. The first occurred in 2006 and concerned endpoints and clinical trial strategies for evaluating new treatments for age-related macular degeneration (AMD) and diabetic retinopathy.[5] The second, as stated, was about clinical trial design and endpoints for evaluating glaucoma treatments.[1] The third concerned the use of patient-reported outcomes (PROs) in medical product development.[6]

Outcome and Importance of the 2008 Meeting

The major understanding established between the vision research community and the FDA at the 2008 meeting was that the **FDA is open to using structural endpoints in clinical trials of new glaucoma drugs provided that the structural measures predict clinically relevant functional change.** (See FDA CDER's *Guidance for Industry: Qualification Process for Drug Development Tools*.[7]) Dr. Chambers suggested that researchers consider (1) whether structural measures will be more consistent and less variable than visual function measures; (2) whether a strong correlation exists between a structural measure and predictability of either current visual function or future visual function; and (3) whether the new approach will be beneficial to patients.

The ophthalmic community as a whole considers it critically important to follow up on this opening from the FDA. New or more endpoints are needed to shrink the cost and duration of clinical trials, to reduce the number of study subjects needed, and, most importantly, to bring better therapies to the public faster.

Measurements of **intraocular pressure** and **standard visual fields** are the endpoints generally accepted by the FDA in evaluations of new therapies for glaucoma. Recently, we have seen that **structural** changes (e.g., in the optic disc measured by stereophotography) predict standard visual field outcomes.[8]

Moreover, improvements in imaging technologies are producing structural data that are more consistent, which should

From the [1]Hamilton Glaucoma Center, University of California, San Diego, La Jolla, California; and the [2]Department of Ophthalmology and Visual Science, University of Wisconsin, Madison, Wisconsin.

Supported by the National Eye Institute of the National Institutes of Health, U.S. Department of Health and Human Services.

The ideas presented in this report represent the opinions and findings of the speakers and discussants at the Clinical Trial Design and Endpoints Symposium.

Submitted for publication April 15, 2011; revised July 27 and August 4, 2011; accepted August 4, 2011.

Disclosure: **R.N. Weinreb**, Alcon Laboratories (C, R), Allergan, Inc. (C, R), Amira (C), Bausch & Lomb (C), Cytokinetics (C, R), Merck Research Laboratories (C, R), Pfizer, Inc. (C, R), Genentech (C), Topcon Medical Systems, Inc. (F), Optovue, Inc. (F, C), Othera (C), Carl Zeiss Meditec (F, C), Heidelberg Engineering (F), Nidek (F), Paradigm (F); **P.L. Kaufman**, Alcon (F, C, R), Allergan, Inc. (F, C, R), Bausch & Lomb (C, R), Cara Therapeutics (F, C), Cytokinetics (C, R), Danube Pharmaceuticals (F, C), Implandata (C), Inspire Pharmaceuticals, Inc. (F, C, R), Neurotech (C), NuLens (F), Cascade (C, R), Pfizer, Inc. (F, C, R), QLT (C, R), Santen, Inc. (F, C), Transcend Medical (F, C), Lens AR, Inc. (aka Frey Research LLC) (R)

Corresponding author: Robert N. Weinreb, Hamilton Glaucoma Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0946; weinreb@eyecenter.ucsd.edu.

DOI:10.1167/iovs.11-7895
Investigative Ophthalmology & Visual Science, October 2011, Vol. 52, No. 11

allow for reliable comparisons in evaluating optic nerve and retinal nerve fiber changes in glaucoma.

Presenters focused on one or more of the three questions posed by the organizers:

1. What are optimal criteria for measuring the rate of tissue loss and defining progressive structural events?
2. What is statistically significant structural change, considering the known variability of each technique?
3. How does structure predict clinically relevant functional outcomes (structure–function relationships)?

The FDA's Perspective on Acceptable Glaucoma Clinical Trial Endpoints

Two FDA leaders, whose offices are responsible for approving ophthalmic drugs and devices, described the past and present use of structural and functional endpoints in glaucoma drug and device clinical trials: Wiley Chambers, MD, Acting Director of the Division of Anti-infective and Ophthalmology Products of the FDA CDER, and Malvina Eydelman, MD, Director of the Division of Ophthalmic, Neurologic and Ear, Nose, and Throat Devices in the FDA Office of Device Evaluation, Center of Devices and Radiologic Health.

According to Dr. Chambers, when studying the safety and efficacy of new ophthalmic drugs, the FDA evaluates patients' **visual function**—such as visual fields, color vision, visual acuity, or contrast sensitivity. The FDA accepts that, if unchecked, degradation of these parameters will predict worsening of functional vision that will affect the patient in the real world. The FDA recognizes inherent limits to threshold methodologies and to functional tests that patients learn easily and improve in performing over time. To reiterate, at this meeting the participants addressed the relationship of structural endpoint measures to visual function and the role of structural endpoints, like functional endpoints, as surrogates for assessing glaucoma therapies in clinical trials.

Dr. Chambers agrees that structural measurements could replace visual function, although no structural endpoints are currently used for glaucoma. **Structural endpoints** may be more consistent than functional endpoints and are not affected by a learning curve. The FDA would like a structural endpoint to show a strong correlation ($R^2 \approx 0.9$) to current vision or future vision (gain or loss). A high R^2 is predictive of future outcomes, indicating that if a given event occurs, then another well-defined event is highly likely to follow.

The FDA allows structural endpoints in other areas of ophthalmology. These areas relate to preventing retinal detachment, to preventing advancement of cytomegalovirus retinitis, and to the three-step progression used in the Early Treatment Diabetic Retinopathy Treatment Study (ETDRS). The FDA is open to considering structural endpoints that involve nerve fiber layer and optic disc changes.

For structural endpoints to be acceptable for the evaluation of medical products for the treatment of glaucoma, a few questions must be addressed:

- Which methodologies best provide reproducible measures of clinically significant changes?
- How much of a change is clinically significant? Which changes correlate highly with current deficits in visual function? What change in vision would be expected to occur due to this structural change?
- How long do the changes have to be present to cause a change in visual function? Which changes correlate highly with or are predictive of deficits or decline in visual function? When would a change in vision be expected to occur due to this structural change?

Participants in the 2010 NEI-FDA Glaucoma Clinical Design and Endpoints Symposium [Commercial Relationship Disclosure]

