Reliability and Consistency of Dark-Adapted Psychophysical Measures in Advanced Eye Disease

Ava K. Kiser, Derek Mladenovich, Fariba Eshraghi, Debra Bourdeau, and Gislin Dagnelie

PURPOSE. To establish reliable vision measures allowing functional monitoring in patients with severe vision loss.

METHODS. Legally blind and normally sighted subjects were enrolled in a repeated-measures study to determine the reproducibility of psychophysical vision measures under scotopic conditions. The tests included dark adaptometry, dark-adapted, full-field flash testing, and dark-adapted macular thresholds, obtained by using a perimeter with 500- and 650-nm targets. Two to five test repetitions were performed on the better eye of each subject at monthly intervals. Subject groups included retinitis pigmentosa (RP; n = 33), macular disease (MD; n = 14), optic nerve disease (ON; n = 4), diabetic retinopathy (DR; n = 5), and other retinal diseases (OR; n = 9), and normally sighted control (CTL; n = 12).

RESULTS. Dark adaptometry timing yielded mean coefficients of variation for subjects across all groups that averaged approximately 20% throughout the test. For dark-adapted perimetry, the coefficients of repeatability (CR.95) were <7 dB (CTL), <8 dB (OR), <6 dB (ON and RP), and <15 dB (MD). Full-field flash test CR.95 by group varied from 5 to 15 dB, and most low-vision groups performed more reliably than CTL subjects.

CONCLUSIONS. Dark-adapted psychophysical tests can provide reproducible vision measures in subjects with severe visual impairments, and these tests would be useful in monitoring outcomes in future clinical trials to reverse, halt, or slow vision loss. The most valuable measure of remaining vision was the dark-adapted, full-field flash test, as it produced repeatable results at all levels of vision loss and for all disease states included in this study. (Invest Ophthalmol Vis Sci. 2006;47: 444–452) DOI:10.1167/iovs.04-1146

Attempts to arrest, slow, prevent, or reverse blinding eye diseases are expanding in nature and number. Some of the more mainstream approaches involve pharmaceutical treatments, pharmacological treatments, gene therapy, macular translocation surgery, retinal prosthetic implants, photodynamic therapy for choroidal neovascularization, and stem or fetal cell transplantation. Given the possible high-risk nature of some therapies, subjects with advanced vision loss are likely to be enrolled in the early phase of clinical trials. In conventional vision tests, subjects with very low levels of vision may demonstrate little to no functional response to these interventions. To compare visual outcomes across therapies and centers, a new and uniform test battery is needed to determine functional vision changes in subjects receiving such treatments. These vision tests must accommodate subjects with visual acuities well below 20/200, small islands of remaining functional retina, and limited ability to fixate and track (even bright) targets. We report on three global or local measures under dark-adapted conditions, which were used to analyze retinal integrity, including rod versus cone function. These measures were collected as part of a project in which we sought to design and validate such a vision test battery. Results from other tests in this same population are presented elsewhere.

Previous studies have evaluated the repeatability of visual acuity and contrast sensitivity measures in both normally sighted and visually impaired individuals, but to our knowledge there have been no reports evaluating the reliability of dark-adapted psychophysical measures. It has been commonly accepted in clinical practice that patients with very low levels of vision produce results that are more variable than those with normal vision, and this was indeed confirmed in a group of legally blind subjects during a parallel study of measures of visual acuity and contrast sensitivity. Nonetheless, these subjects were capable of producing reliable vision measures, and it therefore seems worthwhile to study test reliability in other areas of vision, such as dark-adapted psychophysics. Given the increasing prevalence of adventitious severe vision loss, the determination of repeatability of psychophysical vision measures in legally blind subjects and the availability of a standardized test protocol may be of great help in any future intervention trials for a wide range of ocular diseases.

METHODS

The protocol for the study was approved by the Institutional Review Board (IRB) of the Johns Hopkins University School of Medicine and followed the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study.

