tvst

Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials: Recommendations from the International HOVER Taskforce

Lauren N. Ayton[1,2,*], Joseph F. Rizzo, III[3,*], Ian L. Bailey[4], August Colenbrander[5], Gislin Dagnelie[6], Duane R. Geruschat[6], Philip C. Hessburg[7], Chris D. McCarthy[8], Matthew A. Petoe[9], Gary S. Rubin[10], and Philip R. Troyk[11], for the HOVER International Taskforce

[1] Department of Optometry and Vision Sciences and Department of Surgery (Ophthalmology), The University of Melbourne, Parkville, Australia
[2] Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia
[3] Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA
[4] School of Optometry, University of California-Berkeley, Berkeley, CA, USA
[5] Smith-Kettlewell Eye Research Institute and California Pacific Medical Center, San Francisco, CA, USA
[6] Lions Vision Research and Rehabilitation Center, Johns Hopkins Wilmer Eye Institute, Baltimore, MD, USA
[7] Detroit Institute of Ophthalmology, Henry Ford Health System, Grosse Pointe Park, MI, USA
[8] Department of Computer Science & Software Engineering, Swinburne University of Technology, Melbourne, Australia
[9] Bionics Institute of Australia, East Melbourne, Australia
[10] University College London Institute of Ophthalmology, London, UK
[11] Armour College of Engineering, Illinois Institute of Technology, Chicago, IL, USA

Correspondence: Lauren N. Ayton, Department of Optometry and Vision Sciences, The University of Melbourne, 200 Berkeley St, Carlton VIC 3053, Australia. e-mail: layton@unimelb.edu.au

Translational research in vision prosthetics, gene therapy, optogenetics, stem cell and other forms of transplantation, and sensory substitution is creating new therapeutic options for patients with neural forms of blindness. The technical challenges faced by each of these disciplines differ considerably, but they all face the same challenge of how to assess vision in patients with ultra-low vision (ULV), who will be the earliest subjects to receive new therapies.

Historically, there were few tests to assess vision in ULV patients. In the 1990s, the field of visual prosthetics expanded rapidly, and this activity led to a heightened need to develop better tests to quantify end points for clinical studies. Each group tended to develop novel tests, which made it difficult to compare outcomes across groups. The common lack of validation of the tests and the variable use of controls added to the challenge of interpreting the outcomes of these clinical studies.

In 2014, at the bi-annual International "Eye and the Chip" meeting of experts in the field of visual prosthetics, a group of interested leaders agreed to work cooperatively to develop the International Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials (HOVER) Taskforce. Under this banner, more than 80 specialists across seven topic areas joined an effort to formulate guidelines for performing and reporting psychophysical tests in humans who participate in clinical trials for visual restoration. This document provides the complete version of the consensus opinions from the HOVER taskforce, which, together with its rules of governance, will be posted on the website of the Henry Ford Department of Ophthalmology (www.artificialvision.org).

1

Research groups or companies that choose to follow these guidelines are encouraged to include a specific statement to that effect in their communications to the public. The Executive Committee of the HOVER Taskforce will maintain a list of all human psychophysical research in the relevant fields of research on the same website to provide an overview of methods and outcomes of all clinical work being performed in an attempt to restore vision to the blind. This website will also specify which scientific publications contain the statement of certification. The website will be updated every 2 years and continue to exist as a living document of worldwide efforts to restore vision to the blind.

The HOVER consensus document has been written by over 80 of the world's experts in vision restoration and low vision and provides recommendations on the measurement and reporting of patient outcomes in vision restoration trials.

Contents

translational vision science & technology

translational vision science & technology

Introduction

Lauren Ayton[1] and Joseph Rizzo III[2]

[1]Department of Optometry and Vision Sciences and Department of Surgery (Ophthalmology), The University of Melbourne, Parkville, Australia; Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia (e-mail: layton@unimelb.edu.au)
[2]Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA (e-mail: Joseph_Rizzo@MEEI.harvard.edu)

Restoration of vision to patients with neural forms of blindness is one of the Holy Grails of modern medicine. Large numbers of research teams and companies around the globe have been pursuing a wide range of approaches to achieve this goal, including genetic, prosthetic, optogenetic, stem cell and other transplantation, and sensory substitution strategies. Each approach has advantages and disadvantages, and no approach will likely prove to be well suited for all forms of neural blindness. Given this, robust activity across multiple disciplines would seem to be the best approach in the pursuit of the challenging goal of providing sight to the blind.

The preponderance of preclinical and clinical studies in sight recovery has been conducted with prostheses. The experimental foundation of the field of visual prosthetics was established in 1968 by the work of Brindley and Lewin[1] and later Dobelle and Mladejovsky,[2] who reported that electrical stimulation could produce visual phosphenes in subjects who were severely blind. Since that time, more than 40 research teams have been developing some form of visual prosthesis (Fig. 1), and two devices were commercialized: Argus II (Second Sight Medical Products, Sylmar, CA), which has received both Food and Drug Administration (FDA, United States) and Certification Experts (Conformité Européene, European) mark approval, and Alpha AMS (Retina Implant AG, Reutlingen, Germany), which has received CE mark approval only.

In parallel, in December 2017, the FDA approved the first directly administered gene therapy in the United States, LUXTURNA (voretigene neparvovecrzyl; Spark Therapeutics, Philadelphia, PA) as a treatment for bi-allelic mutations in the RPE65 gene that causes Leber congenital amaurosis. This milestone presages the approval of other genetic therapies as we enter the dawn of a wide range of novel treatment options for the blind. This robust expansion of research across several disciplines brings hope to the millions of blind individuals who may be able to benefit from these sophisticated technologies in the upcoming decades.

The technical challenges faced by the various strategic approaches for visual restoration or augmentation of visual function differ considerably; however, all of these fields must contend with the need to demonstrate safety and efficacy, which are the cornerstones for regulatory approval. This document is focused on the recommended methods to collect evidence to support the latter. The need for special attention to the topic of efficacy is driven by the challenge of obtaining reliable measures of vision function, or functional vision, before and after intervention in subjects who are severely blind and who will be the earliest candidates for intervention. Inaccuracies in measuring endpoints of vision can lead to spurious conclusions of therapeutic benefit when none is present, which could unnecessarily expose patients to risks of injury to their eyes or overall health without a reasonable hope of benefit.

The best outcomes to date from any form of intervention have been meaningful but have not yet reached the level of providing substantial visual improvement for multiple tasks of daily life. Even the best performing subjects implanted with a visual prosthetic have not improved to the level of "legal blindness" on standard measures of visual acuity nor have they been able to perform the majority of assessments routinely used in standard visual testing. As such, out of necessity, the groups that led early human prosthetic testing had to develop novel test methods. An unintended consequence of the use of novel, group-based testing methods is the challenge for scientists, physicians, regulatory agencies, and the corporate sector to readily compare outcomes across groups. Thus, the time-honored scientific principle that places great significance on external (i.e., disinterested third party) confirmation of results has not been possible in this field. For the same reason, it has been challenging to interpret "validation" studies from any group to assess the potential value of a device for end users.

The notion of seeking international consensus on psychophysical testing methods in the emerging disciplines of neural visual restoration was first raised by one of us (JFR) at the inaugural "The Eye and the Chip" conference in 2000. For a variety of reasons, sufficient momentum toward this goal did not materialize until 2014, when we (LA and JFR) catalyzed an initiative by announcing that our respective Australian- and Boston-based teams had agreed to work cooperatively to develop shared testing methods.

This announcement was met with enthusiasm and attracted over 80 researchers to establish a multinational task force to establish the Harmonization of

Figure 1. Active visual prosthetic groups around the world as of November 2019. This map does not include groups that are working on genetic, optogenetic, or transplantation strategies to restore vision to the blind.
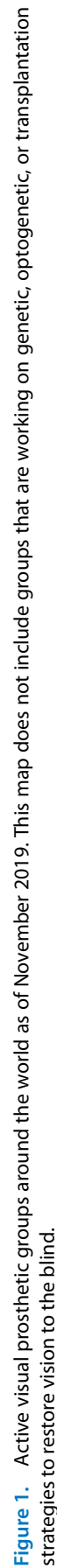
Outcomes and Vision Endpoints in Vision Restoration Trials (HOVER) Taskforce.[3] This taskforce was grounded on the principles of openness, inclusiveness, and collegiality among scientists from around the world, and it sought guidance from recognized experts in visual rehabilitation and visual restoration to develop recommendations for "good practice" for visual psychophysical testing in severely blind humans.

The taskforce operated through a federated structure, where seven working groups were formed in the following areas of interest: visual acuity, electrically evoked device effectiveness, vision processing systems, activities of daily living, orientation and mobility, patient reported outcomes, and psychosocial assessments and ethical considerations. The working group chairs were selected by the taskforce's Executive Committee, but they then had free choice regarding the members of their working groups. The teams represent diversity across nationality and profession and include academics, clinicians, and industry leaders.

The working groups were tasked with developing a consensus document in their area, which often invoked spirited discussions and debate. The majority of this work was done via phone teleconferences and e-mails, with some groups meeting in person when possible. When the group had developed their document and all members were in agreement with the content, the section was reviewed by the broader HOVER Taskforce members (over 80 people who had expressed interest during an initial Special Interest Group at the Association for Research in Vision and Ophthalmology Annual Meeting in 2014). Several reviewers made extraordinary contributions to this process, and they are acknowledged as part of the core HOVER Taskforce in our author list. Final reviews were completed by the Executive Committee. The structure and process of this taskforce are shown in Figure 2.

The Executive Committee and working group leaders met at conferences throughout the process, usually once a year. The process was detailed, thorough, and required a significant time commitment. We cannot thank the contributors enough for their hard work and invaluable expertise.

This initial set of recommendations was developed by experts in the field of visual prosthetics, but our scientific panel includes experts from other sight recovery disciplines (see Acknowledgments) who have agreed to encourage members from their disciplines to provide modifications to our recommendations to better suit their fields of study.

We are proud to share this initial consensus document with our peers, with the knowledge that this will be the first in a series of updates and improvements as the field progresses.

Definition of Terms Relevant to the Psychophysical Assessment of Emerging Visual Restoration Strategies

August Colenbrander[1]

[1]**Smith-Kettlewell Eye Research Institute and California Pacific Medical Center, San Francisco, CA, USA**

Vision is often considered to be the most important source of information about our environment and for our interaction with that environment. This document discusses important issues regarding assessment of outcomes in any type of visual restoration trial. For clarity, the following definitions of terms used in the document are provided.

Assessing visual outcomes can be approached from different points of view. For those who study *how the eye functions* (and, by extension, *how the visual system functions*), the goal of vision seems to be the creation of a visual percept. For those who are interested in *how the PERSON functions*, the goal is more broad—how to facilitate the person's interaction with the environment using visual information.

The two viewpoints are obviously related, but there are distinct differences. The first concept addresses specific *visual function*, such as enabled perception of detail, color, and movement. The second concept addresses *functional vision* and the contribution that vision provides to enhance *task performance*, such as reading, mobility, and activities of daily living. This document specifically addresses individuals who have or will undergo some intervention in the hope of benefiting from improved visual function and functional vision.

The term *vision loss* is a relative term and not an all-or-none phenomenon. Loss of vision can range from mild and moderate to severe, profound, and total. On the other hand, the term *blindness* is not a relative term. Given the dictionary definition of blindness as "to be without light perception," a person cannot be "a little bit blind." However, this term also has vernacular and legal usages that may be used to reference a social category, or a range of disabilities. For the purpose of population statistics, the World Health Organization (WHO) defines blindness numerically as visual acuity "less than 3/60" (which is equivalent to the metrics
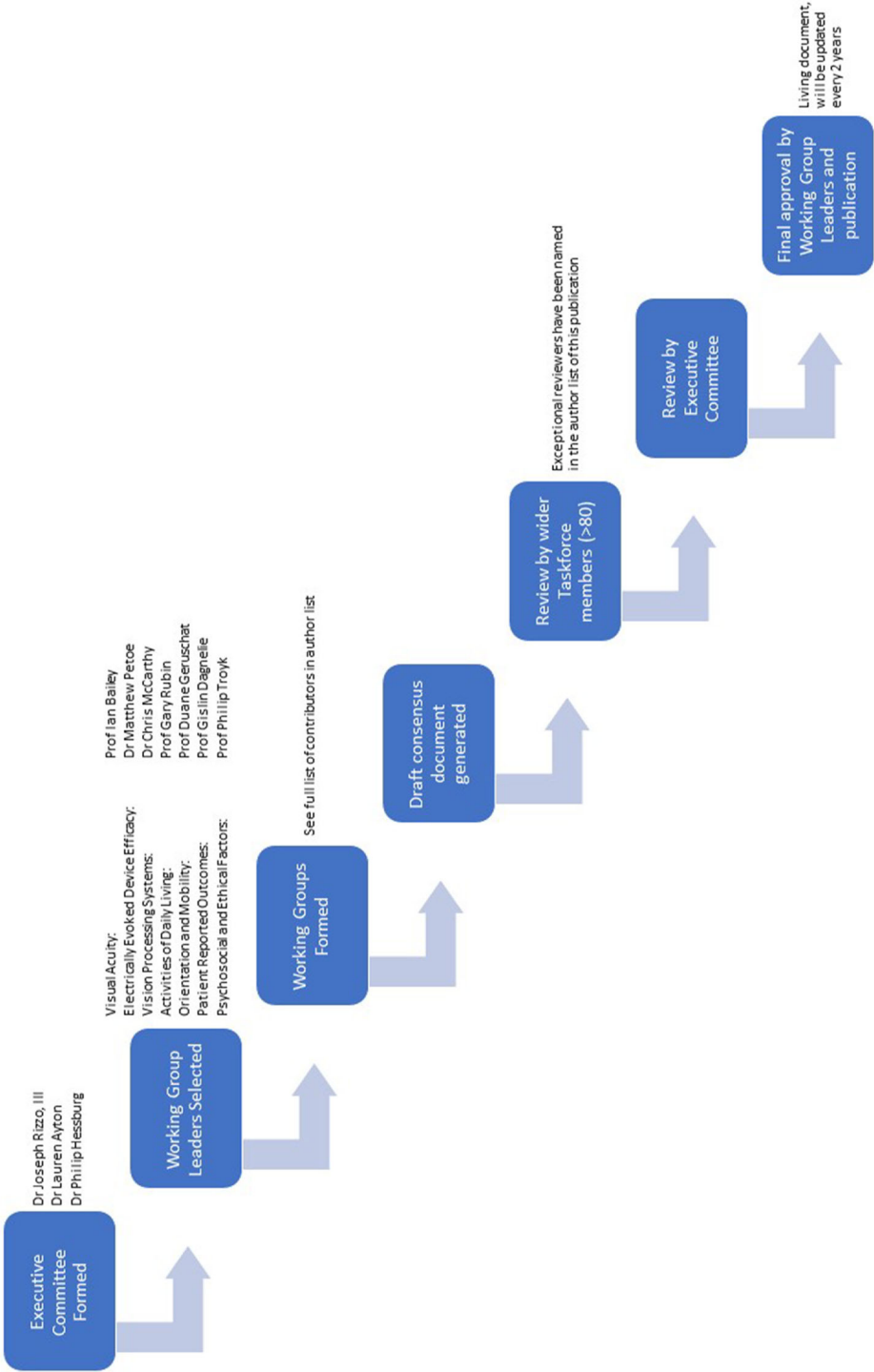
Figure 2. Structure and process flowchart of the HOVER Taskforce.

20/400, 0.05, or 6/120). The International Council of Ophthalmology provided a functional definition that "people may be considered blind when they have no vision or so little vision that they have to rely primarily on *vision substitution* skills (i.e. use of senses other than vision, such as Braille, long cane, text-to-speech software conversion, etc.) to conduct their activities of daily living."[4] In any characterization, it is appreciated that having any residual vision provides a useful adjunct to function.

Traditional Measures of Visual Performance Used by Ophthalmologists

Legally blind—Depends on the defining organization. WHO defines legally blind as 20/400 or worse in the better eye and/or a field of view smaller than 20 degrees.

Count fingers (CF)—Individuals can tell how many fingers the ophthalmologist is holding up.

Hand motion (HM)—Individuals can tell that the ophthalmologist is waving a hand in front of their eyes.

Light perception (LP)—Individuals can tell if the lights in a room are on or off. Roughly equivalent to a normally sighted individuals perception with their eyes closed, and generally assessed using a bright light at between 40 cm and 1 m.

No light perception (NLP)—Individuals cannot tell if the lights in a room are on or off. Generally assessed using a bright light at between 40 cm and 1m.

The term *low vision* refers to having less than normal vision but not being classified as "blind." In these cases, use of any *residual vision* can be improved with various strategies for *vision enhancement*, such as use of magnification, enhanced lighting, or higher levels of contrast.[4]

The term *ultra-low vision (ULV)* refers to having very limited vision but not complete blindness. ULV is so limited that at best only crude shapes can be detected and recognized. Often vision is limited to detection of movement, light projection, or bare light perception. Traditional scales of visual performance, such as those used to describe central visual acuity, are not adequate to convey the potential capability of an individual to perform tasks of daily living, as the experiential level of function depends at least as much on non-visual skills as on the level of vision. Some people are characterized as having ULV because they were sighted and then progressively lost vision, whereas others were blind and then gained vision through various interventions, such as *visual prosthetics* or *genetic therapies*.

Vision rehabilitation aims at improving *how a person functions*, regardless of how the eyes function. To do this, vision rehabilitation may use vision enhancement, vision substitution, or other means. This approach builds upon whatever visual and mental abilities remain.

Vision restoration refers to efforts to restore lost visual abilities, relying on the visual system to convey information. Cataract surgery with implantation of an intraocular lens is by far the most common procedure that restores vision. A variety of emerging approaches, including visual prosthetics, genetic therapies, neurotrophic drugs, stem cells, and optogenetics, represent attempts to restore vision. Techniques that restore function through non-visual means, such as text-to-speech conversion and tactile vision substitution, are potentially valuable to an individual but are not considered to be vision restoration.

In the foreseeable future, use of any of the emerging forms of visual restoration will be restricted to individuals with ULV. Those who undergo such intervention also should have access to other means of visual rehabilitation, such as the use of specialized devices to help with specific tasks. In such cases, it is important to measure performance achieved with visual restoration alone compared to performance when relying on non-restorative rehabilitation. This will provide the greatest insight into how much benefit patients receive from medical interventions.

Assessment of outcomes can be undertaken in many different ways. This document provides guidance on methods that have been peer reviewed and considered acceptable by the HOVER Taskforce.

How the Visual System Functions

Visual acuity assesses the size of the projection of the smallest possible recognizable optotype onto the retinal surface, which, even for a 20/200 letter (0.1, 6/60), corresponds to less than 1 degree of visual angle. Generally, measurement of optotype recognition assumes that only a single fixation of eye position was used to see the optotype. Acuity measurements for larger objects can be more complex to interpret because visual recognition of larger letters may involve the use of searching eye movements and/or scanning with the head (or head-mounted camera). Assistive strategies such as these should be duly noted and recorded when measuring visual acuity. Similarly, reaction times and whether any digital or optical magnification was used should be included as part of the assessment.

The *visual field* is the region of visual space that corresponds to regions of retina that retain a criterion level of visual function. Efficient use of a restricted peripheral field requires conscious scanning techniques, which often requires deliberate training.

This is as important for visual prostheses as it is for disease conditions such as glaucoma and retinitis pigmentosa (RP).

How the Person Functions

How individuals utilize their visual information can be assessed by observation and measurement of their performance with regard to *orientation* and *mobility*, as well as *activities of daily living*. Patient satisfaction can also be assessed subjectively with *patient-reported outcomes*. All three of these areas are covered in detail in this HOVER document.

How the Device Functions

Device effectiveness can be assessed by examining the relationship between stimulation and induced percepts or visual function benefit. Depending on the type of device that is used, device effectiveness can be enhanced by additive technical means, such as preprocessing of the visual or other stimuli to either complement or compensate for the neural processing, or to facilitate perception or function.

In summary, assessment of outcomes can legitimately be undertaken in many different ways depending on the goal of the experimenter. This document provides guidance on methods that have been peer reviewed and considered acceptable by the HOVER Taskforce.

Visual Acuity

Ian Bailey (chair)[1], Michael Bach[2], Rick Ferris[3], Chris Johnson[4], Ava Bittner[5], August Colenbrander[6], and Jill Keeffe[7]

[1]School of Optometry, University of California-Berkeley, Berkeley, CA, USA (e-mail: **ibailey@berkeley.edu**)
[2]Eye Center, University of Freiburg, Freiburg, Germany
[3]National Eye Institute, National Institutes of Health, Bethesda, MD, USA
[4]Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, IA, USA
[5]Nova Southeastern University College of Optometry, Fort Lauderdale, FL, USA
[6]Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA
[7]LV Prasad Eye Institute, Hyderabad, India

Ever since Snellen developed his chart, letter chart acuity has been a mainstay for the assessment of vision. The Snellen or similar type test is relatively easy to perform, generally requires little time, is inexpensive, and can be universally applied. Indeed, its use is so pervasive that the terms "visual acuity" (meaning letter chart acuity that measures foveal, or at least central, vision) and "vision" are often used synonymously. Measuring visual acuity with optotype charts serves many clinical purposes, including identifying and monitoring ocular health, guiding decision-making when correcting refractive errors, and assessing the potential benefit of medical interventions. However, measurement of central vision with optotype charts has limitations, especially when working with individuals who have ULV. Measurement of acuity in these individuals typically requires significant time and effort. It can be difficult to obtain an accurate measurement at any single sitting, complicating valid comparisons across visits. This challenge of obtaining reproducible measurements over time complicates any attempt to assess the benefit of a visual restorative intervention.

For optotype charts, the basic visual task is recognizing objects sequentially to allow an estimate to be made of the minimum angle of resolution that can be reliably reported. Typically, subjective responses for optotypes are sought from larger objects first, and progressively smaller objects are then shown, although the inverse approach is viewed as being advantageous by some, because individuals are often reluctant to acknowledge that they can recognize an optotype if it appears blurry, even with encouragement. At the common presentation distances, the largest letters on most charts have an angular size smaller than 1 degree, which is less than the diameter of the fovea. For most common visual acuity testing, the foveal area of the retina is responsible for the recognition of the optotypes as eye movements systematically shift the attention across and down the chart.

In ULV, the features in the visual acuity targets will generally have to be much larger than 1 degree of visual angle, and there is a higher likelihood of scotomas or other visual field restrictions that can further compromise testing of acuity. Scanning eye movements and sometimes even head movements are likely to be required as the subject inspects and attempts to identify and interpret the features of the test target.

When vision is too poor to perform the visual task of reading a letter chart, then the task should be systematically simplified. Recognizing single optotypes is a simpler task than reading a letter chart, and grating acuity test tasks are simpler than identifying single optotypes.

When considering the consequences of vision disorders, it is often important to assess multiple parameters of ocular function, such as visual acuity, contrast sensitivity, color discrimination, and visual fields. It is also important to identify impairments in these individual functions in each eye separately. The scores from any of the various visual acuity tests should not be assumed to be measures of the person's ability to perform visually guided functional tasks in everyday life.

How the person functions is determined by how the person is able to integrate visual information from the two eyes, as well as information from other sensory systems, into vision-related functioning. This is often referred to as *visual ability*. Visual ability and visual disability should be assessed with both eyes open. Vision and vision-related functioning involve more than just visual acuity; however, when actual ability assessments are not available, "visual acuity of the better eye" is often useful as an estimate or indicator of ability or disability.

Introduction

The measurement of visual acuity in vision restoration trials is of upmost importance, as it is one of the key outcome measures accepted by both regulatory bodies and by the general public as evidence of post-intervention improvement. In standard clinical practice, measurement of visual acuity using a logMAR letter acuity chart, such as the Early Treatment Diabetic Retinopathy Study (ETDRS) chart,[5] is the gold standard for assessment of the minimum angle of resolution that a subject can achieve.[6] However, such charts are only able to measure acuity down to levels of logMAR 1.60 (20/800 or 6/240) and so are not applicable to subjects who cannot achieve this level of acuity. Historically, subjects with vision worse than logMAR = 1.60 had their vision characterized as "count fingers," "hand movements," "light perception," or "no light perception" vision, but these categories have been difficult to standardize and are not sensitive enough for use in vision restoration clinical trials.

Another major challenge with measurement of visual acuity in subjects with ULV is the significant variability (in visual acuity, contrast sensitivity, and visual fields, for example) that exists.[7] Due to the range of visual outcomes in vision restoration trials, and these complicating factors, it is essential to implement a cohort of acuity tests in a standardized and repeatable manner. These tests should be administered before and after the therapeutic intervention.

These guidelines outline recommended methodologies for the testing and reporting of psychophysical results of testing in ULV subjects who participate in clinical therapeutic trials; however, all researchers are free to add other tests or develop new tests that might help identify or quantify other characteristics of vision that may be changing as a result of the interventions. We hope that this HOVER document will provide useful guidance in how to describe testing methods and results, so as to encourage reproducibility by other researchers and clinicians.

Recommended Methodology for Assessment of Visual Acuity

Acuity assessment should involve evaluation of optotype recognition acuities (letters, Landolt rings, and/or tumbling E acuity) and grating acuity. Researchers and clinicians are encouraged to use one of the existing validated tests for these purposes, including the ETDRS chart,[5] the Freiburg Acuity and Contrast test (FrACT),[8] the Berkeley Rudimentary Vision Test (BRVT),[9] the Basic Grating Acuity (BaGA) test,[10] or the Grating Acuity Test (GAT).[11] There are other tests, such as the Basic Assessment of Light and Motion (BaLM) test,[12] that evaluate other aspects of visual function.

General Recommendations for Testing

The examiner should be qualified in the assessment of visual acuity. Ophthalmologists, optometrists, orthoptists, or certified ophthalmic technicians or assistants are all potentially capable to perform this role, provided they are willing to follow a standardized protocol as described below.

The choice of test should depend in part on the level of vision. For subjects with vision of logMAR 1.60 (equivalent to 6/240 or 20/800) or better, the ETDRS test can be used. For subjects with worse vision, the BaLM, BaGA, FrACT, or BRVT tests should be used. New alternative tests for very low vision might be developed in the future.

For pre-intervention measurements, all tests should be completed with the subject's best refraction in place. The refractive correction should be appropriate for the testing distance being used. The pupils should be undilated. If subjects are unable to complete a subjective refraction, the refractive error may be estimated using an auto-refractor or retinoscopy.

Visual acuity measurements for ULV should be made with large targets that have very coarse detail, so that target recognition is robust to moderate magnitudes of optical defocus. If reasonable estimates of refractive error are available, it is generally recommended that the optical correction be worn during

testing. If viewing distances are changed, appropriate adjustments of corrective lens power could be made, although this concern is not consequential if the change of lens power is not substantial. The use of any refractive correction, other than what would be normally used, must be described in any report or publication.

Some interventions may involve lenses or imaging systems that produce magnification (or minification) by causing the perceived image to have a larger (or smaller) angular size than the real-world object. In all reports on such cases, the magnitude of the magnification or minification must be specified, and a detailed description should be given of the lenses or imaging systems used. Quantification of any changes in visual acuity should make distinctions between changes attributed to the imaging system and changes that result from the therapeutic intervention.

For post-intervention measurements, the recommended method depends on the type of intervention used. For prosthetic devices that bypass the optical components of the eye (e.g., camera-based prostheses), refractive correction need not be used for post-intervention measurements if the pre-intervention testing did not yield quantifiable results. However, optimal refractive correction must be used if pre-intervention testing yielded quantifiable results, and it should be used if the subject reports a benefit from optical correction, even if this benefit is not quantifiable. If electronic camera zooming is used, the magnification that was used at the time that testing was performed must be recorded and specified. For photodiode-based prosthetic devices, optimal refractive correction should be used during any acuity measure, and, again, the magnitude of any magnification effects must be specified.