Wiley Chambers, U.S. Food and Drug Administration (FDA) [None]
Malvina B. Eydelman, FDA[None]
Murray Fingeret, VA NY Health Care System [Carl Zeiss Meditec (C, R), Optovue, Inc. (F, R), Topcon Medical Systems, Inc. (C), Heidelberg Engineering (F, R)]
David Friedman, Johns Hopkins University [Bausch and Lomb (C), Merck Research Laboratories (C), Pfizer, Inc. (R), Carl Zeiss Meditec (F)]
David F. Garway-Heath, Moorfields Eye Hospital [Heidelberg Engineering (F, R), Carl Zeiss Meditec (F, C, R), OptoVue, Inc. (F), Moorfields Motion Displacement Test (P)]
Paul L. Kaufman, University of Wisconsin [Inspire Pharmaceuticals, Inc. (F, C, R), Pfizer, Inc. (F, C, R), Alcon Laboratories (F, C, R), Santen, Inc. (F, C, R), Allergan, Inc. (F, C, R), Bausch & Lomb (C, R), Cytokinetics (C, R), QLT (C, R), Neurotech (C), Danube Pharmaceuticals (F, C, R), Cara Therapeutics (F, C), Transcend Medical (F, C), NuLens (F), Cascade (C, R), Lens AR, Inc. (aka Frey Research LLC) (R)]
Robert Lee Kramm, FDA [None]
Christopher Kai-Shun Leung, Hong Kong Eye Hospital [Carl Zeiss Meditec (F, C, R), Optovue, Inc. (F), Tomey (F), Lumenis (F), Alcon Laboratories (F)]
Felipe A. Medeiros, University of California [Carl Zeiss Meditec (F), Pfizer, Inc. (C), Alcon Laboratories (C), Allergan, Inc. (C)]
Eva Rorer, FDA [None]
Joel S. Schuman, University of Pittsburgh [Carl Zeiss Meditec (P), Bioptigen, Inc. (P)]
Mike Sinai, Optovue, Inc. [Optovue, Inc. (E)]
Gianmarco Vizzeri, University of Texas Medical Branch at Galveston [None]
Robert Weinreb, University of California, San Diego [Alcon Laboratories (C, R), Allergan, Inc. (C, R), Amira (C), Bausch & Lomb (C), Cytokinetics (C, R), Merck Research Laboratories (C, R), Pfizer, Inc. (C, R), Genentech (C), Topcon Medical Systems, Inc. (F), Optovue, Inc. (F, C), Othera (C), Carl Zeiss Meditec (F, C), Heidelberg Engineering (F), Nidek (F), Paradigm (F)]
Gadi Wollstein, University of Pittsburgh [Carl Zeiss Meditec (F), Optovue, Inc. (F), Bioptigen, Inc. (P)]
Linda Zangwill, University of California, San Diego [Heidelberg Engineering (F), Carl Zeiss Meditec (F), Topcon Medical Systems, Inc. (F), Optovue, Inc. (F)]

The FDA CDER is prepared to consider structural tests that researchers propose and validate for studying glaucoma.

Dr. Eydelman discussed the regulatory history of diagnostic and therapeutic devices used in glaucoma management. Diagnostic devices regulated by the FDA include tonometers and fundus cameras, standard automated perimeters (SAPs), short-wavelength automated perimeters (SWAPs), frequency-doubling technology (FDT) perimeters, scanning laser polarimeters (SLPs), confocal scanning laser ophthalmoscopes (CSLOs), and optical coherence tomographs (OCTs).

Most therapeutic glaucoma devices currently regulated by the FDA fall into two major categories: lasers and implantable glaucoma drainage devices. Therapeutic glaucoma drainage device trial endpoints are driven by the sponsor's "statement of indications for use"—in other words, the purpose for which and the patient population on which the sponsor intends the device to be used. Glaucoma lasers, such as Nd:YAG, argon, and diode, have indications that include laser trabeculoplasty, iridotomy, and cyclophotocoagulation. Before marketing, lasers must show substantial equivalence to a legally marketed predicate device and be cleared via the FDA 510(k) process.

The most commonly used outcome measure of clinical trials of glaucoma drainage devices is intraocular pressure (IOP).

Structural and functional measures have not been used as primary endpoints. Dr. Eydelman indicated that the main reason functional and structural measures have not been used is because they have not been proposed. The sponsor of a submission to the FDA proposes the indications for use of the device for which the sponsor would like to seek marketing approval or clearance. Glaucoma drainage devices, which have been indicated for reducing IOP in patients with glaucoma, provide a good example. Because sponsors have not pursued treatment of glaucoma *itself* as an indication, they have not had to show that their devices can retard the progression of glaucomatous optic neuropathy. Reducing IOP is a surrogate endpoint of value in glaucoma, but not itself a measure of structural or functional glaucomatous optic neuropathy. Demonstrating with clinical performance data that glaucoma drainage devices retard the progression of glaucomatous optic neuropathy would be required for a sponsor to include this claim in the device indication or label.

Ophthalmic Devices for Assessing Structural Endpoints

Symposium presenters described evidence for using structural endpoints from OCT, stereophotographs, SLP, and CSLO in glaucoma clinical trials as surrogates for changes in visual function.

Advances in ophthalmic imaging technology have improved the ability of researchers and clinicians to measure optic disc and retinal nerve fiber layer (RNFL) changes in eyes of patients with glaucoma. Each technology has advantages and limitations, and each employs different principles to obtain its respective measurements of the optic nerve.

Requirements for a surrogate endpoint would be biological plausibility and strong association between the surrogate and functional outcome, remembering that the gold standard metric for function—standard white-on-white automated perimetry—is itself a surrogate. Further, the effect of treatment on the surrogate endpoint would have to predict the effect on a clinically relevant outcome. Still further, there would have to be a favorable risk–benefit profile associated with the treatment.

Optical Coherence Tomography

OCT measures the intensity of reflected light from interference patterns and calculates tissue thickness based on cross-sectional data. During the past 20 years, OCT technology has advanced from a single A-scan taking ~1.3 seconds[9] to a technology that can perform more than 300,000 scans per second and from the commercially available time-domain (TD)-OCT to the more robust spectral-domain (SD)-OCT. Coupling with advanced software provides reproducible data on the optic nerve, peripapillary RNFL, and macular ganglion cell complex (GCC). OCT images of disease states can be analyzed in three dimensions, compared with normative data, and studied for clinically relevant changes.

An obviously important question, as pointed out by Joel Schuman, MD, is how structural changes measured with OCT correspond to clinically significant functional changes in patients with glaucoma. Several studies show that progressive loss of the RNFL measured by OCT corresponds to progressive visual field loss and, further, that sometimes OCT anatomic changes may be detected earlier than SAP functional changes. In one study, over a period of approximately 5 years, OCT showed progression in 22% of patients, whereas visual field tests revealed progression in only 9%, suggesting a greater sensitivity for detecting glaucomatous progression using OCT.[10] Moreover, OCT of the RNFL discriminates between "progressors," meaning eyes shown by visual field tests and/or stereoscopic optic disc photographs to be worsening, and "nonprogressors," meaning eyes that remained stable. Further, the actual rate of loss of RNFL in progressing eyes was shown to be significantly greater than in nonprogressing eyes.[11] In other words, there was a correspondence between structural and functional measures; and in certain circumstances, OCT can outperform standard technologies and possibly detect glaucoma progression earlier.

As Christopher Leung, MD, aptly described, OCT is useful for detecting progression of glaucoma when the difference between baseline and follow-up measurements exceeds the *variability* of the instrument. With the newer and more sophisticated SD-OCT compared to the older TD-OCT, variability is greatly reduced. As an example, intervisit variability of the average RNFL thickness for the TD- and the SD-OCT has been shown to be approximately 11 and 5 μm, respectively (95% CI).[12] Additional evidence suggests that variability may not differ among the stages of glaucoma.[13,14] A prospective study with multiple OCT measurements obtained from 45 normal individuals and 43 glaucoma patients in various stages showed that there is no association between intervisit RNFL measurement variability and the mean RNFL thickness.[13]

Glaucoma is not the first condition for which Dr. Chambers has been consulted with respect to the use of OCT for measuring retinal changes in clinical trials. He has been consulted about macular edema related to multiple sclerosis. Novartis (Basel, Switzerland) was seeking approval of Gilenya capsules (fingolimod) for delaying disability progression in patients with relapsing forms of multiple sclerosis (MS). FDA supported the use of OCT as a safety measure for monitoring macular thickness in treated patients. However, according to Dr. Chambers, retinal thickness in the fingolimod trial did not by necessity predict visual acuity; in glaucoma clinical trials, the FDA wants structural endpoints to show a strong correlation to current or future visual function.

