THE HARMONIZATION OF OUTCOMES AND VISION ENDPOINTS IN VISION RESTORATION TRIALS (HOVER) CONSENSUS DOCUMENT

Compiled consensus documents - DRAFT

Updated 25 June 2016

**TABLE OF CONTENTS**

Recommendations for Visual Acuity Measurements 3

Preamble 3

Defining Principles 4

Recommended Methodology for Assessment of Visual Acuity 4

General recommendations for testing include: 4

Specific recommendation for post-intervention measurements include: 6

Specific Visual Acuity Test Methodologies 6

Reporting Guidelines 9

Epilogue 9

References 9

Recommendations for Orientation and Mobility Assessments 11

Defining Principles 11

Research Methods for Assessment of Orientation and Mobility 12

Reporting Guidelines 13

Epilogue 14

References 14

Recommendations for Assessment of Activities of Daily Living 16

Defining Principles 16

Recommended Methodology for Assessment of Activities of Daily Living 17

Reporting Guidelines 19

Epilogue 20

References 20

Recommendations for Assessment of Patient Reported Outcomes 22

Defining Principles 22

Recommended Methodology for Patient-Reported Outcomes 23

Reporting Guidelines 29

Epilogue 30

References 31

Recommendations for Measurement and Reporting of Electrically-Evoked Device Efficacy 33

Defining Principles 33

Considerations 33

Recommendations 34

Appendix A - Specific Device Parameters 38

References 40

Recommendations for Implementation and Reporting of Vision Processing Algorithms 42

Preamble 42

Defining principles 42

Reporting Guidelines for Vision Processing 44

Recommendations for assessing vision processing methods 46

References 48

Recommendations for Psychosocial Assessments and Ethical Considerations 49

Defining Principles 49

Considerations 49

Recommendations 51

References 53

# Recommendations for Visual Acuity Measurements

Ian Bailey (chair), Michael Bach, Rick Ferris, Chris Johnson, Ava Bittner, August Colenbrander, Jill Keeffe

## Preamble

Ever since Snellen developed his chart, measurement of letter chart acuity has been a mainstay for the assessment of vision. Its application is easy, cheap and fast. Indeed, its use is so pervasive that the terms “visual acuity” (meaning letter chart acuity) and “vision” are often used as synonyms. When considering extending the reach of visual acuity measurement to ultra-low vision, beyond the range of letter charts, it is important to also consider the limitations of this type of measurement.

Testing with visual acuity letter charts serve many clinical purposes. It can provide a means of identifying and monitoring differences and changes in ocular health in many ocular diseases, and it can guide decision-making when correcting refractive errors.

For letter charts, the visual task is recognizing letters (or other optotypes) sequentially within a chart array in which the letters become smaller from one row to the next. 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 fovea area of the retina is responsible to the recognition of the optotypes as eye-movements systematically shift the attention across and down the chart.

In ultra-low vision, the features in the visual acuity targets will generally need to be much larger than 1 degree, and also, there is a higher likelihood of scotomas or other visual field restrictions. 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 the test target.

When vision is too poor to perform the visual task of reading a letter chart, then the test 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 ocular disorders, it is important to consider how each eye functions. This requires assessment of various parameters of ocular function such as visual acuity, contrast sensitivity, color discrimination, visual fields, etc., and identifying any impairments in these individual functions in each eye. 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 which is vision-related functioning, also 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 estimate or indicator of ability or disability.

## Defining Principles

The measurement of visual acuity in vision restoration trials is of upmost importance, being one of the key outcome measures accepted by both regulatory bodies and 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 ETDRS chart1, is the gold standard for assessment of the minimum angle of resolution that a subject can achieve2. However, such charts are only able to measure acuity down to levels of logMAR 1.6 (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 = 1.6 were assessed as having “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 ultra low vision is the significant variability that exists.3 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 guidelines outline the recommended methodologies for the testing and reporting of psychophysical results of testing in ultra-low vision 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.

## 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, the Freiburg Acuity and Contrast test (FrACT) or the Berkeley Rudimentary Vision Test (BRVT),

## General recommendations for testing include:

The examiner should be qualified in the assessment of visual acuity. Ophthalmologists, optometrists, orthoptists or certified ophthalmic technicians or assistants can all suitable for this role provided they are willing to follow a standardized protocol.

The choice of test should depend in part upon the level of vision. For subjects with vision of logMAR 1.6 (equivalent 6/240 or 20/800) or better, the ETDRS test can be used. For subjects with worse vision, the FrACT or BRVT tests should be used. New alternative tests for very low vision might be developed n 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, refractive error may be estimated using an auto-refractor or retinoscopy. The use of any refractive correction, other than what would be normally used, must be described in any report or publication.

For post-intervention measurements, the recommended method depends upon the type of prosthetic device used. For 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 should be used if the subject reports a benefit even if not quantifiable. For photodiode-based prosthetic devices, optimal refractive correction should be used during any acuity measure.

For ETDRS and BRVT testing, room lighting should be kept at 550 lux ± 10%, respectively. For other analogous testing methods, variations in ambient light intensity, including absence of room lighting, can be used as deemed appropriate. Care should be taken to ensure that the lighting remains the same from one test session to the next.

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. During the monocular tests, the second eye should be patched

Within a test protocol, there should be rules for stopping, for guessing or for allowing subjects to correct or change responses. There should be standard rules and procedures for encouraging guessing and pointing to helping 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. This has been shown to improve performance on simulated prosthetic vision tests4 and in natural low vision, and so must be reported if allowed.

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

In order to quantify residual islands of vision, hand-held optotype charts may be used (either an ETDRS letter chart or a Berkeley Rudimentary Vision Test chart). Such testing should begin at a distance of 4 meters, with the examiner moving the chart into all four quadrants 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 for 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 individual subjects, it will be appropriate to make visual acuity measurements with more than one kind of visual acuity test task. Scores of visual acuity can show wide differences from one test task to another. Grating acuity scores may well be different from scores of single optotype acuity, and single optotype acuity may be different from visual acuity with a logMAR chart.

## Specific recommendation for post-intervention measurements include:

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 and eyes to assess if there is a bilateral effect on visual performance.

The fellow eye should be patched during 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.

## Specific Visual Acuity Test Methodologies

LogMAR letter acuity tests

A standardized logMAR letter acuity is the gold standard acuity measure for subjects with vision greater than logMAR 1.6 (20/800 or 6/240). The most widely used format is the ETDRS chart1.

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 3rd row of letters (the 0.8 logMAR line) is 125 ± 5 cm from the floor and 4 meters away from the subject, regardless of their vision.
2. The right eye is tested first using ETDRS chart #1.
3. The subject is 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 letter5.
5. If the subject is unable to read more than 20 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. At the completion of the left eye testing, remove the occluder from the right eye and repeat 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 ultra-low visual acuity include the Freiburg Acuity and Contrast Test (FrACT) and the Berkeley Rudimentary Vision Test (BRVT) as described below.

Low vision optotype test – The Freiburg Acuity and Contrast Test (FrACT)

The FrACT 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 which is available online without cost ([http://michaelbach.de/fract/](http://michaelbach.de/fract/index.html) ). Full details of the test have been published elsewhere.8-10

The recommended protocol is as follows:

1. Prior to the commencement of testing, the program must be calibrated by:
2. Measuring the observation distance from the eye to the screen, and entering this number into the “observer distance” box on the setup screen; and
3. Measuring the blue calibration line on the computer screen, and entering this value into the “length of blue ruler” box.
4. Once calibrated, the system will calculate the visual acuity range that can be presented from the distance measurements and screen resolution.
5. Acuity can be measured using a range of optotypes, including Sloan letters, Landolt Rings and tumbling E's.
6. Follow the provided checklist and on-screen instructions to complete each opt type test.
7. 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 – The Berkeley Rudimentary Vision Test (BRVT)

The BRVT test was developed for the clinical measurement of visual acuity in subjects with ultra-low vision in the range of logMAR acuity 1.6 and below. The test is administered with three card-pairs, each of which consists of two 25 cm square cards hinged together, which thus provide four panels that can be used as targets. 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 both as 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 the test have been published elsewhere.6,7

The recommended protocol for the BRVT is as follows:

1. Start testing with the Single Tumbling E (STE) acuity test at a viewing distance of 1 meter). At this distance, the STE acuity range is from logMAR = 2.0–1.4 (equivalent to 6/600 to 6/ 150, 20/2000 to 20/500), and can be measured in increments of 0.20 log units.
2. All cards in the BRVT should be presented at least four times to the subject, with the orientation changed randomly each time. For the 4-choice STE task, successful identification is taken as better than 50% correct responses for 6 or more presentations.
3. For the two-choice Grating acuity task, successful identification is taken as 80% or more correct response for 8 or more presentations
4. If the orientation of the largest STE (100 M) cannot be recognized at 1 meter, then the viewing distance is reduced to 25 cm, where the acuity range becomes logMAR 2.6–2.0 (6/2400 to 6/600 or 20/8000 to 20/2000).
5. If the subject is unable to identify the orientation of the 100 M Single Tumbling E at 25 cm, change to the second card pair of square wave gratings. These gratings should be presented at 25 cm, which provides a logMAR range of 2.90–2.30 (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 being a white quadrant on a black background and for the other the card is divided into black and white halves. The subject’s task is to locate the white quad-field or the white hemi-field.
8. If the subject fails the White Field Projection Test, the Black White Discrimination test is administered. The subject’s task is to distinguish the all black card from the all-white card.
9. To assess for light perception, the brightest light delivered by an indirect ophthalmoscope should be used. Room lighting should remain at the same level used for normal acuity testing. In addition to patching, the fellow eye also should be covered with the palm of the subject’s hand to ensure a tight seal around the orbit and bridge of the nose. 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. The subject should report when they see the light. If the examiner is convinced the subject has can identify the onset of the light the light, this response should be recorded as “light perception”. Otherwise, the subject is assumed to have “no light perception’ vision.

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 for 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 vision is classified as “no light perception’ (NLP).

## Reporting Guidelines

Any publication or presentation should contain sufficient information so that another group can replicate the testing completed. In particular, the following information must be stated:

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

1. The name of the test (and version number if available)
2. The type of acuity optotype used (e.g. letter, Landolt Ring, single tumbling E, flanked letters, etc.).
3. Room lighting (illuminance in lux)
4. The luminance (in cd/m2) of any computer screens or back-illuminated charts used in testing
5. The time cutoff for test completion
6. Qualitative control-related information, including differences with respect to whether occluder or patching were used, or scrambled vs. unscrambled stimulation input.
7. 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.
8. Indication of whether subjects used eccentric viewing when taking the tests.
9. Indication of whether subjects made scanning eye movements when taking the tests.
10. Indication of the distant refraction (as measured or estimated) and the power of any corrective lenses used for testing. Use of optical telescopes or other low vision aids, or use of electronic “zooming” must be stipulated.
11. The number of test runs within sessions and the number of sessions used to obtain results.

## Epilogue

If the study has met the above guidelines and has followed the accepted best clinical practices, the following statement may be used in publications and presentations:

“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations for Acuity Measurement of Ultra Low Vision”.

If the study was unable to comply with the guidelines (for example, if the control setting of scrambled phosphenes was unable to be implemented due to device design), then this should be stated in the publication or presentation.

E.g. “This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations for Acuity Measurement of Ultra Low Vision, EXCEPT for…”.

## References

1. Ferris FL, 3rd, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol* 1982;94(1):91-6.

2. 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(1):181-2.

3. Bittner AK, Ibrahim MA, Haythornthwaite JA, et al. Vision test variability in retinitis pigmentosa and psychosocial factors. *Optom Vis Sci* 2011;88(12):1496-506.

4. Chen SC, Hallum LE, Suaning GJ, Lovell NH. Psychophysics of prosthetic vision: I. Visual scanning and visual acuity. *Conf Proc IEEE Eng Med Biol Soc* 2006;1:4400-3.

5. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci* 1991;32(2):422-32.

