

Report From the National Eye Institute Workshop on Neuro-Ophthalmic Disease Clinical Trial Endpoints: Optic Neuropathies

Leonard A. Levin,¹ Mohor Sengupta,² Laura J. Balcer,³ Mark J. Kupersmith,⁴ and Neil R. Miller⁵

¹Departments of Ophthalmology and Visual Sciences, Neurology & Neurosurgery, McGill University, Montreal, Quebec, Canada

²National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States

³Departments of Neurology, Ophthalmology, and Population Health, New York University Grossman School of Medicine, New York, New York, United States

⁴Departments of Neurology, Ophthalmology, and Neurosurgery, Icahn School of Medicine at Mount Sinai and New York Eye and Ear Infirmary, New York, New York, United States

⁵Neuro-Ophthalmology Division, Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

Correspondence: Leonard A. Levin, Montreal Neurological Institute, 3801 University Ave, Room MP116, Montreal, QC H3A 2B4, Canada; leonard.levin@mcgill.ca

Received: June 4, 2021

Accepted: July 13, 2021

Published: November 30, 2021

Citation: Levin LA, Sengupta M, Balcer LJ, Kupersmith MJ, Miller NR. Report from the National Eye Institute Workshop on Neuro-Ophthalmic Disease Clinical Trial Endpoints: Optic neuropathies. *Invest Ophthalmol Vis Sci*. 2021;62(14):30. <https://doi.org/10.1167/iovs.62.14.30>

In association with members of the North American Neuro-Ophthalmology Society and the Neuro-Ophthalmology Research Disease Investigator Consortium (NORDIC), the National Eye Institute (NEI) held a public workshop on Neuro-Ophthalmic Disease Clinical Trial Endpoints, focusing on optic neuropathies, on June 28, 2019. Participants included researchers, clinicians, clinician-scientists, and regulatory authorities, working together to discuss issues relevant to endpoints and outcomes for clinical trials of treatments for optic neuropathies. The workshop was organized by Leonard A. Levin, MD, PhD, Chair of Ophthalmology and Visual Sciences at McGill University; Mark Kupersmith, MD, Chair of NORDIC and Director of the Neuro-ophthalmology Services at New York Eye and Ear Infirmary and Mount Sinai Healthcare System; Neil R. Miller, MD, FACS, Co-Chair of NORDIC and Professor of Ophthalmology, Neurology, and Neurosurgery at Johns Hopkins; Laura J. Balcer, MD, MSCE, NORDIC Chair of Quality of Life Committee; and Roy W. Beck, MD, PhD, Executive Director of the Jaeb Center for Health Research.

The goal of the workshop was to bring together experts in neuro-ophthalmology, previous and current neuro-ophthalmic trials, visual structure and function measurements, outcomes research, quality-of-life (QOL) measures, and regulatory issues and—using both formal presentations and panel discussions—determine the optimum outcome measures for various types of optic neuropathies. The following summarizes the day's presentations, discussions, and recommendations.

Dr. Paul Sieving, Director of the National Eye Institute, welcomed the attendees and provided background on the

rationale for the meeting. Dr. Levin then opened the scientific session, emphasizing that different optic neuropathies often cause different types of visual loss. For example, the major damage in glaucoma usually begins in the peripheral visual field (VF), with preservation of central visual acuity (VA) until late in the disease. Papilledema follows a similar course. In contrast, optic neuritis usually causes rapid central vision loss that progresses rapidly over a week or two and then slowly improves over up to a year. Leber hereditary optic neuropathy (LHON), a maternally inherited disorder associated with mutations in mitochondrial DNA, affects central acuity like optic neuritis, but improvement is uncommon. Hence the necessity that clinically meaningful visual endpoints be chosen specifically for the disease being studied.

LESSONS LEARNED FROM ENDPOINTS IN OPTIC NEUROPATHY TREATMENT TRIALS

Idiopathic Intracranial Hypertension

Dr. Michael Wall summarized the NEI-supported Idiopathic Intracranial Hypertension Treatment Trial (IIHT). Idiopathic intracranial hypertension (IIH) is characterized by increased intracranial pressure associated with papilledema but without an intracranial mass lesion and with normal cerebrospinal fluid content. In the IIHT, the primary outcome measure was perimetric mean deviation (MD) rather than visual acuity. Patients were eligible for the trial if the MD in their worse eye was from -2 dB to -7 dB



(−2 dB to give room for improvement and −7 dB because this was the greatest degree of VF loss for which the standard of care allowed nonsurgical intervention). Participants were randomized to either acetazolamide plus diet or placebo plus diet. Participants were monitored for 6 months, during which they underwent serial monitoring of VF, VA, and optic disc appearance. They also were assessed for QOL and had measurements of lumbar puncture opening pressure.

The results showed that acetazolamide plus diet provided more improvement in MD, reduction in papilledema, and improvement in QOL than diet alone but no major difference in visual acuity. The study confirmed that for patients with mild/moderate papilledema in the setting of IIH, MD was a better parameter to measure than VA or other measures of visual function.

Dr. Wall concluded that MD was an excellent outcome variable, that carrying the last observation forward better reflected the outcome than multiple imputation, and that optic disc volume was superior to Frisén papilledema grade.

Leber Hereditary Optic Neuropathy

Dr. Valerie Biousse discussed several treatment trials for LHON: the RHODOS and LEROS trials assessed the effects of the drug idebenone, as well as three studies, one in China, one sponsored by GenSight, and a third sponsored by NEI, assessing the effects of gene therapy.

Dr. Biousse pointed out that because the VA in almost all patients with LHON is or becomes off-chart when the Early Treatment of Diabetic Retinopathy Study (ETDRS) VA chart is used, the decision was made to assess VA using a customized scale. Because of the profound loss of central vision, assessments of the central VF were not helpful in determining treatment effects. In essence, these findings were the reverse of those in the IIHTT. Dr. Biousse also pointed out the potential for using electrophysiology to measure visual outcome.

Dr. Biousse concluded by stating that despite efforts to find a clinically meaningful measure of visual function in treatment trials for LHON, optimal primary outcome measures have yet to be established. Although historically high-contrast VA changes from baseline are used as outcomes in Food and Drug Administration (FDA)–submitted trials, she suggested that low-contrast VA measurements might be more sensitive because the former usually has a floor effect, as do the results of automated perimetry, color vision testing, and RGC layer OCT. Furthermore, OCT of the peripapillary retinal nerve fiber layer (RNFL) is confounded by acute swelling. Ultimately, better structural and functional outcome measures are needed.

RENEW (Opicinumab in Acute Optic Neuritis) Trial

Dr. Diego Cadavid described the RENEW trial in acute optic neuritis of opicinumab, an anti-LINGO-1 monoclonal antibody that showed improvement of central nervous system (CNS) myelin repair in preclinical models. Because VA in patients with optic neuritis improves spontaneously over time, the challenge is to measure a treatment-associated outcome. The RENEW trial was designed to identify evidence of biological activity primarily on CNS remyelination and investigate any potential trends on clinical benefit and neuroprotection as measured by OCT. It was hypothesized that instead of clinical measurements of visual function

such as VA, color vision, or VF, electrophysiologic testing—specifically, the P100 latency of the visual evoked potential (VEP) at 24 weeks—was better evidence for remyelination. VEP was therefore used for the primary endpoint, with secondary outcomes of OCT of the RGC and RNFL, patient-reported outcomes, and VA. The RENEW trial thus had the potential to assess potential correlations between structure (with OCT), function (with VEP and VA), and patient-reported outcomes. The RENEW study was exploratory and not powered for statistical significance on the primary or secondary endpoints.