In glaucoma, questions remain about the course of change in slow and fast progressors, about whether one group responds to interventions differently than the other, and even about the long-term rate of change in the RNFL in normal subjects.

Stereoscopic Optic Disc Photographs

Stereoscopic optic disc photographs are commonly used in practice to image the optic disc. A recent study of patients with suspected glaucoma showed the capacity of stereophotography to predict visual function. Researchers followed up, for an average of 8 years, 639 eyes of 407 patients, all of whom had normal visual fields at the study's outset.[8] Progressive optic disc changes were detected by stereophotography in 15% (96) of eyes during follow-up. Of these, 66% (63) underwent SAP visual field conversion. In contrast, of the 543 eyes that had no evidence of disc damage, only 6% (32) showed visual field loss during the follow-up period. Progressive optic disc change was strongly predictive and, in fact, was the most important predictive factor for development of visual field loss, with a coefficient of determination (R^2) of 0.79. R^2 for other predictive factors such as IOP, corneal thickness, and other baseline variables ranged from only 0.06 to 0.26. These results support a structural surrogate (optic disc change by stereoscopic photography) as a strong predictor of a functional endpoint (visual field loss by SAP).

Stereoscopic disc photography has limitations. It describes qualitative, not quantitative, changes and, therefore, estimates of rates of change are subjective. Its ability to detect change depends on the quality of the photograph, and reading centers are necessary to standardize readings and reduce the effect of interobserver variability.

Scanning Laser Polarimetry

SLP (GDx; Carl Zeiss Meditec, Dublin, CA) is used to evaluate peripapillary RNFL thickness. It uses polarized light directed into the back of the eye to measure birefringence, a surrogate for RNFL thickness. Microtubules of the ganglion cell axons are the major contributors to birefringence. SLP is able to detect and measure the size and depth of RNFL defects resulting from the decline in ganglion cells axons. The technology is specific to glaucoma and other optic neuropathies.

SLP has been shown to recognize early RNFL loss.[15,16] TSNIT average (equivalent of RNFL thickness)—referring to a temporal-superior-nasal-inferior-temporal scan pattern along a path that begins superiorly and ends temporally—is reported to be the most reproducible SLP parameter with a within-session coefficient of 2.5 mm, and a between-session repeatability of 4.7 mm. This test–retest reliability indicates that a difference between successive measurements exceeding these values can be attributed to true tissue RNFL loss.[16]

Software developed for the GDx provides guided progression analysis (GPA), both globally and by sector. Agreement has been reported between the structural changes detected by SLP and the functional changes detected by perimetry.[17] The cutoff criteria for progression used in this analysis are based on measurement variability derived from the population or from the individual tested eye. The newest version of SLP, GDx ECC (enhanced corneal compensation), demonstrates the structure–function relationship more strongly than the earlier VCC version.[18] Rates of RNFL progression can be estimated.[19]

Confocal Scanning Laser Ophthalmoscopy

CSLO (HRT; Heidelberg Engineering, Heidelberg, Germany) measures the intensity of reflected light at various depths and reconstructs optic disc surface topography. Ideally, progressive glaucomatous structural damage would be measured as a structural change in the neural rim area that is greater than that which occurs with normal aging and that predicts functional visual field damage. Research using CSLO has shown that small age-related changes occur in normal controls at a significantly slower rate than in patients with glaucoma.[20] The rate of change in normal controls could be the backdrop against which significant rate- or event-based change is measured by the CSLO for predicting functional change in glaucoma patients.

We also know, from the Confocal Scanning Laser Ophthalmoscopy Ancillary Study to the Ocular Hypertension Treatment Study (OHTS), that structural change based on a baseline measure can predict future glaucomatous visual field progression.[21] Researchers have shown that the mean rate of rim loss in eyes that develop glaucoma visual field endpoints is five times faster than in eyes that do not develop glaucoma (Zangwill LM, Jain S, Dirkes K, et al., unpublished data, 2011).[21] Unfortunately, Venn diagrams reveal modest overlap in CSLO structural and visual field measures; quite possibly the tests detect different aspects of change that may not be perfectly synchronous. Regardless of the imaging technology used, a structural endpoint would qualify as a surrogate for a functional endpoint in identifying treatment effect only if a change in structure strongly predicts the change in function and, also, if the structure is responsive to the studied treatment.

Another unanswered question regards how structural and functional endpoints would be affected by the severity of glaucoma. In more advanced disease, nonneural tissue may be involved and factor into measurements.

Regulation of Ophthalmic Devices Used in the Management of Glaucoma

It is important for medical device users and manufacturers alike to understand the level of evidence required by FDA in order for that device to reach the U.S. market.

The FDA considers an item to be a medical device if it diagnoses, cures, mitigates, treats, or prevents a disease or condition; affects the function or structure of the body; does not achieve its intended use through chemical action; and is not metabolized. A device is assigned to a class (classes I–III) based on its complexity and the risk associated with its use. The class determines the regulatory path to market.

Class I devices are typically of simple design and low risk and are subject to general controls. Most are exempt from premarket submission. Examples of class I devices are most visual acuity charts, perimeters, and manual surgical instruments.

Class II devices carry a higher risk or are more complex and usually require submission to the FDA of a Premarket Notification, also known as a 510(k), demonstrating substantial equivalence to a predicate device. In addition to general controls, class II devices are subject to additional special controls that may include FDA guidance documents, special labeling requirements, and performance standards. Several examples relevant to glaucoma are slit lamps, tonometers, glaucoma implants for the refractory population, and lasers used for the reduction of IOP. Class II devices also include fundus cameras, direct and indirect ophthalmoscopes, SLO polarimeters, CSLO topographers, and OCTs.

Class III devices carry the highest risk and typically are life-supporting or life-sustaining, are for a use that is of substantial importance in preventing impairment of human health, or present a potential unreasonable risk of illness or injury. They are subject to general controls and require premarket approval (PMA). The application must contain sufficient valid scientific evidence to provide reasonable assurance that the device is safe and effective for its intended use. Examples in glaucoma are glaucoma implants (for the nonrefractory population) and viscoelastics.

The FDA's Dr. Kramm pointed out that there is no specific FDA guidance regarding the content of premarket submissions for GDx, HRT, or OCT; however, there are several applicable performance standards and FDA guidance documents to which a device sponsor can refer.

Additional Considerations in Regulation of Devices for Assessing Structural Endpoints

When considering structural endpoints, it is important to fully understand the capabilities and limitations of devices used for such assessments. The degree to which a structural endpoint may be useful for the evaluation of a therapeutic modality is dependent on the characterization and validation of the performance of the device for that assessment.

Basic validation of an instrument's ability to measure that which the sponsor claims requires nonclinical and/or clinical performance data. The sponsor should show precision and agreement meaning, respectively, variability among repeated measures and comparability of measurements between devices. When designing measurement validation studies, sponsors should refer to ISO 5725-2, which "amplifies the general principles to be observed in designing experiments for the numerical estimation of the precision of measurement methods by means of a collaborative inter-laboratory experiment, provides a detailed practical description of the basic method for routine use in estimating the precision of measurement

methods, provides guidance to all personnel concerned with designing, performing or analyzing the results of the tests for estimating precision."