6. Bailey IL, Jackson AJ, Minto H, et al. The Berkeley Rudimentary Vision Test. *Optom Vis Sci* 2012;89(9):1257-64.

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

8. Bach M. The Freiburg Visual Acuity test--automatic measurement of visual acuity. *Optom Vis Sci* 1996;73(1):49-53.

9. Bach M. The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefes Arch Clin Exp Ophthalmol* 2007;245(7):965-71.

10. 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(3):1236-40

# Recommendations for Orientation and Mobility Assessments

Duane R Geruschat (chair), Sharon A Bentley, Lil Deverell, Marshall Flax, Richard G Long, James D Weiland, Russell L Woods

## Defining Principles

Assessment of orientation and mobility (O&M) performance is essential in evaluating the impact of vision restoration and substitution interventions, as well as vision rehabilitation. *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). It includes functional tasks such as moving efficiently along routes and learning the spatial layouts of new places (Wiener et al., 2010). *Mobility* in this context refers to one’s ability to effectively preview the path ahead, and to move about safely and efficiently. It includes one’s ability to detect and avoid or negotiate obstacles, changes in elevation (e.g., curbs and stairs) and other environmental features (Wiener et al., 2010). There is extensive interplay between orientation and mobility as individuals with vision impairment move about in their homes and communities.

This document provides defining principles, recommended methodologies, and reporting guidelines that will assist researchers across settings and across interventions to assess O&M in similar ways. The focus here is on assessment in the context of vision restoration, substitution and rehabilitation technologies and interventions (e.g., retinal prostheses and gene based therapies). In particular, this document is intended to guide the assessment of interventions for individuals with ultra-low vision (ULV), defined as vision impairment that impacts most daily living activities involving visual shape recognition. With regard to visual acuity, one proposed definition of ULV is the inability to discern the largest Early Treatment of Diabetic Retinopathy Study letters at 0.5 m, i.e., 20/1600 or 6/480 (Geruschat et al., 2012).

The following principles should guide O&M assessment in ULV:

* A variety of methodologies and strategies should be used to assess O&M. Assessments may occur in laboratory settings, in the low vision or hospital clinic, and in the community. Researchers should look for opportunities to relate assessment data within individuals across these settings. In addition to mobility, assessment of orientation specifically is an important consideration, as it is in orientation that some interventions might have the most functional impact.
* A mixed methods approach to data collection is essential. Assessments should include both quantitative and qualitative measurements, and should make use of within-subjects measurement of performance, as individuals may travel with and without the aid of a vision restoration or substitution system, or before and after rehabilitation.
* Widely established, standardized measures of either orientation or mobility performance do not exist at this time. However, the recommended methodologies outlined in Section 2 provide guidance.

Assessments should be completed by O&M professionals who are experienced and skilled in assessing O&M of individuals with low vision or blindness, with support from professionals with knowledge of vision restoration, vision substitution or vision rehabilitation and experienced with the medical care of individuals with low vision or blindness. Safety must be considered and individuals should be familiarized with the assessment task prior to commencement.

Individuals should be free to use their primary mobility aid (e.g., the long cane, dog guide) when participating in the assessment as the task requires. In addition, they may be given the opportunity to be assessed without their primary mobility aid, provided adequate safeguards are in place.

## Research Methods for Assessment of Orientation and Mobility

Assessment of O&M among individuals with ULV should occur before and after the technologies and interventions are applied. Assessments should be sensitive to changes in performance that are likely to occur as a result of implementing the interventions. Assessments may take place in laboratory settings, in clinical settings and in the community. Decisions about venue will depend on the research question. A more specific or granular enquiry can be investigated effectively in a laboratory-based setting using experimental design, whereas community-based research is necessary to evaluate the broader lifestyle impact of the intervention. Researchers and O&M professionals can usually exert more experimental control when conducting laboratory and clinical studies in contrast to community-based studies, where control of potentially confounding variables, such as illumination and obstacles, is more difficult to achieve. However, researchers must carefully select the settings for their assessments so that they strike the desired balance between study validity, experimental control, and a host of other factors influencing research design and anticipated outcomes.

O&M professionals who teach O&M skills typically use a qualitative, individualized approach to assessment for planning programs and evaluating client progress in instruction. This assessment usually includes a combination of self-report and observed performance, and might involve the use of checklists (e.g., specific cane techniques or ability to recall and retrace routes). Case notes are also important in assessing gains and challenges in O&M performance, and allow for recording of clinical observations about the nuances of client behavior, the setting, and other aspects. Usually, assessment is completed in both familiar and unfamiliar areas, and under various environmental conditions (e.g., during the day and at night). However, there is little standardization in clinical O&M practice with regard to client assessment and care must be taken to avoid observer bias.

Researchers typically use a different set of tools than O&M professionals for assessing O&M. At this time, researchers are primarily concerned with measuring changes (often small ones) that are due to specific vision restoration, substitution and rehabilitation interventions. In the reasonably predictable environment of the laboratory and clinic, researchers may evaluate walking speed or percentage of preferred walking speed, or may tally the frequency and type of obstacle contacts (Marron and Bailey, 1982; Haymes et al., 1996; Soong et al., 2000; Geruschat et al., 2012) This is usually done using experimenter-developed obstacle courses (Leat and Lovie-Kitchin, 2006), before and after intervention. Virtual reality and augmented reality have had limited use in the assessment of ultra-low vision, but are worthy of development and use for O&M assessments. In regard to orientation, researchers might evaluate route travel that is guided by maps or verbal directions, and ask individuals to point to places out of view or construct maps and models of explored places. Finally, changes in posture and gait as a result of interventions may also be evaluated in the laboratory or clinic setting (Elliott et al., 1995; Wood et al., 2009).

As researchers move to the more uncontrolled community setting, they are likely to emphasize qualitative data in a mixed methods approach to O&M assessment. Qualitative methods are important because quantitative methods alone cannot capture the ways that ULV, enhanced by technologies and interventions, is useful in daily life as reflected in the ‘lived experience’ of individuals. The use of video, focus groups, interviews, questionnaires and ratings of abilities on various tasks before and after intervention are examples of the many qualitative measurement approaches that may be useful to researchers, particularly those working in community settings, as they seek to capture the context-specificity of O&M in real-world settings.

## Reporting Guidelines

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 (e.g., repeated measures) and a description of the control condition or group.
2. Where the study includes an intervention that can be manipulated (e.g. the image processing for a retinal prosthesis), report the approach and any settings or parameter values (e.g. type of image processing algorithm implemented, modifications to the image such as reverse contrast or electronic “zooming”).  Where appropriate to study aims, compare to a baseline or 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 test(s) with appropriate citation(s). If not previously published, a detailed description of the test(s) and indication of the measurement noise (e.g., measures of agreement or repeatability).
4. The number and duration of testing sessions. If sessions were conducted on different days, this should be noted. 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 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 clusters (e.g., experimenters, locations, sites), number of participants in each cluster, expertise of each cluster. In addition, efforts to standardize procedures between clusters and measures of agreement (e.g., between experimenters or sites).
8. Randomization of procedures and conditions undertaken to avoid the confounding of factors of interest with test order.
9. The methods used to mask participants and/or the experimenter(s) to the status of the intervention at each session, if any.
10. A description of the location(s) 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).
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. The viewing strategy(ies) used to complete the assessment (e.g., eccentric viewing, head scanning, use of other sensory modalities), where applicable.
13. The use of any optical correction or device (e.g., habitual spectacle correction or telescope), including the type and power, where applicable.
14. Whether the O&M tests were conducted with or without other assistive mobility devices (e.g., long cane or dog guide) and a description of any such device.
15. For assessments that include a task, a description of the task, its components and how the task was scored.
16. For observer assessments and participant self-report, a description of the methods and efforts to standardize data collection and analysis
17. For focus groups and interviews, a description of the methods, including the expertise of the interviewer(s) and questions or prompts used.
18. Any changes to the experimental protocol or conduct of the test procedure(s) that occurred during the study, whether to all, some or a single participant (including special accommodations).
19. 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.
20. 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.

## Epilogue

If the study has met the above guidelines, the following statement may be used in publications and presentations:

“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations for Orientation and Mobility Assessments of Individuals with Ultra Low Vision”

## References

1. Elliott DB, Patla AE, Flanagan JG, Spaulding S, Rietdyk S, Strong G, Brown S. The Waterloo Vision and Mobility Study: Postural control strategies in subjects with ARM. Ophthalmic Physiol Opt 1995; 15(6): 553-559.
2. Geruschat DR, Bittner AK, Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. Optom Vis Sci 2012; 89(9): 1308-1315.
3. 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(10): 621-637.
4. Leat SJ, Lovie-Kitchin JE. Measuring mobility performance: Experience gained in designing a mobility course. Clin Exp Optom 2006; 89(4): 215-228.
5. Marron JA, Bailey IL. Visual factors and orientation-mobility performance. Am J Optom Physiol Opt 1982; 59(5): 413-426.
6. 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(5): 279-282.
7. Wiener WR, Welsh RL, Blasch BB. Foundations of Orientation and Mobility. 3rd ed. New York, NY: AFB Press; 2010.
8. 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(1): 482-487.

# Recommendations for Assessment of Activities of Daily Living

Gary Rubin (chair), Mary Lou Jackson, Shane McSweeney, Cynthia Owsley, Robert Finger, Jill Keeffe, Sharon Bentley, Gislin Dagnelie, Joan Stelmack

## Defining Principles

The term “Activities of Daily Living” (ADL) refers to a series of self-care tasks that are essential for maintaining independence. The most common instrument for measuring ADLs is the Katz Index of Independence in Activities of Daily Living (Katz ADL (1)). 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 Katz ADL is undoubtedly the most widely used instrument to assess activities of daily living. The Lawton Instrumental Activities of Daily Living scale (IADL, (2)) extends the assessment to include eight more demanding activities such as shopping, using the telephone and managing personal finances. Curiously, the original instructions for the Lawton IADL scale suggested that women but not men be scored on food preparation, housekeeping and laundering.

The Katz ADL and the Lawton IADL scales fall somewhere between traditional patient reported outcomes (PROs), otherwise known as questionnaires, and performance-based tests (PBTs). The usual method of administration 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. For example the laundry item on the Lawton IADL 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 (eating and dressing) to cognitively demanding tasks (reading and problem solving)(3). Since PROs are being evaluated by a different group, we will limit our discussion to performance-based tests (PBTs). For a comprehensive review of performance-based, vision-related ADL instruments, see Warrian (4)

When designing a new ADL instrument the first question that needs to be answered is what tasks or activities should be included. Many ADLs (in the generic sense of the term as opposed to the Katz ADL Scale) do not require vision. Feeding, spoken communication, memory and problem solving require little if any visual input. On the other hand, some tasks included in the list of ADLs and IADL are very dependent on vision. It may be useful to include both visually dependent and vision independent tasks to help provide an assessment of the participant’s level of adjustment to vision loss.

Very little work has been done to adapt ADL measures for testing patients with ultra low vision. [but see below] The typical response is to simplify the visual requirements of the task by increasing the size and contrast of visual stimuli. For example, the patient is asked to walk along a bold white stripe on a black floor, or to find a black door on a white wall)(5). It is understandable that the test developer wants a test that patients can accomplish. But when the task is modified for the patient with ULV it may no longer represent an ADL. When navigating in the real world we seldom have a high contrast line to follow and most doors present lower contrast than black on white. If our interest is to measure some aspect of visual function like contrast sensitivity, then there are specific tests designed for that purpose – a purpose that must be distinguished from measuring the ULV patient’s ability to do everyday tasks 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.Once calibrated, these items can be used a ability scores for subjects performing the tasks, or a suitable subset. A new set of such tasks for use in ULV functional assessment has recently been developed and is currently undergoing calibration (Geruschat & Dagnelie). It is anticipated that this task battery will become available for ULV assessment in the near future.