Results showed that opicinumab resulted in improvement in P100 latency compared to placebo, which is consistent with recovery of latency in the affected eye. The RENEW study included a multifocal VEP substudy to measure more precisely changes in latency and amplitude in the affected and fellow eye. By chance, there were more participants with severe visual loss randomized to opicinumab than placebo.

Glaucoma Neuroprotection Trials

Dr. Jeffrey Liebmann discussed several challenges posed by glaucoma neuroprotection trials. Previous NEI-FDA endpoints meetings provided some guidance on what should be used as functional endpoints, including the results of automated perimetry and measurements of contrast sensitivity, color vision, and VA. However, there is no consensus on structural endpoints. The FDA has stated that it is open to using structural endpoints in clinical trials of new glaucoma drugs, provided that the structural measures predicting functional change strongly correlate with clinically meaningful functional changes.

Glaucoma progresses from a stage when measurements of the RNFL begin to thin, following which early VF loss can be detected, and eventually, severe VF loss and reduction in VA occur. Thus, whereas structural change occurs throughout the spectrum of disease progression, functional change occurs only in the latter part of the glaucoma disease progression scale. Garway-Heath et al.,³⁵ assessing the efficacy of latanoprost versus placebo in patients with open-angle glaucoma (OAG), enrolled 258 patients per arm. The primary outcome measure was VF progression over 2 years. The study showed that latanoprost-treated eyes had better VF preservation and that treatment effects could be detected in as little as 12 months if VF clustering (multiple tests within a short period of time) measurements were used to detect trend-based and event-based differences between the two groups. The key learning points from this study were that VF clustering increases the ability to detect change and that short-duration glaucoma treatment trials with a manageable number of participants can yield meaningful data.

Dr. Liebmann also discussed two large trials that assessed the efficacy of oral memantine for OAG and failed to meet their primary endpoints. Both studies concluded that memantine did not protect against VF loss, but they did show that the FDA was flexible with respect to the endpoints, that it was possible to complete a neuroprotection trial, that there was enough expertise worldwide to execute such a trial, and that enrollment of highest-risk subjects was possible.

A randomized trial (LoGTS trial) of topical brimonidine versus timolol in subjects with normal-tension glaucoma showed a marked difference in outcome between the groups, with brimonidine being neuroprotective. This trial

was the first to demonstrate that an IOP-independent treatment could be used for glaucoma management. Other key lessons learned were that trend analysis can be a potent tool, that studies with small subject numbers are possible with careful selection, and that minimizing the IOP effect may help identify pressure-independent neuroprotection.

These and other late-phase glaucoma neuroprotection trials show that prestudy planning, endpoint selection, study duration, and study design—particularly target population and risk profiling—are crucial for the successful execution of a trial. Industry-sponsored trials should allow sufficient physician input in monitoring the execution of the trial and avoid an insular approach that could negatively affect trial results. In addition, phase II trials are crucial in determining which phase III endpoints should be chosen.

Panel Discussion on Lessons Learned From Endpoints in Optic Neuropathy Treatment Trials

Q. What is clinically meaningful? Can a 1-dB change in mean deviation in a successful trial be clinically meaningful to FDA?

A. Probably not. We need guidance from the FDA not just to incorporate the totality of the data or patient-supported outcomes but for help in deciding and convincing ourselves and others what really is clinically meaningful.

Q. On the issue of subjective measurements of visual functions, the issue of consistency of baseline visual functional measurement is relevant. Do you want to comment on the use of electronic vision assessment (EVA)?

A. EVA is a computer-based method for evaluating visual acuity at a 3-m distance, which has been well validated.^{1,2} EVA decreases the learning effect—the ability of the participants to know where they are on the chart and read the letters. There is also some advantage to what is shown to be a tighter confidence limit on measurements.

Q. The Regenera trial was designed to ensure that any improvement or worsening was real. Can you comment on that?

A. When one doesn't have a strong procedure to make sure that there is no training effect, there will be a training effect. In this case, it was for visual acuity, but the same is true for visual fields and many other functional outcome measures.

Q. When one is stuck with subjective responses, there is an incredible variability in these issues. What have we learned from glaucoma studies?

A. Clustering, trend analysis, and better use of reliable fields are examples of techniques that have made the detection of change much easier and more robust.

CLINICAL MEANINGFULNESS OF ENDPOINTS

Letter Threshold Changes versus Mean Letter Counts in Best-Corrected VA

Dr. Paul Van Veldhuisen discussed letter threshold changes versus mean letter counts in best-corrected VA (BCVA) as an outcome measure for efficacy of treatments.

The following are examples of threshold outcomes for BCVA: slow progression to blindness: $\leq 20/200$ in the better eye for legal blindness, $\geq 20/40$ after cataract surgery, and a 15-letter improvement from baseline. The last example is used in many trials to meet regulatory requirements. This outcome is on a patient level (i.e., it must be easily

interpretable to both patient and clinician). Some statistics of threshold outcomes include risk difference, relative risk, odds ratio, and time-to-event outcome (Kaplan–Meier, hazard ratios).

In contrast to threshold outcomes, continuous outcomes for BCVA take advantage of the full distribution of data, using mean changes from baseline and median changes if distribution is influenced by outliers. Because these outcomes are measured on a group level, they are less interpretable to a clinician or patient, thus raising the question of if it is better to look at mean changes or thresholds when VA is the primary outcome.

The disadvantages of a dichotomous outcome from BCVA are loss of information, misclassification, and both floor and ceiling effects. Loss of information translates into a loss of statistical power, requiring a larger sample size. Another issue when creating a binary outcome based on responder and nonresponder classification is misclassification caused by measurement errors and systematic biases, potentially resulting in both false positives and false negatives. The issue is magnified when there is more variability of data and when there are a lot of data close to the cut point.

Floor and ceiling effects also tend to occur with VA binary outcome measurements. Although strict eligibility on VA can mitigate floor and ceiling effects, this limits generalizability.

VA letter scores may be more appropriate as primary outcomes, but they have their drawbacks. For example, large trials may show a small difference (not clinically meaningful) in treatment and control, irrespective of being significantly different. Threshold outcomes should be considered secondary unless reaching a threshold is the primary objective of the study. When powered on a continuous measure, studies may not have sufficient sample size to detect differences for thresholds.

Visual Fields: Type and Quantity of Field Loss

Dr. Gustavo De Moraes discussed standard automated perimetry (SAP) to assess the functional status in glaucoma and other optic neuropathies. There is no consensus on how to measure functional changes with SAP. The main challenge is to differentiate signal (progression) from noise (test variability).

Progression can be defined by event analysis, which yields a binary outcome. In this approach, measures of test variability should be known a priori to help inform the determination of progression. If the patient remains within a certain variability range over time, they are considered stable. If variability crosses the lower threshold of this range, the patient is deemed progressing. Progression can also be defined by trend analysis, which yields a numeric outcome. In this approach, sensitivity data are plotted over time, resulting in a slope. Instead of a binary measure, a continuous measure is calculated, where the slope gives the rate (or speed) of progression.