Another issue important to the FDA in the regulation of structural endpoint devices is the inclusion of a normative (reference) database in the software and how it is developed (i.e., sample size, effect of covariates, and a clear clinical definition of normal). Dr. Kramm emphasized that percentile bins do not necessarily equate to clinical decision limits for discrimination between disease and nondisease states—an important point when interpreting the results from devices with a normative database.

The ability of a device's software algorithm to track changes in structural parameters over time to detect disease progression is highly dependent on the precision of the device, and so it must be well characterized. It is important to note that clinically significant change, which does not necessarily equate to statistically significant change, is highly dependent on time lapse and other factors that may change over time and affect measurements.

The Indications for Use (IFU) statement in a premarket submission should be supported by performance data, and the level of data required is dependent on the statement. When a device is said to be an "aid in diagnosis," which is language commonly found in an IFU statement and in labeling, it is meant that the device is to be used in conjunction with other clinical assessments and not in a stand-alone capacity.

If a device is indicated to provide a measurement or is indicated generally as an aid in the diagnosis of diseases that affect the RNFL or optic nerve, performance should be characterized, at a minimum, by a measurement validation study in healthy subjects and those with representative diseases. The performance data should be included in labeling because it helps the user interpret the measurements provided by a device.

If a claim is made that the device can be used as an aid in the diagnosis of a specific eye disease, performance should be characterized via a diagnostic accuracy study. CDRH has guidance regarding statistics for reporting results from studies that evaluate sensitivity and specificity of diagnostic tests.[22]

To ensure safe use of a device, labeling should include appropriate contraindications, warnings, and precautions. Manufacturers should remind device users about the importance of scan quality scores in the interpretation of data, and a minimum acceptable value should be specified. Users should also be informed about the limitations of the reference database study design (including covariates not accounted for), conditions in which measurements may be unreliable, and other variables that may influence the measurement (e.g., media opacity, peripapillary atrophy, extreme optic nerve size, and optic nerve tilt). The broader concern of the FDA in the regulation of these devices is that eye care providers understand the proper role of the device in their diagnostic paradigm and that the measurements not be misinterpreted, which could lead to improper patient management.

To summarize, the regulation of ophthalmic devices for assessing structural endpoints is commonly via the 510(k) path through a substantial equivalence comparison. FDA clearance of a device as a diagnostic tool should not be misinterpreted to mean that the device can be a stand-alone diagnostic modality for specific diseases.

See the sidebar for information about FDA approval of data collected outside the United States.

Comparison and Reconciliation of Imaging Technologies

The imaging technologies OCT, stereophotography, SLP, and CSLO each have specific advantages and limitations in patient management and/or endpoints in clinical trials. OCT measures the intensity of reflected light from an interference pattern and calculates RNFL thickness based on cross-sectional images. SLP measures the thickness of the RNFL based on the birefringence of the layer. CSLO measures the intensity of reflected light at various depths and reconstructs a three-dimensional optic disc surface topographic image. OCT and SLP technologies are primarily assessing nerve fiber layer thickness in the peripapillary region of the optic disc. OCT and CSLO are used for examining optic disc morphology. SD-OCT further segments the retina by measuring the GCC as well.

Variability Measurements: CV, ICC, and RC

To distinguish true biological change (e.g., due to pathology such as glaucoma) from normal measurement variability, the amount of change measured must significantly exceed the normal test–retest variability. Several variability measurements are useful for comparing data about RNFL, optic disc, and GCC obtained from the different imaging technologies, particularly for detecting change over time. Useful variability measurements include coefficient of variability (CV), intraclass correlation coefficient (ICC), and the less well-known reproducibility coefficient (RC) or coefficient of reproducibility.

- The CV is the ratio of the within-subject standard deviation to the overall mean. It is useful for comparing the variability of different variables that are in different units of measure. It is expressed as a percentage; the lower the number, the better, meaning less variability.
- The ICC is a ratio of the between-subject variance to the total variance (within-subject variance and between-subject variance). It describes how much of the variance is due to between-subject factors versus within-subject factors. The higher the number, the lower the variability and the better the performance.
- The RC is a measure of the between-visit variability that can have direct clinical applications. The RC is defined as 2.77 times the average intervisit variability of within-subject standard deviation.[23] It is a useful clinical value because it describes the amount of change necessary to reach statistical significance. In other words, if the measured change exceeds the RC, it is statistically significant. The lower the value, the more sensitive the measure is for detecting significant change. This metric is used in studies as a way of quantifying and comparing test–retest variability.[24–27]

OCT Variability. In studies evaluating the CV among OCT measurements, from patients with and without glaucoma, taken in a single visit, interscan variability in the RNFL and GCC is very low.[24–32] Studies measuring the CV for within-session variability also generally show values that are very low, indicating good repeatability.[26–32] For within session repeatability, Tan et al.[30] found that the CV was 1.72 for normal subjects and 2.86 for glaucoma patients with TD-OCT and was 1.09 for normal subjects and 1.25 for glaucoma patients with SD-OCT. Gonzalez-Garcia et al.[28] showed that the CV was 2.33 and 2.26 in normal subjects and glaucoma patients with TD-OCT and was 1.54 and 1.9 for normal subjects and patients with SD-OCT, respectively. Similar values were found by others.[30,31] SD-OCT is better than TD-OCT, probably because of its faster speed, a higher density of data sampled, and more accurate positioning. Variability in glaucoma patients is generally higher than in normal subjects.[26,31]

TABLE 1. Coefficient of Reproducibility

Study	TD RNFL	SD RNFL	SD GCC
Garas et al.[29]		6.01	5.55
Budenz et al.[25]	9.5		
Leung et al.[26]	11.1	4.86	
Mwanza et al.[32]		3.89	

Studies in which within-session ICCs were used have also shown very good results. ICCs from various studies range from 0.89 for TD-OCT (Stratus; Carl Zeiss Meditec), to 0.99 for SD-OCT.[29–33]

Variability measures between examinations are generally slightly higher than those within re-examinations because of the potential for an additional source of variability (different day). Between-visit variability for ICCs and CVs has also been shown to be very good with OCT, and as for within-session studies, SD-OCT tends to perform better than TD-OCT.[25–34]

The most relevant measure for characterizing the ability of an OCT to detect progression is likely to be the coefficient of reproducibility (CR). Table 1 shows these values from several studies comparing TD-OCT with SD-OCT. First, it can be noted that the TD-OCT values were consistently worse (higher numbers mean more change is needed to be significant) than those obtained with SD-OCT. Second, a comparison of the values shows that the thickness change in the RNFL must be twice as much for TD-OCT as for SD-OCT ($\sim$10 μm for TD-OCT versus $\sim$5 μm for SD-OCT).

In summary, studies measuring OCT variability have found very good results, indicating both good repeatability (within-session variability) and reproducibility (between-session variability). In addition, normal subjects tend to have better results than glaucoma patients, especially for TD-OCT. Also, SD-OCT results tend to be better than TD-OCT results. Not much work has been reported comparing the performance of the various SD-OCT devices. However, Seibold et al.,[27] comparing the RCs for three different SD-OCT systems, found that the RC for the RTVue (Optovue, Fremont, CA) was 6.59 μm, compared with 8.89 for Cirrus (Carl Zeiss Meditec) and 11.72 for Spectralis (Heidelberg Engineering). This single study cannot be taken to indicate a definite advantage for one type of SD-OCT over another, but it does underscore the fact that performance may not be the same, even within a given type of technology, let alone between different technologies.