When developing a test to measure performance on ADLs there is a tension between the need to standardize the testing conditions and the intention to preserve “ecological validity,” to insure that the test is truly representative of everyday activities. Reading tests present a good example. There are several continuous-text reading tests including MNRead (6), Radner Reading Test (7), Colenbrander Continuous Text Reading Cards, the SEE Study reading test (8)and the IReST texts (9). These tests fall along a continuum from the highly standardized and carefully controlled sentences of the Radner Reading Test, with all of the sentences having the same number of words, word length, semantic and syntactic complexity, to the SEE reading test which used paragraphs selected for grade level, but did 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.(10)

These guidelines outline the recommended methodologies for the testing of ADLs and reporting of results in ultra-low vision subjects who participate in clinical therapeutic trials.

## Recommended Methodology for Assessment of Activities of Daily Living

ADL assessment should involve evaluation of simulated everyday visual tasks, tested under standardized conditions.

General recommendations for testing include:

* 1. Testing should follow a written protocol that specifies the test apparatus and conditions (light levels, viewing distance, etc.), procedures, instructions to the participant and scoring method.
  2. Particular attention should be paid to the illumination conditions and to the luminance and contrast of the test stimuli
  3. The examiner should be qualified in the assessment of ADLs. Ophthalmologists, optometrists, orthoptists, certified ophthalmic technicians or associates, and trained research assistants are all suitable for this role.
  4. All tasks should be completed both with and without the prosthetic device turned on, preferably more than once; however, single trial administration is acceptable when administering multiple tests from an ADL battery that has previously been calibrated, e.g., using Rasch analysis. If there are multiple trials for each test these should be undertaken in a counter-balanced order (e.g. ABBA) to control for learning and fatigue effects
  5. Wherever possible, criterion free test procedures, such as forced-choice testing should be used. This is important to minimize bias due to differences between participants in their willingness to guess(11)
  6. All tests should be performed binocularly if the device permits binocular viewing. ADL evaluation aims to quantify the participant’s ability to perform everyday tasks and these tasks are normally performed with both eyes. Artificially restricting vision to one eye may be suitable for some tests of visual function, such as visual acuity, but not for everyday tasks. The same reasoning would suggest that participants be allowed to use any aids or devices they customarily use. It may be desirable to repeat testing without aids, but not to the exclusion of testing with devices; note, however, that this invalidates the calibration of the test activities, which typically has been performed without image enhancement.
  7. Task performance should be recorded in terms of both speed and accuracy. Some participants may choose to race through a task while committing numerous mistakes whereas others may perform the task more deliberately, being careful not to make any mistakes. Where possible, the scoring system should take both speed and accuracy into account. For example, reading performance is typically quantified by recording reading duration and number of words read incorrectly. Speed and accuracy can be taken into account by computing the number of correct words read per minute. It is argued that errors are as or more informative than, speed for some ADLs such as independent navigation (12). 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, since 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 the participant read the sentence correctly or not. 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, the rating scale should allow fractional or decimalized responses. It has been shown (13) that the smaller the step size, the greater the sensitivity to change

Specific ULV ADL Test Methodologies

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

1. ADL Test Batteries
   1. **FLORA (Second Sight, (14)**): The Functional Low-Vision Observer Rated Assessment (FLORA)was developed by Second Sight to evaluate the Argus II retinal prosthesis. FLORA includes both self-reported difficulty and observed performance of for a set of ADLs. In addition, the FLORA collects a narrative case summary from expert observers. So far, 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).
   2. **IADL-VLV (Robert Finger(15)**): Finger and colleagues used Delphi survey techniques to select 25 tasks from an initial set of 296 Bionic Vision Australia retinal prosthesis project. The tasks were performed by 40 participants with ULV and scored for speed and accuracy. Rasch and principal components analyses were used to evaluate the measurement properties of the. A final set of 23 tasks were deemed to have adequate measurement properties
2. Tests of a Specific ADLs
   1. **Picture Recognition (Gary Rubin(16)):** Rubin and colleagues developed a picture recognition test for Pixium Vision’s retinal prosthesis project. 100 photos were taken across five different categories, such as doorways, stairs, footpath obstacles. Each picture displayed an item from the category (e.g. a doorway) on the left or right side. 30 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 uni-dimensional difficulty scale.
   2. **Visually-guided navigation (Gary Rubin**): The navigation task was developed for a gene therapy study involving patients with Lebers Congenital Amaurosis{Bainbridge, 2015 #11;Bainbridge, 2008 #28}.The test takes place on a 75 m2 raised platform. Participants walk along a straight, unobstructed 8 m path to gauge their preferred walking speed, then negotiate a 13 m 8-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.

## 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 name of the test
* A brief description of the task and how it is related to daily activities
* A description of the visual stimuli including their size, color, luminance, contrast, and motion characteristics, if any
* Room lighting measured in lux at the participants eye.
* Viewing conditions – seated/standing, distance to target, monocular/binocular
* The maximum time allowed to complete the task and how time was measured
* Description of scoring procedure – how errors were defined, if relevant
* Randomization number and structure of trial sequence (blocked?)
* Instructions, practice and feedback
* Scoring rules 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.

## Epilogue

If the study has met the above guidelines, the following statement may be used in publications and presentations:

**“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations for Activities of Daily Living Evaluation for Ultra Low Vision”.**

If the study was unable to comply with the guidelines (for example, if the control setting of scrambled phosphenes could not be implemented due to device design), then this should be stated in the publication or presentation:

E.g. **“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations for Activities of Daily Living Evaluation for Ultra Low Vision EXCEPT for…”.**

## References

1. 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;185:914-9.

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

3. Turner-Stokes L, Nyein K, Turner-Stokes T, Gatehouse C. The UK FIM+ FAM: development and evaluation. Clinical Rehabilitation. 1999;13(4):277-87.

4. Warrian KJ, Altangerel U, Spaeth GL. Performance-based measures of visual function. Surv Ophthalmol. 2010;55(2):146-61.

5. Humayun MS, Dorn JD, da Cruz L, Dagnelie G, Sahel JA, Stanga PE, et al. Interim results from the international trial of Second Sight's visual prosthesis. Ophthalmology. 2012;119(4):779-88.

6. Mansfield JS, Legge GE, Bane MC. Psychophysics of reading. XV: Font effects in normal and low vision. Invest Ophthalmol Vis Sci. 1996;37(8):1492-501.

7. 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;213:174-81.

8. West S, Muñoz B, Rubin GS, Fried L, Schein O, Roche K, et al. Visual impairment and functional status in older persons : Methodologies for the Salisbury Eye Evaluation Project. Invest Opthalmol Vis Sci. 1995;36:S419.

9. Trauzettel-Klosinski S, Dietz K, Group IRS. Standardized assessment of reading performance: the New International Reading Speed Texts IReST. Invest Ophthalmol Vis Sci. 2012;53(9):5452-61.

10. 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(4):M209-17.

11. Vaegan, Halliday BL. A forced-choice test improves clinical contrast sensitivity testing. Br J Ophthalmol. 1982;66:477-91.

12. Warrian KJ, Katz LJ, Myers JS, Moster MR, Pro MJ, Wizov SS, et al. A comparison of methods used to evaluate mobility performance in the visually impaired. Br J Ophthalmol. 2015;99(1):113-8.

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

14. Geruschat DR, Flax M, Tanna N, Bianchi M, Fisher A, Goldschmidt M, et al. FLORA™: Phase I development of a functional vision assessment for prosthetic vision users. Clinical and Experimental Optometry. 2015;98(4):342-7.

15. Finger RP, McSweeney SC, Deverell L, O'Hare F, Bentley SA, Luu CD, et al. Developing an Instrumental Activities of Daily Living Tool as Part of the Low Vision Assessment of Daily Activities ProtocolDeveloping the IADL-VLV. Investigative ophthalmology & visual science. 2014;55(12):8458-66.

16. Gulati R, Roche H, Thayaparan K, Hornig R, Rubin GS. The Development of a Picture Discrimination Test for People with Very Poor Vision. Investigative Ophthalmology & Visual Science. 2011;52(14):1197-.

# Recommendations for Assessment of Patient Reported Outcomes

Gislin Dagnelie (chair), Akos Kusnyerik, Anne Ripley, Collette Mann, Eberhart Zrenner, Eduardo Fernandez, Francis Lane, Gary Rubin, Katarina Stingl, Lotfi Merabet, Robert Finger, Takashi Fujikado, Vernon Odom

## Defining Principles

The term Patient-Reported Outcomes, while 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, typically using rating scales to capture the effect of vision loss or a change in well-being or vision status reported by the patient. In this document we will concentrate on this type of instrument; complementary techniques will be touched upon briefly at the end of the document.

Vision-related PRO questionnaires typically explore one or both of the following aspects of vision loss: quality of life (QoL) or visual outcomes. In the case of QoL, the instrument may be designed to explore changes related specifically to vision loss (we refer to these as VRQoL instruments) or to health in general (HRQoL instruments); here we will concentrate on vision-related instruments, but some addess emotional or other health aspects and should thus be referred to as HRQoLs.

In the case of visual outcomes, these instruments are typically referred to as visual function questionnaires (VFQ), with one particular questionnaire developed under the auspices of the National Eye Institute, the NEI VFQ, as the example most familiar to both clinicians and researchers(1). “Classical” VFQs, like many survey instruments used throughout the social and clinical sciences, have three typical features(2):

* 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 partially rely on vision.
* 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 the patient does not perform, for non-visual reasons, and therefore is unable or unwilling to judge. The ratings are usually coded numerically for analysis.
* The questionnaire items are often grouped into subscales, representing different aspects of the construct under study. In the case of VFQs, subscales may be concrete visual domains such as distance vision, near vision, driving, peripheral vision, and color vision, or more abstract domains such as visual information gathering vs. hand-eye coordination; if the concept of subscales is extended to non-visual aspects such as general health, dependency, mental health, and social function, then it is more appropriate to speak of a combined HRQoL/VFQ iinstrument. Items in these instruments may contribute to multiple subscales.

One may note that some questions in the NEI VFQ explore HRQoL as well as visual function, and this is true for a number of the questionnaires that are referred to as VFQs, as we will see below.

Until approximately 2005, VFQ data were typically analyzed as though the ratings were cardinal (values), rather than ordinal (rankings), and the respondent’s ability score was typically obtained by summing the ratings, and comparing the sum to a commonly accepted threshold score(3). The adoption of item response theory (IRT) and Rasch analysis, developed previously for use in psychology and physical rehabilitation research, has allowed the assignment of an item difficulty score to each VFQ item, and of a person ability score to each respondent, on the basis of Rasch analysis of a sufficiently large data set (typically responses from well over 100 respondents)(4). The underlying assumption of IRT is that all items can be ranked along a common “visual difficulty” scale, and all respondents along a common “visual ability” scale; that more able respondents will rate a given item as less difficult than less able ones; and that more difficult items will be ranked as more difficult than less difficult ones by any respondent, regardless of ability.

In the last 10 years a number of new VFQs have been developed, several of them in multiple languages, for use in different low vision populations. Moreover, it has been demonstrated that older questionnaire data lend themselves to re-calibration using Rasch analysis(5).

## Recommended Methodology for Patient-Reported Outcomes

As indicated above, we recommend that PROs be collected through the use of standardized VFQs rather than through an open-ended history. While good clinical practice may require the use of patient history as a tool for diagnosis and clinical decision making, abstracting quantitative information from history transcripts is labor-intensive, error-prone, and sometimes impossible. Here we will concentrate on VFQs, because a number of excellent VFQs has been developed in recent years. Suitable VFQ for low vision can now be found for a broad range of visual ability, for both children and adults, and in multiple languages.

When it comes to ULV, however, most existing VFQs are unable to capture meaningful information about differences in visual ability and possible effects of rehabilitation, because the items are geared towards visual requirements well beyond those of typical ULV+. Just like a bathroom scale is unsuited to the preparation of medication dosages in (milli)grams, so the NEI VFQ cannot distinguish between a person who can only tell whether the room light is on or off and one who can make out where the ceiling light is located.