A review discussed that VF changes may be acceptable as a clinically relevant primary endpoint by the FDA if a between-group difference in field progression is demonstrated. VF progression will be suspected if ≥ 5 reproducible locations have significant changes from baseline beyond the 5% probability based on event analysis. In their review, De Moraes et al.³ used clinical trial data to show that a 3- to 5-mm Hg IOP reduction results in significant differences in progression between groups using event analysis. Such event-based progression corresponds to an MD slope faster

than 0.5 dB/y in an 18-month trial. A 30% decrease in rate of MD progression could be projected to have a significant effect on QOL. A study confirmed that patients whose VF progresses using event analysis also progress faster than 0.5 dB/y. These studies suggest that rates of MD progression could be used as endpoints in neuroprotection clinical trials.

The more often patients are tested, the better. Clustering of measurements at start and end points helps minimize false-positive results. Sample size can be decreased by 20% to 30% if the study paradigm is shifted from evenly spaced to clustering. This setup also increases the likelihood of finding a clinically meaningful outcome.

Finally, in glaucoma trials, the macula (OCT ganglion cell layer and 10-2 VF) should be assessed to identify progression, as glaucomatous damage to macula is common, can occur early in the disease, and significantly affects QOL.

Microperimetry

Dr. Catherine Cukras discussed microperimetry as an outcome measure for macular function. Macular function is not fully characterized by central VA, and testing VA alone disregards paracentral or central scotomata that strongly affect the patient's self-assessment of visual function. Unlike standard perimetry, fundus-guided perimetry (FGP) reflects the visual sensitivity of the system as a whole and not just the retina. Its specific utility is its eye-tracking ability, which allows for real-time compensation in eye movement and precise identification of a field defect by ensuring stability of fixation. FGP also allows correlation of structure and function and has good test-retest reliability for the same test point over time. There are multiple currently available instruments to perform FGP.

Microperimetry has frequently been used to study retinal disease, especially in structural-functional relationships. For example, areas of geographic atrophy with reduced sensitivity can be identified and subsequent reduction in retinal sensitivity in these areas measured and followed.⁴

There are many uses of microperimetry in glaucoma. The structure-function relationship between VF sensitivities (measured with MP-3 or Humphrey field analyser (HFA)) and RGC layer thickness (adjusted for RGC displacement) were analyzed using linear mixed modeling. The MP-3 perimeter had a test-retest reproducibility similar to that of the HFA in patients with glaucoma; however, it offered a significantly stronger structure-function relationship.⁵

There are multiple modalities for macular function testing beyond VA. All have pros and cons. Microperimetry adds to the psychophysical armamentarium (VA, color vision, and conventional perimetry) by offering the ability to test while imaging, to accommodate noncentral fixators, and to investigate structure-function correlations and longitudinal follow-up of nonfoveal points.

Color Vision

Dr. Brett Jeffrey discussed color vision as an outcome measure. Historically, color vision tests (Ishihara, Hardy-Rand-Rittler (HRR), Farnsworth D-15, and anomaloscope) were developed to detect dichromacy. The Cambridge Color Test was subsequently developed to determine the precise location of the abnormality in protan, deutan, and tritan color-blind individuals; the Low Vision Cambridge Color Test (LvCCT) can be used in patients with VA $\geq 20/800$ and has

been found to be useful in assessing color discrimination in patients with dominantly inherited optic atrophy⁶ and also has been used in a clinical trial of achromatopsia⁷ and in a natural history study of ABCA4 retinopathy, which demonstrated progressive color vision loss in some patients despite stable BCVA. Dr. Jeffrey said that LvCCT has the potential to show progressive deterioration in visual function despite stable VA in natural history studies, provides considerable information about the nature of dyschromatopsia, but is time-consuming (30 minutes/eye).

For rapid quantification of color vision, the cone contrast test is often used.⁸ The test design is in letter-chart format with a 100-point numerical score, cone contrasts, and quantitative categorization of color vision. It is a simple test in which a colored letter appears on a screen with varying degrees of background contrast, and the patient responds yes or no to whether he or she can see the letter. A study⁹ of color impairment after an episode of optic neuritis found that test scores were significantly lower in affected eyes than fellow eyes and that there was a positive correlation between test results and RNFL thickness in both eyes.

Few clinical trials have used color vision as an outcome measure, and few studies have considered how QOL is affected by compromised color vision.¹⁰ These studies include very specific groups (e.g., drivers and medical students), and only partial conclusions can be drawn with respect to the general population.

Contrast Sensitivity and Low-Contrast and Low-Luminance Visual Acuity

Dr. Laura Balcer discussed contrast sensitivity, low-contrast letter acuity, and low-luminance VA as outcomes for patients with multiple sclerosis (MS). This research was sparked by patients with MS who complained that their vision was not right, despite being better than 20/20 when tested with high-contrast VA.

The Optic Neuritis Treatment Trial measured contrast sensitivity as an outcome using Pelli-Robson contrast sensitivity; this measure demonstrated persistent abnormalities in >50% of patients at 6 months. Reductions of contrast sensitivity were also associated with worse National Eye Institute 25-item Visual Function Questionnaire (NEI-VFQ-25) performance, even years later.¹¹

In initial MS studies, Balcer and colleagues¹² compared contrast sensitivity, low-contrast letter acuity (Sloan charts), and color vision. Although all tests distinguished patients from controls, low-contrast letter acuity best identified patients with MS versus disease-free controls. A phase III trial in MS showed that low-contrast letter acuity was able to demonstrate treatment effects and to capture sustained visual loss not detected by high-contrast VA. This study showed that, for MS, studies assessing treatment effects should target low-contrast letter acuity.

Low-contrast letter acuity is an excellent measure of visual function in patients with MS, whereas OCT measurements provide useful structural information. The RNFL is the only part of the visual sensory system where unmyelinated axons are present and can be visualized noninvasively. Walter et al.¹³ showed that reduced low-contrast letter acuity was associated with ganglion cell layer thinning by OCT and with worse NEI-VFQ-25 scores. Talman et al.¹⁴ showed that RNFL thinning over time was associated with worsening low-contrast letter acuity in MS, even in the absence of optic neuritis.

Driving, Reading, Mazes, and Other Higher-Order Functions

Dr. Cynthia Owsley said that patients evaluate treatment success not by the number of letters read on a chart but on how well they can engage in visual activities of daily life. Two ways to assess this are visual task performance and patient-reported outcome questionnaires. Dr. Owsley discussed visual task performance measures, including driving, reading, mazes, visual processing speed, and physical activity, as endpoints in observational studies or clinical trials.

A range of driving simulators are available, such as PC-based simulators with a steering wheel and gas pedal, a cab from a real vehicle placed in front of central and peripheral screens, and virtual reality devices with moving bases, vibration, and proprioceptive feedback. Dependent measures include lane boundary crossings, average speed, pedestrian detection, obstacle detection, obeying traffic control devices, and the impact of secondary tasks (e.g., texting).

The FDA allows the use of driving simulation to evaluate drug safety (e.g., to see if the effect of sleeping medications persist in the morning and to differentiate sedating from nonsedating antihistamines). The FDA has not, however, used driving simulation to establish treatment efficacy for vision.

The most commonly used reading task in observational studies and clinical trials is the MNRead Acuity Chart. This chart requires the subject to read a series of sentences written at a third-grade level and provides measures of reading acuity, maximum reading speed, and critical print size. The International Reading Speed Texts (IREST) test requires the subject to read a paragraph written at a sixth-grade level and measures average reading speed. The rationale behind reading a paragraph rather than single sentences is that it requires more sustained reading, which vision-impaired people find more difficult than reading an individual sentence.