SLP Variability. Looking at the CV, ICC, and RC obtained in repeated measurements during a single session, using the GDx ECC technology (the newer generation SLP; Carl Zeiss Meditec), research shows that the within-session repeatability is very good. CVs are between 1.7% in normal subjects and 3.1% in patients with glaucoma.[35] ICCs have also been shown to be very good, with reports of 0.98 in normal subjects to 0.93 to 0.99 in glaucoma patients.[35,36] Comparing CVs in studies of RNFL using SD-OCT and GDx, researchers found that the CV for SD-OCT was significantly lower, meaning more reproducible, than the CV for SLP.[37]

CSLO Variability. With the Heidelberg Retina Tomograph (HRT: Heidelberg Engineering), the between-visit ICC and CV have been shown to be good, especially in rim area and mean cup depth.[37] Strouthidis et al.[38] found that the CV ranged from 7% for rim area to 28% for cup shape, whereas the ICC ranged from 0.86 for cup shape to 0.97 for cup volume. Comparing between-visit variability of optic disc measurements from TD-OCT and HRT showed good reproducibility for both technologies for most parameters. The ICC for the rim area was significantly better for HRT than for TD-OCT (0.946 vs. 0.86; $P < 0.001$).

FDA Approval of Data Gathered Outside the United States

All clinical studies performed *inside* the United States in support of an FDA 510(k) or premarket approval (PMA) must be conducted in accordance with the Investigational Device Exemption (IDE) regulation. In contrast, the FDA does not have jurisdiction over clinical studies performed outside the United States. However, the agency encourages sponsors to follow a uniform protocol at all investigational sites.

U.S. regulations do allow data from research conducted solely outside the United States to be used in support of U.S. approval of drugs and medical devices. The population being tested must be comparable to the treatment population in the United States, and the study design must be consistent with U.S. medical practice. Informed consent must be obtained on all patients in conformance with the Declaration of Helsinki. Furthermore, studies must be performed by clinical investigators of recognized competence. Data must be considered valid without the need for FDA on-site inspection; however, if necessary, the FDA may validate the data through an on-site inspection or other appropriate means.

Confounding factors considered in determining the applicability of foreign data to the U.S. population include demographic factors, clinical factors, population/system-related factors, and protocol-related factors. Any of these can significantly affect the applicability of data. Confounding demographic data can relate to race, sex, ethnicity, age, socioeconomic status, or educational status, for example. Confounding clinical variables could include prevalence of smoking, diabetes or obesity; compliance with medical regimen or follow-up; education level (e.g., ability to understand directions); and language and cultural differences that might affect the collection of information. Other possible considerations are concomitant medication use, differing physician and medical practices, legal factors, and the use of adjunct devices.

Examples of protocol-related confounding variables are inclusion and exclusion criteria, procedural characteristics, and the test materials being used. The FDA considers these and the mentioned confounding factors to determine applicability of foreign data. The statistical methodology for demonstrating applicability of foreign data to the U.S. population and medical practice is to show baseline homogeneity and outcome comparability. Multivariate regression modeling and propensity score analysis can be used to adjust for covariate differences.

The FDA invites and encourages sponsors to request a meeting before starting international clinical studies. Many questions are addressed at http://www.fda.gov/cdrh/devadvice. Additional information can be requested by calling the FDA Division of Ophthalmic, Neurologic, and ENT Devices, at 301-796-5620.

Researchers working with non-U.S. data should also consult the Code of Federal Regulations (21 CFR 312.120) concerning good clinical practice (GCP) and the role of FDA on-site inspections.

Although the FDA does not accept, as primary support for a marketing application, a study that does not meet the conditions just described, it will examine data from such a study.

Correlating SD-OCT with TD-OCT

The correlation between TD-OCT RNFL thickness and SD-OCT RNFL thickness was investigated by Gonzalez-Garcia et al.[28] In general, they found good correlation: $R^2 = 0.81$ in normal subjects and 0.86 in glaucoma patients. The same group found a slightly weaker correlation in optic disc measurements. Several other studies have found a similar result, that the TD- and SD-OCT measurements correlate highly, but there are absolute thickness differences between them, indicating that the measurements are not interchangeable.[27,39–43]

Correlation of Imaging Technologies with Visual Field Outcome

Although all the imaging technologies operate on different principles and measure different structures, they all generally show good correlation with visual field outcome.[26,42–44] Correlation between RNFL thickness and visual fields with TD- and SD-OCT are similar. For example, Sehi et al.[42] found similar correlation to visual field damage (pattern standard deviation; PSD) for TD- and SD-OCT (SD-OCT $r = -0.40$ vs. TD-OCT $r = -0.37$)[42] while Leung et al.[26] showed similar results with mean deviation, MD (SD-OCT $R^2 = 0.577$ vs. TD-OCT $R^2 = 0.621$). In addition, Bowd et al.[43] found that the correlation with visual field sensitivity was better for TD-OCT ($R^2 = 0.38$) compared with either CSLO ($R^2 = 0.25$) or SLP ($R^2 = 0.21$).

To summarize this comparison of imaging technologies, measurements using OCT, SLP, and CSLO to quantify the same structure generally show good reproducibility and correlate well with visual fields. However, the absolute values are very different, and measurements between technologies are not interchangeable. Even within OCT technology, significant differences are found between TD-OCT and SD-OCT. Differences between imaging devices, even in the same structure (e.g., RNFL), suggest that longitudinal studies must use the same technology for appropriate comparisons.

The reproducibility of these instruments suggests that small structural changes can be detected. The challenge is to identify a clinically significant change that is greater than that expected from normal aging and that is associated with future visual function changes.

Other Considerations in Comparing Imaging Technologies

Comparing imaging technologies for detecting change related to glaucoma is challenging, for several additional reasons:

- There is no gold standard against which a new instrument can be held.
- Instruments vary in the structures that they assess (e.g., neural versus nonneural tissue), the criteria each uses to define change, their registration methods and segmentation algorithms, and the quality of images that they record.
- Change is population dependent, meaning it is influenced by factors such as age, disease severity, disc size, axial length, fluctuations in IOP, prior glaucoma-related surgery, and frequency of testing.
- Instruments and software are still evolving (improving), meaning that new structures or values constantly enter the picture.

Linda Zangwill, PhD, stated at the symposium that agreement between analysis strategies within an imaging instrument and across different instruments varies and can be influenced by factors such as image quality, age, axial length, and glaucoma severity. For example, among the technologies for evaluating glaucomatous progression, studies suggest that agreement between HRT topographic change analysis (TCA) and qualitative photographic assessment varies between 56% and 80%.[38,39] Good evidence shows that HRT baseline parameters are interchangeable with photo cup:disc ratio in models designed to predict the development of glaucomatous change.[45–51]

Using different imaging instruments, researchers have found consistent evidence of approximately a five times faster rate of thinning of the nerve fiber layer and rim area in progressing eyes (meaning those that are changing by stereophotograph based optic disc and/or visual field criteria) compared to nonprogressors. For example, TD-OCT revealed a rate of change in the RNFL that was five times faster in progressors than in nonprogressors and a rate of change in the average cup size that was four times faster than in nonprogressors.[21] Similarly, in an analogous patient population, rate of change in the SLP nerve fiber layer thickness, and CSLO rim area loss was about five times faster in eyes with detectable optic disc and visual field damage.