General recommendations for the design of VFQs for ULV:

* When collecting items for the design a VFQ for an unfamiliar population, it is essential to take an inventory of visual activities relevant to members of the population, preferably through the use of focus groups and a structured breakdown of daily acitivities, e.g., the Massof Activity Inventory(6). Specific questions should be asked about how vision is useful in each activity.
* From the activities that benefit from vision, a selection should be made such that both the 4 major functional domains – detail vision, visual information gathering, wayfinding, and hand-eye coordination – and visual aspects – lighting, contrast, size, distance, color, and familiarity, etc. – are represented. Questionnaire items should then be formulated that capture vision use in those activities.
* A preliminary form of the questionnaire should be administered to a representative sample of the target population large enough to perform an initial Rasch analysis (e.g., 50 respondents). Misfitted items (underfit > 4 SDs above the mean) should be eliminated or reworded: The wording may have been ambiguous, or the visual aspects of the activity may be equivocal.
* A 2nd and possibly 3rd administration round should be used to further calibrate the items; note that a 3rd round is not necessary if the item fit reliability is close to 1.0 (typically, > 0.96).
* Once the production version of the new VFQ has been administered to a representative sample of at least 250 respondents, the item measures can be anchored and the VFQ can be used to estimate person measures for new respondents without performing a full Rasch analysis.
* Validity of the VFQ for use in a given population requires that the items span a difficulty range corresponding to the visual ability range of the population.
* The VFQ must be calibrated separately for new populations with different visual characteristics.
* The use of visual domains or subscales may no longer be meaningful in ULV, but the use of principal factor analysis can be helpful to explore whether more than one visual factor is required to account for the variance within the population.

General recommendations for the administration of VFQs:

* Both operator (in person or by phone) or self (on paper or on-line) administered versions of the instrument can be used, provided that the instructions and item presentation are as similar as possible. It is recommended to administer both versions to a subset of respondents to ascertain equivalence.
* Instructions should be explicit as to the use of vision and of low vision aids(7). 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.
* If the VFQ has anchored item measures, then administration of a properly chosen subset of items may be acceptable.

Specific PRO Instruments

Instruments that may be suitable for a ULV population are marked with \*; each instrument is also categorized as “QoL,” “VF,” or “QoL+VF”

* 1. **Activity Level of the Blind (ALB)(8)(QoL+VF)**

A questionnaire to measure the activity level of blind veterans with visual handicaps was developed in which separate components of activity are measured: "independence" and "difficulty" in performing various activities, "loss felt" in not performing the activities, and "motivation to learn" the activities. 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 useful to evaluate both what patients are specifically taught and how that training effects activities not specifically taught.

* 1. **Activity Breakdown Structure (ABS)(9)/Activity Inventory (AI)(6)(VF)**

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 response categories: “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 were: “not difficult,” “somewhat difficult,” “moderately difficult,” “very difficult,” and “impossible.” These tasks represented the visual function domains: 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 for any cause of visual disorders.

* 1. **Visual Activities Questionnaire (VAQ)(10)(VF)**

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

* 1. **Activities of Daily Vision Scale (ADVS)(11)(VF)**

The authors identified 20 visual activities and categorized them into five subscales (distance vision, near vision, glare disability, night driving, and daytime driving) that comprised the Activities of Daily Vision Scale (ADVS). For each of the 20 activities, participants could select from among 5 ordered categories that reflected 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). Subjects were 334 patients scheduled for cataract extraction. Reliability and validity (content, criterion) was assessed. The authors conclude the ADVS is a reliable and valid measure of patient's perception of visual functional impairment.

* 1. **NEI VFQ 51/25 (1, 12)(QoL+VF)**

The 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 the breadth of content in the 51-item NEI VFQ. Eligible participants had to have 1 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 0 to 100 scale, 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 for the general health rating question. The psychometric properties of the NEIVFQ-51 and NEI VFQ-25 are similar.

* 1. **LV Prasad Functional Vision Questionnaire (LVP-FVQ)(13) and LVP-FVQ II(14)(VF)**

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 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, and this was administered to 150 children with VI. 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 validate the LVP-FVQ II.

* 1. **Children’s Visual Function Questionnaire (CVFQ)(15)(QoL+VF)**

Age-specific versions of a Children’s Visual Function Questionnaire (CVFQ) were defined 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 (for example, excellent, very good, and so forth), frequency (for example, never, once in a while, and so forth), agreement (for example, strongly disagree, disagree, and so forth), and difficulty (for example, no difficulty, a little difficulty, and so forth) scales were used.

* 1. **Veterans Administration Low Vision VFQ (VALVVFQ)(16) and VALVVFQ-48/20(17) (VF)**

Veterans Affairs (VA) Low-Vision Visual Functioning Questionnaire (LVVFQ-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 and 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 VA LV VFQ-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 for clinical practice and outcomes research was evaluated. Items were eliminated from the VALVVFQ-48 to reduce redundancy and shorten the instrument. A 20-item short form of the instrument was constructed for use in low vision service delivery.

* 1. **Impact of Vision Impairment Questionnaire (IVI)(18, 19) and IVI-C(hildren)(20, 21) (QoL+VF)**

The Rasch-scaled 28-item IVI 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 visual acuity less than 6/12 or visual field deficit). Identified domains include: Work and leisure, Household and personal care, Mobility, Consumer and social interaction, 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 was administered by trained interviewers to 115 people with impaired vision. Participants were asked how much their eyesight deficiency had interfered with an activity “in the past month.” Responses to items were rated from “not at all” (0), “rarely” (1), “a little” (2), “a fair amount” (3), and “a lot” (4) to “all the time” (5) and for some items, “can’t do because of eyesight” (5) or “don’t do because of other reasons” (8).

The Impact of Vision Impairment for Children (the IVI\_C) was validated as a new vision-specific pediatric instrument designed to assess the affect of impaired vision on quality of life (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 IVI\_C was subjected to Rasch analysis to assess the scale unidimensionality, measurement characteristics, response options, and targeting. 126 students with vision impairment were recruited presenting VA worse than 0.3 logMAR (i.e., 20/40) and/or a restricted visual field of **<**60°. Unlike most adult vision-related questionnaires and both the LVP-FVQ and the CVAQC, which use negative item phrasing, most of the items were positively framed to eliminate negative suggestions about students’ circumstances. All questions had a 5-point scored response, which were, “always,” 5; “almost always,” 4; “sometimes,” 3; “almost never,” 2; and “never,” 1. The IVI\_C was demonstrated as 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 rather than ability/difficulty, and this is reflected in the use of non-visual subscales in the presentation of results.

* 1. **Cardiff Visual Ability Questionnaire for Children (CVAQC)(22)(VF)**

The Cardiff Visual Ability Questionnaire for Children (CVAQC) is a short, psychometrically robust and a self-reported instrument that works to form a unidimensional scale for the assessment of the visual ability in children and young people with a visual impairment. All participants were aged between 5 and 18 years. 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 the focus groups with children and young people. This means that the instrument is highly relevant to this population focusing on the most important activities both in and out of school.

* 1. **Self-Report Assessment of Functional Visual Performance (SRAFVP)(23)(VF)**

The Self-Report Assessment of Functional Visual Performance (SRAFVP) was validated as a measure of the severity of activity of daily living (ADL) limitations in people with homonymous hemianopia (HH). Thirty adults with HH from stroke rated their level of difficulty in completing the SRAFVP. The SRAFVP consists of 38 items addressing reading, writing, communication, financial and health management, feeding, personal hygiene, dressing, clothing care, meal preparation, shopping, functional mobility, and community or social and leisure participation. Three visual subscales/domains were identified (Reading, Eye–Hand Coordination, and Functional Mobility). The ability of the person to perform each item is rated on a 3-point scale. The composite score ranges from 38 to 114; a higher score indicates more independent ADL performance.

* 1. **IVI for Very Low Vision\* (IVI-VLV)(24)(QoL+VF)**

The IVI-VLV is a measure of VRQoL in persons with ULV. It is derived from the original IVI, based on focus group discussions and participant and expert input, and has been developed with two sets of persons with ULV 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 use the same rating scale with the following four response options: Not at all, a little, some of the time, and a lot. In addition, all items have a Don’t do this for other reasons option. It has two subscales: 1. Emotional Wellbeing (EWB) and 2. Activities of Daily Living, Mobility and Safety (ADLMS). The EWB subscale contains 12 items, and the ADLMS subscale 16 items. The IVI-VLV can differentiate between different levels of VRQoL in participants, and measurement is unaffected by almost all levels of general or mental health. It meets all requirements of the Rasch model, and proposed quality criteria for health status questionnaires, such as content validity, internal consistency, reliability, no floor or ceiling effects and good interpretability.

* 1. **Ultra Low Vision Questionnaire\* (ULVQ)(25)(VF)**

150 items were developed from statements about vision use from 45 focus group members with current or prior (now blind) ULV, including 6 Argus II wearers, in response to the full Massof Activity Inventory. Items cover four functional domains (detail vision, visual information gathering, mobility, and hand-eye coordination) and visual aspects such as contrast, lighting, size, and familiarity. The ULVQ was pilot tested in a (prior) ULV/Argus population, followed by Rasch analysis and item adjustments, and re-testing in the same population. Item reliability is 0.97. Versions with 150 and 53 items are available, as well as an adaptive version, presented at ARVO 2015.

Alternate methods of PRO collection

According to our present experience, prosthetic vision can be considered a form of ULV. For example, item measures obtained with the ULVQ in a small sample of Argus II recipients did not differ significantly from those obtained in samples of current or previous ULV individuals(26). As new visual prostheses are introduced, and novel treatments such as gene therapy or stem cell-based vision restoration reach clinical application, it is conceivable that the vision experienced by recipients of such new approaches differs from ULV experienced currently by those with native or prosthetic vision. For that reason, and to keep an open mind about any new ways in which vision may be experienced that is not captured by 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 VFQs to assure 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, provide researchers with invaluable information needed for continued progress.

The working group recognizes that there should always be opportunities for open-ended reports from patients, but wishes to emphasize that such feedback should be elicited in addition to, not instead of, calibrated and validated measures of patient-reported outcomes.

A t the same time, the working group recognizes the need for internationally accepted standards of validity and calibration across individuals and treatment modalities for these patient-reported outcomes methodologies, to meet the needs of the scientific community, regulatory bodies, and health insurance companies in evaluating the safety and efficacy of each treatment.

## Reporting Guidelines

Any publication or presentation reporting the results of PRO using a VFQ should include enough information to allow replication, including:

* The name of the VFQ and, if applicable, the version
* If the VFQ has not been previously validated, amy relevant validation procedure and population information
* If the VFQ has not been previously released, an item list and response scale, including any “not applicable” category
* Method of administration and, if not standardized, instructions provided to the respondents
* Scoring rules and algorithms

## Epilogue

If the study has met the above guidelines, the following statement may be used in publications and presentations:

**“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations on Patient-Reported Outcomes for Ultra Low Vision”.**

If the study did not comply with the guidelines (for example, if the VFQ used had not been previously applied to a ULV population, as is true for the large majority of currently used VFQs), then this should be stated in the publication or presentation:

**“This study complied with the International Task-Force for Vision Restoration Outcomes (ITV) Recommendations on Patient-Reported Outcomes for Ultra Low Vision EXCEPT in the following respect(s)…”.**

## References

1. 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(7):1050-8.

2. Massof R, Rubin G. Visual Function Assessment Questionnaires. Surv of Ophthalmology. 2001;45(6):531-48.

3. Massof RW. Likert and Guttman scaling of visual function rating scale questionnaires. Ophthalmic Epidemiol. 2004;11(5):381-99.

4. Massof RW. Application of stochastic measurement models to visual function rating scale questionnaires. Ophthalmic Epidemiol. 2005;12(2):103-24.

5. Massof RW, Ahmadian L. What do different visual function questionnaires measure? Ophthalmic Epidemiol. 2007;14(4):198-204.

6. Massof RW, Ahmadian L, Grover LL, Deremeik JT, Goldstein JE, Rainey C, et al. The Activity Inventory: an adaptive visual function questionnaire. Optom Vis Sci. 2007;84(8):763-74.

7. Stelmack JA, Babcock-Parziale JL, Head DN, Wolfe GS, Fakhoury NE, Wu SM, et al. Timing and directions for administration of questionnaires affect outcomes measurement. J Rehabil Res Dev. 2006;43(6):809-16.