For ETDRS and BRVT testing, the recommended room illuminance is 500 lux, which is representative of lighting levels in well-lit office environments. Luminance levels from 250 to 1000 lux are acceptable. Within a given program or clinic, it is important that the illuminance levels be kept consistent (to within ±20%) from one test session to the next and from one testing station to another. It is also important to verify that no room lights or other bright surfaces are acting as glare sources to the patient or that their reflections are producing "hot spots" on the stimulus chart.

For other testing methods, variations in ambient light intensity, including absence of room lighting, can be used as deemed appropriate. Again, care should be taken to ensure that the lighting remains the same from one test session to the next. Other illumination levels may be considered if there is reason to believe that illumination levels are having a significant effect on performance.

Specific testing distances should be used. Generally, these will be distances recommended for specific test charts or distances determined from the calibration of screen displays to achieve the desired angular sizes.

Measurements should be made with the right eye and left eye separately and then binocularly, when appropriate. Due to time restraints, this may not always be possible; in this case, testing the study eye is always the priority. During the monocular tests, care must be taken to ensure that the other eye is fully occluded. Within a test protocol, there should be documented rules for stopping, for guessing, or for allowing subjects to correct or change responses. There should also be standard rules and procedures for encouraging guessing and pointing to help the patient locate the test target. Subjects should be allowed to move their head and eyes as they wish to assist in the identification and utilization of any islands of residual vision, as these strategies have been shown to improve performance on simulated prosthetic vision tests[13] and in natural low vision. It should be reported whether or not head and eye movements were allowed.

A time limit should be set for each response (sometimes specified by the test manufacturer).

In order to identify residual islands of vision, hand-held optotype charts may be used (either an ETDRS letter chart or a BRVT chart). Such testing should begin at a distance of 4 meters, with the examiner moving the chart into all four quadrants of the visual field and asking the subject about the preferred location for seeing the chart. If an island of vision is found at this distance, then the testing distance can be changed in order to obtain a more precise measurement of visual acuity. If no island of vision is found at 4 meters, the same procedure should be administered at a distance of 1 meter.

For each stage of evaluation (i.e., subject screening and selection and pre- and post-intervention assessments), acuity measurements should be obtained from a minimum of two test sessions separated by at least 24 hours. The variability of individual subject's responses, both within sessions and between sessions, should be determined. Prior to any intervention, there should be at least two sessions of testing to establish the baseline measurement. Subjects should be excluded from the study if the variability of their responses exceeds a specified criterion.

For some subjects, it will be appropriate to make visual acuity measurements with different visual acuity test tasks. For example, a patient with poor acuity combined with very small visual fields (as is typical with a vision prosthesis) may not be able to trace out and recognize the shape of a large letter, so a grating acuity task may provide a better measure of the

translational vision science & technology

visual resolution ability. The visual resolution task used in the visual acuity tests becomes progressively more complex going from gratings to isolated optotypes to single optotypes with flanking bars and to charts of optotypes in logMAR or other multi-optotype formats. Measurements of visual acuity can show wide differences from one test task to another, and, at this time, there has not been a comprehensive comparison of all of the various methods and how the visual acuity scores may be affected for different pathology groups. Hence, it is vital to define which visual acuity task was used. For monitoring changes in vision over time, it is very important that consecutive test sessions include at least one of the same visual acuity tests, administered under the same conditions and following the same protocol.

Specific Recommendations for Post-Intervention Measurements

If only one eye is treated, visual acuity still should be measured in each eye separately and with binocular viewing. Any visual performance changes in the untreated eye, or changes under binocular viewing should be identified to assess if there is a bilateral effect on visual performance. The fellow eye should be patched, or at least there should be assurance that the fellow eye is occluded, during monocular acuity testing.

In post-intervention measures, the test should be completed in a random order with (1) device on, (2) device off, and (3) when practical, a control condition, which might include strategically scrambled stimulation input from the device. If a control condition is used, specific details should be given about the nature of the control. If a control condition is not used, this should be explicitly stated.

Performance can vary significantly depending on whether the optotype is white on a black background or black on a white background.[14] In some prosthetic devices, contrast reversal is under the patient control. The contrast direction of the optotypes should be reported, and it should be reported whether or not this contrast direction was the choice of the patient or the experimenter. Ideally, acuity measurements should be made with both white-on-black and black-on-white optotypes, but this may be restricted by time limitations.

Specific Visual Acuity Test Methodologies

LogMAR Letter Acuity Tests

A standardized logMAR letter chart acuity is the gold standard acuity measure for subjects with vision better than logMAR 1.60 (20/800 or 6/240). The most widely used format is the ETDRS chart (see Fig. 3).[5]

The published ETDRS guidelines for measuring acuity on a letter optotype chart are summarized as follows:

1. The ETDRS chart should be positioned such that the third row of letters (the 0.80 logMAR line) is 125 ± 5 cm (49 ± 2 inches) from the floor and 4 meters (13 feet) away from the subject, regardless of their vision.
2. The right eye is tested first using ETDRS Chart 1.
3. Subjects are instructed that they should attempt to identify all letters on the chart, from the top line down, and they are encouraged to guess when they are uncertain. The subject should be told that there are no numbers or shapes other than English letters.
4. The examiner records the location and optotype of each correct identification, allocating a score of 0.02 log units per letter.[5]
5. If the subject is unable to read more than 10 letters at 4 meters, then the chart is moved forward to a distance of 1 meter from the subject, where the test is repeated. A +0.75D lens added to the refractive correction is required to maintain clear focus.
6. The left eye then is measured using the ETDRS Chart 2.
7. After completion of the left eye testing, the occluder is removed from the right eye and testing is repeated with both eyes open.
8. If subjects are not able to see the letters on a standard acuity chart, then a low vision optotype recognition test should be used. Recommendations for testing of ULV acuity include use of the FrACT or BRVT, as described below.

Electronic ETDRS Test

The electronic ETDRS (E-ETDRS) test[16] (see Fig. 4) is a computer-based test of visual acuity that uses single letters, each with four flanking bars that are separated from the letter by one letter-width. This test has been shown to provide reliable scores of visual acuity that are well correlated with the scores from the ETDRS letter chart. At the standard test distances of 3 meters (10 feet), the upper limit of the measurable visual acuity range is logMAR 1.60 (0.025, 6/240, 20/800), but this can be extended by reducing the test distance. Computerization facilitates the recording and scoring of responses and allows easy randomization of the sequence of letters. The E-ETDRS test creates efficiencies by making a quick estimate of the acuity before strategically concentrating testing near the threshold level. The main advantages include improved testing efficiency, and randomizing the letter

translational vision science & technology
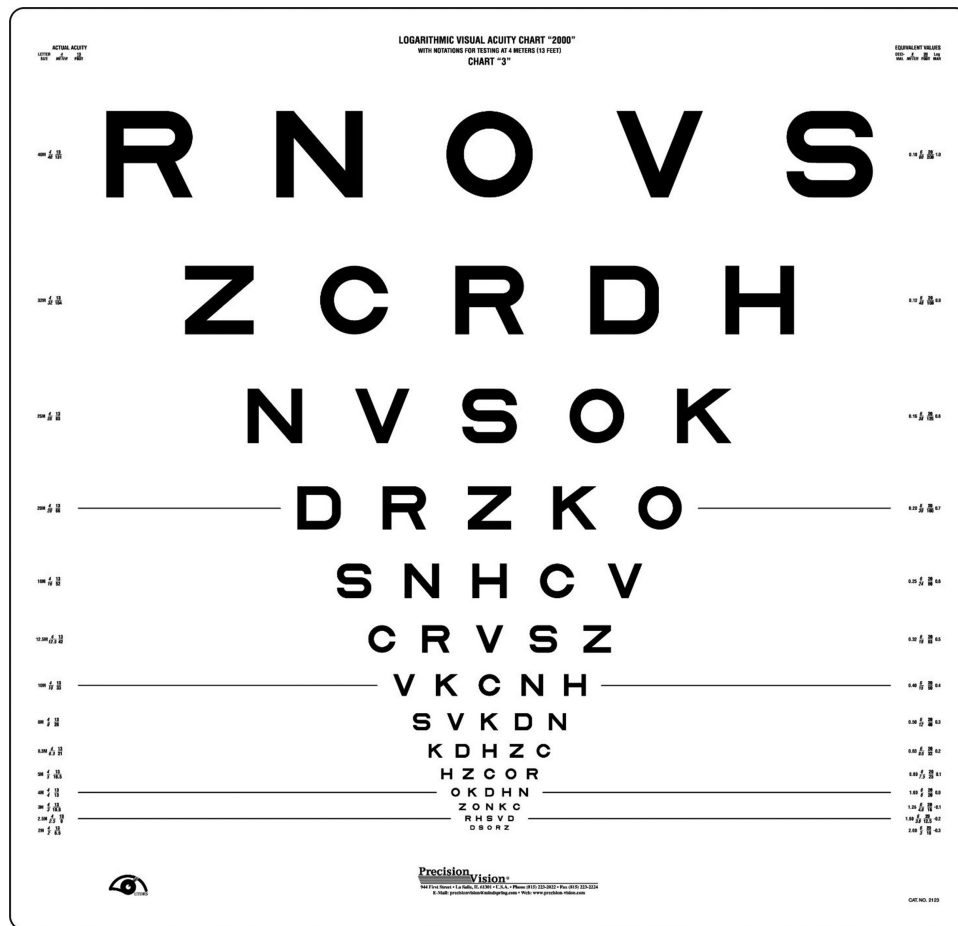
Figure 3. The standard Early Treatment of Diabetic Retinopathy Study (ETDRS) logMAR visual acuity chart.

Figure 4. The electronic Early Treatment of Diabetic Retinopathy Study (E-ETDRS) visual acuity test.

sequences avoids memorization issues when repeated measures are made. Because the E-ETDRS targets are single Sloan letters with flanking bars, the visual task will likely be easier than reading from charts with five letters per row; hence, some differences in visual acuity scores can be expected.

Low-Vision Optotype Test: FrACT

The FrACT[8] was designed for the assessment of low vision covering the entire range of visual acuity measurable with optotypes. The FrACT is a computer-driven program that is available online without cost (http://michaelbach.de/fract/) (see Fig. 5). Full details

of the test have been published elsewhere.[8,17,18] The FrACT also includes tests of contrast sensitivity.

The recommended protocol is as follows:

1. Prior to the commencement of testing, the program must be calibrated by
   a. Measuring the observation distance from the eye to the screen and entering this number into the "observer distance" box on the setup screen.
   b. Measuring the blue calibration line on the computer screen and entering this value into the "length of blue ruler" box.
2. When the system has been calibrated, it will calculate the visual acuity range that can be presented from the distance measurements and screen resolution.
3. Acuity can be measured using a range of optotypes, including Sloan letters, Landolt rings, and tumbling E; normally the optotypes will be black on a white background, but this can be altered in the settings.
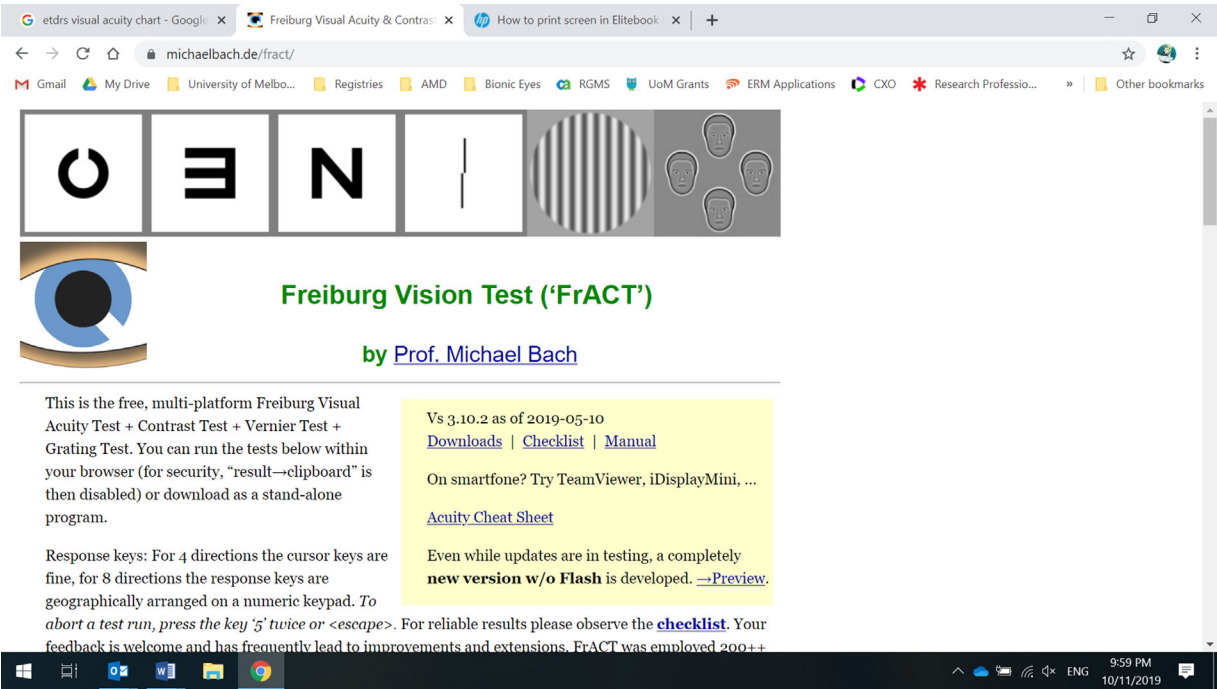
Figure 5. Screenshot of the Freiburg Acuity and Contrast Test (FrACT), available online at http://michaelbach.de/fract/.
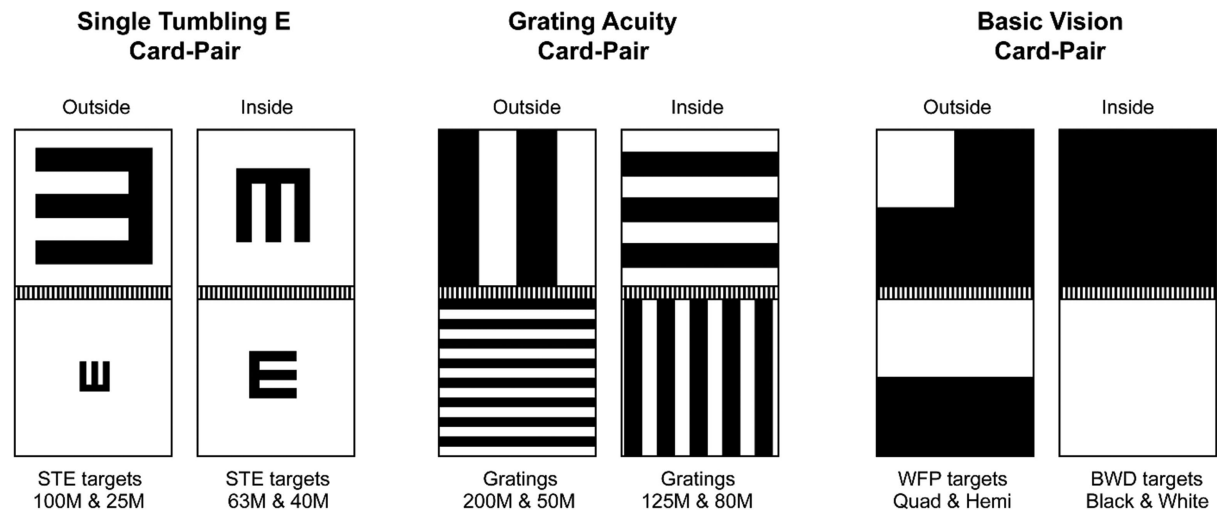
Figure 6. The Berkeley Rudimentary Vision Test (BRVT).

4. The provided checklist and on-screen instructions are followed to complete each optotype test.
5. The size and resolution of the display screen should be chosen to avoid floor and ceiling effects. The screen size should accommodate the largest letters, and the resolution should be sufficient for satisfactory rendition of the smallest letters.

Low-Vision Optotype Test: BRVT

The BRVT test[9] was developed for the clinical measurement of visual acuity in subjects with ULV in the range of logMAR acuity 1.60 and below. The test is administered with three card pairs, each of which consists of two 25-cm-square cards hinged together, thus providing four panels that can be used as targets (Fig. 6). The first card pair consists of single tumbling E (STE) letter optotypes; the second card pair displays square wave gratings to measure spatial resolution; and the third card pair is used as both a discrimination test (using a diffuse white or black card) and a detection task (by identifying the location of a white region on an otherwise back background). Full

details of recommended test methods are published elsewhere.[9,19]

The recommended protocol for the BRVT is as follows:

1. Begin testing with the STE acuity test at a viewing distance of 1 meter. At this distance, the STE acuity range is from logMAR = 2.00 to 1.40 (equivalent to 6/600 to 6/ 150, 20/2000 to 20/500), and it can be measured in increments of 0.20 log units.
2. Present all cards in the BRVT at least four times to the subject, with the orientation changed randomly each time. For the four-choice STE task, successful identification is taken as better than 50% correct responses across six or more presentations.
3. For the two-choice grating acuity task, successful identification is taken as 80% or more correct responses across eight or more presentations
4. If the orientation of the largest STE (100 M) cannot be recognized at 1 meter, then reduce the viewing distance to 25 cm, where the acuity range becomes logMAR 2.60 to 2.00 (6/2400 to 6/600, 20/8000 to 20/2000).
5. If the subject is unable to identify the orientation of the 100 M STE at 25 cm, change to the second card pair of square wave gratings. These gratings should be presented at 25 cm, which provides a grating acuity range of logMAR 2.90 to 2.30 (6/4800 to 6/1200, 20/16000 to 20/4000) in steps of 0.20 log units.
6. If the subject is unable to identify the orientation of the largest grating, change to the third card pair, which has the white field projection and black–white discrimination tests. These cards also should be presented at 25 cm.
7. The white field projection test has two targets. One is a white quadrant on a black background; for the other, the card is divided into black and white halves. The subject's task is to locate the white quadfield or the white hemi-field.
8. If the subject fails the white field projection test, administer the black–white discrimination test. The subject's task is to distinguish the all-black card from the all-white card.

BaGA Test

The BaGA test[10] uses a computer screen display to present a circular field filled with a sine-wave grating. There are four possible orientations for the gratings, two cardinal and two oblique, and four different spatial frequencies (3.3, 1.0, 0.33, and 0.10 cpd). Different viewing distances, field sizes, and gamma values may be selected. The subject responds to the four-alternative forced choice task on a keyboard. Correct responses

and response times are recorded for each of the grating orientations.

GAT

The GAT,[11] another computerized test, presents a square wave grating in a 37.5-cm circular field for presentation at 1.0, 2.0, or 4.0 m using a four-alternative forced-choice paradigm with responses recorded by pressing a button. The spatial frequencies of the gratings are incremented in steps of 0.10 log units, and, for each subject, the testing distance is based on the individual visual acuity at their initial visit.

BaLM Test

The BaLM test was developed specifically for use with prosthetic vision devices.[12] The BaLM test includes tests for perception of light, detection of motion, light localization, and the temporal discrimination of two flashes. Subject responses are delivered through a numeric keypad, and auditory cues are provided to prompt responses at the appropriate time.

Testing for Light Perception

When subjects cannot satisfactorily respond to the various tests of visual acuity and spatial vision, then light perception should be tested.

To assess light perception,

1. The brightest light delivered by an indirect ophthalmoscope should be used.
2. Room lighting should remain at the same level used for normal acuity testing.
3. Each eye should be tested separately. The fellow eye should be patched, and it should also be covered with the palm of the subject's hand to ensure a tight seal around the orbit and bridge of the nose.
4. The indirect ophthalmoscope should be held at 1 meter and the light beam directed into and away from the pupil of the eye at least eight times.
5. The subject should report when they see the light.
6. If the examiner is convinced the subject can identify the onset of the light, this response can be recorded as light perception (LP); otherwise, the level of vision is classified as no light perception (NLP).

Presentation and Analysis of Results

Changes in visual acuity scores should be analyzed using logarithmic scaling. The preferred method for designating the visual acuity scores is to use logMAR. For the charts of optotypes, flanked single optotypes, single optotypes, and grating targets, the minimum angle of resolution (MAR) is the angular size of the

critical detail in minutes of arc. Most optotypes are built on a 5 × 5 grid, so the MAR is assumed to be one-fifth of the height of the optotype. For grating targets with a 50/50 duty cycle, the MAR is given by the size of a stripe width when expressed in minutes of arc. Visual acuity results should be presented and reported in logMAR units, but when authors wish to use Snellen fractions or decimal notations these should be added in parentheses after the logMAR value.

This logarithmic scaling of visual acuity should be used for analysis of differences or changes, as well as for graphical presentation of population data. When datasets include visual acuity measures from different test tasks (gratings, single optotypes, flanked optotypes, or charts of optotypes) or from different viewing conditions or testing procedures, it becomes important to clearly identify which results came from which variant of the visual acuity tests.

Some vision restoration systems will be able to incorporate optical or electronic display systems that allow magnification, minification, or repositioning of some, or all, of the perceived image. The visual acuity score is a measure of the angular size that the critical detail in the test target subtends at the observer's eye. All measurements of visual acuity should be expressed in angular terms, relative to the observer's eye, and full information should be reported about the testing procedures and the test tasks (gratings, single optotypes or charts).

When any optional magnification or minification is used during visual acuity testing, the reports of all visual acuity results should include a detailed description of all image manipulations. In such cases, it may be important to measure and report associated changes in the visual field.

The tests of temporal resolution, motion detection, spatial localization, and light perception in the BaLM and the white field projection and black–white discrimination test in the BRVT are not tests of visual resolution, and the results of such tests should be reported separately.

Reporting Guidelines

Any publication or presentation should contain sufficient information so that another group can replicate the testing methodology.

When reporting the results for testing of visual acuity, the following information must be included:

1. Name of the test (and version number if available)
2. Type of acuity optotype used (e.g., letter, Landolt ring, STE, flanked letters)
3. Room lighting (illuminance in lux), measured from the point of the subject's eye
4. Luminance (in cd/m^2) of any computer screens or back-illuminated charts used in testing
5. Contrast polarity of the optotype (white-on-black or black-on-white); if gray or colored targets or backgrounds are used, the gray levels or color characteristics should be specified
6. Time cutoff for each response
7. Testing distance
8. Information about the angular size of the visual display used
9. Qualitative control-related information, including whether an occluder or patch was used or whether scrambled versus unscrambled stimulation inputs were used
10. Quantitative control-related information, including the number of data points (test and control) acquired for each test condition, with the mean, median, and standard deviations, as appropriate.
11. Indication of whether subjects used eccentric viewing when taking the tests
12. Indication of whether subjects made scanning eye movements when taking the tests
13. Indication of the distant refraction (as measured or estimated) and the power of any corrective lenses that were used for testing; use of optical telescopes or other low vision aids, or any electronic zooming systems, must be described in detail
14. Number of test runs within sessions and number of sessions used to obtain results, including the timing and duration of any scheduled or ad hoc rest sessions

Note that reference to the use of a publicly available detailed refraction and visual acuity protocol can be used to summarize these details.

Electrophysiology

Gislin Dagnelie (chair)[1], Michael Bach[2], David Birch[3], Laura Frishman[4], and J. Vernon Odom[5]

[1]Lions Vision Research and Rehabilitation Center, Johns Hopkins Wilmer Eye Institute, Baltimore, MD, USA (e-mail: gislin@jhu.edu)
[2]Functional Vision Research, Eye Center, Medical Center, Freiburg University, Freiburg, Germany
[3]Retina Foundation of the Southwest, Dallas, TX, USA
[4]College of Optometry, University of Houston, Houston, TX, USA

Figure 7. Example of equipment that can be used for full-field electroretinography, the Espion ColorDome LED-based full-field stimulator.[37] New models of the Epsion system also include software to enable full-field stimulus threshold testing.

[5]**WVU Eye Institute and Blanchette Rockefeller Neurosciences Institute, West Virginia University, Morgantown, WV, USA**

Introduction

Electrophysiological methods are one of the few tools available to researchers and clinicians to obtain information about signal transduction in the visual system that does not depend on the behavior of the patient. Electrophysiology provides detailed timing information, albeit with limited spatial resolution. Electrical response activity along the visual pathway can be recorded non-invasively at the ocular surface (electroretinogram, ERG) and from the occipital scalp (visually evoked potential, VEP). The signals captured by electrodes on the cornea, conjunctiva, or scalp contain small responses embedded in noise that is generated by spontaneous neuronal and muscle activity and by the electrode–tissue interface. Signal amplitudes are in the microvolt range, even in normally sighted individuals, and the interpretation of abnormalities in these small responses requires sophisticated signal analysis techniques such as averaging and spectral (Fourier) analysis.

The International Society for Clinical Electrophysiology of Vision (ISCEV) has published a series of standards that allow similar stimulation and recording methods to be used worldwide; most applicable to vision restoration trials are the standards for ERG[20] and VEP.[21] HOVER electrophysiology standards should adhere to ISCEV standards as much as possible, but, as described below, judicious adjustments to these standards may sometimes be required to obtain meaningful ERG or VEP recordings.

In clinical practice, multiple stimulus types and recordings may be required to confirm or differentiate among clinical diagnoses, but the use of a single well-chosen ERG or VEP method may be sufficient to confirm or rule out functionality of the retina or visual pathway (i.e., efficacy of a vision restoration approach). Moreover, special adaptations may be necessary. To test functionality of a photosensitive electronic implant in the macular area of a patient with age-related macular degeneration (AMD), a VEP elicited by a visible light stimulus is ambiguous, as the response may originate in peripheral retinal photoreceptors, whereas a near-infrared stimulus would be invisible to the native photoreceptors, so a VEP elicited by this stimulus would necessarily come from the implant by virtue

of its near-infrared sensitivity. Generally speaking, the choice of stimulus parameters needs to be adjusted to the particular vision restoration technique and its (potential) benefits, and the same may be true for recording and analysis methods.

Here, we will consider a number of aspects that should allow a trained clinical electrophysiologist to remain close to the ISCEV standards when developing electrophysiological outcome measures for vision restoration trials, yet make judicious adjustments to these standards as called for by the specific conditions and properties of the disorder and treatment.

ERG Stimulation and Signal Analysis Techniques

ERG responses reflect signal processing and homeostatic recovery in the retina and therefore can provide evidence of (restored) retinal function in vision restoration trials, but the extent to which this is feasible depends to a large extent on the therapeutic modality. As an example, major components of the ERG arise in response to rod and cone photoreceptor activity and recovery, so a treatment restoring outer retinal (photoreceptor and/or retinal pigment epithelium) function is much more likely to have a measurable effect on the ERG than a treatment intervening at the retinal ganglion cell level.