Comparisons within the same study population show SLP nerve fiber layer changes that are six times faster in progressing than nonprogressing eyes, yet no significant differences between the groups in CSLO rim area loss. The nonsignificance appears to be related to surgical interventions. When the researchers removed patients who had glaucoma surgery to reduce IOP from the analysis, the rate of HRT rim area loss was, indeed, significantly greater (three times faster) in the progressing eyes than in the nonprogressing eyes. Several studies, using HRT, have shown that topographic changes can occur with the lowering of IOP after trabeculectomy or medical treatment. Therefore, surgery may have led to more variability in the rim area measurements, making it more difficult to detect difference between progressing and nonprogressing eyes. As this example illustrates, in choosing endpoints in a clinical trial, it is not only important to understand the technologies, but also to understand other factors that influence measurements.

In terms of progressive optic disc damage relative to visual field loss, recent photo-based analysis shows that a patient with optic disc progression is nearly 26 times more likely than a nonprogressor to develop visual field loss.[8] Similarly, those with visual field progression, were approximately three times more likely to have HRT-documented topographic change than were those without visual field change.[44] These studies provide strong evidence that structural changes are predictive of future visual field loss in glaucoma.

Furthermore, there is close agreement between estimates of the number of retinal ganglion cell somas derived from visual field sensitivity and estimates of the number of retinal ganglion cell axons from TD-OCT data in experimental normal and glaucomatous monkey eyes and in human normal and glaucomatous eyes. Specifically, the model that Harwerth et al.[52] constructed from clinical studies of aging in normal eyes and in clinical glaucoma showed a strong correlation between retinal ganglion cell estimates from standard clinical perimetry and OCT ($R^2 = 0.94$). The researchers further confirmed a relationship between RNFL thickness and visual sensitivities from clinical perimetry. The challenge is to understand how neural and nonneural tissue affects measurements and how abnormalities in structure and function correlate in progressive stages of glaucoma severity.[43]

In summary, imaging instruments are not interchangeable. Patient and instrument factors influence agreement of instrument measurements. There is a consistent association between structural change and visual field deterioration, although it varies by the technique used, instrument factors (e.g., image and visual field quality and summary parameters evaluated), and patient characteristics such as age, IOP, and disease severity. The relationship between structure and function improves by measuring, characterizing, and reducing sources of variability.

In adopting structural measure as endpoints in clinical trials, it is important to understand the strengths and limitations of techniques and to assure that the quality of the data is good. For optimal results, instruments at all locations for a single study must be calibrated to each other, and the assessment of data must be standardized.

What Imaging Technologies Provide the Most Reproducible, Objective, and Quantifiable Measures? Are Any (or All) Appropriate for Clinical Trials?

Ultimately, the choice of technologies is likely to depend on the structure and structure–visual function relationship being studied. For example, SD-OCT analysis of nerve fiber layer thickness may be best for comparing structure–visual function, whereas CSLO measurements of the optic nerve head may be more important for analyzing long-term structural progression of glaucoma, but not necessarily visual function. Additional data are needed to confirm that one or another instrument is uniquely capable and to determine whether the more useful data will be that which replaces current visual function measurements or that which is complementary to visual function measurements.

David Garway-Heath, MD, found a high correlation between visual field sensitivity predicted from SLP images and visual field sensitivity that actually occurred.[44] The Pearson correlation coefficient comparing visual fields predicted from SLP images with measured visual field sensitivity was 0.81, which agrees very favorably with the correlation coefficient of 0.89 between visual fields measured twice in the same patients. Dr. Garway-Heath feels that additional studies are needed to compare cross-sectional and longitudinal measurements from other instruments to visual field measurements, since extracting correlations from the literature would be impossible.

In the opinion of Felipe Mederios, MD, PhD, the data suggest that all the devices are similar in terms of reproducibility of measurements. Further, he pointed out, complementary data are useful only if the information is beneficial to the patient and, because of the curvilinear relationship between structure and function over time, not all complementary data would be similarly useful at all stages of glaucoma progression. Structural changes are more readily observable in early disease; functional change is more readily observable later in the disease process.

In terms of rates of change of retinal structures in glaucoma, Dr. Leung reported that researchers in his laboratory do not see good agreement between the rate of change in the RNFL and in the optic nerve rim,[53] which suggests that detectable change is not parallel and that the synchronicity of change in the various structural components of the glaucomatous eye and in similar patients would have to be determined.

Dr. Chambers pointed out that (1) what is truly important to a patient is change that affects his or her life and (2) a correlation between patient-reported outcomes and visual field change and/or structural change would be a useful measure. However, many patients in clinical trials have very early disease and weak correspondence, if any, between structure and function. In other words, functional measurements are likely to be fairly flat while structure is undergoing change. Dr. Weinreb suggested that it may sometimes be useful to have clinical trial entry criteria that also include patients with later-stage disease who have a stronger correlation between structure and function. Among other advantages, by enrolling patients at a stage at which glaucoma is more advanced, the timeline and cost of longitudinal studies could be reduced.

Drs. Chambers and Eydelman emphasized the need for agreement to be spelled out and published on what—from a visual field perspective— constitutes disease progression in glaucoma. Over the years, as technology has improved, clinical trial parameters have shifted. For instance, the Early Manifest Glaucoma Trial (EMGT) change criteria were all based on the statistical program for the Humphrey Zeiss Perimeter (Carl Zeiss Meditec), which was not available for earlier trials. Researchers also tend to use different criteria based on clinical scenarios, some using event-based change criteria and others preferring trend criteria. Dr. Zangwill pointed out that researchers are moving toward using the EMGT criteria, made easier by software for calculating visual field event-based and trend-based change. Dr. Eydelman suggested that the glaucoma subspecialty should reach consensus to define and publish a scientifically based definition for progression in subgroups of patients. "We [the FDA] would be happy to look at it and to use it to move forward," continued Dr. Eydelman. That consensus could become the bar for showing a correlation between functional and structural measures in glaucoma, to demonstrate efficacy of a neuroprotective drug or device to the FDA.

A suggestion was made that NEI form a panel to establish standards against which trial data could be compared, much as the National Institute on Aging did in establishing definitions and consensus about progression in Alzheimer's disease.

Dr. Chambers continued: "There is a minimum threshold of change ('progression') that is clinically relevant to most patients [or subgroup of patients]. Once this progression is defined, you then know what manifestation of the disease is best to avoid. If this progression can be defined by a structural measure that is synonymous and/or closely related to this manifestation, then you are dealing with what is essentially a surrogate for that parameter."

Dr. Chambers explained further that, given that visual fields predict clinically significant visual deficits (i.e., visual fields are surrogates), as long as the structural metric is highly correlated with the visual field, it can also be used as a surrogate to predict clinically significant visual field progression. Risk versus benefit is also a primary FDA concern.