Dr. Jean Bennett's group developed a multiluminance mobility maze test for phase III RPE-65 gene therapy trials for Leber congenital amaurosis. It establishes validity, reliability, repeatability, and relationship to vision. A change of at least two light levels was considered a clinically meaningful change. The group worked closely with the FDA to meet endpoint criteria and establish efficacy of the intervention. Another maze test, the Pedestrian Accessibility and Mobility Laboratory, was developed by a group at University College London for their 2008 Leber congenital amaurosis gene therapy trial.¹⁵ The investigators reported that the therapy resulted in "significant improvement in subjective test of visual mobility."

Another outcome measure, visual processing speed, is the amount of time (in milliseconds) required to make a correct judgment about a visual stimulus and involves higher-order visual processing. Processing speed is probed under task demands such as divided attention or distraction. Visual processing speed measures have been associated with health and well-being. For example, poor visual processing speed is correlated with higher collision rates, driving performance problems, performance mobility problems, reduced physical activity, and increased time to perform instrumental activities of daily living, such as finding an object in a room.

Regulatory Perspectives of FDA on Functional Outcomes

Dr. Wiley Chambers discussed regulatory perspectives of the FDA on functional outcomes. Premarket review of drugs

and devices occurs under the Food, Drug and Cosmetic Act and for biologics under the Public Health Service Act. The mission of the Center for Drug Evaluation and Research under the FDA is to ensure that safe and effective drugs are available to the American population. That goal is accomplished by monitoring drug development processes during the investigational stages (confidential), approving new drugs that have been proven safe and efficacious (confidential until approval and then designed to be transparent), and monitoring adverse effects after approval.

There are risks associated with all drug products. Assessment of a drug's risks improves as more individuals receive it. Products are approved based on an assessment of risks and benefits of the product when taken as labeled by the intended population. New drug applications require adequate and well-controlled studies to establish safety and efficacy. Isolated case reports, random experience, reports lacking details, and uncontrolled or partially controlled studies are not acceptable as the sole basis for approval of a product.

An adequate and well-controlled trial is one that has a clear statement of objectives; a study design that permits valid comparison; a subject selection method that provides adequate assurance that the subject group has or will develop the condition; minimum bias in assigning subjects to a group; minimum bias on the part of subjects, observers, and analysts; well-defined and reliable method(s) of assessment; and adequate result analysis to assess the effects of the drug.

There is a strong desire to approve products based on how the product helps subjects. Subjective endpoints are patient-reported outcomes that address single- or multiple-domain questions. Single-domain questions may inquire about itching, pain, ocular irritation, and ocular dryness. Usually, a single-domain question presents a 5-point scale of response and expects at least a 1-point mean change. Multiple-domain questions may include QOL measures. These questions are specific to the intended population and require knowledge of how much weight to give to each QOL domain. No ophthalmology QOL measures are currently validated for drug-evaluation research.

Functional endpoints for ophthalmology drug trials frequently measure visual function, which includes but is not limited to high- and low-contrast VA (doubling of visual angle on ETDRS chart); VF, in which a 7-dB change is usually expected over a predefined area of at least 5 points; contrast sensitivity, which uses doubling of visual angle; and activities of daily living, such as the ability to perform tasks in a low-light setting (endpoint is light level). Other functional endpoints acceptable to the FDA but associated with a high level of variability are reading speed, driving performance, and color discrimination.

While many objective measures have been used to approve new drug products, the ability to measure a difference in these endpoints does not necessarily make them clinically relevant. The following are objective measures that have been used as functional endpoints: intraocular pressure (5- to 7-mm Hg reduction), refractive power (50% slowing of progressive change), pupil size (maintenance of 6-mm diameter under bright light), and tear production (increase by 10 mm by Schirmer score).

Anatomic measures must predict a clinical benefit for patients (e.g., prevention of progression of cytomegalovirus retinitis, diabetic retinopathy, retinal detachment, or photoreceptor loss; resolution of anterior chamber cell and flare or conjunctival redness; and reepithelialization of a previously infected cornea). These measures alone

do not require any visual performance measures to show improvement.

Panel Discussion on Clinical Meaningfulness and Functional Endpoints

Q. How are criteria on doubling of visual angle determined?

A. Doubling of the visual angle was defined in the Early Treatment Diabetic Retinopathy Study as the minor change considered to be clinically significant.

Q. What FDA-guided qualification measures are used to determine an outcome?

A. There is an FDA guidance document specifying procedures for developing a patient-reported outcome that everyone is encouraged to follow. In that way, data can be generated that support patient-reported outcome.

Q. Is there an issue with learning effects and how is that handled?

A. VA, VF, etc. are affected by learning effects, but there are methods to minimize it. Low-contrast VA doesn't have a large practice effect compared with some of the other measures. In the walk-in-maze test, there are built-in measures to minimize practice effects.

Q. My question is about the variability or reliability of visual field where there's a new patient or a patient you've been following. There may be some false-positive measurements. How do you determine the true effects considering these factors and including training effect? How many times do you take readings for a single test and which reading is set as the best reading?

A. There is no substitute for multiple tests. In addition, it is important to have a good technician in the room with the patient during the test. He/she can observe the patient's responses and restart the test if necessary.

Q. When designing QOL domains, is it possible to design a tool that asks patients how much that domain means to them? Is there a way to individualize it?

A. The guidance document that I talked about earlier is designed to create a more targeted patient-reported outcome. This can be achieved by picking a population that is best suited for a particular test. Every test is meant for a specific population and the results do not give an accurate interpretation if the patient population doesn't match with the test.

Q. How do you address the variability of terms used by individual patients when assessing a large number of patients?

A. Instead of asking if there is, for example, about itching or dryness in the eye, patients can be asked what is bothering them the most. In follow-up, patient-reported changes throughout the trial duration can be followed.

Q. How does completing a maze describe an effect on visual function?

A. The goal of the maze testing used in a gene therapy trial was to demonstrate the patient's improved ability to see in low-light conditions. The maze was a modification of an activity of daily living and the test evaluated whether the task could be performed at a particular luminance level. After treatment, the test measured the difference in luminance in which the task could also be performed.

ENDPOINTS ISSUES SPECIFIC TO NEURO-OPHTHALMIC DISEASES

Statistical Issues Related to Rare Diseases

Dr. Maureen Maguire discussed statistical issues related to rare diseases. According to the Orphan Drug Act, "rare" is

a prevalence of <200,000 in the United States or <0.06% of the population. These are serious diseases, usually with no effective treatment, and although each disease is rare individually, collectively there are many of them. Some study designs and approaches work around difficulties posed by small sample sizes.

She discussed strategies for trial design of rare diseases. To ensure that a given percentage of reduction in progression by treatment results in a larger difference in outcome, one could select patients who progress more quickly, select the outcome measures that progress the fastest, or follow patients for longer, so that there is more time to accumulate the difference between the groups. Alternatively, decreasing the variability within treatment and control groups would minimize error. This can be done by enrolling patients more homogeneous in progression rates (e.g., VF loss at baseline) and have less day-to-day variation in response (RNFL versus VF versus microperimetry), by increasing the number of measurements for each patient and analyzing repeat measurements as a cluster, and by decreasing testing variability by standardizing the way the test is done and graded.

Another approach to combatting the very small sample size problem is the use of historical controls (e.g., from natural history studies of the disease). One major concern is that the expected course of the treated patients is different from the course of the historical controls. There may be patient selection and informed consent factors for the historical controls. Moreover, there may be temporal changes and other contemporary factors affecting outcome in the treatment group that do not have an equivalent in the historical control group. FDA guidance on rare diseases states that historical controls may be considered when there is an unmet medical need; a well-documented, highly predictive disease course that can be measured objectively; and the expected treatment effect is large, self-evident, and temporally closely associated with treatment.