In normally sighted individuals, ERG responses to full-field bright flashes administered following dark adaptation can be as large as $\approx 500\,\mu V$, so with properly configured bandpass filtering and amplification clean responses can be obtained without any need for averaging or postprocessing. This large response is due to the combined activity of similar cellular circuitry across the entire retina, but the situation is very different in advanced vision loss and most vision restoration methods, as only a small retinal area may have residual or restored activity. The resulting small responses to repeated flashes can, in some instances, be recovered through signal averaging, possibly in combination with bandpass filtering.[22,23]

Further improvements in signal-to-noise ratio (SNR) can be obtained by using a train of flashes (commonly referred to as "flicker") and analyzing the periodic response components with Fourier decomposition. The amplitude and (especially) phase of the first and sometimes higher harmonics (especially the second harmonic) are used as characteristic indicators for inner retinal function.[24] Cone function is best analyzed with high flash rates ($\sim$30/s), whereas rods have a longer refractory period and thus are best studied with lower rates ($\sim$9–10/s).[25,26] Use of flicker

responses as the only outcome does have inherent risks, however, as the spectral content of the recorded signal should be carefully analyzed to verify that the harmonic (signal + noise) spectral components are significantly larger than the surrounding (noise only) components, confirming the presence of a reliable response signal.[27]

As indicated above, full-field stimuli are typically used. One reason to use full-field stimuli is to elicit larger ERG responses. A second motivation is that assessing local retinal function is fraught with problems such as intraocular reflections causing spurious stray light responses from the peripheral retina.[28] For most vision restoration trials, where the untreated portion of the retina is non-responsive, this is not a concern, but in patients with intact peripheral retinal function (such as in AMD and Stargardt disease) a focal stimulus must be used, preferably with steady peripheral adapting illumination to minimize the possible effects of stray light on peripheral rods and cones.

VEP Stimulation and Signal Analysis Techniques

VEPs reflect cortical processing activity in response to visual stimulation. In native vision, but also in vision restored by electronic retinal implants or other vision restoration methods at the ocular or optic nerve level, VEPs are recorded on the scalp and reflect a mixture of components originating in primary (V1) and nearby higher visual cortical areas. The folding of V1, with only the foveal and parafoveal retina projected onto the occipital pole, and more peripheral areas represented along the medial wall and in the calcarine sulcus of the cerebral hemispheres, causes VEPs to peripheral stimulation to be smaller, more variable between individuals, and therefore more difficult to quantify than those elicited with foveal stimuli.

The favorable location of the foveal projection has allowed researchers to record VEPs to stimulation with retinal implants. Stronks et al.[29] demonstrated a clear difference in waveforms and amplitudes when foveal versus extrafoveal electrodes in the Argus II retinal implant were stimulated. This finding highlights two important applications of the VEP, whether elicited electrically or with light stimuli: (1) to objectively demonstrate integrity of the visual pathway from retinal ganglion cells to primary visual cortex, and (2) to differentiate between retinal origins of the response and thus show integrity of the cortical projection. When using light stimuli, the latter can only be verified by using localized retinal stimulation, utilizing either a

focal flickering stimulus in a steady background or a luminance-balanced pattern reversal stimulus. Pattern stimuli are preferred, as they elicit more characteristic VEPs with less interindividual variability than flash or flicker VEPs. Pattern VEPs also reflect intracortical processing and can be used to determine visual acuity independently of behavioral measures.[30,31]

In advanced disease of the retina or visual pathway, as well as in diseases with a central scotoma, recording reliable VEPs may be problematic. Not only will the VEP amplitudes be small, but treatment effects that do not improve macular function may be difficult to document. Improving SNR through lengthy averaging may be impractical, but use of flicker stimuli or (better) rapid reversing pattern stimuli, combined with Fourier decomposition of the recorded signals, may yield more sensitive outcome measures. Optimal flicker rates for VEPs are likely to be between 10 and 20 flashes/s, whereas pattern reversal rates should be kept under 10 rps (reversals per second).

In normal vision, the optimal pattern size eliciting a robust response will be 10 to 15 arcmin (0.17°–0.25°), but in patients with retinal degeneration or poorly developed native vision optimal sizes may be much larger (1°, 2°, or even 4°).

Small Signal Considerations

As indicated above, ERG and VEP responses are small, even in normally sighted individuals, and smaller in most conditions leading to severe vision loss. This becomes all the more vexing in vision restoration trials, which in most cases treat only a small retinal (or cortical) area. This makes ERG recordings particularly challenging, particularly if certain components of the noise are time locked with the stimulus and therefore will not be reduced by averaging. A good example of this can be found in Stronks et al.[32] The ERG signals recorded in Argus II recipients were heavily contaminated by stimulus artifacts, but also by rapid pulses generated by de-multiplexing electronics in the implant itself, which could only be removed through sophisticated multistage filtering. The underlying retinal activity had extremely small amplitudes, so only through long recording times and signal averaging could these putative responses, with amplitudes well below 100 nV, be extracted—clearly not a practical procedure for assessing visual function in clinical trials.

The same authors demonstrated, however, that VEPs in these same Argus II recipients contained a clear response signature and that foveal versus extrafoveal stimulation resulted in two distinct waveforms with a larger amplitude in response to foveal stimulation, just as would be expected for flash

and flicker VEPs in normally sighted observers.[29] Thus, for interventions affecting the macula, the VEP can be a useful tool, by virtue of the enlarged and exposed projection of the macula on the occipital pole.

Stimulus Considerations

The sensitivity of the visual system in patients who are candidates for vision restoration trials will typically be severely reduced, and even after successful treatment the sensitivity is likely to remain well below normal levels. In general, therefore, high stimulation levels will be required to elicit even a small response, and as a general rule white light may be preferred to obtain the largest response possible. Here, again, though, the choice of stimulus parameters must be guided by the properties of the system. As an example, an intervention aiming to restore rod function will require testing under dark-adapted conditions, leaving the subject in complete darkness for at least 40 minutes to allow for slower than normal adaptation, and using a stimulus wavelength that preferentially stimulates the rods (i.e., wavelengths below 500 nm, or blue light).

As mentioned before, rod and cone function can be differentiated by the temporal frequency eliciting the optimal flicker response in the ERG. Cone photoreceptors have a shorter refractory period than rods, so flicker rates around 30 flashes/s are typically used for cone-mediated flicker ERGs, whereas 9 or 10 flashes/s (in combination with dark adaptation and short stimulus wavelength) will be optimal to elicit a rod flicker ERG. For flicker VEPs, repetition rates ranging from 10 to 20 flashes/s are optimal.

Electrical Pulse Stimulation

Most types of vision restoration aim to improve or re-create light sensitivity, and for these approaches light stimuli will generally be used to ascertain functionality. In some cases (e.g., prior to optogenetic transfection of ganglion cells), electrical stimulation may be useful for testing the integrity of the retinocortical pathway. Electrical stimulation would also be useful if the VEPs to light stimuli are too small to be recorded reliably, and there is concern that the visual pathway may be compromised.

Electronic retinal implants with external cameras typically have a fixed internal frame rate, limiting the allowable stimulus repetition rate to the frame rate or an integer fraction of it. This limits the frequency choices for flicker stimuli and must be considered when designing a protocol using the ERG or VEP as an outcome measure. Moreover, the likelihood of large

artifacts in the recorded ERG signals is high.[32] VEPs elicited by retinal electrical stimulation are much less likely to be contaminated by artifacts, and several successful applications have been shown in the literature.[29]

Electrical stimulation using cortical implants is likely to cause large switching and stimulation artifacts in the VEP, so the potential to study cortical processing with these implants may be limited. This is likely to be an area of study in the next few years, as more cortical visual prostheses enter clinical trials.

Sample Special Cases

To end this section, let us look at a few special cases that illustrate some of the considerations when pursuing evidence of functional vision restoration using electrophysiology.

Central Vision Restoration in AMD

The optimal outcome measure is likely to change as the trial progresses:

- In early-stage (feasibility) studies, it will be important to demonstrate that the intervention is safe (i.e., that function in the intact peripheral retina is not adversely affected by the intervention). Standard ISCEV recordings of the ERG[20] before the intervention and at several follow-up times are the most appropriate tools to monitor overall retinal function, but to rule out more localized adverse effects this should be supplemented by a multifocal ERG,[33] provided the subject can reliably fixate the center of the hexagonal stimulus pattern throughout the recording time, which may be in excess of 10 minutes.
- To establish functionality of the treatment (i.e., restored functionality in the macula), the VEP is a more appropriate outcome measure, by virtue of the large macular representation in the visual cortex. But here a more judicious choice of the appropriate stimulus will have to be made to be able to distinguish macular and peripheral responses, and this choice will depend on the nature of the treatment:
  - For electronic implants, one can make use of the broader wavelength selectivity of electronic imagers, particularly their sensitivity to near-infrared wavelengths; stimuli with wavelengths between 800 and 850 nm will be invisible to the peripheral retina, and any measurable response in the VEP must be mediated by the implant.
  - If the treated and native retinal areas cannot be distinguished by wavelength sensitivity, spatially restricted stimuli must be used, and fixation must be monitored to ascertain that the stimulus falls on the treated retinal area. As mentioned before, the effect of stray light from the stimulated area onto the peripheral retina can be reduced by steady illumination in the peripheral visual field, but a better approach may be to balance stimulus intensity by using a counterphase modulated pattern.

Optogenetics Treatment in the Macula

For safety monitoring, ERG methods can be used as indicated above. For efficacy, however, one can make use of the fact that transgenic light-sensitive channels created by the treatment will have different temporal and dynamic response properties from the native photoreceptors. By varying adaptation level, stimulus amplitude, and flicker frequency one can separate components in the VEP, similar to the techniques used to separate rod- and cone-mediated responses in the ERG.[34] Optimal stimulus and analysis choices will be determined by the specifics of the receptor channels and the native retina.

Recording Through Electronic Implants

As mentioned above, recording ERG or VEP responses to electrical stimulation by retinal or cortical visual prostheses requires extensive averaging and postprocessing due to artifacts arising from the electrical stimulus and implant electronics; the only possible exception is VEPs in response to retinal electrical stimulation.[29] It is conceivable, however, that future electronic implants will have built-in recording capabilities through the implanted electrodes. Not only will this allow signals to be captured in the source tissue rather than at a distant location (cornea or scalp), yielding much larger response amplitudes, but this may also guide proper design of the implant, helping to ensure that recording electronics are isolated during stimulation, eliminating stimulus artifacts and thus the need for arduous postprocessing. This design is common in cochlear implants and has led to improved signal processing and much better performance being achieved with those devices.

Full-Field Stimulus Threshold Test

It may seem strange to include a psychophysical test of dark-adapted threshold sensitivity in this section about electrophysiology, but it is justified by the fact that this test is run on an ERG Ganzfeld stimulator (Espion ColorDome; Diagnosys LLC, Lowell, MA (see

Fig. 7)) and generally administered by technicians most familiar with that equipment (i.e., clinical electrophysiology staff).

When vision has dropped below a level that can be measured psychophysically with common visual function tests (visual field, ETDRS visual acuity, contrast sensitivity), there are few commonly accepted methods available to monitor progression or improvement. The development of the full-field stimulus threshold test (FST)[35] created a psychophysical method to measure the illuminance necessary to be perceived by the most sensitive parts of the retina and thus obtain a quantifiable threshold that could be assessed before and after intervention, even in subjects with extremely low vision. This most sensitive area is tested without knowing its exact location, which is especially useful because treatment effects do not always occur in predictable areas; subretinal injections are a good example of a treatment where benefit may be patchy and are difficult to pick up with full field or even multifocal ERG.

As an additional benefit, this method does not require fixation, making it useful even in subjects with better vision who are unable to reliably perform perimetry due to conditions such as nystagmus. The further development of the method moved the test from a customized hardware base[36] to software on the Espion ColorDome LED-based full-field stimulator.[37] This software is now commercially available as the Diagnosys full-field stimulus threshold test (D-FST) in the software library of the Espion system. Thus, a technique for obtaining full-field thresholds has become a de facto standard that has been used as an outcome measure in the Argus II retinal prosthesis trial (Dagnelie G, unpublished observations) and Luxterna gene therapy trials.[38] It is currently being explored as an outcome measure in numerous other early-phase prevention and restoration trials for inherited retinal disease.

There is not currently a standard protocol for this test that is comparable to the ISCEV standards, but the expectation is that such a standard may be published in the near future. Until then, multiple publications using the FST are available in the literature.

Reporting Guidelines

Any publication or presentation reporting the results of electrophysiological recordings in vision restoration trials should include sufficient information to allow replication of the work, including the following:

- The name of the test and, if applicable, any changes made to the corresponding ISCEV standard
- If the test does not correspond to a previously validated (ISCEV) standard, an explanation why this particular test was selected for the population studied and information regarding the validation procedure used to ascertain accuracy and precision in normal vision, as well as in the study population
- If the test has not been previously published or is not available to the general public, accuracy and precision data in normally sighted observers of similar age and gender as the study population
- Any non-standard equipment or settings required to run the tests
- Normative values and confidence intervals in normally sighted observers age and gender matched to the study population

Given the availability of a number of validated and well-calibrated ISCEV standards for assessment of retinal and higher visual pathway function, the working group expresses as its strong opinion that all clinical trials seeking regulatory approval should adhere to these normative tests as closely as possible, with clear arguments why the adaptations used are most appropriate for the study population and the intervention.

Electrically Evoked Device Effectiveness

Matthew Petoe (chair)[1], Daniel Rathbun[2], Ethan Cohen[3], Ione Fine[4], and Ralf Hornig[5]

[1]**Bionics Institute of Australia, East Melbourne, Australia (e-mail: mpetoe@bionicsinstitute.org)**
[2]**Werner Reichardt Centre for Integrative Neuroscience and Institute for Ophthalmic Research, University of Tuebingen, Tuebingen, Germany**
[3]**Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA**
[4]**Department of Psychology, University of Washington, Seattle, WA, USA**
[5]**Pixium Vision SA, Paris, France**

Introduction

This section applies to visual restorative approaches that utilize a device, such as a visual prosthesis, to induce vision by electrically or perhaps chemically stimulating neural tissue. This section is also relevant
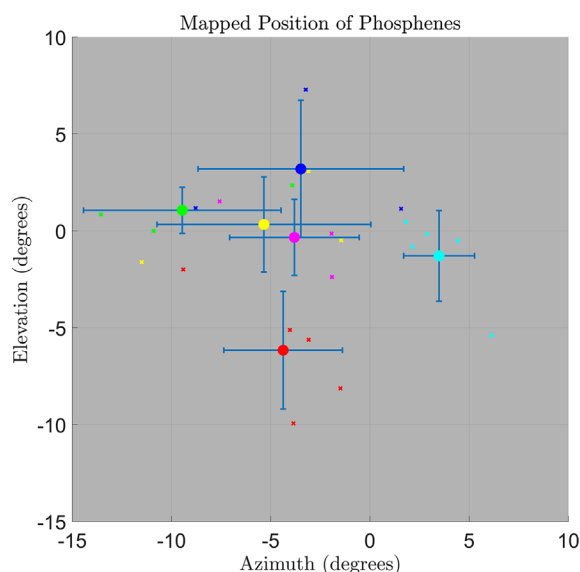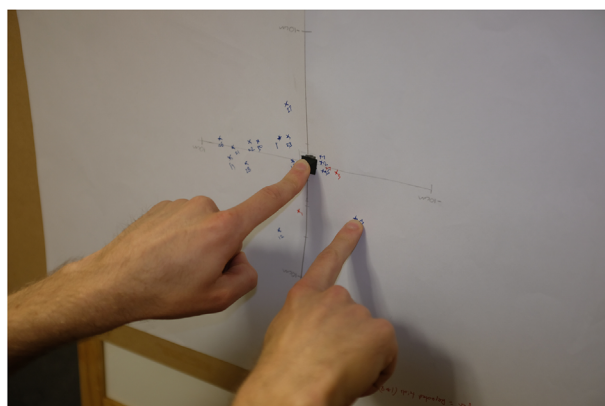
Figure 8. A method of phosphene mapping using an easel. (*Left*) The participant is instructed to place their left and right index fingers on a tactile marker positioned within a large sheet of paper mounted on an easel. After a short stimulus, the participant moves their right index finger to the remembered position and holds it in place while the researcher marks the paper. (*Right*) Multiple measurements ("x") give an indication of each phosphene position, with the average position indicated by a solid colored circle. The bars indicate ±1 SD of phosphene position measurements. Data courtesy of Bionic Vision Technologies, Australia.

for optogenetic and photo-switch approaches that use a device to deliver photic input to the genetically modified neurons.[39] The goal of this section is to provide guidance for those who conduct human psychophysical experiments, in terms of both methodology and reporting.

Differing Technologies

There are many variations in the design of prosthetic systems, especially with respect to how visual images are captured (with external cameras or implanted photodiode arrays, for example) and the number and density of implanted electrodes.

Visual prosthetic devices all require a camera or photosensor to dynamically capture the visual scene. Translation of this visual information into induced neural activity can be mediated by image processing algorithms (for camera-based systems). The resulting output must then be converted into either a light (for optogenetics and photo-switches) or electrical stimulation protocol. This stimulation pattern is then processed by the patient's remaining visual system. For many devices, testing can be carried out either via direct stimulation of the device electrodes or under naturalistic conditions where input is provided by the camera. For others, such as photodiode arrays, only naturalistic stimulation is possible.

The location of the stimulating array varies among devices. For visual prosthetic devices, the interface with neural tissue can be in the suprachoroidal space, the retina (either epi- or sub-retinal surface), optic nerve, lateral geniculate nucleus (LGN), occipital lobe, or higher visual cortical association areas.

Sensory substitution devices (i.e., those that take visual images and provide sensory input to the patient via some non-visual sensory means) may position a stimulating array on the tongue, forehead, corneal surface, or lower back, among other possibilities.

The commonality across all these methodologies is that recipients will perceive a representation of the visual environment as a sensory construct. For devices that provide visual input, the evoked percepts are referred to as "phosphenes." Use of the term phosphene conveys that some visual percept was induced, but by itself this term does not convey the quality, detail, or perceptual value of the percept.

Behavioral Measures

The following basic aspects of phosphene generation should be considered and reported when possible:

1. *Phosphene (or perceptual) threshold*—A measure of the electrical charge and/or light intensity (or tactile force, in the case of a sensory substitution, tactile

prosthesis) required to produce a percept. It is necessary to know the specific details of the technical methods used to achieve a given threshold, as thresholds vary in accordance with many variables, including pulse duration, frequency, inter-pulse spacing, duration of spike train, polarity, waveform, electrode geometry, material and coating, use of one or more electrodes simultaneously or interleaved, distance from the retina or retinal neurons, location of stimulation (e.g., in the macula or periphery of the retina), the status of the neuronal substrate (as might be assessed by optical coherence tomography for retinal devices), residual visual capacity of the subject, and duration of blindness.[40-43] There are a variety of methods for estimating thresholds, so the method used should be described, including whether the threshold for each individual electrode or diode was independently measured or whether electrode thresholds were interpolated across the surface of the device (as might be done to speed the process of threshold determination for high electrode count devices). In addition, methods to validate the reliability and precision of the thresholds should also be reported.

2. *Phosphene brightness/size*—Both phosphene size[44] and phosphene brightness[44,45] as a function of stimulation properties can generally be estimated quickly and reliably using subjective magnitude ratings. For electrical devices, useful stimulation properties to test and report include the relationships among current amplitude, pulse duration, frequency, and subjective brightness/size. For photodiode devices, the equivalent of current amplitude is external luminance, although the luminance–diode current relationship should also be reported whenever possible.

Brightness estimates can be misleading if untested assumptions are made about the brightness of a just-detectable phosphene. A threshold phosphene can appear either dim or bright, depending on the location of the stimulating electrodes and the spatial and temporal properties of the current waveform. Therefore, consistent procedures for assigning subjective size and brightness ratings should be developed and validated ahead of time to maximize comparability across subjects. When possible, brightness scales should include the ratings given to just-detectable phosphenes.

Size estimates are strongly dependent on perceived distance. Size ratings should be related to the size of a familiar object of known size held at a known distance within arm's length. For example, a US quarter coin held 2 feet from the eye subtends a little over 2 degrees of visual angle.

1. *Phosphene persistence*—In many current devices, there is a desensitizing phenomenon[42,46] whereby the phosphene gradually fades over time.[47] One way of measuring this desensitization is by asking subjects to report the duration that a phosphene persists in relation to the duration of stimulation. Variables that should be considered and reported include how quickly the brightness of a phosphene fades during the stimulus and whether there is an additional offset response or persistence after the stimulus ceases.[47] The effects of stimulation parameters, such as stimulus amplitude or frequency, can also be of interest.

2. *Phosphene shape*—It has been shown that subjects can reliably draw the shape of phosphenes, for example on a touch screen monitor.[48] The effect of variation in stimulation profile and electrode location on the features of phosphenes should be studied.

3. *Phosphene stability*—One critical factor is whether threshold and phosphene shapes remain stable over time. Repeating a standard set of measurements over months or years is therefore recommended.[42]

Additionally, the following text addresses the discriminability and localization of phosphenes. There may be technical limitations to these investigations, but researchers should attempt to provide an indication of device functionality with regards to these issues:

1. *Relative phosphene position*—A map of phosphene position relative to the corresponding electrode position within the electrode array. Even for visually normal subjects, estimating the location of a spot of light without visual references is challenging.[49] In severely blind subjects, the eyes often are misaligned. When blindness has been severe since early in life, patients may not be able to volitionally reposition their eyes into a primary position, and nystagmus may be present so that there is no constant position. These confounding variables make it extremely difficult to determine the absolute positions of phosphenes in either the visual field or in real-world coordinates. Methods for examining relative phosphene position are described in more detail below and an example experimental setup is shown in Fig. 8.

2. *Phosphene discrimination*—The ability to discriminate phosphenes from one another. This is a much more complex topic than is implied by the term "discrimination." The phosphenes elicited by stimulation from neighboring electrodes A and B may look similar or different individually. Differences may be due to different retinal circuitry, different

connectivity between electrode and retina, different responses to increases in stimulation level, or different degrees of retinal degradation. If the phosphenes are similar, they may or may not be identifiable as different when they are turned on at different times. When they are turned on simultaneously, they may be resolvable (two-point resolution) or not, or the combined phosphene may look entirely different than either phosphene alone. These complexities are all increased exponentially when electrodes C, D, etc., are added to the sequence. Basic tests of phosphene discrimination should follow a basic mapping of single-electrode phosphene shapes, sizes, and locations and should include two-point resolution for at least a representative sample of working electrode pairs. Testing should explore the ability to distinguish two phosphenes in relation to use of specified electrode geometries, distributed at a given pitch (i.e., center-to-center spacing) and with respect to a given stimulus profile (e.g., amplitude, duration). Qualitative descriptions might include whether phosphenes overlap and whether they are of similar size, geometry, and brightness. Researchers will appreciate that there will be a spatiotemporal interaction among neighboring electrodes,[50] so the effect of interleaved versus simultaneous stimulation,[51] or any other variation in the stimulus profile used across electrodes, on the ability to discriminate phosphenes should be stated.

3. *Phosphene field*—A measurement or indication of the spatial extent of the visual field that is occupied by elicited phosphenes.

Other qualitative descriptions might include the color and sharpness of induced percepts, especially how these features vary with stimulation. In addition to responses given to the specified questions to assess the perceptual parameters described above, patients should also be encouraged to provide their own comments to enhance understanding of what they have seen.

Neurophysiological Measures

In addition to the above psychophysical results, it can be useful to also report physiological results. Quantitative measures of retinal images obtained with optical coherence tomography provide useful information about the anatomical status of the retinae being studied. Results obtained by electroretinography[52,53] or visual evoked potentials[54] may also be useful in assessing the relationship between neural responses in the visual pathway and perceptual outcomes.

Psychophysical Procedures

General Considerations

Visual prostheses can utilize a wide range of stimulation parameters and strategies to effect neural responses. Stimulation and processing parameters used in clinical trials should be detailed enough to provide a clear understanding of the relationship among stimulus parameters, the resulting retinal stimulation, and perceptual outcomes.

Reporting of perceptual results should include information about the waveform, amplitude, polarity (e.g., cathodic or anodic first), duration of a single biphasic pulse, inter-pulse interval, frequency, and total duration of a single pulse train (which includes numerous individual biphasic pulses). A full list of reportable parameters is provided in the Appendix. Test stimuli should not exceed the conventional limits on charge stimulation safety (as is relevant for the specific electrode geometry and material being used).

However, there is no need to provide detailed specifications on the full range of stimulation parameters that can potentially be assessed with a given device if this information could reveal proprietary information and is not germane to understanding the results of the published studies. An appendix of reportable device parameters is included. These reporting guidelines likely will be revised to incorporate future innovations in best practices and new technological developments.

Standard Psychophysical Paradigms: Protocols

Two-interval, forced choice—Two temporal intervals occur (generally signaled by an auditory cue), and subjects are asked which interval contained a particular percept, such as which interval contained a phosphene, which interval contained the larger percept, or which interval contained the brightest percept. The advantage of this method is that it avoids subject criterion effects. Disadvantages are that, because there are two intervals, chance performance is 50%, so a fairly large amount of data must be collected to find an accurate threshold.

n-Interval, forced choice—This is similar to the two-alternative forced choice, but the subject is asked which of three or more intervals has the brightest (for example) stimulus. Because chance performance is now 33%, this method is considerably more efficient (even though there are now three intervals). Because there is a slightly larger memory component, it may not be suitable for subjects with memory loss or cognitive difficulties.

Two-alternative, forced choice—A single stimulus is presented and subjects must report whether (for

example) whether or not a stimulus was presented or whether there were one or two stimuli. This is an efficient method, but it is susceptible to subject bias; for example, one subject may say there is a single phosphene unless they were confident there were two distinct phosphenes. A different subject (or the same subject on a different day) might report two phosphenes whenever they see a complex shape. Thus, the same perceptual experience might result in very different patient reports. In the case of detection tasks, catch trials (in which a null stimulus is presented at random intervals) should be used in 10% to 20% of the total number of trials.

Rating—A classical brightness rating procedure was described by Stevens.[55] Subjects are first presented with a visual stimulus with an agreed reference brightness (e.g., 10) and then asked to numerically rate the brightness of a second stimulus in relation to the first. For example, a subject would assign a value of 20 if the second stimulation appeared to be twice as bright as the initial percept. Stimuli should always be presented in a random order. The reference stimulus need not be provided every trial but should be provided regularly, such as at the start of the session and perhaps at a minimum of every five trials. The reference can also be included as a member of the test set, as this provides a useful way of assessing subject rating accuracy. Subjects show surprising reliability on this task across an incredibly wide variety of domains,[55] including rating the brightness[44,56] and size[44] of phosphenes.

Stimulation Methods

For devices with camera-based systems that have wired or wireless access to stimulating electrodes, psychophysical experiments should be carried out by direct control of the electrodes whenever possible.

For devices with no direct access to single electrodes (e.g., photodiode devices), whole-array psychophysics may be conducted using a calibrated full-field stimulus source, such as a Ganzfeld flash stimulator.[57] This approach has been shown to be sufficient to determine global parameters such as activation threshold and amplifier gain.[52,58] Photodiode devices may preferentially use red or infrared light rather than white light to avoid confounding effects of photic stimulation of surviving photoreceptors.[59,60] Photodiode devices should also include control tasks comparing white and red/infrared light to differentiate contributions from residual vision. To confine activity to a subset of electrodes, focused light can be projected directly onto photodiodes by automated tracking of the fundus or an adaptive optics scanning laser ophthalmoscope.[58] If head-worn display goggles are used to drive a photodi-

ode array, the optical display specifications (e.g., power spectrum of emitted light, total brightness of emitted light, transparency, field-of-view) should be reported.

For systems or research environments that are unable to utilize the above methods or for tasks that want to include the effects of eye or head motion on visual performance, visual stimuli can be delivered with a computer screen positioned at a specified distance from the patient's eyes. A typical distance is 57 cm, so 1 cm on the screen corresponds to 1 degree in the visual field. Using this approach for a photodiode-based system, the Basic Assessment of Light and Motion test [27] is a suggested method to assess light perception, temporal resolution, object localization, and movement detection. Room lighting should be uniform, controllable, and either dark or dimmed (e.g., below 300 lux) to avoid confounding perceptual experiences for patients that have some level of residual vision, even if only bare light perception.