8. Becker SW, Lambert RW, Schulz EM, Wright BD, Burnet DL. An instrument to measure the activity level of the blind. International journal of rehabilitation research Internationale Zeitschrift fur Rehabilitationsforschung Revue internationale de recherches de readaptation. 1985;8(4):415-24. Epub 1985/01/01.

9. Massof R. A systems model for low vision rehabilitation. II. Measurement of vision disabilities. Optom Vision Sci. 1998;75(5):349-73.

10. Sloane ME, Ball K, Owsley C, Bruni JR, Roenker DL. The visual activities questionnaire: developing an instrument for assessing problems in everyday visual tasks. Noninvasive Assessment of the Visual System (OSA Techncial Digest Series); Santa Fe, NM: Optical Society of America (Washington, DC); 1992. p. 26-9.

11. Mangione CM, Phillips RS, Seddon JM, Lawrence MG, Cook EF, Dailey R, et al. Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. Medical care. 1992;30(12):1111-26. Epub 1992/12/01.

12. 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(11):1496-504. Epub 1998/11/21.

13. 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(9):4131-9.

14. 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(11):1601-10. Epub 2012/10/17.

15. Felius J, Stager DR, Sr., Berry PM, Fawcett SL, Stager DR, Jr., Salomao SR, et al. Development of an instrument to assess vision-related quality of life in young children. Am J Ophthalmol. 2004;138(3):362-72. Epub 2004/09/15.

16. Stelmack JA, Szlyk JP, Stelmack TR, Demers-Turco P, Williams RT, Moran D, et al. Psychometric properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. Invest Ophthalmol Vis Sci. 2004;45(11):3919-28.

17. Stelmack JA, Massof RW. Using the VA LV VFQ-48 and LV VFQ-20 in low vision rehabilitation. Optom Vis Sci. 2007;84(8):705-9.

18. 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(11):4732-41. Epub 2006/10/27.

19. Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. Invest Ophthalmol Vis Sci. 2002;43(4):927-35. Epub 2002/03/30.

20. 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(2):114-20. Epub 2008/04/25.

21. 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(3):1632-40. Epub 2010/12/24.

22. 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(6):730-5. Epub 2010/05/29.

23. Mennem TA, Warren M, Yuen HK. Preliminary validation of a vision-dependent activities of daily living instrument on adults with homonymous hemianopia. The American journal of occupational therapy : official publication of the American Occupational Therapy Association. 2012;66(4):478-82. Epub 2012/06/30.

24. 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(10):6150-8. Epub 2014/09/06.

25. Dagnelie G, Jeter PE, Adeyemo K, Rozanski C, Nkodo AF, Massof RW. Psychometric properties of the PLoVR ultra-low vision (ULV) questionnaire. Invest Ophthalmol Visual Sci. 2014;55(ARVO E-abstr):2150.

26. 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(ARVO E-abstr):1838.

# Recommendations for Measurement and Reporting of Electrically-Evoked Device Efficacy

Matthew Petoe (chair), Daniel Rathbun, Ethan Cohen, Ione Fine

## Defining Principles

Visual prostheses can utilize a wide range of stimulation parameters and strategies. Small changes in these parameters can have a profound effect on the quality and stability of a visual percept. Stimulation and processing parameters used in clinical trials should be communicated in such a way that clarity can be provided to the wider research community and ultimately to end-users.

A guidance document has been previously issued by the FDA, describing pre-clinical tests to characterize device safety prior to clinical testing (IDE Guidance for Retinal Prostheses, Food and Drug Administration, 2013). The present document outlines recommended methodologies for the psychophysical testing and reporting of visual performance in ultra-low vision subjects who participate in visual prosthesis clinical research trials. It is recognized that these guidelines cannot be static and must be regularly revised to reflect innovations in best practices as well as new technological developments in prosthetic devices. However these guidelines are a work in progress, and we encourage researchers in the field to help shape these guidelines as the field progresses.

## Considerations

The aim of this document is to guide the researcher in faithfully reporting the translation of technical device capabilities into behavioral and perceptible outcomes. There are many differences in visual prostheses, with regard to surgical placement, the use of externalized cameras, and the density of implanted electrodes, to name a few. The commonality is that recipients will perceive a representation of the visual environment as evoked visual percepts (termed ‘phosphenes’).

Differing Technologies

Prosthetic vision devices generally require a camera or photosensor to dynamically capture the visual scene and an encoding method to translate this input 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. Thus, for many devices, testing can be carried out either via direct stimulation of the device or under ‘naturalistic’ conditions where input is provided by the camera.

The location of the stimulating array varies between devices – targeting the retina, optic nerve, lateral geniculate nucleus (LGN), occipital lobe, or any number of sites along the visual pathway. Devices incorporating sensory substitution or augmentation may position their stimulating array on the tongue, forehead, corneal surface, or lower back, among other possibilities. The commonality is that recipients will perceive a representation of the visual environment as evoked percepts (which, for visual percepts, are termed ‘phosphenes’).

Behavioral Measures

We expect that the following basic aspects of phosphenes can be measured and reported:

1. Phosphene Threshold. An indication of the electrical charge and/or light intensity[[1]](#footnote-1) (and/or tactile force, in the case of tactile prostheses) required to produce a percept and the method used to establish these levels. Considerations include whether estimates or interpolation were used (e.g. to speed the process for high electrode counts), and whether these thresholds remain stable over time.
2. Phosphene Brightness. Whether the delivery of additional charge and/or luminance above threshold produces a brighter phosphene, and the extent to which this brightness growth is controllable.
3. Phosphene Persistence. The temporal profile of phosphenes; whether their brightness fades throughout stimulation, and whether there is an additional offset response or persistence following stimulus cessation.

Additionally, the following aspects address 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 in these aspects:

1. Phosphene Size. A measurement or indication of the typical size and extent of phosphenes, and whether size changes with additional charge and/or luminance.
2. Phosphene Position. Recipients may prefer that phosphene activity be centered in the foveal region, but this may not always be achievable, so an indication of typical offset relative to fovea is warranted. Similarly, an indication of whether the position of evoked phosphenes adhere to the geometry of the electrode layout (or whether some disorganization is inherent) and any effect of eye gaze on phosphene position.
3. Phosphene Discrimination. A corollary to size and position is the question of phosphene overlap and at what distance apart electrodes are discriminable from each other. Researchers will appreciate that there will be a spatiotemporal interaction between neighboring electrodes (Horsager et al., 2010), and so the effect of interleaved versus simultaneous stimulation on discriminability is of interest.

Finally, consideration can be given to more qualitative measures, such as shape, color, and sharpness.

## Recommendations

Stimulus Source

For devices with direct access to the stimulating electrodes, such as retinal or cortical prostheses, psychophysical experiments will be more deterministic if electrode parameters are controlled directly; temporarily bypassing any attached video camera.

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 (Dagnelie, 2008). This has been shown to be sufficient to set global parameters such as activation threshold and amplifier gain (Zrenner et al., 2011). Photodiode devices may use red or infrared light preferentially over white light, to avoid contributions from any remaining natural photoreceptors (Stingl et al., 2013, Lorach et al., 2015). To confine activity to a sub-set of electrodes, focused light can be projected directly onto photodiodes with the aid of fundus control or, less-commonly, an adaptive optics scanning laser ophthalmoscope (Zrenner et al., 2011).

Finally, for devices or researchers unable to use the above equipment, a computer screen at a calibrated distance can be used (typically 57 to 60 cm, so 1 cm on the screen corresponds to 1 degree in the visual field). Light detection, basic temporal resolution, object localization and movement detection have been trialed in photodiode devices using the ‘basic light and motion test’, BaLM (Bach et al., 2010) presented using a computer screen.

Room lighting should be uniform, controllable, and slightly dimmed (e.g. 100 to 300 Lux) to avoid contributions from residual bare light perception.

Stimulus Parameters

Functional electrical stimulation is likely to be mediated by pulse trains, which generally require lower stimulation thresholds than single-pulse stimulation (Horsager et al., 2011). For direct-stimulation devices, the evoking stimulus should be a pulse train, 0.5 – 1 second in duration, for each of the behavioural measures in these guidelines. Thresholds to a single pulse may also be reported, since these measures are more directly comparable across studies.

Inter-trial intervals no less than 2 seconds should occur between presentations, to mitigate perceptual fading observed at higher presentation rates (Horsager et al., 2009). Intra-trial intervals (e.g. for two-interval, forced-choice) should be at least 1 second, unless temporal interactions are being specifically explored.

A description of the pulse waveforms should include leading polarity (e.g. cathodic-first) and timing parameters. A full list of reportable parameters is in Appendix A.

Auditory Cueing

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

Fixation Monitoring

Many of the psychophysical tests will require good fixation, owing to relationship between percept and eye position. Thus, fixation should be monitored, preferably using an eye-tracker incorporating an eye-facing camera. When this option is not available, careful observation of the eye position during the psychophysical task is recommended. In the case of deficient ocular motility (e.g. nystagmus), participants may use a tactile fixation target to reduce gaze eccentricity.

Minimum reporting should indicate whether subjects used eccentric viewing and whether eye movements were observed.

Phosphene Threshold

The methods of obtaining a perceptual threshold for each electrode are likely to differ between groups, but it is important to acknowledge that thresholds may vary over time, even within a single test session (Humayun et al., 2003, Velikay-Parel et al., 2013). An ideal method would be robust and repeatable, balancing rapid convergence with fault tolerance, to monitor the basic functionality of the implanted electrodes (Dagnelie, 2008).

A staircased procedure, in which stimulus intensity is increased for incorrect responses and decreased for correct responses (or a series of consecutively correct responses), is a preferred implementation of a Threshold procedure. Decreasing stimulus intensity only for consecutively correct responses increases the reliability of detection. For example, the 1-up/2-down variant of the Transformed Up/Down Method (Levitt, 1971) targets a detection threshold of 70.71%, assuming the threshold estimate is calculated as the average across multiple reversals.

The sensitivity index, d-prime, should be reported to allow comparison across studies, regardless of the choice of threshold method[[2]](#footnote-2). To enable this, catch trials (in which a null stimulus is presented at random intervals) should be used in 10 – 20% of the total number of trials. The stimulus strength that gives d-prime = 1 is a common definition of discrimination threshold (Klein, 2001).

In the case of current-steering (or ‘virtual electrodes’) it is not possible to use threshold values of single electrode stimulation to infer the expected threshold for paired stimulation. Paired stimulation has either a facilitating or suppressive effect depending on inter-pulse delay (Cicione et al., 2014, Horsager et al., 2011) and so new thresholds must be established at each current-steering ratio.

Phosphene Brightness

Two key psychophysical tasks relating to brightness include the brightness-rating task (to determine an effective dynamic range) and a brightness-matching task (to balance subjective brightness throughout some or all of the dynamic range). Subjects may find brightness rating difficult and unsatisfactory, and so Greenwald et al (2009) suggest that the brightness-matching task provides more reliable differentiation between brightness levels.

**Brightness Rating**

A typical brightness-balancing procedure is described by Stevens (1957), in which subjects are first presented a stimulation with an agreed reference brightness (e.g. ‘10’) and then asked to numerically rate the brightness of a second stimulation in relation to the first, e.g. they may state ‘20’ if the second stimulation appeared to be twice as bright. Test stimuli should span the previously determined safe dynamic range, presented in random order. The reference stimulus need not be provided every trial, but should be provided regularly, e.g. at the start of the session and a minimum of every five trials.

If percepts are not uniformly bright, subjects may estimate brightness based on average brightness or the brightest part of the percept, or an increase in percept size - so a repeatable criterion must be agreed on.

Perceptual dynamic range should be determined as the point at which reported brightness ratings asymptote, i.e. when subjective brightness stops increasing with increased charge delivery.

Reporting should include the dynamic range of representative electrodes on the array, e.g. foveal, para-foveal, and peripheral.

**Brightness Matching**

Brightness matching can be performed to establish the point of subjective equality (PSE) in brightness across spatially separate electrodes, or between different stimulation strategies on the same electrode.