In the crossover design, a patient receives treatments in a randomized manner. A washout period is included between treatments so that the effect of the first has ended before the second one is started. Advantages of crossover studies are that each person serves as their own control, which reduces variability; a smaller sample size is needed compared with traditional parallel groups; and all patients receive the treatment being tested, which is more acceptable to study participants. Disadvantages are that the approach applies only to chronic, stable, and incurable conditions, such as pain or seizures; it is appropriate only for treatments that do not produce chronic effects; the washout period must be known; response must occur within the treatment window; and carryover and period effects can complicate interpretation.

In *N*-of-1 studies, treatment is given to patients in a randomized order over several periods. Again, a drug with a rapid response and effect loss is required. With longer periods of study, an *N*-of-1 study can help reach conclusions about the tested drug in specific patients. It is possible to combine the results from different patients using meta-analysis techniques. This type of study offers generalizability with a small number of patients and may be the most useful when treatment effect sizes are expected to differ across patients. *N*-of-1 studies have similar advantages and disadvantages to crossover studies.

In a randomized placebo-phase design, patients are randomized to receive treatment or placebo, as in traditional studies. However, the placebo group gets the experimental treatment after a certain amount of time. The assumption is that, if effective, treatment produces a lasting effect, and

patients treated early will respond sooner. The advantage is that all patients get the new treatment, but a disadvantage is difficulty in determining how long the placebo phase should be.

In randomized withdrawal design, all patients receive treatment; those who apparently “respond” are randomized to continuation or withdrawal of treatment (placebo given) and followed. Patients who apparently do not respond are removed from the trial. An advantage of randomized withdrawal design is that it enriches the patient population in the trial with those most likely to respond to the treatment. Disadvantages are that it is difficult to fix the duration of the initial treatment phase, it is only applicable to treatments with no or few lasting effects, and it offers limited generalizability to the general patient population.

Adaptive designs are prospective plans to use data collected during the trial to change aspects of the study design, which usually aim to reduce the study sample size. Adaptive designs require interim monitoring that may lead to a change in randomization ratio based on observed results, change in the randomization design to increase balance on key covariates among treatment groups, or recalculation of sample size or follow-up time based on degree of variability.

In response adaptive randomization, a type of adaptive design, the goal is to maximize the number of patients assigned to the more effective treatment while minimizing the overall *N*. Responses from previously assigned patients are used to adjust the allocation ratio of treatments to a higher probability of the more effective treatment. Limitations are that it requires a very quick response, can effectively unmask investigators to next assignment, and disrupts balance over time, which means that it is vulnerable to temporal drift.

Endpoints Specific to Severe and/or Temporary Visual Loss

Dr. Kay Dickersin emphasized that to be successful, a clinical trial should answer well-framed and defined questions that are based on the planned participants, intervention, comparator, and outcomes. Today, the results of multiple clinical trials (that assessed similar outcomes) are often merged in a systematic review and meta-analysis.¹⁶ Outcomes have five elements, however, and often a few elements are different across trials or are not reported: (1) domain (e.g., IOP, VA), (2) specific measurements (e.g., slit-lamp examination), (3) specific metrics (e.g., value at a time point), (4) method of aggregation (e.g., proportion), and (5) time point(s) of measurement (e.g., 1 month).¹⁷ For example, a systematic review of gabapentin trials revealed that for 21 trials and four prespecified outcome domains (sleep disturbance, QOL, mood, and pain intensity), there were 214 defined effectiveness outcomes (making it impossible to merge the data from all 21 trials).¹⁸ The number of effectiveness outcomes was large because all five elements were considered for each outcome domain. Furthermore, the results for pain intensity varied, depending on which data source was used (e.g., all reports, clinical study reports only, FDA data only).

Dr. Dickersin pointed out that one of the limitations of glaucoma trials found by Lê et al.,¹⁹ Law et al.,²⁰ and Saldanha et al.²¹ is that patient-centered outcomes are rarely measured. These outcomes, which could include the

patients' limitations in performing vision-dependent activities of daily living, problems with visual function or perception, and the burden of medical treatment, could be obtained from qualitative interviews with the patients and could help guide the development, evaluation, and labeling of new interventions for glaucoma. For example, a survey of patients with dry eye syndrome found a significant discrepancy between outcomes described in published research studies and outcomes that patients with dry eye considered important.²²

Core outcome sets (COSs) involve stakeholders, including researchers, patients, practitioners, funders, and policymakers, in the development of the outcomes so that the outcomes are relevant to all. COSs ensure inclusion and do not limit outcome choice. Certain areas of health care have developed COSs, and this has made a major positive impact on their fields (e.g., OMERACT) (see www.comet.initiative.org).

At a 2018 meeting concerning core visual outcomes in AMD, glaucoma, and dry eye, the following suggestions were made: (1) Make COS for each field to assess comparability of trials. (2) COS developers should identify five elements for each outcome. (3) The same practical considerations for developing a COS exist for each disease entity: disease severity, preferred measurement instruments, feasibility of measuring each outcome, and the opinion of patients, clinicians, and other stakeholders in gauging which outcomes are important.²³

If trials are to affect evidence-based health care, their results need to be included in systematic reviews and meta-analyses. Yet trialists and reviewers studying the same disease often do not report the same outcomes. Dickersin asked, how do we decide which outcomes are important for clinical decision making without the information we need?

Clinically Available Structural Endpoints Relevant to Optic Neuropathy

Dr. Randy Kardon explained that clinical trials are concerned with disease monitoring, and there are commercially available tools to analyze linear progression over time. These tools also provide a measure of change from baseline, not only compared with normal but also the interval change among follow-ups. For example, OCT can assess RNFL progression and ganglion cell layer/inner plexiform layer (GCL/IPL) area difference between diseased and age-matched healthy retinas. Changes in structure–function relationships over time also can be detected this way. Group analysis can look at the response to treatment of the group as a whole, using general estimating equations, but may miss individual responses to treatment. Dr. Kardon then demonstrated examples of progressive thinning of RNFL and GCL/IPL over time for a single patient and noted that the thickness of these layers, the probability of change, and the actual change from baseline could all be measured over the study period.

The OCTiMS study measured RNFL thickness and change from baseline for 36 months in patients with MS, finding a significant decrease in the MS group versus controls. In another study, loss of retinal neurons after traumatic brain injury (TBI) over time was measured in normal versus veterans with TBI. The results were mapped for individual participants and showed that RNFL thickness declined in about 12% of the TBI veterans²⁴ and that patients showing degeneration could be specifically identified.

Dr. Kardon also noted that detection of RNFL thickness change results in more artifacts than total retina volume, as found by studies in which OCT quantification of papilledema was performed. He demonstrated a case in which increasing papilledema resulted in Bruch's membrane deformation toward the vitreous, offering an additional feature in detecting changes in intracranial pressure, even in the setting of acute shunt failure or if there is optic atrophy.

Dr. Kardon concluded by stating that structure and function measures can be combined to improve detection and measurement of progression.²⁵ The next steps are applying structure–function to monitoring treatments via artificial intelligence.