Stimulus Parameters

For direct-stimulation devices (i.e., those with either wired or wireless capability to address and drive electrodes), individual electrode thresholds should be reported, as this measurement is most directly comparable across various devices and groups.

For real-world applications, electrical stimulation is likely to be performed by pulse trains, which generally yield lower thresholds than single-pulse stimulation.[50] In general, when measuring detection for pulse trains, relatively short pulse trains (e.g., 0.5 seconds) are recommended.[42] This duration is on the shorter end of the stimulus durations that are traditionally used in light psychophysics (0.5–1 second). Remaining at the shorter end of this range may help reduce confounding effects of desensitization over multiple trials (see below).

Detection (Threshold) Tasks

The methods of obtaining a perceptual threshold for each electrode are likely to differ among groups, but in any case the methods should be robust and repeatable, balancing rapid convergence with fault tolerance, to monitor the basic functionality of the electrodes and the psychophysical experience of each subject.[57]

As described in the inset box, the method of constant stimuli using two interval-forced choice is the most reliable method for estimating most psychometric functions, including a detection threshold. However, this method is time consuming, especially when the range of expected threshold values is not well known. One approach, if highly precise measurements are desirable, is to use a staircase procedure

to crudely estimate the psychometric function and then use constant stimuli. But, for most purposes, the accuracy provided by two-alternative and staircase procedures (as described below) should be more than adequate.

As described below, a threshold should be thought of as the current amplitude required to reach a given performance level; as such, the threshold value will differ depending on what performance level is defined as the threshold (e.g., 50% vs. 75% detection) and what task is used (e.g., two-interval vs. two-alternative). Consequently, the sensitivity index d'[62] should be reported to allow comparison across studies, where possible. For tasks where subject criteria play a role this will require a measurement of subjects' criteria (e.g., false-positive rates in a yes/no task). Sample MATLAB code for calculating d' in psychometric functions can be found as an appendix in Fine and Jacobs.[166]

Psychophysical Procedures: Stimulus Choice

Method of constant stimuli—The observer is presented with a fixed, predetermined set of stimuli of which some are above and others are below threshold. The stimulus set is presented in a random order. Advantages are that this method prevents the observer from being able to predict what the next stimulus will be and minimizes the effect of fatigue on estimated thresholds. One disadvantage is that this approach is time consuming, especially when the range between the above and below thresholds is not well known so a large number of stimulus intensities must be included.

Staircase—In staircase procedures, stimulus intensity (e.g., current amplitude) is adaptively increased for incorrect responses and decreased if there is a series of consecutively correct responses. This provides an efficient way to focus trials on stimulus intensities that are near threshold. The number of consecutively correct responses that are required to decrease the stimulus intensity determines the point on the psychometric function (describing probability of detection as a function of current amplitude) that is targeted. For example, the 1-up/2-down variant of the transformed up/down method[61] will converge toward presenting current amplitudes that result in a detection performance of 71%.

ML staircases—These methods use maximum likelihood algorithms to select the stimulus intensity for each trial that is expected to provide the maximal amount of information about the threshold, given the previous history of trials. Although highly efficient in theory, keypress errors early in the staircase can result on the staircase taking a very long time to converge

or, if the number of trials is limited, converging to an incorrect threshold. Methods susceptible to keypress errors are best used with highly experienced and reliable observers.

Regardless of whether data are collected with a method of constant stimuli or a staircase, the resulting data can be fit with a psychometric function to find the current amplitude that results in a specified level of performance (which need not be the convergence point of the staircase). The stimulus strength that gives $d' = 1$ is a common choice of discrimination threshold.[63] If enough trials are collected in a staircase procedure, the threshold for the convergence performance level is reasonably well approximated by averaging current amplitude across the last five or so staircase reversals.

No matter how the thresholds are determined, it is recognized that threshold values may vary over time or even within a single test session.[41,64,65] One recommendation is to select a subset of tests that can be repeated on a regular basis to examine the stability of these measurements over time.

Discrimination Tasks

These tasks focus on whether two stimuli can be successfully discriminated. Analogous to detection tasks, discrimination tasks can be carried out using two-interval forced choice or n-alternative forced choice and stimulus difference can manipulated using either method of constant stimuli or a staircase procedure.

One example of a commonly used discrimination task is brightness discrimination. The standard stimulus has a fixed current amplitude, and the test stimulus is varied in amplitude using a staircase procedure.[45] Subjects are asked which interval contains the brighter stimulus. The psychometric function describing the ability to discriminate the two stimuli as a function of the amplitude difference between them gives insight into how many discriminable brightness levels can be generated based on varying amplitude. Analogous experiments could be carried out examining how brightness varies as a function of other properties of the pulse train, such as frequency.

Matching Tasks

The goal of matching tasks is to find the point of subjective equality (PSE) between two stimuli. For example, in the case of brightness matching, PSE might represent the stimulus intensity at which the same pulse train on two spatially separate electrodes,[45] two different pulse trains on the same electrode,[23] or

different pulse trains of different electrodes[50] appear equally bright.

The recommended procedure is generally a two-interval, forced-choice protocol, in which each interval contains either a reference stimulus or a test stimulus (in randomized order). The test stimulus should be modulated (e.g., by amplitude or frequency) using, for example, a 1-up/1-down staircase procedure based on the subjects' report of which interval contained the brighter stimulus in the previous trial. The PSE describes the test stimulus intensity where the subject is equally likely to say that the test or the reference stimulus is brightest. This is the 50% point in the psychometric function. The PSE can either be estimated from the psychometric function or be estimated as the average across the last five or so multiple reversals in a 1-up/1-down staircase.

If a time-saving procedure is desired, with less precision, the reference and test stimuli can be presented in a single interval (e.g., interleaved) and the device setting modulated based on the subjects' report of whether the test stimulus is brighter or duller than the reference stimulus.

Rating Tasks

Rating experiments can be used to examine a variety of qualia, including brightness, size, and flicker. Indeed, rating experiments provide an excellent way of asking patients to directly report their perceptual experiences and are also extremely efficient time-wise. Rating experiments are particular informative in the case of brightness, where they can help determine an effective dynamic range and can also help brightness balance across multiple electrodes.

It is important with rating tasks to be very specific about what is to be rated. For example, in the case of size, subjects could be asked, "How much paint would you need to cover the phosphene? If you would need twice as much paint, then report 20." In the case of brightness matching, if the subjects report that the phosphenes elicited by an electrode are non-uniformly bright, the degree of brightness can be estimated based on either the average brightness or perhaps the brightest part of the percept. In such cases, the records should include a comment about experimenter instructions and any comments about strategy that patients make during the test. The subjects should be always advised to distinguish a change in brightness (or whatever qualia they have been asked to judge) from a concomitant change in the size or other irrelevant properties of the percept. Whereas it is possible to ask patients to report more than one qualia (e.g., size and brightness on a single trial), rating tasks are extremely time

efficient, so it is probably preferable to report on a single qualia on a given trial to avoid response interactions.

Performance on almost all rating tasks follows a power law.[55] Thus, the relationship between stimulus intensity and brightness can be described using the equation $B = aC^b$, where B is the brightness rating of the subject and C is the stimulus intensity.[26]

The perceptual dynamic range can be inferred by comparing the brightness rating just above threshold and the brightness rating at a stimulus intensity at which either safety limits are reached, or the level of brightness reaches an asymptote (i.e., when subjective brightness stops increasing with increased stimulus intensity). The reporting of the dynamic range should include a representative selection of electrodes across the array, with specification as to the location of the electrodes, such as foveal, parafoveal, or peripheral (ideally reported with specific retinal coordinates, such as 2 mm temporal to the fovea along the horizontal meridian).

It is important to be aware that there is no guarantee that a constant reference will produce the same brightness across repeated trials, especially if the standard current is applied repeatedly, as adaptation can occur. If this is a concern, then it should be noted that subjects can estimate magnitudes reliably without a reference. The subject is instructed to respond "zero" if no phosphene is visible and a number proportional to the brightness of any visible phosphene. Different subjects may choose different scaling factors, but these can be normalized by assigning proportionality constants that equate the mean estimates of the different subjects.

Brightness Balancing

To produce a brightness-balanced map of image intensity to subjective brightness, the least sensitive electrode (the one with the lowest maximum brightness rating) should be used as a reference. The dynamic range of more sensitive electrodes can then be attenuated to match that of the least sensitive electrode. If brightness rating experiments suggest that a similar power function exponent can be applied across all electrodes, then brightness balancing can be carried out using the equation $C_T = aC_R$, where $a = SI_R/SI_T$. SI_R can be either the stimulus intensity on the reference electrode at the point of a brightness match or the stimulus intensity that produces a particular brightness rating (e.g., 20). Similarly, SI_T is either the stimulus intensity on the test electrode at the point of that brightness match or the stimulus intensity on the test electrode that produced a brightness rating of 20 (if power functions differ significantly across electrodes[55]).

If this procedure is followed, then the researcher can report that brightness-balanced maps were used for subsequent vision testing.

Phosphene Shape Tasks

Where possible, phosphene shape and size should be recorded at a range of supra-threshold levels using patient drawings (e.g., on a touchscreen, using finger-tracking).

Reportable outcomes should include a description of shape and an indication of size in degrees on major and minor axes.[44] The effect on phosphene size and shape of modulating amplitude, frequency, time-course, or other stimulating strategy should be described.[44]

Phosphene Position Tasks

Reliable and repeatable assessment of phosphene position remains problematic, although it is clear that a combination of absolute and relative phosphene position measurements gives complementary results.[66]

One method of recording absolute phosphene position was described in 2015 (Kaskhedikar GP, et al. *IOVS*. 2015;56:ARVO E-Abstract 4315). These researchers presented a stimulus pulse train while the subject attempted to fixate centrally. At stimulus cessation, the subject was instructed to shift their gaze to the remembered phosphene position. In this manner, a basic map of phosphene position can be inferred from eye gaze position, with minimal angular distortion and acceptable radial distortion.

Another procedure mapping phosphenes in absolute coordinates was developed by Brindley and Lewin[1] and later implemented in an optic nerve prosthesis by Brelén et al.[67] The subject pointed with the left hand at a central point and was then instructed to fixate their gaze on this point and use the right hand to point to the location of the phosphene on the inside of a hemispherical surface. Through the application of this map to images captured by a camera, the subjects were able to perform simple pattern recognition tasks.

Examining relative position may provide a more sensitive measure, especially for phosphenes that are close to each other in space. This can be done by asking the subject to maintain their head in a straight-ahead position, look straight ahead, and report the location of two rapidly presented induced phosphenes in relative space (i.e., the location of the first and second phosphenes with respect to one another). Information can then be collected on whether the relative positions of evoked phosphenes correspond retinotopically to the geometry of the electrode layout. If eye gaze is monitored, trials in which there is a shift of gaze can

be excluded from analysis. Dagnelie[57] proposed asking subjects to report using a clock hour system with 37 presentation points, 1 centrally located and 36 in rings of 12 at increasing eccentricity from center. In this paradigm, the patient can respond with "central" or with a clock hour followed by "close," "middle," or "far." An alternative, simpler verbal response is the eight directions of a compass.

Phosphene mapping is a particularly critical factor for prosthetic electrode placements in which electrode configuration may not map systematically on to the visual field, such as optic nerve and thalamic devices, as well as cortical implants that may be positioned to lie along the banks of multiple gyri. In non-retinal arrays, phosphenes may vary in size, threshold, and position due to the magnification of the foveal representation. Therefore, an accurate reporting of the location, eccentricity, and proximity of the stimulation electrode array to the target tissue is recommended.

If assessments indicate that phosphene position is topographic, then it is possible to use electrode layout as a basis for representing visual space. In this event, the researcher should state that electrode geometry was used as a basis for image sampling in subsequent vision testing.

Other Factors to Consider

Desensitization and Persistence

The time course of phosphene brightness should be investigated, as it is often reported that repetitive electrical stimulation can lead to brightness fading.[2,42,46,47,68]

Researchers can examine phosphene persistence over prolonged durations using continuous sampling of subjective brightness, such as using a joystick to describe the brightness profile,[47] using a potentiometer, or having the subject trace the brightness with their finger (analyzed using videorecording). A reportable metric is the average time taken for phosphene brightness to fade to below 50% of its initial brightness. In some instances, an increase in brightness over time may be observed, especially after stimulus offset.[47] Measures of persistence tend to be relatively variable, so it is important to collect enough trials to have an estimate of trial-to-trial variability. Less subjective and variable methods for measuring and modeling desensitization have only been applied to desensitization over very short time scales (∼1 second).[42]

There are a number of situations where desensitization is a particular concern. First, when carrying out a two- or n-alternative forced choice experiment, there may be interactions between the stimula-

tion intervals, such that the second interval is desensitized. The effect of interval order should therefore be included in the analysis—for example, in a brightness discrimination experiment, did subjects show a systematic tendency to be less likely to report that the second interval was brighter? Given that the amount of fading seems to be correlated with input charge,[42] avoiding a two-interval forced choice procedure (or increasing the inter-interval duration) is recommended when using long pulse trains, long pulse widths, or high amplitude stimulation or when the amount of charge presented in the two intervals is significantly unbalanced.

Second, and likely more importantly, sensitivity may decline over the duration of a single session.

Interactions Between Electrodes

Interactions between electrodes occur at multiple levels:

1. The current fields of two electrodes that are stimulated simultaneously differ from the additive sum of the current field of each individual electrode. These interactions can easily be eliminated by making sure electrodes are stimulated in a rastered sequence.
2. Spatiotemporal interactions also occur in stimulating neurons that lie intermediate between two electrodes.[50] In the simplest case, the pair of pulses received by a neuron lying between two electrodes will result in a different response than would be obtained by stimulation by either electrode independently. These effects can either be facilitatory or suppressive, depending on inter-pulse delay.[51,69] For example, a neuron that lies between two electrodes might be more sensitive to a pair of pulses generated by two electrodes than to the pulses generated by either electrode alone. Alternatively, the neuron might desensitized by prolonged stimulation on electrode A, thereby becoming less responsive to electrode B. Neuronal interactions can take place over considerable distances on the retinal surface due to phosphenes generated by axonal stimulation[39,44,51,70] and are likely to occur over a time scale of many milliseconds.[42]

In general, given the complexity of spatiotemporal interactions, the best strategy is to raster electrodes so as to maximize the temporal separation between pulses on different electrodes and to be aware that spatiotemporal interactions are still likely to be concern.

It should be noted that the potential to exploit these interactions is a promising future direction. The use of current-steering across some fixed number of physical electrodes has the potential to create percepts whose centroid location is intermediate between two electrodes.[50,71,72] When reporting current steering data, it is important to include the normal description of stimulus parameters used at each electrode as well as how current steering was carried out. Thresholds for individual as well as for virtual electrodes should be reported, because including anodic shaping currents may lead to significant drops in sensitivity.[73]

Auditory Cueing and Feedback

The use of auditory cueing, including auditory feedback, should be reported.

It should be noted that, especially if feedback is provided, a subject is capable of learning to perform a given task using any perceptual information that is available. For example, a subject may be able to successfully report which interval contains two phosphenes without actually ever seeing two phosphenes in either interval, simply by choosing the brighter interval (see section below). Despite this, auditory feedback should be used whenever possible, especially for low-level tasks (such as detection and simple discrimination tasks) because patients perform surprisingly accurately in a stimulus range where their subjective experience is one of guessing. With feedback, subjects remain motivated in this stimulus range.

Artifactual Cues

When subjects are given feedback, it is important to remember that any perceptual quality can be used to perform the task. For example, as described above, if subjects are asked to detect which of two intervals contains two phosphenes versus one, they may be able to perform significantly better than chance simply by selecting the interval containing the brighter percept. Subjects are somewhat less likely to rely on artifactual cues when performing a two-alternative forced choice test, in which only one variety of stimulation is presented from two possible choices and subjects are asked to give a quantitative response (e.g., one or two phosphenes). However, with feedback subjects are still likely learn to rely on artifactual cues over time.

The best way of dealing with artifactual cues is by adding variation in the stimulus along perceptual dimensions that you want to subject to ignore. Take, for example, a task where subjects are asked to identify whether there was one or two phosphenes. Before the experiment begins, one would carry out a preliminary brightness matching experiment so as to roughly match brightness across the two stimulation conditions.[51] One would then vary current amplitude (or frequency) around these brightness matched levels, so that brightness is no longer a reliable cue for the task. Data can then be analyzed to confirm that the proba-

bility of the subject reporting that there were two phosphenes cannot be predicted by the current amplitude of the stimuli.

Control Trials

Whenever possible, a significant percentage of control trials should always be included in a test session. Whenever possible, these control trials should be interleaved randomly with experimental trials.

Depending on the psychophysical test being performed and the device being tested, these control trials will vary widely. In detection tasks catch trials are generally those in which no current is used. In trials using video, input catch trials may require scrambling the input-to-electrode mapping. For systems that use goggles, control trials might include using very low intensities of light that had been shown by prior tests to be subthreshold or scrambling the visual input. For photodiode systems that do not use goggles and do not have direct access to electrodes, control trials can be performed by reducing the supplemental power delivered to the photodiode array to a level that had been determined to be subthreshold by prior testing.

Eye Movements

It has been established that the perceived location of percepts in real-world coordinates correlates with the position of the eye in the orbit.[74] Thus, psychophysical tests that might be influenced by eye position should be performed while asking the subject to look straight ahead, and preferably the eye position should be recorded.

Although this is technically challenging, it may be useful to assess the effect of a change in eye position on the perceived location of the percept to explore, for example, when a single electrode is stimulated twice with an intermediate eye-movement, whether there is a corresponding displacement of the second phosphene. However, it should be noted that these measurements will have to be made in the absence of eye-tracker calibration.[75] Thus, although it is relatively straightforward to see whether the reported position of a phosphene in visual space shifts (e.g., rightward with a rightward eye-movement) and whether the magnitude of the eye movement and the corresponding perceptual shift is correlated, it is very difficult (although not impossible[75]) to examine if the size of the shift is predicted by the magnitude of the eye movement.

Nystagmus (i.e., adventitial, rhythmic eye movements that have a slow phase) is not known to affect the perception of where a phosphene is located in space, but the presence of nystagmus should be noted and described with traditional terminology, including spontaneous or gaze-evoked; vector (i.e., horizontal, vertical, torsional, or mixed); conjugate (i.e., movement of the eyes in the same direction); dysconjugate (i.e., movement of the eyes in the same direction but with different amplitudes); or dysjunctive (i.e., movement of the eyes in toward one another [convergence] or away from one another [divergence]).

With regard to phosphene position, as much information as possible should be provided about how ocular position relates to the perceived location of phosphenes and to the relationship to camera angle. Minimum reporting should indicate whether subjects used eccentric viewing and whether eye movements were observed.

Appendix. Specific Device Parameters

There is a wide range of parameters that can vary among visual prostheses. Some of these may profoundly affect perception, others less so. We list these parameters below to highlight those of interest to the field and recommend that researchers report values for as many of these as is practical.

The FDA has prepared a guidance document (IDE Guidance for Retinal Prostheses, Food and Drug Administration, 2013) which, when adapted slightly, may serve as a basis for reporting on prostheses in other visual areas, although some of the parameters are more concerned with preclinical risk assessment or device safety. The adapted electrical specifications identified to be of relevance to the current document are as follows.

Electrode Specifications

1. Dimensions of the entire array
2. Number and spacing of electrodes
3. Layout and configuration (e.g., distal or colocated return, bipolar)
4. Material composition, coatings, and/or treatment (e.g., nanostructuring)
5. Description of macro-, micro-, and nanogeometry (e.g., planar, rounded, textured)
6. Geometric surface area, accounting for any methods of roughening, shaping, or nanostructuring
7. Surgical placement and anatomical position relative to the fovea, optic nerve head, optic nerve, LGN, or visual cortex, as applicable, with special attention given to the visuotopic azimuth, elevation, rotation, and distance from target cells, as measured from optical coherence tomography or equivalent.

Electrical Specifications

1. Whether the pulses are current or voltage regulated
2. Recordings or description of the current and/or voltage waveforms delivered by each pulse
3. Whether the stimulation is monopolar, bipolar, or some other configuration
4. The charge per phase delivered
5. The pulse charge density in mC/cm^2 per phase
6. The pulse sequence and polarities (e.g., monophasic or biphasic, cathodic-first or anodic-first)
7. Whether stimulation is simultaneous or interleaved and any inherent limit to the instantaneous number of electrodes that can be used to describe an image
8. For pulse trains, intra- and inter-pulse intervals, duration, and frequency (if an atypical envelope is described, then an example recording or image should be provided)
9. For the charge recovery method, whether the pulses are capacitively coupled, charge-balanced, or asymmetric
10. Configuration of unused electrodes during stimulation (i.e., whether they are shorted or floating)

Vision Processing Systems

Chris McCarthy[1] (chair) and Vincent Bismuth[2]

[1]Department of Computer Science & Software Engineering, Swinburne University of Technology, Melbourne, Australia (e-mail: cdmccarthy@swin.edu.au)

[2]Pixium Vision SA, Paris, France

Introduction

The role of vision processing that is embedded into visual prosthetic systems is to transfer information from a real-world visual scene to some alternative, artificial display. In its simplest form, data are continuously sampled from camera-captured images and transferred to a system-specific display. The display can provide a modified visual image or stream of data that can evoke visual percepts, or the display of artificial vision can be induced by delivering electrical stimulation to the visual pathway. An alternative approach is to embed these functions into implanted electronics, such as photodiode arrays, wherein local processing is used to convert incoming information into a stimulation protocol.

Increasingly sophisticated vision processing techniques that are designed to maximize the throughput of information from the real world to the user are being employed. These vision processing techniques are capable of targeting specific functional outcomes. For example, vision processing schemes can include intermediate filtering processes that seek to accentuate task-relevant features in the incoming data stream,[76] more advanced scene augmentation algorithms that alter or entirely replace sampled values with alternative encodings of scene structure,[76–78] or symbolic representations.[79] Driving these algorithms are fast-paced improvements in sensing technologies, computing capacity, and power resources, all of which are external components that can be swapped or upgraded with relative ease. These continuing advances in vision processing technologies offer exciting possibilities for the users of visual prosthetic systems. However, implementation of these increasingly complex image processing methods also presents new challenges for the open and informative reporting of relevant system details when publishing clinical results.

This section proposes a set of guidelines for reporting vision processing in scientific or clinical publications. To be effective and unambiguous, any set of proposed guidelines must be as specific as possible; however, to be meaningful and relevant, such guidelines must also encompass the inherent capabilities of the wide range of software and hardware options that are, and will be, available in the various visual prosthetic systems. To be practical, these guidelines also must respect the commercial sensitivities that surround aspects of many current and future vision-processing systems. The following guidelines aim to strike a balance among these conceptual boundaries.

The functions of a vision processing component exist as a conceptual layer within a visual restoration system. A visual processing component can include software, hardware, or both. Herein, a visual processing component is defined as a layer that includes everything between input sensors (e.g., camera, inertial sensors) at the front end and stimulation by the neural/sensory interface at the back end.

For the purpose of the proposed guidelines, the visual processing component is divided into four components:

1. Input streams and data capture
2. Digital processing and augmentation
3. Image representation
4. Human interaction

Each component is detailed below.

Visual Processing Components

Input Streams and Data Capture. This stage encapsulates all aspects of the system relating to the capture of input data from which the final display is determined—that is, all sensors that gather data about the local environment and/or the movement or location of the device wearer in the scene. In most cases, the embodiment consists of a body-worn camera; however, additional sensors providing range information, thermal information, acceleration, gravity, and eye tracking, among other factors, may also be included. Included also in this stage are implanted sensor arrays (e.g., photodiode arrays) that capture light entering the eye and process visual information in situ. This stage also encapsulates all key parameter settings that determine the characteristics of input capture, such as the temporal sampling frequency, spatial resolution of information provided (or information throughput per unit time), and field of view captured.

Digital Processing and Augmentation. The processing and augmentation layer relates to how input streams are filtered, merged, combined, transformed, or selected in preparation for the resulting display of information that is delivered to device users. These functions can include the direct sampling of input values (e.g., grayscale intensities) or the use of algorithms that can decode, highlight, select, or infer features about the scenes or environments that are then provided in a modified form to the user.[76] Processing may also incorporate the output of previous processing cycles such as through temporal filtering or as additional inputs from previous processing cycles. Processing may occur in hardware and/or software and may be initiated, altered, or stopped with or without human intervention. The output of the processing stage may also be influenced by a range of parameter settings, mode selections, and environmental conditions that may be set prior to or during operation.

Image Representation. The representation component encompasses the methods used to reconfigure and encode information from the visual scene into the display or into the stimulus parameters that activate neural visual pathways. For example, standard vision processing systems employed with visual prosthetic systems typically encode the grayscale scene luminance as levels of perceived intensity, or "brightness," in evoked phosphenes. However, numerous other representations of the visual scene are possible and should be understood to be relevant under this conceptual heading of "Image Representation."

Human Interaction. The function of all previously described components may be changeable over time. Such changes may be actuated by automatic adjustments in response to measured or inferred changes in the environment (e.g., lighting, structure), or through manual intervention (from investigator, clinician, or user) during operation. The dynamics that are relevant to these proposed guidelines are those that impact the resulting appearance and/or meaning of the display that is presented to the user. Examples of these interactions include, but are not limited to, camera-capture settings (e.g., brightness, contrast, exposure time), image filter/algorithm choices, contrast inversion, modality of operation, zoom control, on-demand processing, and image sampling locations.

Reporting Guidelines for Vision Processing

The following section provides detailed guidelines for reporting on the design and function of each of the visual processing components. To assist in the practical application of these guidelines, the recommendations have been divided into those that are considered to be essential for the reporting of clinical outcomes and those that are considered to be desirable. These details either should be described with sufficient detail so that someone skilled in the art can understand their basic design or should reference prior publications.

Input Streams and Capture. Reporting on sensor and input capture information should include statements on the following:

1. A list of all sensors and input stream modalities that contribute to the appearance of the display or to the delivered electrical stimulation, including red–green–blue (RGB)/grayscale images, contrast, and motion/inertial measurements
2. The spatial and temporal resolution of the input capture device when it is a possible limiting factor on the device display; specifically, this reporting should specify:
  a. An input temporal resolution that is less than or within a factor of 3 of the device refresh rate
  b. An input spatial sampling resolution that is less than or within a factor of 3 of the device display spatial resolution
3. The size of the sampling window, stated as a ratio of the input to display (i.e., the physical display area) viewing angle when the input stream provides spatial coverage of the visualized scene. For example, a ratio of 1.0 would indicate that the input sampling window matches the field of view (in degrees of visual angle) occupied by the

physical display. A ratio of 2.0 would indicate that the capture window is twice the visual angle of the display device. In instances in which the physical dimensions of the display device are not easily determined (e.g., sensory substitution), then the field of view used for input capture should be explicitly stated as degrees of visual angle along the horizontal and vertical planes.

4. The focal length and depth of focus of the lens through which light passes to the visual capture device, as well as any adjustability of these properties and any other properties of the lens that may significantly impact perception.

5. A statement indicating the physical position and orientation of the sensor used to acquire the input stream with respect to the viewer's forward-facing head, such as "The camera was positioned centrally on the forehead, just above the eyes of the patient" or "The camera was head-mounted and forward-facing, with a downward tilt of approximately 30 degrees."

When practical, the following additional information should be reported:

1. The specific model and manufacturer of the device used to capture the input stream. If the device is custom made or if it is a modified version of an existing device, then a statement should be included to explain how the device acquires the input stream.