Role of Patient-Reported Outcomes in Assessment of Glaucoma

A patient-reported outcome is a measurement of any aspect of a patient's health status that is reported directly by the patient, free of interpretation by a physician, researcher, or other person. It is an account of how the patient functions or feels relative to a health condition or therapy.—Varma et al., The 2009 NEI/FDA Clinical Trial Endpoints Symposium.[6]

Eva Rorer, MD, of the FDA Center for Devices and Radiologic Health discussed the agency's guidance document on the use of patient-reported outcome (PRO) measures in medical product development to support labeling claims.[54] The purpose of the PRO guidance is to emphasize that, when appropriate, the FDA does recognize the importance of the patient's perspective. It explains how the FDA reviews evidence that a PRO instrument measures the concept represented by a treatment benefit claim. (A 2009 NEI/FDA Endpoints Symposium specifically addressed the use of patient-reported outcomes in medical product development in ophthalmology.[6])

A PRO instrument (i.e., a questionnaire) is a means of capturing PRO measurement data plus all the information and documentation that support its use. The instrument requires a development process to ensure that it is well defined and reliable. The "concept" being measured, such as a symptom or effect on a particular function, represents an aspect of how patients function or feel in relation to a health condition or its treatment.

The evaluation of a PRO instrument used in a clinical trial to support claims in medical product labeling includes consideration of (1) qualitative evidence demonstrating the extent to which the instrument measures the concept of interest in the intended population (content validity) and (2) how closely the instrument measures matches the concept underlying the targeted labeling claim. Other measurement properties consid-

ered in the review of a PRO instrument include construct validity, reliability, and ability to detect change.

In general, PROs are not currently used as primary effectiveness endpoints in clinical trials to support marketing of ophthalmic devices; however, they are often used in premarket studies as indicators of safety. They are also sometimes used as endpoints in postmarket studies.

With regard to new implantable glaucoma devices, the PROs of most interest to the FDA are potential side effects (e.g., pain or discomfort, foreign body sensation, droopy eyelid, dry eye, tearing, and red eye). However, to date, no PRO instrument has been fully validated to evaluate PROs related to implantable glaucoma devices. As an initial step toward the validation process, the FDA has entered into a collaborative agreement with the University of Michigan to analyze the relationships among the items of the Symptoms and Health Problems Chart questionnaire of the Collaborative Initial Glaucoma Treatment Study (CIGTS).[55]

Dr. Rorer emphasized that the FDA recommends including PRO measures in clinical trials. The FDA encourages industry and academia to collaborate in developing well-defined and reliable PRO instruments that focus on the impact of treatment with glaucoma products. PRO instrument development should begin with an end in mind—the targeted labeling goals—and should begin early in medical product development. Finally, PRO instrument development documentation should include empiric evidence supporting content validity and other measurement properties in the intended clinical trial target population. Specific questions about PRO measures to support labeling claims should be directed to the appropriate product review branch of FDA.

Summary

At this daylong NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium planned collaboratively by the NEI, the FDA, and glaucoma experts, the researchers and clinicians described OCT, stereophotography, SLP, and CSLO criteria for identifying clinically significant structural change related to glaucoma progression and treatment that is greater than that expected from normal aging and is associated with future visual function changes. The FDA presented its position on using the structural metrics for detecting progression of glaucoma in clinical trials of glaucoma drugs and devices. The position of the FDA is that it is the responsibility of the glaucoma research community to establish definitions of glaucoma progression and consensus about structural–functional relationships that characterize early, moderate, and late stages of the disease. With that, the FDA would be willing to accept a structural parameter as the basis for an approval of a drug or device to treat glaucoma.

Dr. Chambers emphasized that any definition presented by the glaucoma community must have clinical relevance for the patient. Once researchers establish definitions empirically, the FDA will evaluate them. The FDA is also willing to consider drugs or devices for subgroups of patients with glaucoma. The agency reiterated its position from an earlier NEI/FDA clinical trial design and endpoints symposium that it will consider PROs as endpoints in clinical trials in ophthalmology. The key to using PROs is developing well-defined and reliable instruments. Similarly, the key to using structural endpoints in clinical trials of glaucoma is establishing an association with visual field measurements that the FDA already accepts as a surrogate functional metric that in turn strongly predicts a functional change that will be important in a patient's everyday life.

It is anticipated that using optic disc and retinal structural characteristics in clinical studies of drugs and devices for detecting and treating glaucoma progression will reduce the time and cost of clinical trials and the burden of vision loss related to glaucoma.

Acknowledgments

The authors thank all program participants and the NEI Office of the Clinical Director, Frederick L. Ferris III, MD, which organized this 2010 NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium; the Association for Research in Vision and Ophthalmology (ARVO) for managing the meeting; and Elaine A. Richman, PhD, of the bioscience communications firm Richman Associates, LLC, (Baltimore, MD), for her contribution to the writing and clarity of this document.

References

1. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci.* 2009;50:1497–1505.
2. *The Economic Impact of Vision Problems: The Toll of Major Adult Eye Disorders*, *Visual Impairment and Blindness on the U.S. Economy*. Chicago: Prevent Blindness America; 2007.
3. www.glaucoma.org. Accessed September 22, 2011.
4. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol.* 2006;90:262–267.
5. Csaky KS, Richman EA, Ferris FL. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci.* 2008;49:479–489.
6. Varma R, Richman EA, Ferris FL III, Bressler NM. Use of patient-reported outcomes in medical product development: a report from the 2009 NEI/FDA Clinical Trial Endpoints Symposium. *Invest Ophthalmol Vis Sci.* 2010;51:6095–6103.
7. *Guidance for Industry: Qualification Process for Drug Development Tools.* Draft Guidance. Washington, DC: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; October 2010
8. Medeiros FA, Alencar LM, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Prediction of functional loss in glaucoma from progressive optic disc change. *Arch Ophthalmol.* 2009;127:1250–1256.
9. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science.* 1991;254:1178–1181.
10. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol.* 2005;123:464–470.
11. Medeiros FA, Zangwill LM, Alencar LM. Detection of glaucoma progression with stratus OCT retinal nerve fiber layer, optic nerve head, and macular thickness measurements. *Invest Ophthalmol Vis Sci.* 2009;50:5741–5748.
12. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology.* 2009;116:1257–1263.
13. Leung CK, Cheung CY, Lin D, et al. Longitudinal variability of optic disc and retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci.* 2008;49:4886–4892.
14. Rao HL, Leite MT, Weinreb RN, et al. Effect of disease severity and optic disc size on diagnostic accuracy of RTVue spectral domain optical coherence tomograph in glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52:1290–1296.
15. Weinreb RN, Medeiros FA. Is scanning laser polarimetry ready for clinical practice? *Am J Ophthalmol.* 2007;143:674–676.
16. Mohammadi K, Bowd C, Weinreb RN, et al. Retinal nerve fiber layer thickness measurements with scanning laser polarimetry predict visual field loss. *Am J Ophthalmol.* 2004;138:592–601.
17. Alencar LM, Zangwill LM, Weinreb RM, et al. A comparison of rates of change in neuroretinal rim area and retinal nerve fiber layer thickness in progressive glaucoma. *Invest Ophthalmol Vis Sci.* 2010;51:3531–3539.
18. Mai TA, Reus NJ, Lemij HG. Structure–function relationship is stronger with enhanced corneal compensation than with variable