Direct Optical Readout of RGC Function at a Cellular Scale

Dr. Juliette McGregor discussed imaging the retina at the individual cell level. Label-free imaging individual RGCs in the living eye has been demonstrated using adaptive optics scanning laser ophthalmoscopy (AOSLO)²⁶ and adaptive optics OCT.²⁷ The latter method has been used to study glaucomatous eyes and showed a reduction in cell density that was correlated with visual function loss.²⁸

Dr. McGregor then discussed high-resolution ophthalmic imaging of extrinsic fluorophores to assess function in individual RGCs. By expressing calcium indicator proteins such as GCaMP6 in RGCs, it is possible to optically read out activity-dependent changes in intracellular calcium as changes in fluorescence using AOSLO. Retinal circuitry,²⁹ loss of function, and therapeutically restored activity³⁰ can be studied in the fovea of the living primate using this approach.

Dr. McGregor described how two-photon excited fluorescence can be used to measure cone function from intrinsic signals.³¹ She also described how electrically active RGCs undergo a change in optical path length that can be detected using phase-sensitive OCT and used as a label-free optical readout of RGC function.³² This has yet to be achieved on a cellular scale, but work is ongoing.

Panel Discussion

Dr. Kupersmith commented that he performed a prospective study of patients with past optic neuritis.³³ Following an attack of optic neuritis, at 6 months, GCL/IPL thickness correlated with 10-2 MD threshold perimetry ($r = 0.43$) but not with low-contrast visual acuity (LCVA). The amount of GCL/IPL thinning correlated strongly with the 10-2 MD ($r = -0.60$) and moderately with LCVA ($r = -0.46$). He said that GCL/IPL thickness appears to be the best structural measure and 10-2 MD may be the most sensitive functional measure for determining residual deficits due to optic neuritis.

Q. Although OCT tests are objective, several trials showed a small effect size that is not significant. Function may need to have the major role because we don't know what the change in structure may actually mean.

A. The effect depends on structure–function correlation, which, in turn, depends on where one starts.

Q. What about using the fellow eye to measure outcomes and treatment effect?

A. We discourage using the fellow eye as subjective endpoints. However, the fellow eye may be used as an objective endpoint and to assess natural history. The same patient can be used as a control, thus increasing efficiency. The

biggest problem with fellow eye assessment is any kind of floor effect.

LEARNING FROM OTHER FIELDS

Endpoints in Regenerative Medicine Clinical Trials

Dr. Brad Kolls discussed how measures of functional recovery are different for different diseases. No single endpoint sufficiently captures structural and functional change or QOL measures. Most of the knowledge about recovery outcome from ischemic stroke is from rehab literature. These define the natural history of physiotherapies and differences in recovery trajectory, among other parameters. Even though the natural history is the same, functional recovery from baseline after stroke will differ for every outcome measure, such as sensory, motor, or visual functions.

The Modified Rankin Score (mRS) is considered the best primary endpoint for stroke therapy, having been repeatedly validated in the past 3 to 4 decades. It is fast and easy, using a 0 to 6 scoring system, where 0 is absence of a symptom and 6 is death. It thus reflects the level of disability and not only if there is impairment. The Stem Cell Therapy in Chronic Stroke Disability Trial focused on trying to improve function in chronically disabled stroke patients 6 to 12 months after stroke with mRS scores of 3 or 4 through intracranial injection of stem cells into the injured hemisphere. After reviewing numerous options for endpoints, the FDA strongly encouraged the mRS to be the primary endpoint because this scale would provide information on “clinically meaningful changes in function.”

The Validated Phone Survey Study is another example of mRS implementation in practice. The study found good agreement between in-person and phone ratings over the whole range of mRS scores (82% agreement), with the scores correlating better with the physical than the mental components of the Short Form 12v2.³⁴

In regenerative medicine, we often are repairing either an established condition or an acute injury and not preventing a specific disease or clinical event. In considering an endpoint to a study, we need to choose a domain outcome that measures anticipated effects (structure, activity, or participation), consider if the endpoint has been used and validated in the study design and disease, and make sure that the outcome has clinical relevance to patients.

Multiple Sclerosis Outcome Assessments Consortium

Dr. Nicholas La Rocca stated that the Multiple Sclerosis Outcome Assessments Consortium (MSOAC) was formed to obtain regulatory approval for a set of endpoints that would accurately reflect outcomes in MS. MSOAC is a consortium of 6 MS patient advocacy groups, 27 academic centers, and 10 pharmaceutical companies. MSOAC was managed by C-Path with funding from the National MS Society and industry. FDA, European Medicines Agency (EMA), and National Institute of Neurological Disorders and Stroke (NINDS) are liaisons. The purpose of this consortium is to obtain approval from the FDA and EMA for a set of measures reflecting MS disability in trials. Currently, the MS disability measure most commonly used in trials is the Expanded Disability Status Scale (EDSS), which ranges from 0 (normal neurologic exam) to 10 (death due to MS). Limitations of the EDSS include that it is a rating scale rather than a

performance measure, it is not an equal interval scale, it is relatively insensitive to change, it has a bimodal distribution; underlying meaning of scores vary along its scale, and it is too dependent on ambulation.

MSOAC is seeking regulatory approval for a set of measures to serve as a coprimary or secondary endpoints for future trials. The new measure will reflect disease progression, be useful for demonstrating clinical change, and will assess domains important to daily activities.

As part of the project, C-Path created a data standard for MS to facilitate pooling and secondary analysis of data from MS trials. Using these data, the consortium analyzed outcomes focused on relapses, progression, walking, manual dexterity, vision, cognition, and quality of life. The consortium also made available a data set from the placebo arms of several MS trials that is available to other researchers.

NORDIC

Dr. Kupersmith described NORDIC, consisting of clinical sites with experienced investigators, coordinators, and technicians. Some sites serve as reading centers for VFs, OCTs, fundus photos, and MRIs. NORDIC is a renewable network for conducting clinical trials sponsored by government agencies, nonprofits, and industry. Investigators associated with NORDIC rely on network leadership to have the network conduct valid and meaningful studies. The network shares results, forms writing committees, and augments careers. Investigators associated with NORDIC develop projects and shape study design, given their breadth of experience in various disorders. Depending on the disorder being studied, NORDIC is prepared to add new investigators from both academic centers and private practices to its network.

Dr. Kupersmith pointed out that it is a challenge to conduct studies in the many disorders where neuro-ophthalmologists are vital. NORDIC maintains flexibility to serve as an academic research organization (ARO), providing skilled sites and reading centers accommodating industry needs. Specifically, it provides contracts; protocols for manual of procedures (MOP) creation, advancement, and approval; biostatisticians; and site performance metrics and corrective actions. It also manages study committees.

A challenge faced by NORDIC is management of vital resources. Many neuro-ophthalmic disorders that cause an optic neuropathy are uncommon or considered rare. This leads to competition for studies on the same disorder. There are a limited number of neuro-ophthalmologists who have the setup to conduct clinical trials.

Pediatric Eye Disease Investigator Group and Diabetic Retinopathy Clinical Research Network

Dr. Michael Repka discussed the Pediatric Eye Disease Investigator Group (PEDIG), which was funded by the NEI in 1997 to conduct a congenital esotropia observational study and an amblyopia treatment study. Subsequently, NEI funded PEDIG as a network, and since then, PEDIG has conducted 47 studies, 10 of which are ongoing. PEDIG includes both community and academic sites in the United States, Canadian provinces, Mexico (although not currently), and the United Kingdom

PEDIG has its own institutional review board and coordinating center. Its sites include private practice ophthalmologists, ophthalmologists in academic medical centers, optometrists at optometry schools, and a few private-

practice optometrists. The endpoints adopted by PEDIG include high- and low-contrast VA, angle of strabismus, control of strabismus, tear film, RNFL thickness using OCT, refractive error, and patient-reported outcomes. To measure visual acuity, EVA, mean of absolute VA outcome (used in randomized amblyopia studies), lines or letter of improvement, threshold acuity, and comparison with normal have been used. QOL measures include EyeQ, PedsQL, and disease- or treatment-specific questionnaires.