2. A listing of relevant parameter settings for the input capture device; for example, a standard camera typically provides adjustable settings for brightness gain, contrast gain, and exposure time, among others. If these parameters are automatically adjusted by the device during operation, then the range of functionality should be stated. If, on the other hand, these settings are manually set prior to operation, then a statement regarding how each setting is determined should be included.

3. Further detail on the throughput of the input stream, including the following:

   a. Dimensions of each frame of captured data
   b. Dynamic range of the input data
   c. Temporal sampling frequency (e.g., frames per second) during normal operation of the full system
   d. Relevant spatial (and, if applicable, temporal) window from which the input stream is sampled—for example, the field of view of a camera (expressed as horizontal and vertical angles) and/or the operating range of a depth sensor (as minimum and maximum distance from the sensor).

Processing and Augmentation. Reporting of processing and augmentation components should include statements on the following:

1. *The primary objective of processing and/or augmentations performed on each input stream.* This objective should be expressed with specific reference to the perceptual and/or functional outcomes being evaluated; for example: "A contrast enhancement filter was applied on sampled images in order to increase the prominence of intensity edges in the final display" or "An obstacle detection algorithm was applied to identify and enhance the visibility of potential obstructions on the ground plane."

2. *How each input stream contributes to the resulting stimulation pattern and the nature of how this is done*; for example: "Corresponding range estimates from the depth sensor were used to filter objects in the distance" or "Inertial measurements were used to stabilize the image capture prior to sampling" or "Sampling locations in the image were moved in accordance with eye-gaze tracking."

3. *What (if any) environmental and/or operational limitations or constraints are imposed by the processing algorithms applied*; for example: "The object detection algorithm assumes the background is black" or "Due to the processing time between display updates, stimuli motion was kept below 3 degrees per second."

When practical to do so, details on the underlying algorithm design should be included, or appropriate reference made to literature that describes the algorithm and/or prior evaluation of its use for prosthetic vision or other relevant applications.

Representation. Reporting on representation should include statements on the following:

1. *The information encoded in the parameters of individual display elements*; for example: "Sampled image intensity values are mapped to a linear scale of perceived intensity levels in the final display."

2. *The information, if any, conveyed via the spatial arrangement of display elements.* This includes the spatial coverage of the physical space (i.e., the field of view) represented in the display and/or any specific groupings of display elements used to collectively convey a property of the scene; for example: "Each detected letter was mapped to a corresponding pattern of phosphene activation."

3. *The information, if any, encoded in temporal patterns of display elements, or the overall display pattern*; for example: "Phosphene stimuli associated with the

detected object of interest were rapidly oscillated between on and off to cue the presence of the object to the participant."

4. *When known, the number of discrete levels of perceivable difference assumed to be available in the display to encode information.* This may be stated as an approximate upper limit.

5. *The update/refresh rate of the representation on the display.*

System Calibration, Interaction, and Dynamics.

1. *A clear description of any procedures that are used to calibrate any aspect of the vision processing system that can affect the patient's perception for a given task should be provided*; for example, "Camera alignment was adjusted to match the perceived location of phosphenes in the visual field. This was performed using a high contrast marker in random locations, which the patient was asked to locate and touch using a single finger."

2. *Reporting should address the following details about human interactions with the vision processing system during operation*:
   a. Any manually adjustable controls or modes of operation that were available during trials:
      i. To the device wearer
      ii. To the experimenter/clinician/engineer
   b. The primary purpose of each adjustable setting:
      i. Any automatically adjusted controls or modes of operation that were active during operation, and the primary purpose of the automatic setting adjustment
      ii. The primary determinant of the setting adjustment or mode selection; for example: "Patients had the option to invert the contrast of the captured image when in bright outdoor settings."
   c. If practical to do so, the following additional information should be provided:
      i. A list of all vision processing devices (e.g., cameras) and the processing and display settings that required calibration prior to operation, as well as for what purpose; for example: "After camera fitting, sampling locations in the image were calibrated to align with the patient's subjective reporting of evoked percepts in the visual field" or "The image contrast gain was manually adjusted to ensure that the reference object was discernible from the background."

Recommendations for Assessing Vision Processing Methods

Vision Processing Assessment and Human Trials. The evaluation of vision-processing methods generally requires human trials over multiple phases of development. Issues concerning assessment, research design, data collection, and analysis should reflect the recommendations set out in the relevant sections of these guidelines. The use of computer-based and/or simulated prosthetic vision using blindfolded, normal-sighted participants is a commonly applied methodology for evaluating vision processing strategies during development. Simulations of prosthetic vision in human trials are recommended to evaluate and ascertain potential functional benefits of proposed vision processing methods prior to testing with the clinical population. When reporting outcomes of a simulation or computer-based study of vision processing methods, it is recommended that this information be described in a manner consistent with the above guidelines. A full description of the simulation model and the perceptual experience it seeks to emulate should also be provided.

Benchmark Vision Processing. Benchmarking refers to the need to measure and assess the basic vision function provided by vision restoration devices, relative to the theoretical maximum visual performance permitted by the device design. Although strategies of vision processing offer significant potential benefits to enhance functional outcomes, basic measures of visual function (e.g., basic light localization, visual acuity, contrast acuity, motion detection) without signal augmentations or enhancements should also be reported.

Benchmark Vision Processing for Lab-Based, Controlled-Lighting Conditions.

Input streams and capture:

- A single, head-mounted sensor should provide the device's primary modality of input capture, at a capture rate equal to or greater than the refresh rate of the final display.
- The sensor should be rigidly mounted with respect to the head.
- Ideally, the input field of view should match the theoretical field of view of the output display device. Where impractical to do so, the difference in the visual extent of capture and display windows should be reported as a ratio, as per the reporting guidelines.
- Input capture settings should be calibrated for each individual viewer under the controlled conditions of testing. If it is not possible to disable auto-adjustment settings (e.g., the camera's software does not provide the ability to manually set or turn off auto-adjustment), then all reasonable efforts to

maintain the constancy of these settings should be made and reported.

Processing:

- For the purpose of a benchmark vision processing system, processing applied to the captured image stream should serve only to transfer sensor captured data to stimulation. That is, no deliberate augmentation or enhancement of the signal should be performed beyond that which seeks to remove noise and/or faithfully approximate/reconstruct the captured input stream (e.g., regional averaging, Gaussian filtering, Lanczos2 filtering). Where spatial filtering is applied, it is recommended that the filtering window size be set to ensure that the frequency cutoff of the filter is as close to the Nyquist band limit as possible. This size will depend on the filter chosen and the distance between display elements.

Representation:

- The final display should present to the user the system's primary modality of visual representation; for example, if the system employs a conventional camera, then the visual representation would likely encode scene luminance information captured in the grayscale input. When appropriate, it is recommended that viewer-specific scaling be performed on intensity levels to ensure maximal contrast discrimination in the evoked perception.

Benchmark Vision Processing Under Less Controlled Conditions. The assessment of orientation and mobility and other tasks of daily living is necessarily performed under less controlled conditions than is typical for lab-based vision function testing. Thus, in these cases, benchmark vision processing can reasonably include the automatic management of image-capture settings by the camera. Other settings and configurations should follow the guidelines as above.

Masking System-Off. The evaluation of vision function with prosthetic vision devices requires adequate control conditions to enable a thorough comparison of potential benefit with and without the prosthetic system. Traditionally this has been achieved through the use of a "system-off" condition, in which trials are conducted with the vision restoration device switched off. However, the use of system-off as a sole control condition is problematic for masking experimental conditions from study participants, leading to

recent clinical[80–82] and simulation[83,84] studies that have proposed the use of synthetically generated visual representations as an alternative control condition. Common variants include so-called scrambling or shuffling, in which the mapping of image sample points to stimulation locations in the visual field are randomly redistributed, either at the beginning of a trial or at regular intervals during a trial run (e.g., every 5 seconds). However, scrambling does not provide a system-off equivalent, because the captured video input still provides some information about the visual world, despite individual stimulation locations being scrambled. Random stimulation generation has also been employed. Although it provides no meaningful information to patients, it has been generally found to be less effective for masking; the resulting perception is in obvious conflict with patient expectations, so patients sometimes quickly realize that random stimulation is a control condition.

In addition to the conventional system-off condition, the assessment of basic visual function with vision restoration devices should include a synthetically generated stimulation pattern as a masked system-off condition. The recommended properties for an appropriate synthetically generated masked system-off condition are described below.

A synthetically generated masked system-off condition should

1. Have no meaningful connection with the within-device spatiotemporal properties of the imaged scene that would have been utilized under normal operating conditions.
2. Provide a net stimulation per display update that represents a plausible approximation to the expected conditions of the task.
3. Approximate the expected extent of temporal variation of stimulation for the task.

The implementation of the above properties will vary with the needs and flexibility of different devices and with the nature of the task; for example, synthetic stimulation patterns may be generated from prerecorded/synthetic video sequences and/or pre-prepared stimulation patterns or random pattern generation can be used under constraints that meet the above criteria. The method employed should be detailed in full, with explicit statements addressing how the synthetic stimulation meets the above criteria for the task.

Despite the importance of masking the system-off condition, these guidelines recommend only the sparing use of synthetically generated system-off masking with patients. Long exposure of patients

Figure 9. Examples of activity of daily living tasks that could be used in vision restoration trial, from the IADL-VLV[98]; (*left*) sorting socks and (*right*) kitchen object identification.

to the system-off condition reduces its effectiveness as a masking agent, but, more importantly, it also may reduce confidence and impair learning with the device.

Activities of Daily Living

Gary Rubin (chair)[1], Mary Lou Jackson[2], Shane McSweeney[3], Cynthia Owsley[4], Robert Finger[5], Jill Keeffe[6], Sharon Bentley[7], Gislin Dagnelie[8], and Joan Stelmack[9]

[1]**University College London Institute of Ophthalmology, London, UK (e-mail: g.rubin@ucl.ac.uk)**

[2]**Department of Ophthalmology & Vision Sciences, University of British Columbia, Vancouver, Canada**

[3]**Diversified Rehab Pty Ltd, Melbourne, Australia**

[4]**Department of Ophthalmology, The University of Alabama at Birmingham, Birmingham, AL, USA**

[5]**Department of Ophthalmology, University of Bonn, Bonn, Germany**

[6]**LV Prasad Eye Institute, Hyderabad, India**

[7]**Australian College of Optometry and Department of Optometry and Vision Sciences, University of Melbourne, Melbourne, Australia**

[8]**Lions Vision Research and Rehabilitation Center, Johns Hopkins Wilmer Eye Institute, Baltimore, MD, USA.**

[9]**Department of Ophthalmology, University of Illinois College of Medicine, Chicago, IL, USA**

Introduction

The term *activities of daily living* (ADLs) refers to a series of self-care tasks that are essential for maintaining independence. The most widely used instrument for measuring ADLs is the Katz Index of Independence in Activities of Daily Living (Katz ADL).[85] The Katz ADL includes only six activities: bathing, dressing, toileting, transferring, continence, and feeding. Each item is scored 1 if the patient can do the task independently and 0 if dependent on others or needs help to complete the task. Originally developed in the 1960s for studies of aging populations, the Katz ADL has been adapted to a wide range of patient groups. The Lawton Instrumental Activities of Daily Living (IADL) Scale[86] extends the assessment to include eight more demanding activities such as shopping, using the telephone, and managing personal finances.

The Katz ADL and the Lawton IADL scales fall somewhere between traditional patient reported outcomes (PROs), otherwise known as questionnaires, and performance-based tests. The usual method of administration of the Katz and Lawton scales is to ask the participant if and at what level they can do the task without assistance, and the participant gets a score of 1 for each task that can be completed to an adequate standard, without assistance. The laundry item on the Lawton IADL, for example, gives 1 point if the respondent does personal laundry completely or launders small items independently, but no points if all laundry must be done by others.

There are other ADL/IADL instruments that are based entirely on task performance, such as the Functional Independence Measure (FIM) and Functional Assessment Measure (FAM), which cover

a wide range of daily activities ranging from self-care (e.g., eating, dressing) to cognitively demanding tasks (e.g., reading, problem solving).[87] Because PROs are being evaluated by a different group within this guidance document, the discussion in this section will be limited to performance-based tests (PBTs). For a comprehensive review of performance-based, vision-related ADL instruments, see Warrian et al.[88]

When designing a new ADL instrument, the first question that needs to be answered is what tasks or activities should be included. Many ADLs, including feeding, spoken communication, memory, and problem solving, can be performed with little if any visual input. On the other hand, some tasks included in the list of ADLs and IADL are highly dependent on vision. Thus, it can be useful to include both vision-dependent and vision-independent tasks to provide a more complete assessment of a participant's level of adjustment to vision loss.

Notably, very little work has been done to adapt ADL measures for testing patients with ultra-low vision. The typical approach to enabling testing of more severely blind patients is to simplify the visual requirements of the task by increasing the size and contrast of visual stimuli; for example, the patient may be asked to walk along a bold white stripe on a black floor or to find a black door positioned within a white wall.[89] Although it is understandable that the examiner wants to conduct a test that a patient can accomplish, these modified versions of the test may no longer represent a realistic assessment of an ADL. Specifically, when navigating in the real world, there are seldom high-contrast lines to follow, and most doors present lower contrast than a black-on-white paradigm. If the goal is to measure some aspect of visual function such as contrast sensitivity, for example, then specific tests that were designed for that purpose should be used. But, an outcome on that type of specialized test should not be used to make an assertion about the ability of a patient with ULV to perform tasks of daily living required for independent living.

Just as can be done for PROs, it is possible to design a unidimensional set of ADL tests for people with ULV using Rasch analysis. When they have been calibrated, these items can be used as ability scores for subjects performing the tasks or a suitable subset of the tasks. A new set of such tasks for use in the functional assessment of patients with ULV is under development by Geruschat and Dagnelie.

When developing a test to measure performance on ADLs, there is a tension between the need to standardize testing conditions and the intention to preserve ecological validity to ensure that the test is truly representative of everyday activities. Reading tests present a good example. There are several continuous-text reading tests, including MNRead,[90] Radner Reading Test,[91] Colenbrander Continuous Text Reading Cards, Salisbury Eye Evaluation (SEE) Project reading test,[92] and the International Reading Speed Texts (IReST).[93] These tests fall along a continuum from the highly standardized and carefully controlled Radner Reading Test, with all sentences having the same number of words, word length, and semantic and syntactic complexity, to the SEE reading test, which uses paragraphs selected for grade level but which does not control for other linguistic features. Despite this lack of standardization (or perhaps because of it), the SEE reading test was found to be highly predictive of everyday reading performance under natural conditions in the home.[94]

The guidelines below outline recommended methodologies for testing of ADLs and for reporting results in ULV subjects who participate in clinical therapeutic trials.

Recommended Methodology for Assessment of ADLs

An evaluation of ADLs should assess the function of simulated everyday visual tasks under standardized conditions.

General recommendations for testing include the following:

1. Adherence to a written protocol that specifies the test apparatus and conditions (e.g., light levels, viewing distance), procedures, instructions to the participant, and scoring criteria.
2. Standardization of ambient illumination. Results should include measurements of room lighting, luminance of the test stimuli, and contrast of the test stimuli compared to the background.
3. Examination of the patient by an examiner who is qualified in the assessment of ADLs. Ophthalmologists, optometrists, orthoptists, certified ophthalmic technicians or associates, disability and rehabilitation researchers, and trained research assistants are all suitable for this role.
4. Assessment of all tasks with and without input of the prosthetic devices and in a scrambled device-on condition, if possible. (See the above section on Vision Processing Systems). Preferably, the assessments should be performed more than once; however, single trials are acceptable when administering multiple tests from an ADL battery that had previously been calibrated with, for example, a Rasch analysis. If multiple trials are performed for each test, the tests should be undertaken in a

counterbalanced (e.g., A–B–B–A) or random order to control for learning and fatigue effects.

5. Use of criterion-free test procedures, such as forced-choice testing. This is important to minimize biases due to differences among participants in their willingness to guess.[95]

6. Binocular testing. ADL evaluation aims to quantify the participant's ability to perform everyday tasks, and these tasks are normally performed using both eyes. Artificially restricting vision to one eye may be suitable for some tests of visual function, such as visual acuity, but this is not suitable for assessment of everyday tasks. The ADL assessment should be performed by allowing patients to use visual aids or devices that they customarily use, although it is informative to repeat testing without the aids. However, separate calibration with and without vision aids would be required, as calibration of the test activities typically is performed without image enhancement.

7. Task performance must be recorded in terms of both speed and accuracy. Some participants may impatiently hurry through a task while committing numerous mistakes, whereas others may perform the task more deliberately and thus more carefully. When possible, the scoring strategy should take into account both speed and accuracy; for example, reading performance is typically quantified by measuring both the time required to complete a reading task and the number of words that were read incorrectly. The interplay between speed and accuracy can be reflected by reporting the number of words that were read correctly per minute. Notably, it has been argued that the frequency of errors is equally or more predictive than speed for assessing performance on some ADLs, such as independent navigation.[96] As an alternative to recording both time and accuracy, subjects may be instructed to be as accurate as possible or as quick as possible; however, this reduces the ecological validity of the test, as subjects may not normally operate under such a constraint.

8. Performance should be measured on a continuous scale where possible; for example, reading speed is preferred over a pass/fail score indicating whether or not the participant read the sentence correctly. If the scale must be quantized, the quantization steps should be made a small as possible; for example, if task performance is rated on a numerical scale, then the rating scale should allow fractional or decimalized responses. It has been shown that the smaller the step size, the greater the reliability.[15]

Specific ULV ADL Test Methodologies

This section includes examples of the performance-based ADL test batteries that have specifically developed for people with ULV.

ADL Test Batteries. FLORA (Second Sight[97]): The Functional Low-Vision Observer Rated Assessment (FLORA) was developed by Second Sight to evaluate the Argus II retinal prosthesis. The FLORA includes both self-reported difficulty and observed performance for a set of ADLs. In addition, the FLORA collects a narrative case summary from expert observers. So far, the FLORA has been used only by Second Sight for relatively small groups of participants with ULV (e.g., 26 participants in the above-cited study).

IADL-VLV (Finger et al.[98]): Finger and colleagues used Delphi survey techniques to select 25 tasks from an initial set of 296 in the Bionic Vision Australia retinal prosthesis project. The tasks were performed by 40 participants with very low vision (VLV) and scored for speed and accuracy. Rasch and principal components analyses were used to evaluate the measurement properties of the tasks. A final set of 23 tasks were deemed to have adequate measurement properties.

Tests of Specific ADLs. Picture recognition (Rubin et al.[99]): Rubin and colleagues developed a picture recognition test for Pixium Vision's retinal prosthesis project. One hundred photos were taken across five different categories, such as doorways, stairs, and footpath obstacles. Each picture displayed an item from the category (e.g., a doorway) on the left or right side. Thirty normally sighted observers viewed the pictures through a head-mounted display that simulated the phosphene structure of a retinal prosthesis and indicated whether the object appeared on the left or right. Rasch analysis was used to calibrate the difficulty of each picture and to select pictures arrayed along an underlying unidimensional difficulty scale.

Visually guided navigation (Bainbridge et al.[100,101]): The navigation task was developed for a gene therapy study involving patients with Leber's congenital amaurosis. The test takes place on a 75-m^2 raised platform. Participants walk along a straight, unobstructed 8-m path to gauge their preferred walking speed and then negotiate a 13-m, eight-segment maze followed by another 8-m straight path with two foam obstacles representing curbstones. The walk is repeated, with different maze configurations, at a series of calibrated light levels ranging from daylight (240 lux) to nighttime residential street lighting (2.5 lux). Speed and accuracy are recorded by a trained observer who also protects the participant from injury.

Dining table scenery (Wilke et al.[10]): This table-top search task was developed for the Retina Implant AG subretinal prosthesis project and requires subjects to count, identify, and locate common dining utensils on a table (e.g., plate, cup, fork). The objects are high contrast (i.e., white on a black background).

Reporting Guidelines. Any publication or presentation reporting the results of ADL assessments should include enough information to allow the test to be replicated, including the following:

- Name of the test
- Brief description of the task and how it is related to daily activities
- Description of the visual stimuli, including their size, color, luminance, contrast, and motion characteristics, if any
- Room lighting measured in lux at the participant's eye
- Viewing conditions (e.g., seated/standing, distance to target, monocular/binocular)
- The maximum time allowed to complete the task and how time was measured
- Description of scoring procedure and how errors were defined, if relevant
- Randomization number and structure of trial sequence
- Instructions, practice, and feedback
- Scoring criteria and algorithms

If a detailed description of the assessment is available in the peer-reviewed literature, a reference to that publication and an abbreviated summary of the test and conditions may suffice.

Orientation and Mobility Assessments

Duane Geruschat[1] (chair), Sharon Bentley[2], Marshall Flax[3], Richard Long[4], James Weiland[5], and Russell Woods[6]

[1] **Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA (e-mail: dgeruschat@jhmi.edu)**

[2] **Queensland University of Technology, School of Optometry and Vision Science, Queensland, Australia**

[3] **McPherson Eye Research Institute, University of Wisconsin-Madison, Madison, WI, USA**

[4] **Western Michigan University, Kalamazoo, MI, USA**

[5] **University of Michigan, Ann Arbor, MI, USA**

[6] **Schepens Eye Research Institute and Harvard Medical School, Boston, MA, USA**

Introduction

Assessment of orientation and mobility (O&M) performance is essential to evaluate the impact of visual loss and any form of visual rehabilitation, including visual substitution/restoration approaches. *Orientation* in this context refers to an individual's ability to establish and maintain self-to-object and object-to-object spatial relationships (i.e., distances and directions to perceived or remembered places). The ability to orient includes the ability to move efficiently along routes and to learn the spatial layout of new places.[102] *Mobility* in this context refers to one's ability to effectively preview the path ahead and to navigate safely and efficiently through that path. Effective mobility requires detection and avoidance of obstacles, changes in elevation (e.g., curbs, stairs) and other environmental features that may be present along a path.[102] The orientation and motor skills required for mobility involve extensive interplay between visual and cognitive demands as visually impaired individuals move about in their homes and communities.[103] O&M is essential to safe and efficient *wayfinding*, which is the purposeful and directed movements necessary to reach a predetermined destination.[104,105] O&M is greater than the sum of the parts; it is "an integrated set of behaviors occurring in complex and changeable travel environments."[106]

The intent of this section on O&M is to focus on research as it pertains to visual restoration, substitution, and rehabilitation technologies and interventions (e.g., retinal prostheses, gene-based therapies). The goal is to recommend how interventions will be evaluated for individuals with ULV, defined as vision impairment that impacts most daily living activities involving visual shape recognition. ULV has also been described as "very limited vision, but not total blindness," likely between 20/1600 (6/360) to light perception[107] (refer to section titled Definitions of Terms). For people with ULV, residual vision is predominantly used for orientation. This is because ULV is not sufficient for the detection of most obstacles and certainly not for the detection and navigation of changes in elevation (stairs, curbs). Functionally, for people with ULV, mobility needs are usually addressed through rehabilitation aids, in particular the long cane and guide dog. Traditional scales of visual performance (such as visual acuity) are not adequate to describe O&M with ULV. However, there are no established validated measures of orien-

translational vision science & technology

tation or mobility, or standardized environments in which they should be measured, which is a serious limitation in this field. Here, we propose laboratory measures that have been used in research to date; however, it must be noted that laboratory measures are limited in that they do not address the needs and uses of vision in O&M in the real world. The ultimate goal of any intervention is to demonstrate a benefit in the natural environment for activities important to the person with ULV. We acknowledge that this is difficult to assess, due to the lack of experimenter control; nevertheless, in the future, we expect to see valuable advances in O&M assessments in real-world and patient-relevant settings and the increased use of validated qualitative methods in assessment of O&M.

Background

Broadly, there are two approaches to O&M research: (1) clinical or laboratory trials and (2) functional or real-world assessment. Clinical or laboratory O&M studies are usually conducted under conditions that are controlled and repeatable. Clinical or laboratory O&M researchers often select a specific travel environment (e.g., indoor vs. outdoor) to limit variation in performance due to environmental factors and then choose and measure highly specific variables of interest to test their hypotheses. Repeated trials are often used to increase statistical power. A controlled environment adds to scientific rigor but often limits the assessment to smaller, indoor areas, which do not capture the variability encountered during real-world wayfinding. Thus, these tests can be replicated in different locations and can answer a specific question, such as whether or not the user can find a door-shaped object that is placed on a wall 4 meters away) (Fig. 10), but may not be the best predictors of success in wayfinding under more realistic conditions. Currently, no validated measures of orientation or mobility exist; however, assessment procedures have been described in the literature that form a useful basis from which to begin. These include the FLORA tool, which provides a model for functional vision and O&M assessment in the context of vision restoration research.[97] The main limitation currently is the lack of standardization. Orientation ability of individuals with ULV has been successfully measured with tasks, including locating the source of a bright light, walking along a line,[88,96] determining the direction of movement of a person, and locating a door.[89] These four examples are isolated tasks that have been found to be useful orientation metrics in laboratory settings.

For the measurement of mobility, researchers have evaluated walking speed or percentage of preferred walking speed and tallied the frequency and type of obstacle contacts.[107,110–112] This is usually done using complex experimenter-developed obstacle courses[113,114] or in outdoor real-world settings[115] before and after intervention. In addition, there have been some mobility courses developed for the evaluation of interventions that aim to increase the ability to see in low light conditions (e.g., RPE65-LCA gene therapy). These approaches use highly constrained physical spaces with stimuli that are barely recognizable as objects relevant to mobility and control of lighting that allows titration of the ability to perform the task at specific levels of illumination. These are not true measures of mobility; instead, the purpose of such tests is to measure changes in sensitivity in low illumination conditions with a performance task.

Virtual reality and augmented reality formats so far have had limited use in the assessment of mobility with ULV but have much potential and are worthy of further development and use for O&M assessments, as they can present a large number of visual environments, can be titrated in complexity, and are easy to replicate across multiple testing sites.

Qualitative functional or real-world O&M assessments are completed by O&M professionals who are experienced and skilled in the travel skills of individuals with low vision or blindness, with support from relevant professionals who have knowledge of vision restoration, vision substitution, or vision rehabilitation strategies and experience with the medical care of individuals with low vision or blindness.[97] Real-world assessments typically use checklists and observer ratings. Safety must be considered, and individuals should be familiarized with the tasks prior to beginning the assessment. An approach to provide quantitative data in such settings is to employ checklists and rating scales often broken down into components.

Whether to use or exclude the primary mobility aid (e.g., long cane, dog guide) that is normally used by an individual during an assessment is a critical decision. If subjects are required to walk without their preferred mobility aids, then the assessment does not truly measure their functional ability. Excluding preferred mobility aids may be appropriate for some clinical or laboratory studies, but it should be acknowledged that this is not an assessment of functional mobility.[109] To assess functional mobility, individuals should be free to use their primary mobility aid.

Preliminary Recommendations

Orientation measurements of individuals who have ULV should be made before and after the

translational vision science & technology

Figure 10. Example of the "find the door" task, used by research groups including Second Sight (USA) and Bionic Vision Technologies (Australia) as an orientation task. Image courtesy of the Centre for Eye Research Australia and Bionic Vision Technologies (Australia).

intervention or rehabilitation is applied. Evaluations should be sensitive to changes in performance that are likely to occur as a result of implementing the intervention or rehabilitation. Specifically, assessment items for orientation should include the following:

- Detection of the location of a standard source of light (windows, ceiling or wall lights)
- Detection of the location of a door (real or simulated) against a contrasting wall
- Detection of moving people
- Follow a contrasting line on the ground

The critical features of the environment will include contrast and illumination. Each assessment item should be tested multiple times. Tasks (e.g., detection of light, door, or moving people) with deterministic correct/incorrect outcomes should be repeated multiple times, and accuracy reporting should include binomial estimates of standard error. Accuracy or signal detection metrics are recommended as the primary measure of accuracy. Should it be of interest, the time required to accomplish the task can also be recorded and for some tasks may be a preferred metric. Evaluations should take place in laboratory settings but may also be conducted in real-world settings; decisions about venue will depend on the research question. Researchers must carefully select the settings for their assessments so that they strike the desired balance among study validity, experimental control, and other factors influencing research design and anticipated outcomes.