corneal compensation in scanning laser polarimetry. *Invest Ophthalmol Vis Sci.* 2007;48:651–658.
19. Medeiros FA, Zangwill LM, Alencar LM, et al. Rates of progressive retinal nerve fiber layer loss in glaucoma measured by scanning laser polarimetry. *Am J Ophthalmol.* 2010;149:908–915.
20. See JL, Nicolela MT, Chauhan BC. Rates of neuroretinal rim and peripapillary atrophy area change: a comparative study of glaucoma patients and normal controls. *Ophthalmology.* 2009;116:840–847.
21. Weinreb RN, Zangwill LM, Jain S, et al. Predicting the onset of glaucoma: the confocal scanning laser ophthalmoscopy ancillary study to the Ocular Hypertension Treatment Study. *Ophthalmology.* 2010;117:1674–1683.
22. *Guidance for Industry and FDA Staff. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests.* Washington, DC: U.S. Department of Health and Human Services. Food and Drug Administration. Center for Devices and Radiological Health; March 13, 2007.
23. Bland JM, Altman DG. Measurement error. *BMJ.* 1996;313:744.
24. Budenz DL, Fredette MJ, Feuer WJ, Anderson DR. Reproducibility of peripapillary retinal nerve fiber thickness measurements with Stratus OCT in glaucomatous eyes. *Ophthalmology.* 2008;115:661–666.
25. Budenz DL, Chang RT, Huang X, Knighton RW, Tielsch JM. Reproducibility of retinal nerve fiber thickness measurements using Stratus OCT in normal subjects and glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2005;46:2440–2443.
26. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology.* 2009;116:1257–1263.
27. Seibold LK, Mandava N, Kahook MY. Comparison of retinal nerve fiber layer thickness in normal eyes using time-domain and spectral domain optical coherence tomography. *Am J Ophthalmol.* 2010;150:807–814.
28. Gonzalez-Garcia AO, Vizzeri G, Bowd C, Medeiros FA, Zangwill LM, Weinreb RN. Reproducibility of RTVue retinal nerve fiber layer thickness and optic disc measurements and agreement with Stratus OCT measurements. *Am J Ophthalmol.* 2009;147:1067–1074.
29. Garas A, Vargha P, Hollo G. Reproducibility of retinal nerve fiber layer and macular thickness measurement with the RTVue-100 optical coherence tomography. *Ophthalmology.* 2010;117:738–746.
30. Tan O, Chopra V, Lu AT, et al. Detection of macular ganglion cell loss in glaucoma by Fourier domain optical coherence tomography. *Ophthalmology.* 2009;116:2305–2314.
31. Kim JS, Ishikawa H, Sung KR, et al. Retinal nerve fiber layer thickness measurement reproducibility improved with spectral domain optical coherence tomography. *Br J Ophthalmol.* 2009;93:1057–1063.
32. Mwanza JC, Chang RT, Budenz DL, et al. Reproducibility of peripapillary retinal nerve fiber layer thickness measured with Cirrus HD-OCT in glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2010;51:5724–5730.
33. Kim JS, Ishikawa H, Gabriel ML, et al. Retinal nerve fiber layer thickness measurement comparability between time domain optical coherence tomography (OCT) and spectral domain OCT. *Invest Ophthalmol Vis Sci.* 2010;51:1896–1902.
34. Mori S, Hangai M, Sakamoto A, Yoshimura N. Spectral-domain optical coherence tomography measurement of macular volume for diagnosing glaucoma. *J Glaucoma.* 2010;19:528–534.
35. Sehi M, Guaqueta DC, Greenfield DS. An enhancement module to improve the atypical birefringence pattern using scanning laser polarimetry with variable corneal compensation. *Br J Ophthalmol.* 2006;90:749–753.
36. Mai TA, Reus NJ, Lemij HG. Retinal nerve fiber layer measurement repeatability in scanning laser polarimetry with enhanced corneal compensation. *J Glaucoma.* 2008;17:269–274.
37. Garas A, Toth M, Vargha P, Hollo G. Comparison of repeatability of retinal nerve fiber layer thickness measurement made using the RTVue Fourier domain optical coherence tomography and the GDx scanning laser polarimeter with variable or enhanced corneal compensation. *J Glaucoma.* 2010;19:412–417.
38. Strouthidis NG, White ET, Owen VMF, et al. Factors affecting the test-retest variability of Heidelberg retina tomography and Heidelberg retina tomography II measurements. *Br J Ophthalmol.* 2005;89:1427–1432.
39. Lin D, Leung CK, Weinreb RN, Cheung CYL, Li H, Lam DSC. Longitudinal evaluation of optic disc measurement variability with optical coherence tomography and confocal scanning laser ophthalmoscopy. *J Glaucoma.* 2009;18:101–106.
40. Sung KR, Kim DY, Park SB, Kook MS. Comparison of retinal nerve fiber layer thickness measured by Cirrus HD and Stratus optical coherence tomography. *Ophthalmology.* 2009;116:1264–1270.
41. Knight OJ, Chang RT, Feuer WJ, Budenz DL. Comparison of retinal nerve fiber layer measurements using time domain and spectral domain optical coherent tomography. *Ophthalmology.* 2009;116:1271–1277.
42. Sehi M, Grewal DS, Sheets CW, Greenfield DS. Diagnostic ability of Fourier domain vs time domain optical coherence tomography for glaucoma detection. *Am J Ophthalmol.* 2009;148:597–605.
43. Bowd C, Zangwill LM, Medeiros FA, et al. Structure-function relationships using confocal scanning laser ophthalmoscopy, optical coherence tomography, and scanning laser polarimetry. *Invest Ophthalmol Vis Sci.* 2006;47:2889–2895.
44. Zhu H, Crabb D, Schlottmann G, et al. predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci.* 2010;51:5657–5666.
45. Kourkoutas D, Buys YM, Flanagan JG, et al. Comparison of glaucoma progression evaluated with Heidelberg retina tomograph II versus optic nerve head stereophotographs. *Can J Ophthalmol.* 2007;42:82–88.
46. Chauhan BC, Nicolela MT, Artes PH. Incidence and rates of visual field progression after longitudinally measured optic disc change in glaucoma. *Ophthalmology.* 2009;116:2110–2118.
47. Chauhan BC, McCormick M, Weinreb RN, et al. Optic disc and visual field changes in a prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography with conventional perimetry and optic disc photography. *Arch Ophthalmol.* 2001;119:1492–1499.
48. Bowd C, Balasubramanian M, Weinreb RN, et al. Performance of confocal scanning laser tomograph topographic change analysis (TCA) for assessing glaucomatous progression. *Invest Ophthalmol Vis Sci.* 2009;50:691–701.
49. O'Leary N, Crabb DP, Mansberger SL, et al. Glaucomatous progression in series of stereoscopic photographs and Heidelberg retina tomograph images. *Arch Ophthalmol.* 2010;128:560–568.
50. Alencar LM, Bowd C, Weinreb RN, et al. Comparison of HRT-3 glaucoma probability score and subjective stereophotograph assessment for prediction images. *Arch Ophthalmol.* 2008;49:1898–1906.
51. Medeiros FA, Zangwill LM, Bowd C, Vasile C, Sample PA, Weinreb RN. Agreement between stereophotographic and confocal scanning laser ophthalmoscopy measurements of cup/disc ratio: effect on a predictive model for glaucoma development. *J Glaucoma.* 2007;16:209–214.
52. Harwerth RS, Wheat JL, Fredette MJ, Anderson DR. Linking structure and function in glaucoma. *Prog Retin Eye Res.* 2010;29:249–271.
53. Leung CK, Liu S, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a prospective analysis with neuroretinal rim and visual field progression. *Ophthalmology.* 2011;118:1551–1557.
54. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf
55. Janz NK, Wren PA, Lichter PR, Musch DC, Gillespie BW, Guire KE. Quality of life in newly diagnosed glaucoma patients: The Collaborative Initial Glaucoma Treatment Study. *Ophthalmology.* 2001;108:887–897.