Dr. Repka then discussed the Diabetic Retinopathy Clinical Research (DRCR) network. This network's mission is to improve the lives of individuals with retinal pathology by performing high-quality, collaborative, clinical research that leads to a better understanding of retinal diseases and advances their treatment. Principal importance is placed on clinical trials, but epidemiologic outcomes and other research also may be supported. The DRCR uses EVA and macular OCT thickness as endpoints.

Panel Discussion

Q. What's the development and acceptance of home tests for vision testing in MS?

A. A validated home-testing method that provides consistency is not yet in practice, but it is needed. In pediatric cases, sometimes there is no improvement during a time period and then a huge improvement in the succeeding time period in which tests are conducted. Home testing would give us a much better projection of the trajectory. It also might be helpful for longitudinal assessment of the outcomes.

Q. In neuro-ophthalmology studies, we use functional- and structure-based studies rather than the modified ranking scale (EDSS). Is there an equivalent of such type of scales that could be used in large simple trials or registry-based randomization?

A. In the 10-point EDSS scale, many different things are measured. If one looks at the overall score in a large sample, two-thirds of the variance is accounted for by function, so this measure tends to obscure rather than delineate what's happening. The advantage of the mRS is that it not only combines motor function assessment but also includes elements of disability, thus addressing QOL. This makes it an attractive endpoint to use, as well as being fast and easy.

Q. There is a dichotomy between functional and structural outcomes and patient-reported outcomes. How do we bring these together?

A. We don't ask patients via a questionnaire what they think about their improvement during treatment. Getting a driver's license in the United States is completely dependent on VA. This is also important for the patient, and therefore is one factor that can be used to assess patient-reported outcomes objectively. There is always a problem in addressing patient responses as an outcome because in many cases, patient reports don't match objective test results.

CLOSING REMARKS ON FUTURE DIRECTIONS

- Neuro-ophthalmologists should develop a validated scale to measure vision-specific QOL.
- The community should develop ways to report outcomes and choose which quantifiable measures are best for specific conditions.
- Subdomain analyses can be added to ongoing studies in order to develop new outcomes.

- R21 or small R01 programs should be developed by the NIH for exploratory development of new trial methodologies, outcomes, and analyses.
- Definitions should be harmonized to ensure that data collection and interpretation are consistent across trials when there are evolving methods of measurement and output.

SUMMARY

This all-day meeting covered a large number of topics in depth. The following summarizes some of the main conclusions relevant to carrying out clinical trials related to various optic neuropathies.

Papilledema (e.g., idiopathic intracranial hypertension): Because visual acuity and color vision are not affected until severe damage has occurred, mean deviation or other measures of progression on automated perimetry are optimal outcome measures for monitoring the effects of papilledema on the visual system.

Leber hereditary optic neuropathy: Because of the profound loss of central vision in this condition, assessments of the central visual field are not helpful in determining treatment effects. Although the use electrophysiologic measures can be considered, optimal primary outcome measures have yet to be established.

Dominant optic atrophy: The Low Vision Cambridge Color Test can be used in patients with visual acuity >20/800 and has been found to be useful in assessing color discrimination.

Optic neuritis: Given that the visual acuity in patients with optic neuritis improves spontaneously over time, clinical measurements of visual function such as central acuity, color vision, or visual field can be less helpful. Electrophysiologic testing—specifically, the P100 latency of the visual evoked potential—may provide evidence for remyelination, and optical coherence tomography of the retinal ganglion cell/inner plexiform layer and peripapillary retinal nerve fiber layer can serve as secondary outcome measures.

Glaucoma: The results of automated perimetry—with clustering the timing of visual field examinations in order to increase the ability to detect change—can be used as primary outcome measures. Measurements of contrast sensitivity, color vision, and visual acuity can be used but may be less sensitive or specific. There is no consensus on structural endpoints, in part because the degree of correlation with clinically meaningful functional changes is not yet sufficient.

Clinical trial issues applicable to multiple optic neuropathies: A variety of trial designs can be used for rare optic neuropathies, in which the number of participants is likely to be small. Similarly, diseases in which there is severe visual loss can benefit from outcome measures that include patient quality of life. A variety of structural measures continue to be developed, which may eventually serve as primary outcomes once there is evidence for sufficient strength of the association with a clinically meaningful outcome.

Acknowledgments

Speaker Affiliations and Commercial Relationships:

Laura Balcer, New York University Grossman School of Medicine, New York, NY, USA, None;

Valerie Biousse, Emory University School of Medicine, Atlanta, GA, USA, GenSight (C), Neurophoenix (C);

Diego Cadavid, University of Massachusetts, Worcester, MA, USA, X4 Pharmaceuticals (E);

Wiley Chambers, US Food and Drug Administration, Silver Spring, MD, USA, None;

Catherine Cukras, National Eye Institute, NIH, Bethesda, MD, USA, None;

C. Gustavo De Moraes, Columbia University, New York, NY, USA, National Eye Institute/NIH (F), Centers for Disease Control and Prevention (F), Carl Zeiss (R, C, F), Topcon (F), Heidelberg Engineering (F), Novartis (F, C), Reichert (C, F), Galimedix (C), Belite (C);

Kay Dickersin, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, None;

Brett G. Jeffrey, National Eye Institute, Bethesda, MD, USA, None;

Randy H. Kardon, University of Iowa, Iowa City, IA, USA, Heidelberg Engineering (F), Santen (R), Novartis (C), FaceX LLC (I), MedFace LLC (I), National Eye Institute/NIH (F), Department of Veterans Affairs (RR&D Division) (F);

Brad J. Kolls, Duke University School of Medicine, Durham, NC, USA, Corticare (C), Reneuron (F), research on stroke systems of care and stroke recovery (F);

Mark J. Kupersmith, Icahn School of Medicine at Mount Sinai, New York, NY, USA, National Eye Institute/NIH (F), Regenera (C), Palestro Foundation (F), New York Eye and Ear Infirmary Foundation (F);

Leonard A. Levin, McGill University, Montreal, Canada, Canada Institutes for Health Research (F), Aerie (C), Eyevensys (C), Galimedix (C), Genentech (C), Perfuse (C), Quark (C), Regenera (C), Santen (C), Wisconsin Alumni Research Foundation (P); Jeffrey M. Liebmann, Columbia University Irving Medical Center, New York, NY, USA, Aerie (C), Allergan (C), Carl Zeiss Meditec (F), Heidelberg Engineering (F), Genentech (C), Thea (C), Novartis (R);

Nicholas LaRocca, National Multiple Sclerosis Society, New York, NY, USA, None;

Maureen Maguire, University of Pennsylvania, Philadelphia, PA, USA, Genentech/Roche (C), Regenera (C), Foundation Fighting Blindness (F);

Juliette E. McGregor, University of Rochester, Rochester, NY, USA, National Eye Institute/NIH (F);

Neil R. Miller, Johns Hopkins University School of Medicine, Baltimore, MD, USA, National Eye Institute/NIH (F), Invex Therapeutics (C);

Cynthia Owsley, University of Alabama at Birmingham, Birmingham, AL, USA, National Eye Institute/NIH (F), National Institute on Aging (F), Centers for Disease Control and Prevention (F), Research to Prevent Blindness (F), Greater Baltimore Medical Center Educational Foundation Inc. (F);

Michael X. Repka, Johns Hopkins University, Baltimore, MD, USA, National Eye Institute/NIH (F), Alcon (C, F), Luminopia (C), American Academy of Ophthalmology (F);

Paul A. Sieving, National Eye Institute, NIH, Bethesda, MD, USA, NIH Intramural Program (F);

Paul C. VanVeldhuisen, The Emmes Company, LLC, Rockville, MD, USA, National Eye Institute/NIH (F);

Michael Wall, University of Iowa, Carver College of Medicine, Iowa City, IA, USA, None.