Reporting Guidelines

The guidelines that follow take into account the need to evaluate future inventions where more complex assessments, as yet not standardized or validated, might be required. Any publication or report of O&M assessment should contain sufficient information so that others can replicate the test procedures and determine study design.

As a guide, the following study information should be reported:

1. The study design, including whether the study is assessing clinical laboratory O&M outcomes (e.g., repeated measures of selected variables) or functional real-world O&M outcomes (e.g., assessment in a participant's home and community), and a description of the control condition or group.

2. The approach and any available settings or parameter values of the intervention device (e.g., type of image processing algorithm, any modifications to the image such as reverse contrast or electronic zooming) included in the test. The specific settings that were used for a given assessment must be specified. Generally, the outcomes should be compared to a baseline or to a standard condition (e.g., no zoom, minimal processing).

3. Each test procedure and how and when it was conducted. If previously published, the name of the tests with appropriate citations should be provided. If not previously published, a detailed description of the tests, and an indication of the measurement noise (e.g., measures of agreement or repeatability) must be offered.

4. The number and duration of testing sessions. It should be reported if sessions were conducted on different days. In any case, the time interval, days or hours, between sessions should be reported.

5. The safety procedures employed, including information provided to participants and any preparation prior to O&M testing (e.g., verbal or tactile familiarization with the task).

6. Maximum time allowed for test completion, where applicable, including any procedures adopted to account for possible interactions between time to complete and quality-of-response scores.

7. Allocation of testing clusters (e.g., experimenters, locations, sites), number of participants in each testing cluster, and expertise of each testing cluster. In laboratory studies, efforts to standardize procedures between testing clusters and measures of agreement (e.g., between experimenters or sites) should be described. In functional studies, processes that describe the tasks and venues should be included.

8. In laboratory studies, randomization of procedures and conditions undertaken to avoid the confounding of factors of interest with test order. In functional or real-world studies, evidence of sufficient assessment in multiple and varied contexts to identify patterns and anomalies in individual performance should be described.

9. In laboratory studies, the methods used to mask participants and/or experimenters to the status of the intervention at each session, if any. In functional or real-world studies, descriptions of potential observer bias and methods to mitigate bias should be provided.

10. A description of the location in which each test procedure was conducted. Depending on the test and location, the description may include physical dimensions, a map, lighting (e.g., average and variation of illuminance in a room, descriptions of the range of conditions experienced outdoors) and other environmental conditions (e.g., noise, distractions, level of pedestrian interaction, complexity of street crossings).

11. For assessments that deploy or use obstacles, the total number, as well as the location, size, contrast (relative to background), and movement (if any) of each.

12. Any visual behaviors (e.g., eccentric viewing, head scanning, eye-scanning) and other sensory behaviors (e.g., use of echolocation, reaching to touch and confirm vision) evident during assessments.

13. Use of any optical correction or device (e.g., habitual spectacle correction, telescope), including the type and power, where applicable.
14. Whether the O&M measures were conducted with or without other assistive mobility devices (e.g., long cane, dog guide, human guide) and a description of any such device.
15. A description of the O&M tasks, their components, how each task was scored, and methods and efforts to standardize data collection and analysis.
16. For focus groups and interviews, a description of the methods and questions or prompts used.
17. Any changes to the experimental protocol or conduct of the test procedures that occurred during the study, whether they were for all, some, or just for a single participant (including any special accommodations that might have been made).
18. For each test or assessment, a description of the data processing and data analysis. For quantitative data, include methods used to account for repeated measures and data clustering, if appropriate, and methods used to evaluate the impact of covariates on the analysis. For qualitative data, include methods used to code the data and determine themes.
19. The quantity of data obtained, including the number of data points (test and control, numbers of participants) acquired for each test condition. In addition, descriptive statistics (e.g., mean, median, standard deviation, quartile ranges, confidence intervals), as appropriate, for each measured test procedure should be provided.

Patient-Reported Outcomes

Gislin Dagnelie (chair)[1], Akos Kusnyerik[2], Collette Mann[3], Eberhart Zrenner[4], Eduardo Fernandez[5], Francis Lane[6], Gary Rubin[7], Katarina Stingl[4], Lotfi Merabet[8], Robert Finger[9], Takashi Fujikado[10], and J. Vernon Odom[11]

[1]**Lions Vision Research and Rehabilitation Center, Johns Hopkins Wilmer Eye Institute, Baltimore, MD, USA (e-mail: gislin@jhu.edu)**

[2]**Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland**

[3]**Monash Vision Group, Monash University, Melbourne, Australia**

[4]**Department of Ophthalmology, University of Tuebingen, Tuebingen, Germany**

[5]**The Miguel Hernandez University of Elche, Alicante, Spain**

[6]**Illinois Institute of Technology, Lewis College of Human Sciences, Chicago, IL, USA**

[7]**University College London Institute of Ophthalmology, London, UK**

[8]**Schepens Eye Research Institute and Harvard Medical School, Boston, MA, USA**

[9]**Department of Ophthalmology, University of Bonn, Bonn, Germany**

[10]**Graduate School of Medicine, Osaka University, Japan**

[11]**WVU Eye Institute and Blanchette Rockefeller Neurosciences Institute, West Virginia University, Morgantown, WV, USA**

Introduction

The term patient-reported outcomes (PROs), although originally used to encompass all effects of a clinical intervention reported by the patient, has more recently been applied almost exclusively to outcome data collected through standardized questionnaires. The questionnaires employ rating scales to assess the impact of blindness or changes in well-being, status of vision, or visual function. This review concentrates on this type of instrument, although complementary techniques will be briefly discussed, as well.

Vision-related PRO questionnaires typically explore one or both of the following aspects of vision loss: quality of life (QoL) or visual outcomes, which should be referred to as *functional vision questionnaires* (FVQs), rather than *visual function questionnaires*, as visual function is typically measured in the clinic, with physical instruments. One version of FVQ, developed under the auspices of the National Eye Institute (NEI), has become the example most familiar to both clinicians and researchers.[118] However, it is important to note that this has not been validated in the population that is eligible for implantation of vision restoration devices (ULV). Using the NEI FVQ, the assessment of QoL may be designed to explore changes related specifically to vision loss, including vision-related quality of life (VRQoL), or to health in general (HRQoL). This review concentrates on vision-related instruments, although some also address emotional or other health aspects and thus should be referred to as HRQoLs.

Like many survey instruments used throughout the social and clinical sciences, classical FVQs have three typical features[119]:

1. Respondents are asked to indicate how difficult a specific activity is, given their current level of vision; the activity may be strictly visual, or it may be more complex and only partially reliant on vision.
2. Ratings are given on a Likert scale, ranging from "very easy" to "impossible," in some cases including an additional rating such as "not applicable" for activities that the patient does not perform because of non-visual reasons. The ratings are usually coded numerically for analysis.
3. The items of the questionnaire may be grouped into subscales, representing different aspects of the construct under study. In the case of FVQs, subscales can be concrete visual domains (such as distance vision, near vision, driving, peripheral vision, and color vision) or more abstract domains (such as the ability to gather visual information, and hand–eye coordination). Items in these instruments may contribute to multiple subscales.

Subscales also can be extended to explore non-visual factors such as general health, mental health, and emotional well-being, dependency, and social function. These broader instruments can be considered to be a combined HRQoL/FVQ tool. This designation is appropriate for the NEI FVQ tool, which contains questions that explore QoL rather than visual function. Several other questionnaires, are similarly broad in scope (see below).

Regulatory bodies such as the FDA and European Medicines Agency (EMA) have recognized the importance of PROs as a useful criterion for the evaluation of the effectiveness of therapeutic interventions, including medications and medical devices. In 2009, the FDA issued a Guidance for Industry entitled "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims,"[120] which lists acceptable criteria for the assessment of outcomes collected through FVQ and HRQoL instruments. In 2006, the EMA tackled the issue of outcome measures for studies of orphan diseases that have small sample sizes.[121] Both documents emphasize that QoL data can provide supportive evidence only; in other words, they should not be used as the primary outcome measure in a feasibility study or clinical trial to assert a claim of efficacy. The FDA document recognizes that PROs can serve as the primary outcome measure in a study aimed at improving symptoms alone. The EMA document does not specifically mention PRO questionnaires.

Until approximately 2005, FVQ data were typically analyzed as though the ratings were cardinal (values) rather than ordinal (rankings). A respondent's ability score was typically obtained by summing the ratings, and comparing the sum to a commonly accepted threshold score.[122] The adoption of item response theory (IRT) and Rasch analysis, originally developed for educational testing and for psychological and physical rehabilitation research, has allowed the assignment of an item difficulty score to each FVQ item and of a person's ability score to each respondent by applying a logistic model to a sufficiently large dataset (typically responses from well over 100 respondents).[123] The underlying assumptions of IRT are that all items can be ranked along a common visual difficulty scale and that all respondents can be ranked along a common visual ability scale. Further assumptions are that more able respondents will rate any given item as being less difficult than less able respondents and that more difficult items will be ranked as being difficult than less difficult items by any respondent, regardless of ability.

It is noteworthy that the aforementioned FDA guidance document does not make reference to any of these new psychometric approaches to PRO data analysis, even though such methods were well established several years before the document was issued. Additional input from multidisciplinary groups, including this HOVER document, can make it more likely that high-quality data collection and analysis through PRO instruments become the norm in the development of novel vision restoration treatments.

It is also important to note that any FVQ used to obtain PROs in a clinical trial intended to support a US marketing application must comply with US validation requirements (available from the FDA).

In the past decade or so, several new FVQs, some published in multiple languages, have been developed for use with various population of low vision subjects. It has also been demonstrated that data collected with older questionnaires can be recalibrated using Rasch analysis.[124] These advances provide the research community with more comprehensive and more advanced tools for the study of PROs than ever before.

Recommended Methodology for Patient-Reported Outcomes

As indicated above, we recommend that PROs be collected with standardized FVQs rather than with unstructured interviews. Although good clinical practice may require the use of a patient history as a tool for diagnosis and clinical decision making, abstracting quantitative information from transcripts of a medical history is labor intensive, subjective (unless based on a verbatim transcript analyzed by high-quality content analysis software), and sometimes impossible. Herein, the focus is on FVQs because a number of excellent FVQs have been developed in

translational vision science & technology

recent years, and suitable FVQs for low vision are available to study a broad range of visual ability, for both children and adults.

The study of ULV is more nuanced. ULV is so limited that at best only crude shapes can be detected and recognized, and often the vision of individuals with ULV is limited to detection of movement, light projection, or bare light perception. Most existing FVQs are unable to capture meaningful information about differences in visual ability or about possible effects of rehabilitation for subjects with ULV because the included items address visual activities that are well beyond the capabilities of ULV subject. The NEI VFQ, for instance, cannot distinguish between a subject who can barely tell whether the room lights are on or off versus one who can discern where in the room each lamp is located.

General Recommendations for the Design of FVQs for ULV.

1. When collecting items for the design of a FVQ for an unfamiliar population, it is essential to take an inventory of visual activities relevant to members of that specific population, preferably through the use of focus groups and structured interviews with members of the affected population and the treating professionals, guided by a systematic inventory of daily activities, using the Massof Activity Inventory (AI)[125] or some similar structured approach. Specific questions should be asked to elucidate how vision is used in each activity.

2. For activities that benefit from vision, a selection should be made such that both the four major functional domains (i.e., detailed vision, visual information gathering, wayfinding, and hand–eye coordination) and visual aspects that determine what makes an activity visually manageable (e.g., lighting, contrast, size, distance, color, familiarity) are represented. Questionnaire items should be formulated to capture how vision is used by the target population across these domains and conditions.

3. A preliminary form of the questionnaire should be administered to a representative sample of the target population that is large enough to perform an initial Rasch analysis (∼50 respondents). Misfitted items (underfit >4 SDs from the mean, implying that scores from different respondents are highly inconsistent and the estimate is likely to be biased) should be eliminated or reworded. Such inconsistencies suggest that the wording was perhaps ambiguous or the visual aspects of the activity might have been equivocal.

4. Differential item functioning (DIF) should be examined for relevant subgroups in the population (e.g., age groups, gender, other demographic or relevant strata).

5. A second and possibly third administration round should be used to further calibrate the items; note that a third round is not necessary if the item fit reliability is close to 1.0 (typically, >0.96) and the variance explained by the model exceeds 60% to 70%.

6. When the production version of the new FVQ has been administered to a representative sample of respondents (perhaps 50–250), depending on the accessible population and the precision of the resulting item estimates, the item measures can be anchored and the FVQ can be used to estimate person measures for new respondents without performing a full Rasch analysis.

7. Validity of the FVQ for use in a given population requires that the items span a difficulty range corresponding to the range of visual abilities of the population being studied.

8. The FVQ must be calibrated separately for new populations with different visual characteristics.

9. The use of visual dimensions or subscales may no longer be meaningful in ULV, but this should be ascertained through principal factor analysis, determining whether more than one factor is required to account for the variance of functional vision within the population.

General Recommendations for the Administration of FVQs.

1. Both operator-administered (in person or by phone) and self-administered (distributed or online, in accessible format) versions of an instrument can be used, provided that the instructions and item presentations are as similar as possible. Both modes of administration should be given to a subset of respondents to ascertain the equivalence of the methods.

2. Instructions for the use of vision and of low vision aids should be explicit.[126] In other words, the respondent should be told (and reminded) that each question refers to the difficulty of performing an activity visually, and whether it is (or is not) acceptable to think of this activity as being performed with customary visual adaptive equipment.

3. If the FVQ has anchored item measures, then administration of a properly chosen subset of items may be acceptable.

4. Administration by proxy (e.g., when the respondent is a child or has a mental limitation) should

Table. English-Language PRO Instruments Assessing Functional Vision and/or Vision-Related Quality of Life

PRO Name	Type	Target Population	ULV Appropriate?	Psychometric Status
ALB	FVQ + QoL	Legally blind veterans	No	Not calibrated
ABS/AI	FVQ	Low to very low vision	No	Calibrated
VAQ	FVQ	Aging/low vision	No	Not calibrated
ADVS	FVQ	Cataract	No	Not calibrated
NEI VFQ 51/25	FVQ + QoL	Low vision	No	Calibrated
LVP-FVQ (II)	FVQ	Pediatric low vision	No	Calibrated
CVFQ	FVQ + QoL	Pediatric low vision	No	Not calibrated
VALVVFQ (-48/-20)	FVQ	Low to very low vision	No	Calibrated
IVI/IVI-C	FVQ + QoL	(Pediatric) low vision	No	Calibrated
CVAQC	FVQ	Pediatric low vision	No	Calibrated
FVQ_CYP	FVQ	Pediatric low vision	No	Calibrated
PedEyeQ	FVQ + QoL	Pediatric low vision	No	Calibrated
IVI-VLV	FVQ + QoL	Very low vision	Yes	Calibrated
ULV-VFQ	FVQ	Ultra-low vision	Yes	Calibrated

Note: The working group advocates use of calibrated instruments only in future clinical trials.

be limited as much as possible and results interpreted with caution. If there is doubt about the reliability of the respondent's answers, clarification should be obtained from a caregiver, but the original responses should be taken into account unless they are clearly unreliable.

Recommendations for the Creation of Adaptive FVQs and Item Banking. When a FVQ has been administered to a sufficiently large and representative sample of a target population and the standard errors on the item measures have been reduced to ∼0.2 logits or less, the item measures are sufficiently precise to consider the items anchored. This designation has several important consequences:

1. Anchored item measures can be used to create a simple spreadsheet with item score weights that allow the person measure for any new respondent to be derived immediately from his/her raw scores.
2. A FVQ can be administered in adaptive form, in which a Bayesian algorithm is used to select the next item, based on previous responses, that is most likely to be at the center of the respondent's ability range and thus provide the best information about the respondent's precise ability. This may limit administration of the FVQ to a dozen items or less and still yield an accurate person measure for the respondent.
3. The administration of the FVQ can be limited to items that are relevant for the respondent.

4. The items may qualify to be included in an item bank,[127] provided it is possible to calibrate them against items already present in the bank; this recalibration is necessary because the item measure (in logits) of an item depends on its difficulty relative to other items as well as on the population in which it has been calibrated.

Specific PRO Instruments

The Table lists currently available FVQ and VRQoL instruments, including indications as to whether they may be suitable for a ULV population and whether the psychometric properties of the instrument have been studied. Note that this overview is limited to general-purpose instruments, leaving out the several dozen other instruments that have been developed to assess the vision-related impact of specific ocular and systemic disorders.

Activity Level of the Blind (QoL + FVQ). The Activity Level of the Blind (ALB) is a questionnaire designed to measure the activity level of blind veterans with visual handicaps. It measures separate components of activity: independence and difficulty in performing various activities, loss felt in not performing the activities, and motivation to learn the activities.[128] For each of the activity items, the following components were measured: frequency, difficulty, satisfaction in performance, and motivation to learn to perform better. One hundred and sixty rehabilitated blind veterans were used to test whether the items conformed to the requirements of a Rasch scale. The questionnaire contained 70 general and 33 travel-specific items. The

questionnaire can be regarded as being useful to evaluate both what patients are specifically taught and how that training generalizes to activities not specifically taught.

Activity Breakdown Structure/Activity Inventory (FVQ).
The Activity Inventory (AI) is an adaptive visual function questionnaire that consists of 459 tasks nested under 50 goals that in turn are nested under three objectives. Each goal is probed for importance, with the response categories of "not important," "somewhat important," "moderately important," and "very important." If a goal has non-zero importance, the tasks nested under that goal are probed for difficulty with the response categories of "not difficult," "somewhat difficult," "moderately difficult," "very difficult," and "impossible." These tasks represent the visual function domains of reading, mobility, visual motor, and visual information processing. Rasch analysis was performed to obtain person ability and item difficulty measures. The calibration sample for the AI consisted of individuals with habitual binocular visual acuity ranging from 20/14 to no light perception; all types of visual disorders were included. The AI can be considered the "Cadillac" of FVQs, but it is comprehensive and too time consuming to be included in most clinical trial protocols.[129]

Visual Activities Questionnaire (FVQ). This instrument was designed to assess the extent to which an individual has problems in everyday visual tasks. The Visual Activities Questionnaire (VAQ) is especially designed for older adults, who are at a higher risk for ocular disease and visual impairment than younger adults. The VAQ was shown to have good reliability and reasonable validity given the complexity of self-report judgments about health and behavior problems, and it is relatively quick to administer as it contains only 33 items. Data indicate that older adults who report visual difficulties in response to the VAQ tend to have visual deficits measurable by visual functional tests. Therefore, the VAQ may prove to be a useful instrument in clinical and epidemiological vision research.[130]

Activities of Daily Vision Scale (FVQ). The authors identified 20 visual activities and categorized them into five subscales (distance vision, near vision, glare disability, night driving, and daytime driving) that comprise the Activities of Daily Vision Scale (ADVS). For each activity, the study subjects (334 patients scheduled for cataract extraction) selected from among five ordered categories reflecting the degree of difficulty. These categories ranged from no difficulty to so difficult that the subject no longer performed the activity for visual reasons. Each subscale in the ADVS was scored between 100 (no visual difficulty) and 0 (inability to perform the activity because of visual difficulty). The reliability and validity (including content and criterion) of each activity were assessed. The authors concluded that the ADVS was a reliable and valid measure of a patient's perceptions of visual functional impairment.[131]

National Eye Institute Visual Function Questionnaire 51/25 (QoL + FVQ). This 51-item field test version of the NEI VFQ was based on the ADVS and designed to also capture the influence of vision on multiple dimensions of HRQoL, such as emotional well-being and social functioning. The 25-item version of the NEI VFQ was constructed to maintain, in condensed form, the breadth of content in the 51-item NEI VFQ. Eligible participants had to have one of a variety of eye conditions; 859 persons contributed data for the item reduction analyses. The NEI VFQ-25 subscale scores are an average of the items in the subscale transformed to a scale from 0 to 100, where 100 represents the best possible score on the measure and 0 represents the worst. The composite NEI VFQ-25 score is an unweighted average of the responses to all items except the general health rating question. The psychometric properties of the NEIVFQ-51 and NEI VFQ-25 are similar.[118,132]

Several attempts have been made to improve the psychometric properties of the NEI VFQ-25. Pesudovs and colleagues[127,133] performed a Rasch and factor analysis for a large cohort of respondents who underwent cataract surgery and concluded that eight items aligned with a visual functioning dimension and 10 questions with a socioemotional dimension; therefore, they advocate separate analysis of these two subsets of the NEI VFQ-25. Massof et al.,[134] using a similar approach to analyze data from a general low vision population, found that 18 items in the NEI VFQ-25 aligned with a single dimension representing visual ability, and they published a method to estimate person measures in logits through an Excel spreadsheet. Even with these calibrations, the precision on the person measures is less than for PROs developed with strict adherence to psychometric principles.

LV Prasad Functional Vision Questionnaire (FVQ). The LV Prasad Functional Vision Questionnaire (LVP-FVQ) was developed using Rasch analysis to assess self-reported difficulties in performing daily tasks in school children with visual impairment (VI) in India. The second version of the LVP-FVQ (LVP-FVQ II) was developed and validated by extracting items from other similar questionnaires (albeit developed for Western populations) and focus group discussions of children

with VI and their parents that resulted in a 32-item pilot questionnaire. Overall, six items from the LVP-FVQ were retained. The questionnaire underwent pilot testing in 25 children, following which a 27-item LVP-FVQ II emerged that was administered to 150 children with VI. The response to each item was rated on a three-category scale (1, no difficulty; 2, some difficulty; and 3, a lot of difficulty). Rasch analysis was used to calibrate the LVP-FVQ II.[135,136]

Children's Visual Function Questionnaire (QoL + FVQ). Age-specific versions of a Children's Visual Function Questionnaire (CVFQ) were developed for ages < 3 years and > 3 years, with 50 and 55 items, respectively. The instrument was applied to 403 consecutive patients with a wide range of ophthalmological diagnoses. Subscales for general health, general vision, competence, personality, family impact, and treatment were defined. All responses were measured on Likert-type scales with either five or six response choices. Quality scales (e.g., excellent, very good, and so forth), frequency (e.g., never, once in a while, and so forth), agreement (e.g., strongly disagree, disagree, and so forth), and difficulty (e.g., no difficulty, a little difficulty, and so forth) were used, making a combined psychometric analysis of its properties challenging.[137]

Veterans Administration Low-Vision VFQ (FVQ). The Veterans Administration Low-Vision VFQ (VALVVFQ-48) was designed to measure the difficulty of visually impaired persons performing daily activities and to evaluate low-vision outcomes. The VALVVFQ-48 was administered by telephone interview to subjects with visual acuity ranging from near normal to total blindness at five sites in the VA system and in the private sector. The VALVVFQ-48 includes four rating categories (not difficult, slightly/moderately difficult, extremely difficult, and impossible). Rasch analysis with the Andrich rating scale model was applied to difficulty ratings from 367 subjects to evaluate measurement properties of the instrument. The VALVVFQ-48 is valid and reliable and has the range and precision necessary to measure visual ability of low-vision patients with moderate to severe vision loss across diverse clinical settings. A short form version of the VALVVFQ-48 questionnaire designed for clinical practice and outcomes research also was evaluated. Items were eliminated from the VALVVFQ-48 to reduce redundancy and to shorten the instrument. A 20-item short form of the instrument was constructed for use in low-vision service delivery.[138,139]

Impact of Vision Impairment Questionnaire (QoL + FVQ). The Rasch-scaled 28-item Impact of Vision Impairment (IVI) questionnaire demonstrates a justifiable scale for measuring perceived restriction of participation in daily activities for individuals with impaired vision. The eligibility criteria for the study included best presenting visual acuity less than 6/12, or visual field deficit. Identified domains included work and leisure, household and personal care, mobility, consumer and social interaction, and emotional reaction to vision loss. Each item is rated on a six-level scale from "no difficulty" to "can't do because of vision." The IVI questionnaire was administered by trained interviewers to 115 people with impaired vision who were asked how much their eyesight deficiency had interfered with an activity "in the past month." Responses to items were rated as "not at all," "very rarely," "a little of the time," "a fair amount of the time," "a lot of the time," or "all the time," with additional response categories of "can't do because of eyesight" or "don't do because of other reasons".[140–143]

The Impact of Vision Impairment for Children (IVI_C) questionnaire was validated as a new vision-specific pediatric instrument designed to assess the effect of impaired vision on QoL in children. The IVI_C was administered to vision-impaired and normally sighted students, 8 to 18 years of age. Reliability and validity were tested, and the data were subjected to Rasch analysis to assess the scale dimensionality, measurement characteristics, response options, and targeting. A total of 126 students with visual acuity worse than 0.3 logMAR (i.e., 20/40) and/or a restricted visual field of <60° were recruited. Unlike most adult vision-related questionnaires, and both the LVP-FVQ and the Cardiff Visual Ability Questionnaire for Children, which use negative item phrasing, most of the IVI_C items were positively framed to eliminate negative suggestions about students' circumstances. All questions had a 5-point scored response: "always," 5; "almost always," 4; "sometimes," 3; "almost never," 2; and "never," 1. The IVI_C was demonstrated to be a reliable tool across administration modes, over time, and between observers. It can also effectively discriminate between normally sighted and vision-impaired groups.

Note that the response categories used in these instruments may represent aspects of functioning other than ability or difficulty, and this is reflected in the use of non-visual subscales in the presentation of results.

Cardiff Visual Ability Questionnaire for Children (FVQ). The Cardiff Visual Ability Questionnaire for Children (CVAQC) is a short, psychometrically robust, self-

translational vision science & technology

report instrument that forms a unidimensional scale for the assessment of the visual ability in children and young people with visual impairment. All participants were between 5 and 18 years of age. The 25-item CVAQC is a valid and a reliable instrument that was developed using Rasch analysis to ensure good content validity, construct validity, and temporal stability. The item selection was based on the information provided by focus groups with children and young people, which makes this instrument highly relevant for this population, and it provides a focus on the most important activities both in and out of school.[144]

Functional Vision Questionnaire for Children and Young People (FVQ). The Functional Vision Questionnaire for Children and Young People (FVQ_CYP) aims to collect an age-appropriate measure of visual ability in the school-age population through ratings of everyday activities both in and out of school. This instrument consists of 36 items with good psychometric properties as determined by Rasch and principal component analyses, derived from a 56-item draft that was developed through interviews with school children and youth ages 10 to 17 years throughout the United Kingdom. Ratings are on a 4-point difficulty scale, with a "not applicable" option. All items map along a unidimensional ability or difficulty scale. The authors claim that it has wider geographic validity than the LVP and Cardiff instruments.[145]

Pediatric Eye Questionnaire (FVQ + QoL). The Pediatric Eye Questionnaire (PedEyeQ) is a recently developed set of questionnaires for children of different age groups (0–4, 5–11, and 12–17 years) that come in self-reported, proxy, and parent versions.[146] For the youngest age group, there is no self-reported version, and only one parent version exists; the proxy instrument therefore comes in three versions, and the self-reported instrument comes in two versions. Each instrument consists of up to 10 questions in three to five domains (functional vision, being bothered by vision loss, social impact, frustration/worry, and eye care), with concerns about these areas forming the domains in the parent questionnaire. The development of these questionnaires followed a process very similar to the recommendations we have formulated above. Rasch analysis and item reduction were used to limit all items within each domain to a single dimension and to optimize the psychometric properties of the instrument. These questionnaires strike a balance between striving for precision and limiting administration time, thus the relatively small set of items per domain. The downside of this choice is that person measures will be less precise and therefore less sensitive to change than would be the case with a larger number

of items. This is partially avoided by the item selection, which for most of the versions spans less than 3 logits, but this is likely to limit the sensitive range of the questionnaire to moderate low vision. Only a limited study including patients with severe vision loss has been performed thus far[147]; a larger study in children with low vision is under way (Birch E, personal communication).