The recommended procedure is a two-interval, forced-choice, in which each interval contains either a reference stimulus or test stimulus (in randomized order). The test stimulus should be modulated (e.g. by amplitude or frequency) using a 1-up/1-down staircase procedure based on the subjects’ report of which interval contained the brighter stimulus in the previous trial (Greenwald et al., 2009). The device setting that produces a matched brightness can be taken as the average across multiple reversals.

To produce a brightness-balanced map of image intensity to subjective brightness, the maximum brightness of the electrode with the smallest Brightness Rating (i.e. the least sensitive electrode) should be used as a starting reference. The dynamic range of more sensitive electrodes can then be attenuated to match that of the least sensitive electrode. If this procedure has been followed, then the researcher can report that brightness-balanced maps were used for subsequent vision testing.

Assuming that brightness growth follows a power law (Greenwald et al., 2009) then isobrightness curves can be established by brightness matching further supra-threshold regions of the dynamic range (e.g. at 25, 50, and 75%). If this procedure has been followed, then the researcher can report that brightness-balanced maps, with balanced brightness growth, were used for subsequent vision testing.

Phosphene Persistence

The time course of phosphene brightness should be investigated, as it is well reported that repetitive electrical stimulation without saccadic compensation can lead to brightness fading (Fornos et al., 2012, Dobelle and Mladejov.Mg, 1974).

Researchers can report indicative phosphene persistence over a 30 second presentation using continuous sampling of subjective brightness – e.g. describing the brightness profile using a joystick (Fornos et al., 2012), potentiometer, or having the subject trace the brightness with their finger (analyzed using video recording). A reportable metric is the average time taken for phosphene brightness to fade to below 70% of its initial brightness.

Phosphene Size

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

Reportable outcomes should include a description of shape, and an indication of size in degrees on major and minor axes (Nanduri et al., 2008). The effect on phosphene size and shape of modulating amplitude, frequency, or other stimulating strategy should be qualified (Nanduri et al., 2012).

Phosphene Position

Reliable and repeatable assessment of phosphene position remains problematic, although it is clear that a combination of absolute and relative phosphene position measurements give complimentary results (Stronks and Dagnelie, 2011).

One method of recording absolute phosphene position is described by Kaskhedikar et al. (2015). The researcher presents a stimulus train whilst the subject fixates their eye centrally. At stimulus cessation, the subject should be 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.

Radial distortion can be further reduced by determining relative phosphene position. Whilst monitoring the patient’s gaze, pairs of phosphenes (particularly those suspected to be perceptually clustered) should be sequentially activated and the patient instructed to describe their relative position. Dagnelie (2008) proposes a clock hour system with 37 presentation points, 1 centrally 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 8 directions of a compass (Dagnelie, 2008).

It is possible to use the electrode layout as a basis for vision processing, if preliminary assessment indicates phosphene position is retinotopic. In this event, the researcher should state that electrode geometry was used as a basis for image sampling in subsequent vision testing.

Phosphene Discrimination

Discrimination of individual phosphenes is affected by both spatial and temporal interactions (Horsager et al., 2010).

To evaluate temporal interactions, the recommended procedure is a two-interval, forced-choice, in which each interval contains either one or two phosphenes. The time interval between the paired phosphenes is to be modulated using a 1-up/1-down staircase procedure based on whether the subject can correctly report which interval contains two phosphenes. For a range of supra-threshold intensities, the researcher should report the minimum time interval at which two phosphenes can be discriminated.

A similar technique should be used to evaluate spatial interactions between electrodes at or near the foveal region. Two intensities should be examined: near-threshold, and supra-threshold. Researchers should report the minimum electrode spacing required to ensure d-prime = 1 on a one versus two phosphene discrimination task.

## Appendix A - Specific Device Parameters

There is a wide range of parameters that can vary between 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 as many of these as is practical - as a means to which other researchers in the field can replicate or improve the parameter selection so that the efficacy of a device can be understood.

The US Food and Drug Administration have prepared a guidance document (IDE Guidance for Retinal Prostheses, Food and Drug Administration, 2013) that may serve as a basis for reporting, although some of the parameters are more concerned with pre-clinical risk assessment or device safety. The electrical specifications identified to be of relevance to the current document are:

Electrode Specifications

1. Dimensions
2. Number and spacing,
3. Material composition,
4. Description of macro-, micro-, and nano-geometry (planar, rounded, textured, etc.),
5. Surface area,
6. Coatings/Treatment, and
7. Surgical placement and anatomical position relative to the fovea, optic nerve head, optic nerve, LGN, or visual cortex, as applicable, with special attention paid also 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 of both the current and voltage waveforms delivered by each pulse,
3. Whether the stimulation is monopolar, bipolar, or some other configuration,
4. The charge/phase delivered,
5. The pulse charge density in mC/cm2 per phase,
6. The pulse sequence and polarities, for example, monophasic or biphasic, and cathodic-first or anodic-first,
7. The number of current or voltage sources, and 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. The frequencies of pulse or pulse train stimulation,
9. For pulse trains: intra- and inter-pulse intervals,
10. The duration/phase of the pulses/pulse trains, and whether the second phase is symmetrical to the first. If an atypical waveform is described, then an example recording should be provided,
11. If waveforms deviate beyond linear scaling as pulse amplitude changes, these waveforms should also be shown,
12. The compliance limit of the stimulator device and the maximal voltage delivered per pulse and any clipping protocols that deviate from a simple compliance limit,
13. The charge recovery method: Whether the pulses are capacitively coupled, charge-balanced or asymmetric,
14. The configuration of unused electrodes during stimulation – whether they are shorted or floating,
15. The impedance of the electrodes, and how this was measured,
16. The leakage resistance of the electrodes to the stimulator case, if applicable.

## References

BACH, M., WILKE, M., WILHELM, B., ZRENNER, E. & WILKE, R. 2010. Basic Quantitative Assessment of Visual Performance in Patients with Very Low Vision. *Investigative Ophthalmology & Visual Science,* 51**,** 1255-1260.

CICIONE, R., FALLON, J. B., RATHBONE, G. D., WILLIAMS, C. E. & SHIVDASANI, M. N. 2014. Spatiotemporal Interactions in the Visual Cortex Following Paired Electrical Stimulation of the Retina. *Investigative Ophthalmology & Visual Science,* 55**,** 7726-7738.

DAGNELIE, G. 2008. Psychophysical Evaluation for Visual Prosthesis. *Annual Review of Biomedical Engineering,* 10**,** 339-368.

DOBELLE, W. H. & MLADEJOV.MG 1974. Phosphenes Produced by Electrical-Stimulation of Human Occipital Cortex, and Their Application to Development of a Prosthesis for Blind. *Journal of Physiology-London,* 243**,** 553-&.

FINE, I. & JACOBS, R. A. 2002. Comparing perceptual learning across tasks: A review. *Journal of Vision,* 2**,** 190-203.

FOOD AND DRUG ADMINISTRATION. 2013. *Investigational Device Exemption Guidance for Retinal Prostheses* [Online]. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm341954.htm [Accessed May 2016].

FORNOS, A. P., SOMMERHALDER, J., DA CRUZ, L., SAHEL, J. A., MOHAND-SAID, S., HAFEZI, F. & PELIZZONE, M. 2012. Temporal Properties of Visual Perception on Electrical Stimulation of the Retina. *Investigative Ophthalmology & Visual Science,* 53**,** 2720-2731.

GREENWALD, S. H., HORSAGER, A., HUMAYUN, M. S., GREENBERG, R. J., MCMAHON, M. J. & FINE, I. 2009. Brightness as a Function of Current Amplitude in Human Retinal Electrical Stimulation. *Investigative Ophthalmology & Visual Science,* 50**,** 5017-5025.

HORSAGER, A., BOYNTON, G. M., GREENBERG, R. J. & FINE, I. 2011. Temporal Interactions during Paired-Electrode Stimulation in Two Retinal Prosthesis Subjects. *Investigative Ophthalmology & Visual Science,* 52**,** 549-557.

HORSAGER, A., GREENBERG, R. J. & FINE, I. 2010. Spatiotemporal Interactions in Retinal Prosthesis Subjects. *Investigative Ophthalmology & Visual Science,* 51**,** 1223-1233.

HORSAGER, A., GREENWALD, S. H., WEILAND, J. D., HUMAYUN, M. S., GREENBERG, R. J., MCMAHON, M. J., BOYNTON, G. M. & FINE, I. 2009. Predicting Visual Sensitivity in Retinal Prosthesis Patients. *Investigative Ophthalmology & Visual Science,* 50**,** 1483-1491.

HUMAYUN, M. S., WEILAND, J. D., FUJII, G. Y., GREENBERG, R., WILLIAMSON, R., LITTLE, J., MECH, B., CIMMARUSTI, V., VAN BOEMEL, G., DAGNELIE, G. & DE JUAN, E. 2003. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Research,* 43**,** 2573-2581.

KASKHEDIKAR, G. P., YANG, L., BOUCHER, T., TROYK, P. & DAGNELIE, G. 2015. Development of Mapping Methods with Simulated Phosphenes for Implementation in Intracortical Visual Prosthesis Recipients. *Investigative Ophthalmology & Visual Science,* 56**,** 4315-4315.

KLEIN, S. A. 2001. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics,* 63**,** 1421-1455.

LEVITT, H. 1971. Transformed up–down procedures in psychophysics. *J. Acoust. Soc. Am,* 49**,** 467-477.

LORACH, H., GOETZ, G., SMITH, R., LEI, X., MANDEL, Y., KAMINS, T., MATHIESON, K., HUIE, P., HARRIS, J., SHER, A. & PALANKER, D. 2015. Photovoltaic restoration of sight with high visual acuity. *Nat Med,* 21**,** 476-482.

NANDURI, D., FINE, I., HORSAGER, A., BOYNTON, G. M., HUMAYUN, M. S., GREENBERG, R. J. & WEILAND, J. D. 2012. Frequency and Amplitude Modulation Have Different Effects on the Percepts Elicited by Retinal Stimulation. *Investigative Ophthalmology & Visual Science,* 53**,** 205-214.

NANDURI, D., HUMAYUN, M. S., GREENBERG, R. J., MCMAHON, M. J. & WEILAND, J. D. 2008. Retinal Prosthesis Phosphene Shape Analysis. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-8***,** 1785-1788.

STEVENS, S. S. 1957. On the Psychophysical Law. *Psychological Review,* 64**,** 153-181.

STINGL, K., BACH, M., BARTZ-SCHMIDT, K. U., BRAUN, A., BRUCKMANN, A., GEKELER, F., GREPPMAIER, U., HORTDORFER, G., KUSNYERIK, A., PETERS, T., WILHELM, B., WILKE, R. & ZRENNER, E. 2013. Safety and efficacy of subretinal visual implants in humans: methodological aspects. *Clinical and Experimental Optometry,* 96**,** 4-13.

STRONKS, H. C. & DAGNELIE, G. 2011. Phosphene Mapping Techniques for Visual Prostheses. *Visual Prosthetics: Physiology, Bioengineering, Rehabilitation***,** 367-383.

VELIKAY-PAREL, M., IVASTINOVIC, D., GEORGI, T., RICHARD, G. & HORNIG, R. 2013. A test method for quantification of stimulus-induced depression effects on perceptual threshold in epiretinal prosthesis. *Acta Ophthalmologica,* 91**,** E595-E602.

ZRENNER, E., BARTZ-SCHMIDT, K. U., BENAV, H., BESCH, D., BRUCKMANN, A., GABEL, V. P., GEKELER, F., GREPPMAIER, U., HARSCHER, A., KIBBEL, S., KOCH, J., KUSNYERIK, A., PETERS, T., STINGL, K., SACHS, H., STETT, A., SZURMAN, P., WILHELM, B. & WILKE, R. 2011. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proceedings of the Royal Society B-Biological Sciences,* 278**,** 1489-1497.