Disclosure: **L.A. Levin**, See Commercial Relationships above; **M. Sengupta**, None; **L.J. Balcer**, None; **M.J. Kupersmith**, See Commercial Relationships above; **N.R. Miller**, See Commercial Relationships above

References

- Beck RW, Moke PS, Turpin AH, et al. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol*. 2003;135(2):194–205.
- Jolly JK, Juenemann K, Boagey H, et al. Validation of electronic visual acuity (EVA) measurement against standardised ETDRS charts in patients with visual field loss from inherited retinal degenerations. *Br J Ophthalmol*. 2020;104(7):924–931.
- De Moraes CG, Liebmann JM, Levin LA. Detection and measurement of clinically meaningful visual field progression in clinical trials for glaucoma. *Prog Retin Eye Res*. 2017;56:107–147.
- Meleth AD, Mettu P, Agron E, et al. Changes in retinal sensitivity in geographic atrophy progression as measured by microperimetry. *Invest Ophthalmol Vis Sci*. 2011;52(2):1119–1126.
- Matsuura M, Murata H, Fujino Y, et al. Evaluating the usefulness of MP-3 microperimetry in glaucoma patients. *Am J Ophthalmol*. 2018;187:1–9.
- Majander A, Joao C, Rider AT, et al. The pattern of retinal ganglion cell loss in OPA1-related autosomal dominant optic atrophy inferred from temporal, spatial, and chromatic sensitivity losses. *Invest Ophthalmol Vis Sci*. 2017;58(1):502–516.
- Zein WM, Jeffrey BG, Wiley HE, et al. CNGB3-achromatopsia clinical trial with CNTF: diminished rod pathway responses with no evidence of improvement in cone function. *Invest Ophthalmol Vis Sci*. 2014;55(10):6301–6308.
- Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. *Invest Ophthalmol Vis Sci*. 2011;52(2):816–820.
- Levin N, Devereux M, Bick A, et al. Color perception impairment following optic neuritis and its association with retinal atrophy. *J Neurol*. 2019;266(5):1160–1166.
- Stoianov M, de Oliveira MS, Santos Ribeiro Dos, Silva MCL, et al. The impacts of abnormal color vision on people's life: an integrative review. *Qual Life Res*. 2019;28(4):855–862.
- Cole SR, Beck RW, Moke PS, et al. The National Eye Institute Visual Function Questionnaire: experience of the ONTT. Optic Neuritis Treatment Trial. *Invest Ophthalmol Vis Sci*. 2000;41(5):1017–1021.
- Balcer LJ, Galetta SL, Calabresi PA, et al. Natalizumab reduces visual loss in patients with relapsing multiple sclerosis. *Neurology*. 2007;68(16):1299–1304.
- Walter SD, Ishikawa H, Galetta KM, et al. Ganglion cell loss in relation to visual disability in multiple sclerosis. *Ophthalmology*. 2012;119(6):1250–1257.
- Talman LS, Bisker ER, Sackel DJ, et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. *Ann Neurol*. 2010;67(6):749–760.
- Bainbridge JW, Smith AJ, Barker SS, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med*. 2008;358(21):2231–2239.
- Lê JT, Qureshi R, Rouse B, et al. Development and content of a database of systematic reviews for eyes and vision [published online April 6, 2021]. *Eye (Lond)*.
- Saldanha IJ, Dickersin K, Wang X, Li T. Outcomes in Cochrane systematic reviews addressing four common eye conditions: an evaluation of completeness and comparability. *PLoS One*. 2014;9(10):e109400.
- Mayo-Wilson E, Fusco N, Li T, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol*. 2017;86:39–50.
- Le JT, Viswanathan S, Tarver ME, et al. Assessment of the incorporation of patient-centric outcomes in studies of minimally invasive glaucoma surgical devices. *JAMA Ophthalmol*. 2016;134(9):1054–1056.
- Law A, Lindsley K, Rouse B, et al. Missed opportunity from randomised controlled trials of medical interventions for open-angle glaucoma. *Br J Ophthalmol*. 2017;101(10):1315–1317.
- Saldanha IJ, Lindsley K, Do DV, et al. Comparison of clinical trial and systematic review outcomes for the 4 most prevalent eye diseases. *JAMA Ophthalmol*. 2017;135(9):933–940.
- Saldanha IJ, Petris R, Han G, et al. Research questions and outcomes prioritized by patients with dry eye. *JAMA Ophthalmol*. 2018;136(10):1170–1179.
- Saldanha IJ, Le JT, Solomon SD, et al. Choosing core outcomes for use in clinical trials in ophthalmology: perspectives from three ophthalmology outcomes working groups. *Ophthalmology*. 2019;126(1):6–9.
- Gilmore CS, Lim KO, Garvin MK, et al. Association of optical coherence tomography with longitudinal neurodegeneration in veterans with chronic mild traumatic brain injury. *JAMA Netw Open*. 2020;3(12):e2030824.
- Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res*. 2017;57:46–75.
- Rossi EA, Granger CE, Sharma R, et al. Imaging individual neurons in the retinal ganglion cell layer of the living eye. *Proc Natl Acad Sci USA*. 2017;114(3):586–591.
- Liu Z, Kurokawa K, Zhang F, et al. Imaging and quantifying ganglion cells and other transparent neurons in the living human retina. *Proc Natl Acad Sci USA*. 2017;114(48):12803–12808.
- Liu Z, Saeedi O, Zhang F, et al. Quantification of retinal ganglion cell morphology in human glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2021;62(3):34.
- McGregor JE, Yin L, Yang Q, et al. Functional architecture of the foveola revealed in the living primate. *PLoS One*. 2018;13(11):e0207102.
- McGregor JE, Godat T, Dhakal KR, et al. Optogenetic restoration of retinal ganglion cell activity in the living primate. *Nat Commun*. 2020;11(1):1703.
- Sharma R, Schwarz C, Williams DR, et al. In vivo two-photon fluorescence kinetics of primate rods and cones. *Invest Ophthalmol Vis Sci*. 2016;57(2):647–657.
- Pfaffle C, Spahr H, Kutzner L, et al. Simultaneous functional imaging of neuronal and photoreceptor layers in living human retina. *Opt Lett*. 2019;44(23):5671–5674.
- Klineova S, Kupersmith M. Promising recovery biomarkers after first event acute demyelinating optic neuritis. *Mult Scler Relat Disord*. 2020;45:102400.
- Bruno A, Akinwuntan AE, Lin C, et al. Simplified modified Rankin scale questionnaire: reproducibility over the telephone and validation with quality of life. *Stroke*. 2011;42(8):2276–2279.
- Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): A randomised, multicentre, placebo-controlled trial. *Lancet*. 2015;385:1295–1304.