Impact of Vision Impairment for Very Low Vision (QoL + FVQ). The Impact of Vision Impairment for Very Low Vision (IVI-VLV) is a measure of VRQoL in persons with VLV. This instrument is derived from the original IVI, based on focus group discussions and participant and expert input, and it was developed with two sets of persons with VLV using Rasch analysis, reducing the original item pool from 76 to 28 items. All items of the IVI-VLV are preceded by "How much does your eyesight …," and each uses the same rating scale with the following four response options: "not at all," "a little," "some," or "a lot." In addition, all items have a "don't do this for other reasons" option. Two subscales are used: (1) Emotional Wellbeing (EWB), which consists of 12 items; and (2) Activities of Daily Living, Mobility and Safety (ADLMS), which consists of 16 items. The IVI-VLV can differentiate between different levels of VRQoL in participants, and its measurements are unaffected by almost all levels of general or mental health. This instrument meets all requirements of the Rasch model and the proposed quality criteria for health status questionnaires, such as content validity, internal consistency, reliability, no floor or ceiling effects, and good interpretability. It should be noted that the VLV population includes individuals that have limited form vision ("count fingers"), whereas the ULV population lacks limited form vision.[148]

Ultra-Low Vision Visual Functioning Questionnaire (FVQ). The Ultra-Low Vision Visual Functioning Questionnaire (ULV-VFQ) includes 150 items that were developed from statements about vision use from 45 focus group members with current or prior (now blind) ULV, including six Argus II wearers, in response to the full Massof Activity Inventory. The items cover four functional domains (detail vision, visual information gathering, mobility, and hand–eye coordination) and visual aspects such as contrast, lighting, size/distance, movement, and familiarity. The ULV-VFQ was pilot tested in a ULV/Argus population, followed by Rasch analysis and item adjustments and retesting in the same population. The item reliability was 0.97. Versions with 150, 50,

and 23 items are available, as well as an adaptive version.[149–151]

Alternative Methods of PRO Collection

According to the current levels of outcomes, prosthetic visual restoration can be considered a form of ULV. For example, item measures obtained with the ULV-VFQ in a small sample of Argus II recipients did not differ significantly from those obtained in samples of current or previous ULV individuals,[152] and both Argus II and Brainport[153] users performed similarly to individuals with native ULV on a set of calibrated activities of daily living (ULV-ADL).[154] As new visual prostheses are introduced and as novel treatments such as gene therapy or stem-cell-based vision restoration reach clinical application, it is conceivable that the vision gains experienced by recipients of such new approaches will differ from ULV experienced currently by those with native or prosthetic vision. For that reason, and to keep an open mind about nuances in the visual experience that would not be captured by currently available standardized questionnaires, open-ended interviews and a careful clinical history remain crucial tools in the early stages of assessment and rehabilitation. Ultimately, though, the findings from such free-form information gathering should be incorporated into new standardized FVQs to ensure that such instruments retain both face and content validity and to allow calibrated assessments across treatment types, study sites, and individuals.

The Working Group on Patient-Reported Outcomes wishes to recognize the crucial contributions of patient volunteers in the development of visual prostheses and other new vision restoration technologies. The active participation and feedback of these volunteers, their descriptions of visual experiences elicited by the therapeutic intervention and subsequent rehabilitation, and their suggestions for further improvements have provided our research community with invaluable information that enhances progress in this field of study.

This working group also encourages the use of open-ended reports from patients and their caregivers but not at the exclusion of feedback elicited with calibrated and validated measures of patient-reported outcomes. The working group also emphasizes the need for internationally accepted standards of validity and calibration when assessing patient-reported outcomes across individuals and treatment modalities, as these standards are designed to meet the needs of the scientific community, study sponsors, regulatory bodies, and health insurance companies in evaluating the safety and effectiveness of each treatment.

Reporting Guidelines

Any publication or presentation reporting the results of PROs using a FVQ should include sufficient information to allow replication of the work, including the following:

- The name of the FVQ and, if applicable, the version
- If the FVQ has not been previously validated, any relevant validation procedure and population information
- If the FVQ has not been previously published or is not available to the general public, an item list and response scale, including any "not applicable" category
- Method of administration and, if not standardized, the verbal instructions provided to the respondents
- Scoring rules and algorithms
- Provided the instrument has been validated in an appropriate population, results of the Rasch analysis, especially item measures, item and person error estimates, and item in-fit statistics.

Given the availability of a number of validated and well-calibrated PRO instruments for assessment of visual ability and quality of life, the working group expresses as its strong opinion that all clinical trials seeking regulatory approval should include one or more of the calibrated instruments, with clear arguments why the selected instruments are most appropriate for the study population and the intervention.

Psychosocial Assessments and Ethical Considerations

Philip Troyk[1] (chair) and Frank Lane[2]

[1]Armour College of Engineering, Illinois Institute of Technology, Chicago, IL, USA (e-mail: **troyk@iit.edu**)
[2]Lewis College of Human Sciences, Illinois Institute of Technology, Chicago, IL, USA

Introduction

Large-scale clinical trials like those used to test new drugs are not appropriate for visual prostheses. Because of the more limited ability of bench and animal studies to establish the safety and effectiveness of new visual prosthesis designs, human testing typically begins with a limited feasibility study of no more than a few

subjects. Here, we consider how to ethically inform and select subjects for initial clinical studies of visual prostheses. Our considerations may also apply to other emerging forms of visual rehabilitation, such as genetic manipulation, stem cell therapies, and optogenetics.

Beneficence, nonmaleficence, autonomy, and justice are regarded as the basic building blocks of modern day bioethics.[155,156] Each of these principles can be distinctly applied to the design, development, and translation of visual prostheses for human use; however, the decision to develop a program structure that respects each principle can be challenging. A well-intentioned desire for nonmaleficence, or protectionism, on the part of a project leader or medical practitioner can conflict with respect for autonomy and the right of self-determination of study subjects—deciding that participation in an experimental study is too risky for a particular person can collide with that person's right to knowingly place themselves at risk. The development of an informed consent document uses ethical principles as building blocks, and the scope and interplay between these guiding factors should be carefully considered.

Visual prosthetic technology is complex, and the information necessary to conduct a proper informed consent process can be equally complex. The need to present all relevant information can distort the original intent of the informed consent process, making the document so complex that it no longer accomplishes the primary goal of education. Consent without authentic education counteracts the fundamental purpose of the informed consent process. Too often forms and protocols developed for the informed consent process appear to provide more protection to the sponsoring institution than to the volunteer.[157,158]

Considerations

Involvement of potential participants in visual prosthesis projects can be structured at various stages of the system development and deployment. Following a traditional model, the trend seems to be that technical development precedes recruitment and involvement of potential prosthetic recipients. The basis for this segmented approach is that technical feasibility should be established before involving human test subjects.

One rationale for this approach is that narrowing the technical approaches to those considered suitable to deliver a safe implantable system before involving potential recipients avoids confusing recipient perceptions about safety. However, an equally compelling rationale is that decisions made about how to shape the technology during the earliest stages of development, even before safety has been demonstrated, can substan-

tially benefit from the input of future recipients. Soliciting user viewpoints about the function and form of a developing visual prosthesis can provide unexpected and significant guidelines for the development of the native technology and can avoid unexpected disappointments in the later stages of system deployment.[158–160] For example, prospective participants who participated in a series of focus groups and individuals who received a specific device stated that the cosmetics of the device was a factor that would determine whether or not they would consent to participate in a visual prosthetic clinical trial.[160] This perspective from prospective users, which modified future considerations in design development, was not fully appreciated or anticipated by the technical development team.

Motivation to be an experimental trial volunteer can vary widely depending on the person's history, current state, and support system. Considered here are restoration of vision, altruism, and adventurism.

Restoration of Vision. Perhaps the most obvious, and potentially dangerous, motivator for experimental trial participation is the expectation of regaining visual function. To understand the nature of the need for restored vision, a series of focus groups was conducted to determine how much benefit a visual prosthesis would have to provide to a prospective recipient to motivate their willingness to become a recipient of the technology. Individuals with severe blindness reported that they regarded any restoration of vision to be beneficial; for example, even minimal light perception could enable an individual to detect stationary or moving objects and possibly improve spatial orientation and navigation. However, other feedback from the group exposed a lack of clarity or understanding about the quality of the restored vision that might be achieved. But, in reality, well-meaning predictions about the utility of emerging visual prosthesis systems can be nothing more than thoughtful estimates prior to performing experimental tests with volunteers. This fundamental and irreconcilable uncertainty exposes a potential uneasy boundary between the principles of beneficence and nonmaleficence.

Altruism. Some investigators have concluded that the strongest psychological benefit from participation in a clinical trial is the experience of altruism.[161,162] Altruism has repeatedly emerged as an important factor for potential participants in focus group studies[159,160] and in retrospective reporting by some recipients of cortical visual prostheses (Lane FJ, et al. *IOVS*. 2013;54:ARVO Abstract 5317). In these studies, altruism was often expressed by potential and actual recipients of visual prostheses in terms of wishing to advance the field of vision restoration for the potential benefit of other

blind individuals, whether in their families or not. Although such altruism can seem compelling, altruism also can be motivated by pathological factors, such as psychotic altruism, where it is the individual's delusion that is motivating the individual, or pseudoaltruism, which appears as altruism but is an underlying motivation to engage in sadomasochistic activity.[162] If the selection of a trial participant is influenced by their expression of altruism, an informed assessment by a trained, multidisciplinary medical team, including a psychiatrist or psychologist, is needed to reduce the risk of including a potential recipient who is inappropriately motivated. In addition, if altruism, coupled with the desire on the part of the participant to add to scientific knowledge, is a significant motivator, then the researchers have an implicit obligation to design and carry out a scientifically credible study that incorporates accepted elements of commonly understood scientific methodologies.

Adventurism. Adventurism is driven by the prospects for excitement and trail-blazing, and it can be a significant motivator for those considering participation in an experimental clinical trial. Altman[163] even proposed that self-experimentation by physicians or other scientific researchers can be motivated by adventurism. A composite self-image of being a pioneer, being first, exploring the unknown, or even achieving a science-fiction-like persona may attract some to experimentation. Kilgore and colleagues[164] reported this influence for recipients of spinal cord implants, and Lane et al.[158–160] found similar trends for recipients of visual prostheses. A danger of relying upon adventurism as a qualifying motivation for trial participation is that risks, most of which are known to the researchers and, it is hoped, conveyed appropriately to potential participants, can be too easily dismissed by accepting the principles of autonomy and self-determination as overriding compensations. This dynamic can produce a complex interaction between nonmaleficence and respect for autonomy on the part of the researcher. If an overwhelming sense of adventurism dominates the motivation to participate, the medical team might be lulled into reducing its attention to responsibly balance the assessment of risk, safety, and efficacy for each potential participant.

Decision Making

The decision-making process used by potential participants in an experimental clinical trial has not been extensively researched and is not well understood. Most likely, there are strong cultural and social group aspects to the decision to be a recipient of an experimental visual prosthesis. The factors used by the decision maker can include being deeply personal, having strong family or friend influences, trusting researchers or health practitioners, or clerical/religious influences. These factors can include strong pressures from others that may contradict personal desires. It is also important to recognize that an informed consent process developed in the West may reflect the cultural importance of independent decision making in that region and be less culturally in tune with the multifactorial decision processes elsewhere. In non-Western societies, for instance, family and even community leaders may be directly involved in the decision-making process for an individual.[165] When using an informed consent process to facilitate the individual's decision making, caution should be exercised regarding whether that process was primarily designed for institutional legal protection or as an authentic aid to the participant's decision to participate through an emphasis upon education.

Managing Expectations

Managing the expectations of volunteers before, during, and after the experimental trial should be an essential component of the oversight of a clinical trial. Prior to the trial, and culminating in the participant's decision process, the motivational factors, as discussed above, play a major role in shaping the participant's expectations. Assessing and weighing those motivations, at least as they are construed by the medical team, and appropriate structuring of the informed consent process to account for these motivations are the primary means of assisting the potential participant in managing their expectations. During the trial, other effects of expectations, whether previously known or not, may emerge. Disappointment with the outcomes as experienced by the participant may result if pre-trial expectations are unrealistic. Despite best efforts in structuring and implementing the informed consent process, misunderstanding of the nature or capabilities of the technology may linger or develop as the trial progresses.

Recommendations

The motivating factors, decision making by prospective participants, and the management of expectations are important considerations for any visual-restoration clinical trial; however, to focus on any one of these, at the exclusion of another, can have a negative impact on a participant. Each of the factors discussed below must be regarded as being equally important, and particular attention should be given to potential interac-

tions among them. In addition to factors specific to visual restoration trials, other factors that pertain to the screening of individuals for any clinical trial must also be considered. The following recommendations elucidate the importance of a comprehensive assessment when identifying appropriate individuals for a visual restoration trial.

Components of a comprehensive assessment include the following:

1. A trained mental health practitioner must be involved in the initial screening and decision-making process of all prospective participants in experimental clinical trials of these types. A comprehensive mental health screening must be a part of that oversight. The evaluation must include a multiparametric assessment of an individual's intellectual capacity, including that person's ability to process and incorporate complex information into their decision-making. The mental health/cognitive screening should be expansive and include an assessment of the individual's personality, emotional state, ability to adjust to disappointment or challenges, and the presence or absence of psychopathology.

2. The comprehensive mental health screening must include an assessment of factors relevant to the potential participant's adjustment to blindness and their current quality of life as a person living with blindness. The term "adjustment" should be regarded as contextual in nature so that each of the considerations described above is interpreted within the larger context of the participant's daily lifestyle, needs, ambitions, etc. The ability of an individual to adjust to unanticipated outcomes in the trial must be considered with respect to how that person adjusted to their loss of vision in the past or perhaps to other earlier adverse life events. Invasive interventions designed to restore vision inherently carry a risk of further loss of vision, and how an individual coped with vision loss in the past likely will be reflective of how they might respond to any inadvertent negative outcome in the clinical trial. Negative outcomes also can include non-visual complications, such as chronic pain, and hence a broader assessment of emotional/psychological equanimity is important.

3. A trained mental health practitioner must be involved in the assessment of all participants throughout the duration of the trial. The trained practitioner should not only assess changes in the individuals emotional functioning and adjustment to gains or losses in visual acuity but also assist the participant in the understanding of complex infor-

mation that is relevant to the trial. It is permissible for the mental health specialist to function in the capacity of an advocate to the potential participant by listening intently and being sensitive, empathic, and supportive of the challenges of the decision-making process. Although this broader role should not be forced, any level of support should be made available to participants if and when they choose to accept and utilize such services.

4. Developing a genuine understanding of the relevant technical information, the risks, and potential benefits is essential to conducting an authentic informed consent process that primarily serves the potential subject. An effective process requires more than just developing an intellectual understanding of the information, and it should include an assessment and consideration of how participation in an experimental trial might affect the participant's life and possibly change their quality of life. These factors are essential to the participant's well-being and perhaps the technical success of the trial.

5. In this regard, one model that the research team might consider incorporating is used by the National Aeronautics and Space Administration (NASA) for the selection of astronauts. It is possible that involvement of the potential participant in an early stage of the planning of a trial can significantly benefit the process of informed consent. NASA's approach is to form groups of multiple potential participants at an early stage in the selection process; for medical interventions, this stage could occur while the requisite engineering or biological tests are being performed and then packaged for FDA scrutiny as the group seeks regulatory approval for the human intervention. This group approach may provide emotional support, enhance the understanding of complex information, and facilitate the ultimate selection of subjects into the program. A mental health practitioner can provide ongoing group facilitation, which may provide a notable support structure for those who eventually participate in the trial. However, due to the constraints imposed by a study's institutional review board (IRB), having a study's potential participants play an authentic role in the formative stages of clinical trials may be difficult to implement. Despite such impediments to involving study participants at an early stage of trial planning, researchers should strive for authentic input from potential, or actual, trial participants at the earliest stages of trial planning, while remaining within the parameters established by the IRB.

Considering and exploring every factor that may be grounds for inclusion, or exclusion, from a visual restoration trial is beyond the scope of this document. As various trials progress around the world, sharing of information among groups about the emotional and mental well-being of participants will be an important guide for the future development of visual rehabilitative field at large. Herein, we have proposed use of the seemingly best information, but we acknowledge that we cannot be certain that we know all constructs that must be assessed and all instruments that should be used, much less how this body of information should best be considered in the process of selection of study participants. In this regard, the considerations and recommendations presented here should be viewed as formative, and not prescriptive.

The nature of humans is complex and human behavior cannot be predicted with accuracy. Although psychometric instruments have been developed with high validity and reliability to measure intelligence, personality, and such, the instruments are not perfect and interpretation of such instruments is subjective. Even a seemingly comprehensive and assiduous mental health assessment can fail to predict a person's response to a poor outcome. A "failure" may be especially impactful on a subject given that the intervention offers the inescapable allure of visual improvement. Nonetheless, we must strive to minimize risk and harm to those who so willingly offer themselves as subjects for clinical experimentation. In this context, the guidelines offered here are intended to increase awareness of what to include in a comprehensive mental health assessment and the challenges and nuances of such assessments. Our ability to provide a more enlightened informed consent process and to better select participants will evolve as our scientific community accrues greater experience with the outcomes of clinical trials.

Conclusions

Lauren Ayton[1] and Joseph Rizzo[2]

[1]Department of Optometry and Vision Sciences and Department of Surgery (Ophthalmology), The University of Melbourne, Parkville, Australia (e-mail: layton@unimelb.edu.au)

[2]Harvard Medical School and the Massachusetts Eye and Ear Infirmary, Boston, MA, USA

This document offers guidelines for the testing and reporting of visual outcomes and, when relevant, device function for visual restoration studies. The desired outcome of this consensus document is to improve the consistency of methods that are used to assess the efficacy of therapeutic interventions. The authors acknowledge that the wide variety of therapeutic approaches, even within the field of visual prosthetics alone, may preclude adherence to some recommendations. Thus, from a practical standpoint, testing methods should attempt to reflect the spirit, if not the specifics, of these guidelines. The recognition of the need to accept some flexibility in testing methods is reflected in our efforts not to be prescriptive. Our attempt to make broad recommendations should enable more groups to adopt our guidelines, which will encourage harmony across the various disciplines that are all dedicated to improving the quality of life for the blind.

Having said this, our international consensus group has developed the following guidelines as a minimal set of expected outcome measures in clinical trials, depending on treatment category:

We expect our guidelines to be a "living document" as we benefit from new knowledge from future testing. Our guidelines and the rules of governance for decision making will be posted on the website of the Henry Ford Department of Ophthalmology (Detroit). This site (www.artificialvision.org) was chosen in recognition of the consistent support that Phillip Hessburg and the Board of Directors of the Detroit Institute of Ophthalmology (which merged with the Henry Ford Department of Ophthalmology) have so generously and selflessly provided to the field of visual prosthetics by hosting biannual meetings, known as the "The Eye and the Chip," which have fostered collegiality and enhanced progress. This website also will post a list of all human psychophysical testing in the fields of visual prosthetics, gene therapy, optogenetics, sensory substitution, and transplantation (stem cell, neural tissue, or retinal pigment epithelium).

The authors of studies in any of these fields who choose to adhere to our guidelines are encouraged to include the following statement in their Abstract and Methods section: "This study complied with the Recommendations of the Task Force for the Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials (HOVER)." If a study did not comply with one or a couple of guidelines, by, for example, using a VFQ that had not been previously applied to an ULV population, as is true for the large majority of currently used VFQs, then this limitation should be specified as such: "This study complied with the Recommendations of the Task Force for the Harmonization of Outcomes and Vision

Outcome Measure	Recommended Tests	Vision Prostheses	Gene Therapy	Stem Cells
Visual acuity	Visual acuity testing methodologies will vary with acuity levels, as follows: For patients with a letter chart visual acuity better than logMAR 1.60 (1/40, 0.025, 6/240, 20/800), we recommend the ETDRS high-contrast acuity chart. For patients with a letter chart visual acuity worse than logMAR 1.60 (1/40, 0.025, 6/240, 20/800), simpler test tasks, such as single letter acuity and grating acuity, should be used, such as the BRVT or FrACT. If no visual acuity measurements can be made, tests of light perception, location, or projection can be used, such as BaLM or BRVT.	X	X	X
Electrophysiology	ISCEV standard electrophysiology tests (available here: https://iscev.wildapricot.org/standards). Note that the FST has not yet been documented in an ISCEV standard but is a recommended test to quantify retinal sensitivity in patients with poor fixation or ULV.	X	X	X
Electrically evoked device effectiveness	The minimum testing or device effectiveness in a vision prosthesis should include the following: Phosphene threshold—Single phosphene detection thresholds using 0.5-s pulse trains Phosphene discrimination—A measure or indication of the ability to discriminate phosphenes from one another Phosphene field—A measure or indication of the spatial extent of the visual field that is occupied by elicited phosphenes	X		
Vision processing	Reports of vision processing algorithms should include the following: A list of all sensors and input stream modalities that contribute to the appearance of the display or to delivered electrical stimulation (e.g., RGB/grayscale images, contrast, motion/inertial measurements) The spatial and temporal resolution of the input capture device when it is a possible limiting factor on the device display The size of the sampling window, stated as a ratio of the input to display (i.e., the physical display area) viewing angles when the input stream provides spatial coverage of the visualized scene The focal length and depth of focus of the lens through which light passes to the visual capture device, as well as any adjustability of these properties and any other properties of the lens that may significantly impact perception	X		
Activities of daily living	Validated activity of daily living assessments for use in vision restoration trials include the following: Functional Low-Vision Observer Rated Assessment (FLORA) Very Low Vision Instrumental Activities of Daily Living (IADL-VLV) Picture recognition and visually guided navigation tasks (Rubin) Note that the choice of which tool to use may depend on the goals of the particular technology, and a customized assessment may be required. If so, this should be developed using the principles outlined in this guidance document.	X	X	X

Endpoints in Vision Restoration Trials (HOVER) except in the following respect(s) … ." The website will distinguish studies that did or did not follow our guidelines, which should improve the consistency of methods that are used and the ability of potential patients, physicians, scientists, regulatory specialists, and the commer-

Continued

Outcome Measure	Recommended Tests	Vision Prostheses	Gene Therapy	Stem Cells
Orientation and mobility	Basic assessment of orientation should include the following: Detection of the location of a standard source of light (windows, ceiling, or wall lights) Detection of the location of a door (real or simulated) against a contrasting wall Detection of moving people Following a contrasting line on the ground	X	X	X
Patient reported outcomes	PRO selection will depend on the population being assessed and, in particular, the level of vision that they have at baseline. The PRO selected should have been validated, calibrated, and, if possible, assessed using Rasch analysis. A list of possible PROs is provided in the main document and could include the following: For low vision, National Eye Institute Visual Function Questionnaire (NEI VFQ); Impact of Vision Impairment (IVI) For ultra-low vision, Ultra-Low Vision Visual Functioning Questionnaire (ULV-VFQ); Impact of Vision Impairment for Very Low Vision (IVI-VLV)	X	X	X
Psychosocial assessments	Psychosocial assessments should include the following: A trained mental health practitioner must be involved in the initial screening and decision-making process of all prospective participants in experimental clinical trials of these types. The comprehensive mental health screening must include an assessment of factors relevant to the potential participant's adjustment to blindness and their current quality of life as a person living with blindness. A trained mental health practitioner must be involved in the assessment of all participants throughout the duration of the trial. The trained practitioner should not only assess changes in the individuals emotional functioning and adjustment to gains or losses in visual acuity but also assist the participant in gaining an understanding of complex information that is relevant to the trial. Involvement of the potential participant in an early stage of the planning of a trial can significantly benefit the process of informed consent.	X	X	X

cial industry to compare and contrast outcomes across the various disciplines.

Acknowledgments

represent US Food and Drug Administration policy or guidelines.

HOVER International Taskforce Working Group Members: Visual Acuity: Ian Bailey (chair), Michael Bach, Rick Ferris, Chris Johnson, Ava Bittner, August Colenbrander, Jill Keeffe; Electrophysiology: Gislin Dagnelie (chair), Michael Bach, David Birch, Laura Frishman, J. Vernon Odom; Electrically-Evoked Device Effectiveness: Matthew Petoe (chair), Daniel Rathbun, Ethan Cohen, Ione Fine, Ralf Hornig; Vision Processing Systems: Chris McCarthy (chair), Vincent Bismuth; Activities of Daily Living: Gary Rubin (chair), Mary Lou Jackson, Shane McSweeney, Cynthia Owsley, Robert Finger, Jill Keeffe, Sharon Bentley, Gislin Dagnelie, Joan Stelmack; Orientation and Mobility: Duane Geruschat (chair), Sharon Bentley, Marshall Flax, Richard Long, James Weiland, Russell L Woods; Patient Reported Outcomes: Gislin Dagnelie (chair), Akos Kusnyerik, Collette Mann, Eberhart Zrenner, Eduardo Fernandez, Frank Lane, Gary Rubin, Katarina Stingl, Lotfi Merabet, Robert Finger, Takashi Fujikado, J. Vernon Odom; Psychosocial Assessments and Ethical Considerations: Philip Troyk (chair), Frank Lane.

Expert Peer Reviewers (within the Taskforce): Carla Abbott, Penelope Allen, Amir Amedi, Nick Barnes, Tamara Brawn, Patrick Degenaar, Jessy Dorn, Thomas Edwards, Cordelia Erickson-Davis, Long-Sheng Fan, Takashi Fujikado, Peter Gabel, David Gamm, David Goldman, Robyn Guymer, Archana Jalligampala, Jasleen Jolly, Bernard Lepri, Yossi Mandel, Fleur O'Hare, Janine Walker, Peter Walter.

References

1. Brindley GS, Lewin WS. The sensations produced by electrical stimulation of the visual cortex. *J Physiol*. 1968;196:479–493.

2. Dobelle WH, Mladejovsky MG. Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. *J Physiol*. 1974;243:553–576.

3. Rizzo JF, Ayton LN. Psychophysical testing of visual prosthetic devices: a call to establish a multi-national joint task force. *J Neural Eng*. 2014;11:020301.

4. International Council of Ophthalmology. Visual standards - aspects and ranges of vision loss with emphasis on population surveys. Available at: http://www.icoph.org/resources/10/Visual-Standards—Aspects-and-Ranges-of-Vision-Loss.html. Accessed May 22, 2020.

5. Ferris FL, 3rd, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol*. 1982;94:91–96.

6. Ferris FL, 3rd, Bailey I. Standardizing the measurement of visual acuity for clinical research studies: guidelines from the Eye Care Technology Forum. *Ophthalmology*. 1996;103:181–192.

7. Bittner AK, Ibrahim MA, Haythornthwaite JA, et al. Vision test variability in retinitis pigmentosa and psychosocial factors. *Optom Vis Sci*. 2011;88:1496–1506.