# Recommendations for Implementation and Reporting of Vision Processing Algorithms

Chris McCarthy (chair), Vincent Bismuth

## Preamble

Fundamentally, vision processing in visual prosthetic devices is concerned with the transfer of information from scene to display. In its simplest form, data is continuously sampled from camera-captured images and passed to a device-specific, viewer-calibrated display system. Increasingly, however, more sophisticated vision processing techniques designed to maximize the throughput of information are being employed; often targeting specific functional outcomes. Vision processing schemes currently range from intermediate filtering processes seeking to accentuate task-relevant features in the incoming data stream, to more advanced scene augmentation algorithms that alter or entirely replace sampled values with alternative encodings of scene structure, or symbolic representations. Driving these algorithms are fast-paced improvements in sensing technologies, computing capacity and power resources, all of which can be swapped or upgraded with relative ease. These continuing advances in vision processing technologies offer exciting possibilities for device wearers. However, they also present 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 clinical publications. To be effective and unambiguous, such guidelines must be as prescriptive and specific as possible. However, to be meaningful and relevant, such guidelines must also encompass the inherent flexibility of vision processing choices available, in terms of hardware and software, as well as future, as yet unseen, future devices. To be practicable, these guidelines must also respect the commercial sensitivities that surround aspects of many current and future vision processing systems. The following guidelines aim to strike a balance between all of these factors.

The vast array of possible vision processing technologies available to prosthetic vision devices also poses challenges for benchmarking and evaluating prosthetic vision devices in common vision function testing. While functional improvements due to vision processing should not be stifled, it is also important to gauge other aspects of the visual prosthetic device without the influence of advanced vision processing. This requires an agreed upon baseline vision processing method for controlled clinical evaluations of vision restoration devices. To this end, we outline a generally applicable scheme for performing standard image down sampling in the Benchmark Vision Processing. Coupled with the need for a “System-On” control condition is the need to ensure unbiased controls upon which to gauge System-On performance. In the final section we outline the implementation of a standard Scrambled vision processing condition as an additional control condition that preserves condition masking with patients.

## Defining principles

We regard vision processing as a conceptual layer of a vision restoration system, consisting of software and hardware. We define this layer as beginning with the input sensors (e.g., camera, inertial sensors etc.), and ending at the neural/sensory interface, where the conceptual target perception is encoded in display-specific parameters. As such, vision processing is largely agnostic to the underlying display mechanisms, though may be parameterised and configured to best utilize device-specific properties.

For the purpose of these guidelines, we divide the system into four components:

1. Input streams and capture
2. Processing and Augmentation
3. Representation
4. Interaction and dynamics

We detail each of these components below

Input streams and 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 scene, and/or the movement or location of the device wearer in the scene. In most cases, this primarily consists of a body-worn camera, however, additional sensors providing range information, acceleration, gravity, eye-tracking, among others, may also be included. This component also encapsulates all key parameter settings that determine the characteristics of input capture, such as the temporal sampling frequency of each sensor, the spatial resolution of information provided (or information throughput per unit time), and the field of view captured.

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 delivered to device users. This may include the direct sampling of input values (eg., greyscale intensities), or more complex scene understanding algorithms inferring properties about the scene. 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.

Representation

The representation component encompasses how scene information is encoded in the final display. That is, *the meaning* assigned to display elements with respect to the imaged scene. For example, standard vision processing systems employed with retinal prostheses typically encode the greyscale scene luminance as levels of perceived intensity in evoked phosphenes. However, alternative representations are possible, and indeed are being used in current vision restoration devices.

User Interaction and system dynamics

All previously described components may be changeable over time. This may be due to automatically adjusted settings responding to inferred environmental change (e.g., lighting, structure etc), or through manual intervention (from investigator, clinician or user) during operation. The system interactions of interest to these guidelines are those that impact the resulting appearance and/or meaning of the display presented to the device wearer. These interactions may be available to the investigator/clinician, the device wearer, or both. Example interactions include, but are not limited to: camera-capture settings (e.g., brightness, contrast, and exposure time), image filter/algorithm choices, contrast inversion, modality of operation, zoom control, on-demand processing, camera offset).

## Reporting Guidelines for Vision Processing

In the following section we provide detailed guidelines for reporting on each component of vision processing outlined in the previous section.

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 streams used to acquire external information (e.g., RGB or grey-scale images, inertial sensing, gravity, depth information, auditory, etc.)
2. For each input stream that contributes to the resulting display:
   1. The modality of input provided (e.g., RGB pixels, depth, acceleration, etc.).
   2. The specific model and manufacturer of the device used to capture the input stream. If the device is custom-made, or is a modified version of an existing device then this should be noted, along with a statement explaining how the device acquires the input stream.
   3. 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 this should be noted. If these settings are manually set prior to operation then a statement regarding how each setting is determined should be included.
   4. The throughput of the input stream. This includes:
      1. the dimensions of each frame of captured data,
      2. the dynamic range of the input data, and
      3. the temporal sampling frequency (e.g., frames-per-second) during normal operation of the full system .
   5. The relevant spatial (and if applicable, temporal) window from which the input stream is sampled. For example, a camera’s field of view (expressed as horizontal and vertical angles) and/or a depth sensor’s operating range (as minimum and maximum distance from the sensor).
   6. The physical position and orientation of the sensor used to acquire the input stream with respect to the viewer’s forward-facing head.

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” etc.
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 three degrees per second.”

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 (ie., 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: “Phosphenes 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. The dynamic range of each display parameter used to encode information (i.e., how many discrete levels of intensity are available to encode information?)
5. The update/refresh rate of the representation on the display

System calibration, interaction and dynamics

Reporting should address the following details about the calibration of vision processing:

1. A list of all vision processing devices (i.e., cameras etc.), processing and display settings that required calibration prior to operation, and 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 the reference object was discernible from the background”*.
2. The procedure for calibrating each display setting.

Reporting should address the following details about human interactions with the vision processing system during operation:

1. Any manually adjustable controls or modes of operation that were available:
   1. To the device wearer during trials.
   2. To the experimenter/clinician/engineer during trials.
2. The primary purpose of each adjustable setting.
   1. Any automatically adjusted controls or modes of operation that were active during operation, and the primary purpose of the automatic setting adjustment;
   2. the primary determinant of the setting adjustment or mode selection.

## Recommendations for assessing vision processing methods

The Benchmark Vision Processing Strategy

While vision processing offers significant scope for enhanced functional outcomes, it is often the case that basic measures of visual function without the confounding influence of vision processing enhancements are required (e.g., basic light localization, visual acuity, contrast acuity, motion detection, etc.). This requires the establishment of an agreed-upon *base level* vision processing strategy that provides basic pixel to stimulation transfer to support controlled and repeatable vision function testing, and is readily applicable to any vision restoration device utilizing an external digital camera and modifiable vision processing software. Below we define the components of a base level vision processing system for vision function testing, and for the benchmarking of alternative vision processing methods. We describe the benchmark system with respect to the conceptual components of vision processing outlined above.

Benchmark vision processing for lab-based, controlled-lighting conditions

* *Input streams and capture:* 
  + A single head-mounted camera providing greyscale images at a capture rate equal to or greater than the update rate of the final display.
  + The camera should be rigidly mounted and located as close to eye-level as possible.
  + The input field-of-view should match the theoretical field-of-view of the output display device. Where the field-of-view of the final display is not sensibly defined (e.g., sensory-substitution devices), then the visual angle between regular sampling points in the image should be reported).
  + If testing is to be conducted in controlled lighting conditions, all camera capture settings should be manually set for optimal performance in the conditions (and all automatic adjustments disabled). 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 should be made to minimize the range of possible values for each auto-adjusted setting. All camera settings effecting image capture should be reported, and if auto-adjustments remain enabled, the methods and procedures used to minimise the impact of this should be reported.
* *Processing*:
  + The use of spatial down sampling filters is standard in image processing and should be included as part of a benchmark vision processing system to avoid the effects of aliasing. There exists a large number of filters appropriate for down sampling, each offering various trade-offs between spatial frequency cut-offs and the effects of aliasing. For the purposes of a benchmark vision processing system, acceptable down sampling filters are those that attempt *only to approximate the original* *signal* using only the sampled values (e.g., Regional Averaging, Gaussian filtering, Lanczos filtering). It is recommended that thee spatial size of the filter window is set to ensure the frequency cut-off of the filter is as close to the Nyquist band limit as possible, thus ensuring optimal reconstruction of the signal with minimal aliasing effects. This size will depend on the filter chosen, and the distance between display elements.
* Representation:
  + The final display should only convey scene luminance information captured in the greyscale input.
  + For tasks involving “black-and-white” stimuli only, viewer-specific adjustments of capture and display settings should be made to ensure maximum perceived contrast in the output display.
  + For tasks involving the distinction of multiple grey levels (e.g., contrast acuity), then lowest and highest input stimuli brightness levels should be mapped to appropriate brightness levels for the specific viewer, and all intermediate levels mapped to reflect a linear scale of *perceived contrast* within these upper and lower intensities. Further details and reporting guidelines for this are provided in the Recommendations *for Psychophysics* section.

Benchmark vision processing under less controlled conditions

The assessment of orientation and mobility, and other tasks of daily living are generally considered under less controlled conditions than is typically achievable for lab-based vision function testing. In this case 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.

Scrambling vision processing

The evaluation of vision function with prosthetic vision devices requires adequate control conditions upon which to compare performance. Traditionally this has been achieved through the use of the “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. These guidelines recommend the inclusion of a “Scrambled” vision processing condition when assessing basic visual function with a vision restoration device. The Scrambled condition aims to deliver a patterned display device wearers that provides little or no meaningful information, but in a way that is non-obviously less useful than the normal System-On condition. Recent clinical studies [cite] have included variants of a basic scrambling strategy, with strong evidence to support the preservation of masking from patients. We therefore draw from these studies to outline guidelines for a standardized implementation of the scrambled condition.

Implementation of Scrambled

The Scrambled condition is implemented using the same input capture and processing scheme outlined for Benchmark Vision Processing. That is:

* Image sampling and processing is performed identically to the Benchmark Vision Processing condition (as outlined above)
* All settings and configurations for image capture and processing are identical to those used for the Benchmark Vision Processing Condition.

Once sampling and processing is complete, and the resulting output values are ready to be assigned to stimulation output locations (e.g., electrodes), a random re-assignment of output values to output locations should be performed such that:

1. Output is delivered using the same number of display elements as would have been used under normal use of the Benchmark Vision Processing condition;
2. The same net per-frame stimulation is delivered as would have been delivered under normal use of the Benchmark Vision Processing condition.
3. At a minimum, the randomised redistribution of output values to display elements is re-computed at the beginning of each trial of the Scrambled condition.

The first two conditions ensure the spatial structure of the resulting display is entirely disconnected from the scene. However, the temporal structure of the stimulation pattern is preserved in order to maintain the display pattern’s plausibility with respect to deliberate head movements by the participant. Thus, the Scrambled condition does not remove all useful information, and cannot therefore be considered equivalent to System-Off. It does, however, provide a robust System-Off-like control where spatial discrimination of the delivered stimulation pattern is central to success in the task (e.g., letter recognition [cite], edge directionality [cite].The third condition alleviates any possibility of the participant learning the re-mapping. The stated recommendation of refreshing this mapping once per trial is a minimum requirement. If re-mapping of display value locations is possible during operation, then it is recommended that random re-distributions occur every five seconds.

## References

# Recommendations for Psychosocial Assessments and Ethical Considerations

Philip Troyk (chair), Francis Lane

## Defining Principles

Visual Prostheses, as a subclass of neural prostheses, pose specialized problems when defining the structure and implementation of development, clinical testing, and deployment. While the regulatory and approval infrastructures commonly used for formulating clinical trials which involve large numbers of human subjects for the testing of pharmaceuticals are well established, visual prosthetic research has typically used a more metered approach with single-subject studies being used to establish feasibility rather than formalized clinical testing. This tendency to structure a research program around single-subject studies perhaps stems from the difficulties in using animal models to investigate both safety and efficacy for emerging visual prosthesis designs. Here, we consider aspects of human participation in experimental studies of visual prostheses.