8. Bach M. The Freiburg Visual Acuity test–automatic measurement of visual acuity. *Optom Vis Sci*. 1996;73:49–53.

9. Bailey IL, Jackson AJ, Minto H, et al. The Berkeley Rudimentary Vision Test. *Optom Vis Sci*. 2012;89:1257–1264.

10. Wilke R, Bach M, Wilhem W, et al. Testing visual functions in patients with visual prostheses. In: Humayun M, Weiland J, Chader G, Greenbaum E, eds. *Artificial Sight: Basic Research, Biomedical Engineering and Clinical Advances*. New York, NY: Springer; 2007:91–111.

11. Bittner AK, Jeter P, Dagnelie G. Grating acuity and contrast tests for clinical trials of severe vision loss. *Optom Vis Sci*. 2011;88:1153–1163.

12. Bach M, Wilke M, Wilhelm B, et al. Basic quantitative assessment of visual performance in patients with very low vision. *Invest Ophthalmol Vis Sci*. 2010;51:1255–1260.

13. Chen SC, Hallum LE, Suaning GJ, Lovell N. Psychophysics of prosthetic vision: I. Visual scanning and visual acuity. *Conf Proc IEEE Eng Med Biol Soc*. 2006;1:4400–4403.

14. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt*. 1976;53:740–745.

15. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991;32:422–432.

16. 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;135:194–205.

17. Bach M. The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefes Arch Clin Exp Ophthalmol*. 2007;245:965–971.

18. Schulze-Bonsel K, Feltgen N, Burau H, et al. Visual acuities "hand motion" and "counting fingers" can be quantified with the Freiburg Visual Acuity Test. *Invest Ophthalmol Vis Sci*. 2006;47:1236–1240

19. Bailey IL, Lovie-Kitchin JE. Visual acuity testing. From the laboratory to the clinic. *Vision Res*. 2013;90:2–9.

20. McCulloch DL, Marmor MF, Brigell MG, et al. ISCEV standard for full-field clinical electroretinography (2015 update). *Doc Ophthalmol*. 2015;130:1–12.

21. Odom JV, Bach M, Brigell M, et al. ISCEV standard for clinical visual evoked potentials (2016 update). *Doc Ophthalmol*. 2016;133:1–9.

22. Birch DG, Sandberg MA. Submicrovolt full-field cone electroretinograms: artifacts and reproducibility. *Doc Ophthalmol*. 1996;92:269–280.

23. Dagnelie G, Massof RW. Sub-microvolt electroretinograms: negotiating the pitfalls of electricity and noise. In: *Vision Science and Its Applications 1994*. Technical Digest Series, Vol. 2. Washington, DC: Optical Society of America; 1994:354–357.

24. Viswanathan S, Frishman LJ, Robson JG. Inner-retinal contributions to the photopic sinusoidal flicker electroretinogram of macaques. Macaque photopic sinusoidal flicker ERG. *Doc Ophthalmol*. 2002;105:223–242.

25. Fishman GA, Birch DG, Holder GE, Brigell MG. *Electrophysiologic Testing in Disorders of the Retina, Optic Nerve, and Visual Pathway*. 2nd ed. San Francisco, CA: Foundation of the American Academy of Ophthalmology; 2001.

26. Schatz A, Wilke R, Strasser T, Gekeler F, Messias A, Zrenner E. Assessment of "non-recordable" electroretinograms by 9 Hz flicker stimulation under scotopic conditions. *Doc Ophthalmol*. 2012;124:27–39.

27. Bach M, Meigen T. Do's and don'ts in Fourier analysis of steady-state potentials. *Doc Ophthalmol*. 1999;99:69–82.

28. Shimada Y, Horiguchi M. Stray light-induced multifocal electroretinograms. *Invest Ophthalmol Vis Sci*. 2003;44:1245–1251.

29. Stronks HC, Barry MP, Dagnelie G. Electrically elicited visual evoked potentials in Argus II retinal implant wearers. *Invest Ophthalmol Vis Sci*. 2013;54:3891–3901.

30. Strasser T, Nasser F, Langrova H, et al. Objective assessment of visual acuity: a refined model for analyzing the sweep VEP. *Doc Ophthalmol*. 2019;138:97–116.

31. Bach M, Farmer JD. Evaluation of the "Freiburg Acuity VEP" on commercial equipment. *Doc Ophthalmol*. 2020;140:139–145.

32. Stronks HC, Barry MP, Dagnelie G. Electrically evoked electroretinograms and pupil responses in Argus II retinal implant wearers. *Doc Ophthalmol*. 2016;132:1–15.

33. Hood DC, Bach M, Brigell M, et al. ISCEV standard for clinical multifocal electroretinography (mfERG) (2011 edition). *Doc Ophthalmol*. 2012;124:1–13.

34. Park JC, Cao D, Collison FT, Fishman GA, McAnany JJ. Rod and cone contributions to the dark-adapted 15-Hz flicker electroretinogram. *Doc Ophthalmol*. 2015;130:111–119.

35. Roman AJ, Schwartz SB, Aleman TS, et al. Quantifying rod photoreceptor-mediated vision in retinal degenerations: dark-adapted thresholds as outcome measures. *Exp Eye Res*. 2005;80:259–272.

36. Roman AJ, Cideciyan AV, Aleman TS, Jacobson SG. Full-field stimulus testing (FST) to quantify visual perception in severely blind candidates for treatment trials. *Physiol Meas*. 2007;28:N51–N56.

37. Klein M, Birch DG. Psychophysical assessment of low visual function in patients with retinal degenerative diseases (RDDs) with the Diagnosys full-field stimulus threshold (D-FST). *Doc Ophthalmol*. 2009;119:217–224.

38. Drack AV, Chung D, Russell S, et al. Results of phase III clinical trial subretinal gene therapy for RPE65-mediated Leber congenital amaurosis (LCA). *JAAPOS*. 2016;20:e4.

39. Fine I, Boynton GM. Pulse trains to percepts: the challenge of creating a perceptually intelligible world with sight recovery technologies. *Philos Trans R Soc Lond B Biol Sci*. 2015;370:1677.

40. Ahuja AK, Yeoh J, Dorn JD, et al. Factors affecting perceptual threshold in Argus II retinal prosthesis subjects. *Transl Vis Sci Technol*. 2013;2:1.

41. de Balthasar C, Patel S, Roy A, Freda R, Greenwald S, Horsager A, et al. Factors affecting perceptual thresholds in epiretinal prostheses. *Invest Ophthalmol Vis Sci*. 2008;49:2303–2314.

42. Horsager A, Greenwald SH, Weiland JD, et al. Predicting visual sensitivity in retinal prosthesis patients. *Invest Ophthalmol Vis Sci*. 2009;50:1483–1491.

43. Shivdasani MN, Sinclair NC, Dimitrov PN, et al. Factors affecting perceptual thresholds in a suprachoroidal retinal prosthesis. *Invest Ophthalmol Vis Sci*. 2014;55:6467–6481.

44. Nanduri D, Fine I, Horsager A, et al. Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation. *Invest Ophthalmol Vis Sci*. 2012;53:205–214.

45. Greenwald SH, Horsager A, Humayun MS, Greenberg RJ, McMahon MJ, Fine I. Brightness as a function of current amplitude in human retinal electrical stimulation. *Invest Ophthalmol Vis Sci*. 2009;50:5017–5025.

46. Freeman DK, Fried SI. Multiple components of ganglion cell desensitization in response to prosthetic stimulation. *J Neural Eng*. 2011;8:016008.

47. Perez Fornos A, Sommerhalder J, da Cruz L, et al. Temporal properties of visual perception on electrical stimulation of the retina. *Invest Ophthalmol Vis Sci*. 2012;53:2720–2731.

48. Nanduri D, Humayun MS, Greenberg RJ, McMahon MJ, Weiland JD. Retinal prosthesis phosphene shape analysis. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:1785–1788.

49. White JM, Levi DM, Aitsebaomo AP. Spatial localization without visual references. *Vision Res*. 1992;32:513–526.

50. Horsager A, Greenberg RJ, Fine I. Spatiotemporal interactions in retinal prosthesis subjects. *Invest Ophthalmol Vis Sci*. 2010;51:1223–1233.

51. Horsager A, Boynton GM, Greenberg RJ, Fine I. Temporal interactions during paired-electrode stimulation in two retinal prosthesis subjects. *Invest Ophthalmol Vis Sci*. 2011;52:549–557.

52. Stingl K, Bartz-Schmidt KU, Braun A, et al. Transfer characteristics of subretinal visual implants: corneally recorded implant responses. *Doc Ophthalmol*. 2016;133:81–90.

53. Stronks HC, Barry MP, Dagnelie G. Electrically evoked electroretinograms and pupil responses in Argus II retinal implant wearers. *Doc Ophthalmol*. 2016;132:1–15.

54. Stronks HC, Barry MP, Dagnelie G. Electrically elicited visual evoked potentials in Argus II retinal implant wearers. *Invest Ophthalmol Vis Sci*. 2013;54:3891–3901.

55. Stevens SS. On the psychophysical law. *Psychol Rev*. 1957;64:153–181.

56. Greenwald SH, Horsager A, Humayun MS, Greenberg RJ, McMahon MJ, Fine I. Brightness as a function of current amplitude in human retinal electrical stimulation. *Invest Ophthalmol Vis Sci*. 2009;50:5017–5025.

57. Dagnelie G. Psychophysical evaluation for visual prosthesis. *Annu Rev Biomed Eng*. 2008;10:339–368.

58. Zrenner E, Bartz-Schmidt KU, Benav H, et al. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc Biol Sci*. 2011;278:1489–1497.

59. Lorach H, Goetz G, Smith R, et al. Photovoltaic restoration of sight with high visual acuity. *Nat Med*. 2015;21:476–482.

60. Stingl K, Bartz-Schmidt KU, Besch D, et al. Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc Biol Sci*. 2013;280:20130077.

61. Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971;49(suppl 2):467.

62. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. Huntington, NY: R.E. Krieger Publishing Company; 1974.

63. Klein SA. Measuring, estimating, and understanding the psychometric function: a commentary. *Percept Psychophys*. 2001;63:1421–155.

64. Humayun MS, Weiland JD, Fujii GY, et al. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Res*. 2003;43:2573–2581.

65. Velikay-Parel M, Ivastinovic D, Georgi T, Richard G, Hornig R. A test method for quantification of stimulus-induced depression effects on perceptual threshold in epiretinal prosthesis. *Acta Ophthalmol*. 2013;91:e595–e602.

66. Stronks HC, Dagnelie G. Phosphene mapping techniques for visual prostheses. In: Dagnelie G, ed. *Visual Prosthetics: Physiology, Bioengineering, Rehabilitation*. New York: Springer; 2011:367–283.

67. Brelén ME, Duret F, Gerard B, Delbeke J, Veraart C. Creating a meaningful visual perception

in blind volunteers by optic nerve stimulation. *J Neural Eng*. 2005;2:S22–S28.

68. Wilke R, Gabel VP, Sachs H, et al. Spatial resolution and perception of patterns mediated by a subretinal 16-electrode array in patients blinded by hereditary retinal dystrophies. *Invest Ophthalmol Vis Sci*. 2011;52:5995–6003.

69. Cicione R, Fallon JB, Rathbone GD, Williams CE, Shivdasani MN. Spatiotemporal interactions in the visual cortex following paired electrical stimulation of the retina. *Invest Ophthalmol Vis Sci*. 2014;55:7726–7738.

70. Beyeler M, Rokem A, Boynton GM, Fine I. Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. *J Neural Eng*. 2017;14:051003.

71. Dumm G, Fallon JB, Williams CE, Shivdasani MN. Virtual electrodes by current steering in retinal prostheses. *Invest Ophthalmol Vis Sci*. 2014;55:8077–8085.

72. Yan Y, Lu Y, Li M, et al. Electrically evoked responses in the rabbit cortex induced by current steering with penetrating optic nerve electrodes. *Invest Ophthalmol Vis Sci*. 2016;57:6327–6338.

73. Bierer JA. Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration. *J Acoust Soc Am*. 2007;121:1642–1653.

74. Sabbah N, Authie CN, Sanda N, Mohand-Said S, Sahel JA, Safran AB. Importance of eye position on spatial localization in blind subjects wearing an Argus II retinal prosthesis. *Invest Ophthalmol Vis Sci*. 2014;55:8259–8266.

75. Caspi A, Roy A, Dorn JD, Greenberg RJ. Retinotopic to spatiotopic mapping in blind patients implanted with the Argus II retinal prosthesis. *Invest Ophthalmol Vis Sci*. 2017;58:119–127.

76. Feng D, McCarthy C. Enhancing scene structure in prosthetic vision using iso-disparity perturbance maps. *Conf Proc IEEE Eng Med Biol Soc*. 2013;5283–5286.

77. McCarthy C, Walker J, Lieby P, Scott A, Barnes N. Mobility and low contrast trip hazard avoidance using augmented depth. *J Neural Eng*. 2014;12:016003.

78. Parikh N, Itti L, Weiland J. Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues. *J Neural Eng*. 2013;10:1–13.

79. Horne L, Alvarez JM, Salzmann M, Barnes N. Semantic labelling for prosthetic vision. *Comput Vis Image Underst*. 2016:149:113–125.

80. Caspi A, Dorn JD, McClure KH, Humayun MS, Greenberg RJ, McMahon MJ. Feasibility study of a retinal prosthesis: spatial vision with a 16-electrode implant. *Arch Ophthalmol*. 2009;127:398–401.

81. da Cruz L, Coley BF, Dorn JD, et al. The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. *Br J Ophthalmol*. 2013;97:632–636.

82. Kotecha A, Zhong J, Stewart D, da Cruz L. The Argus II prosthesis facilitates reaching and grasping tasks: a case series. *BMC Ophthalmol*. 2014;14:71.

83. McCarthy C, Barnes N. Importance weighted image enhancement for prosthetic vision: an augmentation framework. In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2014:45–51.

84. Dagnelie G, Keane P, Naria V, Yang L, Weiland J, Humayun M. Real and virtual performance in simulated prosthetic vision. *J Neural Eng*. 2007;4:S92–S101.

85. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged: The index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914–919.

86. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179–186.

87. Turner-Stokes L, Nyein K, Turner-Stokes T, Gatehouse C. The UK FIM+ FAM: development and evaluation. *Clin Rehab*. 1999;13:277–287.

88. Warrian KJ, Altangerel U, Spaeth GL. Performance-based measures of visual function. *Surv Ophthalmol*. 2010;55:146–161.

89. Humayun MS, Dorn JD, da Cruz L, et al. Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology*. 2012;119:779–788.

90. Mansfield JS, Legge GE, Bane MC. Psychophysics of reading. XV: font effects in normal and low vision. *Invest Ophthalmol Vis Sci*. 1996;37:1492–1501.

91. Radner W, Willinger U, Obermayer W, Mudrich C, Velikay-Parel M, Eisenwort B. A new German reading chart for the simultaneous evaluation of reading acuity and reading speed. *Klin Monatsbl Augenheilkd*. 1998;213:174–181.

92. West S, Muñoz B, Rubin GS, et al. Visual impairment and functional status in older persons:

methodologies for the Salisbury Eye Evaluation Project. *Invest Ophthalmol Vis Sci*. 1995;36:S419.

93. Trauzettel-Klosinski S, Dietz K, IReST Study Group. Standardized assessment of reading performance: the New International Reading Speed Texts IReST. *Invest Ophthalmol Vis Sci*. 2012;53:5452–5461.

94. West SK, Rubin GS, Munoz B, Abraham D, Fried LP. Assessing functional status: correlation between performance on tasks conducted in a clinic setting and performance on the same task conducted at home. The Salisbury Eye Evaluation Project Team. *J Gerontol A Biol Sci Med Sci*. 1997;52:M209–M217.

95. Vaegan, Halliday BL. A forced-choice test improves clinical contrast sensitivity testing. *Br J Ophthalmol*. 1982;66:477–491.

96. Warrian KJ, Katz LJ, Myers JS, et al. A comparison of methods used to evaluate mobility performance in the visually impaired. *Br J Ophthalmol*. 2015;99:113–118.

97. Geruschat DR, Flax M, Tanna N, et al. FLORA: phase I development of a functional vision assessment for prosthetic vision users. *Clin Exp Optom*. 2015;98:342–347.

98. Finger RP, McSweeney SC, Deverell L, et al. Developing an instrumental activities of daily living tool as part of the low vision assessment of daily activities protocol. *Invest Ophthalmol Vis Sci*. 2014;55:8458–8466.

99. Gulati R, Roche H, Thayaparan K, Hornig R, Rubin GS. The development of a picture discrimination test for people with very poor vision. *Invest Ophthalmol Vis Sci*. 2011;52:1197.

100. Bainbridge JW, Mehat MS, Sundaram V, et al. Long-term effect of gene therapy on Leber's congenital amaurosis. *N Engl J Med*. 2015;372: 1887–1897.

101. 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;358:2231–2239.

102. Wiener WR, Welsh RL, Blasch BB. *Foundations of Orientation and Mobility*. 3rd ed. New York: AFB Press; 2010.

103. Deverell L, Bentley SA, Ayton LN, Delany C, Keeffe JE. Effective mobility framework: a tool for designing comprehensive O&M outcomes research. *IJOM*. 2015;7:74–86.

104. Darken RP, Peterson B. Spatial orientation, wayfinding, and representation. In: Stanney KM, ed. *Handbook of Virtual Environments: Design, Implementation, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:493–518.

105. Mast F, Zaehle T. *Spatial Reference Frames Used in Mental Imagery Tasks. Blindness and Brain Plasticity in Navigation and Object Perception*. New York: Lawrence Erlbaum Associates; 2008.

106. Long RG, Rieser JJ, Hill EW. Mobility in individuals with moderate visual impairments. *J Vis Impair Blind*. 1990;84:111–118.

107. Geruschat DR, Bittner AK, Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. *Optom Vis Sci*. 2012;89:1308–1315.

108. Deverell EA. *Functional Vision Research: Measuring Vision-Related Outcomes in Orientation and Mobility - VROOM*. Melbourne, Australia: University of Melbourne; 2016. Thesis.

109. Finger RP, Ayton LN, Deverell L, et al. Developing a very low vision orientation & mobility test battery (O&M-VLV). *Optom Vis Sci*. 2016;93:1127–1136.

110. Marron JA, Bailey IL. Visual factors and orientation-mobility performance. *Am J Optom Physiol Opt*. 1982;59:413–426.

111. Haymes SA, Guest DJ, Heyes AD, Johnston AW. Mobility of people with retinitis pigmentosa as a function of vision and psychological variables. *Optom Vis Sci*. 1996;73:621–637.

112. Soong GP, Lovie-Kitchin JE, Brown B. Preferred walking speed for assessment of mobility performance: sighted guide versus non-sighted guide techniques. *Clin Exp Optom*. 2000;83:279–282.

113. Leat SJ, Lovie-Kitchin JE. Measuring mobility performance: experience gained in designing a mobility course. *Clin Exp Optom*. 2006;89:215–228.

114. Nau AC, Pintar C, Fisher C, Jeong JH, Jeong K. A standardized obstacle course for assessment of visual function in ultra low vision and artificial vision. *J Vis Exp*. 2014;84:e51205.

115. Zebehazy KT, Zimmerman GJ, Bowers AR, Luo G, Peli E. Establishing mobility measures to assess the effectiveness of night vision devices: results of a pilot study. *J Visual Impair Blind*. 2005;99:663–670.

116. Wood JM, Lacherez PF, Black AA, Cole MH, Boon MY, Kerr GK. Postural stability and gait among older adults with age-related maculopathy. *Invest Ophthalmol Vis Sci*. 2009;50:482–487.

117. Elliott DB, Patla AE, Flanagan JG, et al. The Waterloo Vision and Mobility Study: postural control strategies in subjects with ARM. *Ophthalmic Physiol Opt*. 1995;15:553–559.

118. Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function

Questionnaire. *Arch Ophthalmol*. 2001;119:1050–1058.

119. Massof R, Rubin G. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531–548.

120. USFDA. *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Washington, DC: Center for Drug Evaluation and Research, Center for Devices and Radiological Health, Center for Biologics and Evaluation and Research, U.S. Food and Drug Administration; 2009.

121. European Medicines Agency. *Guideline on Clinical Trials in Small Populations*. Amsterdam, The Netherlands: European Medicines Agency; 2006.

122. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2004;11:381–399.

123. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. *Ophthalmic Epidemiol*. 2005;12:103–124.

124. Massof RW, Ahmadian L. What do different visual function questionnaires measure? *Ophthalmic Epidemiol*. 2007;14:198–204.

125. Massof RW, Ahmadian L, Grover LL, et al. The Activity Inventory: an adaptive visual function questionnaire. *Optom Vis Sci*. 2007;84:763–774.

126. Stelmack JA, Babcock-Parziale JL, Head DN, et al. Timing and directions for administration of questionnaires affect outcomes measurement. *J Rehabil Res Dev*. 2006;43:809–816.

127. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci*. 2010;87:285–293.

128. Becker SW, Lambert RW, Schulz EM, Wright BD, Burnet DL. An instrument to measure the activity level of the blind. *Int J Rehab Res*. 1985;8:415–424.

129. Massof R. A systems model for low vision rehabilitation. II. Measurement of vision disabilities. *Optom Vision Sci*. 1998;75:349–373.

130. Sloane ME, Ball K, Owsley C, Bruni JR, Roenker DL. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. *Tech Dig Noninvas Assess Vis Sys*. 1992;1:26–29.

131. Mangione CM, Phillips RS, Seddon JM, et al. Development of the "Activities of Daily Vision Scale." A measure of visual functional status. *Med Care*. 1992;30:1111–1126.

132. Mangione CM, Lee PP, Pitts J, Gutierrez P, Berry S, Hays RD. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Arch Ophthalmol*. 1998;116:1496–1504.

133. Pesudovs K, Gothwal VK, Wright TA, Lamoureux E. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*. 2010;36:718–732.

134. Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. *Optom Vis Sci*. 2007;84:689–704.

135. Gothwal VK, Lovie-Kitchin JE, Nutheti R. The development of the LV Prasad-Functional Vision Questionnaire: a measure of functional vision performance of visually impaired children. *Invest Ophthalmol Vis Sci*. 2003;44:4131–4139.

136. Gothwal VK, Sumalini R, Bharani S, Reddy SP, Bagga DK. The second version of the L. V. Prasad-Functional Vision Questionnaire. *Optom Vis Sci*. 2012;89:1601–1610.

137. Felius J, Stager DR, Sr., Berry PM, et al. Development of an instrument to assess vision-related quality of life in young children. *Am J Ophthalmol*. 2004;138:362–372.

138. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. *Invest Ophthalmol Vis Sci*. 2004;45:3919–3928.

139. Stelmack JA, Massof RW. Using the VA LV VFQ-48 and LV VFQ-20 in low vision rehabilitation. *Optom Vis Sci*. 2007;84:705–709.

140. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2006;47:4732–4741.

141. Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci*. 2002;43:927–935.

142. Cochrane G, Lamoureux E, Keeffe J. Defining the content for a new quality of life questionnaire for students with low vision (the Impact of Vision Impairment on Children: IVI_C). *Ophthalmic Epidemiol*. 2008;15:114–120.

143. Cochrane GM, Marella M, Keeffe JE, Lamoureux EL. The Impact of Vision Impairment for Children (IVI_C): validation of a vision-specific pediatric quality-of-life questionnaire using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2011;52:1632–1640.

144. Khadka J, Ryan B, Margrain TH, Court H, Woodhouse JM. Development of the 25-item

Cardiff Visual Ability Questionnaire for Children (CVAQC). *Br J Ophthalmol*. 2010;94:730–735.

145. Tadic V, Cooper A, Cumberland P, Lewando-Hundt G, Rahi JS, Vision-Related Quality of Life Group. Development of the functional vision questionnaire for children and young people with visual impairment: the FVQ_CYP. *Ophthalmology*. 2013;120:2725–2732.

146. Hatt SR, Leske DA, Castañeda YS, et al. Development of pediatric eye questionnaires for children with eye conditions. *Am J Ophthalmol*. 2019;200:201–217.

147. Leske DA, Hatt SR, Castañeda YS, et al. Validation of the pediatric eye questionnaire in children with visual impairment. *Am J Ophthalmol*. 2019;208:124–132.

148. Finger RP, Tellis B, Crewe J, Keeffe JE, Ayton LN, Guymer RH. Developing the impact of Vision Impairment-Very Low Vision (IVI-VLV) questionnaire as part of the LoVADA protocol. *Invest Ophthalmol Vis Sci*. 2014;55:6150–6158.

149. Jeter PE, Rozanski C, Massof RW, Adeyemo O, Dagnelie G, PLOVR Study Group. Development of the Ultra Low Vision-Visual Functioning Questionnaire (ULV-VFQ) as part of the Prosthetic Low Vision Rehabilitation (PLoVR) curriculum. *Transl Vis Sci Tech*. 2017;6:11.

150. Dagnelie G, Jeter PE, Adeyemo O, PLOVR Study Group. Optimizing the ULV-VFQ for clinical use through item set reduction: psychometric properties and trade-offs. *Transl Vis Sci Tech*. 2017;6:12.

151. Dagnelie G, Barry MP, Adeyemo O, Jeter PE, Massof RW, PLoVR Study Group. Twenty Questions: an adaptive version of the PLoVR ultra-low vision (ULV) questionnaire. *Invest Ophthalmol Vis Sci*. 2015;56:497.

152. Adeyemo K, Jeter PE, Rozanski C, Nkodo AF, Dagnelie G. Comparison of retinal prosthesis wearers to individuals with severe vision loss in the development of an ultra-low vision questionnaire. *Invest Ophthalmol Visual Sci*. 2014;55:1838.

153. Grant P, Spencer L, Arnoldussen A, et al. The functional performance of the BrainPort V100 device in persons who are profoundly blind. *J Vis Impair Blind*. 2016;110:77–88.

154. Dagnelie G, Geruschat D, Massof RW, Jeter PE, Adeyemo O. Developing a calibrated ultra-low vision (ULV) assessment toolkit. *Optom Vis Sci*. 2015;92.

155. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. 1978. Bethesda, MD: U.S. Government Printing Office.

156. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 5th ed. Oxford, UK: Oxford University Press; 2001.

157. Bhutta ZA. Beyond informed consent. *Bull World Health Org*. 2004;82:771–778.

158. Lane FJ, Nitsch KP, Huyck M, Troyk PR, Schug K. Perspectives of optic nerve prostheses. *Disabil Rehabil Assist Technol*. 2016;11:301–309.

159. Lane FJ, Huyck MH, Troyk PR, Schug K. Responses of potential users to the intracortical visual prosthesis: Final themes from the analysis of focus group data. *Disabil Rehabil Assist Technol*. 2012;7:304–313.

160. Lane FJ, Huyck MH, Troyk PR. Planning for the first human intracortical visual prosthesis by using pilot data from focus groups of potential users. *Disabil Rehabil Assist Technol*. 2011;6:139–147.

161. Seelig BJ, Dobelle WH. Altruism and the volunteer: psychological benefits from participating as a research subject. *ASAIO J*. 2001;47:3–5.

162. Seelig BJ, Rosof LS. Normal and pathological altruism. *J Am Psychoanal Assoc*. 2001;49:933–959.

163. Altman LK. *Who Goes First?* New York: Random House; 1986.

164. Kilgore KL, Scherer M, Bobblitt R, et al. Neuroprosthetics consumers' forum: consumer priorities for research directions. *J Rehabil Res Dev*. 2001;38:655–660.

165. Marshall PA, Adebamowo CA, Adeyemo AA, et al. Voluntary participation and informed consent to international genetic research. *Am J Public Health*. 2006;96:1989–1995.

166. Fine I, Jacobs RA. Comparing perceptual learning across tasks: a review. *J Vision*. 2002;2:190–203.