As an extension of the seminal Belmont Report (1978), Beauchamp and Childress (2001) have defined beneficence, nonmaleficence, respect for autonomy, and justice as defining principles of bioethics. Each of these can be distinctly applied to the design of visual prosthesis developments and translation to human use. However, practically they intertwine when making decisions about visual prosthesis program structures and the involvement of human volunteers. A well-intentioned desire for nonmaleficence, on the part of a project leader or medical practitioner, can collide with respect for autonomy and the right of self-determination – deciding that participation in an experimental study is too risky for a particular person can collide with the right to knowingly place oneself at risk without a forced paternalistic influence.

In contrast, the process of *informed consent* is often advanced as the basis and justification for the participation of a human volunteer in a proposed experimental or feasibility trial. Yet, owing to the complexity of emerging visual prosthesis technology, the “informed” portion of the consent process can be easily distorted such that it no longer accomplishes the goal of education, but rather is shrouded in a misplaced attempt to simplify the developmental details of what is inherently complex information. Consent without authentic education counteracts the fundamental purpose of the informed consent process. Too often the informed consent forms and protocols appear to provide more protection to the sponsoring institution, rather than to the volunteer (Bhutta, 2004; Lane, Nitsch, Huyck, Troyk, &, Schug, 2014).

## Considerations

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

established first before involving human test subjects.

One rationale is that narrowing the technical approaches to those suitable for a safe implantable system before involving potential recipients avoids confusion. However, an equally compelling rationale is that decisions made about how to shape the technology during the earliest stages of development can substantially benefit from the perspective of the 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 avoid disappointments in the later stages of system deployment (Lane, Huyck, Troyk, 2011; Lane, Huyck, Troyk, & Schug, 2012; Lane, Nitsch, Huyck, Troyk, & Schug, 2014). As an example, potential recipients of a cortical visual prosthesis, interviewed through focus groups, identified the cosmetic nature of the extracorporeal hardware to be a notable factor that would influence their decision to participate in a future trial (Lane, Huyck, Troyk, 2011).

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

1. Restoration of vision

Perhaps the most obvious, and potentially dangerous, motivator for experimental trial participation is the expectation of regaining visual function. To elucidate the influence of this motivator, a series of focus groups were conducted to determine the threshold of functionality that an individual’s visual prosthesis would have to provide. 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 space orientation and navigation (Lane, Huyck, & Troyk, 2011; Lane, Huyck, Troyk, & Schug, 2012). However, other examples provided by the group participants open to question their extent of understanding about the quality of the restored vision that might be achieved. In this regard, predictions about the utility of emerging visual prosthesis systems may not be answerable until they are tested on humans. This uncertainty contrasts the principles of beneficence and nonmaleficence.

2. Altruism

Some studies have concluded that the strongest psychological benefit from clinical trial participation is altruism (Seelig & Dobelle, 2001; Seelig & Rosof, 2001). Altruism has repeatedly emerged in focus group studies (Lane, Huyck, & Troyk, 2011; Lane, Huyck, Troyk, & Schug, 2012) and as retrospectively reported by some cortical visual prosthesis recipients (Lane, 2012 September; Lane, Nitsch, Huyck, Troyk, 2013, May; Lane, Nitsch & Troyk, 2015). In these studies it was often expressed by potential and actual visual prosthesis recipients to advance the field of vision restoration owing to their concern for other individuals who are blind. Although the expression of altruism can seem compelling, altruism can take various forms, including more than one form of altruism that is considered pathological. If the selection of a trial participant is based upon their altruism, an informed assessment by a trained clinician must be made.

3. Adventurism

The prospects for excitement and trail-blazing in the form of adventurism can also be a significant motivator for those considering participation in an experimental clinical trial. Altman (1986) proposed that self- experimentation by physicians or other scientific researchers may 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 (Kilgore et al., 2001) reported this factor for recipients of spinal cord implants, while Lane and colleagues (Lane, Huyck, & Troyk, 2011; Lane, Huyck, Troyk, & Schug, 2012; Lane, Nitsch, Huyck, Troyk, & Schug, 2014) have found similar trends for visual prostheses. A danger of relying upon adventurism as a qualifying motivation for trial participation is that risks, known to both the researchers and potential participant, can be dismissed using the principle of autonomy and self-determination as overriding compensations. This 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 participant’s motivations, then it becomes too easy for the researcher to dismiss the necessary balance between risks, safety, and efficacy.

4. Decision Making

The 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 culmination of a participant’s decision to be a recipient of an experimental visual prosthesis. The factors used by the decision maker can range from being deeply personal, to having strong family or friend influences, to trusting researchers or health practitioners, or may include clerical influences. Any of these factors can dominate such that the final decision can ignore, or embrace, strong pressures from others, sometimes contradicting personal desires. The nature of the informed consent process may be too influenced by a western view of autonomy as expressed by respect for independent decision-making. In non-western societies, family and even community leaders may be directly involved in the decision-making process (Marshall, et al, 2006). When using an informed consent process to facilitate the individual’s decision making, caution should be exercised about whether the process is designed for institution legal protection or as an authentic aid to the participant’s decision to participate through an emphasis upon education.

5. Managing Expectations

Managing the expectations of volunteers before, during, and after the experimental trial is an essential component of a 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. Assessment and weighing of those motivations, and appropriate structuring of the informed consent process are the primary means of assisting the participant in managing 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, or even involving the participant at an early stage of the prosthesis development, misunderstanding of the nature or capabilities of the technology may be present as the trial progresses.

## Recommendations

The motivating factors, decision-making by prospective participants and the management of expectations are important considerations for vision trials. However, to focus on any one at the exclusion of another can have a negative impact on a participant. Presented differently, each of the factors discussed here must be regarded as equally important with particular attention paid to their interaction. In addition to those specific to visual prosthesis trials, 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 prosthesis trial.

Components of a comprehensive assessment:

1. A trained mental health practitioner must be involved in the initial screening and admission decision-making of all prospective participants in an experimental clinical trial and conduct a comprehensive mental health screening. The evaluation must involve a multi-parametric assessment of an individual’s intellectual functioning, including her or his capacity for processing complex information and utilizing that information for decision-making. In addition, there must also be a comprehensive assessment of the individual’s personality, emotional functioning and presence or absence of pathology.
2. Specific to a visual prosthesis trial, the trained mental health practitioner should evaluate factors peculiar to the nature of the system to be tested, including the nature of the individual’s motivation to participate in a vision trial, expectations for vision restoration, and the decision-making process being used to through the incorporation of complex information.
3. The comprehensive mental health screening must include an assessment of factors relevant to the potential participant’s adjustment to blindness and the current quality of life being experienced as an person with blindness. The term “adjustment” should be regarded as contextual in nature so each of the multitude of considerations described above is interpreted within the larger context of their daily lifestyle. Adjustment should also be considered in the present stemming from how well the individual adjusted to their loss of vision in the past. Owing to the structure of some visual prosthesis trials, visual sensation can be restored and lost within a trial, and how an individual coped with vision loss in the past should be considered.
4. A trained mental health practitioner must be involved in the ongoing 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 information. The mental health individual may also be allowed to function in the capacity of a participant advocate. While this should not be forced on an individual, it should be made available to participants if and when they choose to utilize such services.
5. Involvement of the potential participant in an early stage of the trial planning can be significant with respect to the process of informed consent. Genuine understanding of the nature and significance of complex technical information related to the trial risks, and potential benefits, is essential to creating the authenticity of the informed consent process. Not being merely a matter of intellectual understanding, assimilation of how the participation in an experimental trial will affect the participants life, and change their quality of life, is essential to the participants’ well-being, and perhaps technical success of the trial. In this regard, one model that might be considered is that used by NASA for the selection of astronauts in which multiple potential participants were formed as a group at an early stage in the technological development. The group approach allows for emotional support, enhancement of understanding complex information, and facilitation of trial participant selection. A mental health practitioner can provide on-going group facilitation which may provide a notable support structure for actual trial participants.
6. Considering and exploring every factor that may be grounds for inclusion, or exclusion, from a vision prosthesis trial is beyond the scope of this document. As trials progress, around the world, sharing of information about the course of potential, and actual, participants’ mental health well-being will be an important guide for the field. In this regard, the suggested considerations and recommendations presented here should be viewed as formative, and not prescriptive. While it is too early for standardized measures of subject selection and participant assessment to be defined, application of the emerging fundamental principles, presented here, can form the basis for sharing of psychosocial outcomes from visual prosthesis trials, towards the goal of those and future systems improving the quality of life for those with blindness.

## References

1. United States. (1978). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. Bethesda, Md.: The Commission.
2. Beauchamp, T. L. & Childress, J. F. (2001). *Principles of biomedical ethics, 5th ed.* Oxford: Oxford University Press.
3. Bhutta, Z. A. (2004). Beyond informed consent. *Bulletin of the World Health organization, 82*, 771-778.
4. Lane, F. J., Huyck, M. H, Troyk, P. R. & Schug, K. (2012) Responses of potential users to the intracortical visual prosthesis: Final themes from the analysis of focus group data. *Disability and Rehabilitation: Assistive Technology*, *7*, 304-313.
5. Lane, F. J, Huyck, M. H., & Troyk, P. R. (2011) Planning for the first human intracortical visual prosthesis by using pilot data from focus groups of potential users. *Disability and Rehabilitation: Assistive Technology, 6*, 139-147.
6. Lane, F. J., Nitsch, K. P., Huyck, M. H., & Troyk, P. R. (2013, May). *Participants’ experience in a clinical trial for vision restoration: Motivation to participate, visual perception and functional use, and experience of loss following termination*. Poster presented at the annual conference of the Association for Research in Vision and Opthamology (ARVO), Seattle, Washington, May 8, 2013.
7. Lane, F. J. , Nitsch, K. P,, Huyck, M. , Troyk, P. R., &, Schug, K. (2014) Perspectives of optic nerve prostheses,” *Disability and Rehabilitation: Assistive Technology*: 1–9.
8. Lane, F. J. (2014 September). *Analysis and results of interviews of recipients of a cortical visual prosthesis.* Paper presented at the Eye and the Chip, World Congress on Artificial Vision, Detroit, Michigan, September 30, 2014.
9. Lane, F. J., & Troyk, P. R. (2010, September). *Methods and results from interviews of prospective candidates and recipients of visual prostheses: What really matters?* Invited
10. presentation to the 10th Annual Eye and the Chip: World Congress on Artificial Vision, Detroit, Michigan, September 13, 2010.
11. Lane, F. J., Nitsch, K., Troyk, P. R. (2015) *Participant perspectives from a cortical vision implant study: Ethical and psychological implications.* Proceeds from the 7th international conference on Neural Engineering, Montpellier, France, April 22, 2015.
12. Lane, F.J. (2012, September). *Methods and results from interviews of eleven recipients of a visual cortex implant: An analysis of their experiences.* Paper presented at the Eye and the Chip, World Congress on Artificial Vision, Detroit, Michigan, September 11, 2012.
13. Lane, F.J., Nitsch, K.P., Huyck, M.H., Troyk, P.R. (2013, May). *Participants’ experiences in a clinical trial for vision restoration: Motivation to participate, visual perception and functional use, and experience of loss following terminaton.* Poster presented at the annual conference of the Association for Research in Vision and Opthamology (ARVO), Seattle, Washington, May 8, 2013.
14. Seelig, B. J., & Dobelle, W. H. (2001). Altruism and the volunteer: Psychological benefits from participating as a research subject. *ASAIO Journal 47*(1), 3-5.
15. Seelig, B. J., & Rosof, L. S. (2001). Normal and pathological altruism. *Journal of the American Psychoanalytic Association, 49*, 933-959.

1. Typically luminance (cd/mm2) or irradiance (mW/mm2) [↑](#footnote-ref-1)
2. Sample Matlab code for calculating d-prime in psychometric functions can be found as an appendix in FINE, I. & JACOBS, R. A. 2002. Comparing perceptual learning across tasks: A review. *Journal of Vision,* 2**,** 190-203. [↑](#footnote-ref-2)