Subjects

The results presented in this study were obtained from the better eye in 77 subjects (40 women and 37 men). The better eye was determined based on the results of Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity (Lighthouse International, New York, NY) and Pelli-Robson letter contrast sensitivity (Metropia, Ltd., Essex, UK) testing. All subjects with low vision were legally blind, defined by either best corrected visual acuity <20/200 or visual field diameter <20°. The subjects were divided into 10 different groups, defined on the basis of the type of eye disease and severity of vision loss. Subjects were grouped by the following diagnoses: optic neuropathy (ON), diabetic retinopathy (DR), retinitis pigmentosa (RP), macular degeneration (MD), including Stargardt dystrophy and exudative or nonexudative age-related macular degeneration, and other retinopathies (OR). Table 1 gives the number of eyes and subjects tested in each group, including assignments to subgroups on the basis of visual acuity. RP-IV subjects could not read any letters on the ETDRS chart at 0.5 m, and their acuity level would be judged as “hand motions” or “light perception” by most clinicians. The causes of ON include toxic, traumatic, and Leber’s congenital ON. The “other retinal disease” (OR) group included one subject each with one of the following diagnoses: albinism, sickle cell retinopathy, retinal detachments, congenital aniridia, pathologic myo-

From the Lions Vision Center, Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland.

Supported by a center grant module from the Foundation Fighting Blindness.

Submitted for publication September 27, 2004; revised February 23, May 24, July 7, and August 2, 2005; accepted November 15, 2005.

Disclosure: A.K. Kiser, None; D. Mladenovich, None; F. Eshraghi, None; D. Bourdeau, None; G. Dagnelie, None

The publication costs of this article were defrayed in part by charge payment. This article must therefore be marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Ava K. Kiser, Lions Vision Center, 550 N. Broadway, 6th Floor, Baltimore, MD 21205-2020; abittne1@jhmi.edu.
Reported elsewhere.15 The set order of the tests that were performed during these visits. Results of these tests are presented in Table 1. Subject Data.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subjects (n)</th>
<th>Female (n)</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-I</td>
<td>8</td>
<td>6</td>
<td>&gt;20/40</td>
</tr>
<tr>
<td>RP-II</td>
<td>8</td>
<td>4</td>
<td>20/40 to 20/199</td>
</tr>
<tr>
<td>RP-III</td>
<td>12</td>
<td>6</td>
<td>20/200 to 20/1000</td>
</tr>
<tr>
<td>RP-IV</td>
<td>5</td>
<td>2</td>
<td>&lt;2/1000</td>
</tr>
<tr>
<td>MD-I</td>
<td>12</td>
<td>6</td>
<td>20/200 to 20/500</td>
</tr>
<tr>
<td>MD-II</td>
<td>2</td>
<td>2</td>
<td>&lt;2/500</td>
</tr>
<tr>
<td>ON</td>
<td>4</td>
<td>1</td>
<td>&lt;2/200</td>
</tr>
<tr>
<td>OR</td>
<td>9</td>
<td>5</td>
<td>&lt;2/200</td>
</tr>
<tr>
<td>DR</td>
<td>5</td>
<td>2</td>
<td>&lt;2/200</td>
</tr>
<tr>
<td>CTL</td>
<td>12</td>
<td>6</td>
<td>&gt;20/25</td>
</tr>
</tbody>
</table>

Total of 77 subjects.

Psychophysical Measures in Advanced Eye Disease

Test Procedures

After light-adapted testing, subjects’ eyes were dilated with 1% tropicamide and 2.5% phenylephrine hydrochloride ophthalmic solutions, and the subjects were given a break of at least 15 minutes before dark-adaptation testing. This test was performed on the subject’s better seeing eye only, while the other eye was covered with a black patch. A standardized procedure using a scotopic sensitivity test model SST-1; LKC Technologies, Gaithersburg, MD) was used to measure the dark adaptation curve by recording the time needed to perceive lights of decreasing intensity. First, the subject was light adapted in the SST monocular Ganzfeld with moderate intensity (1000 cd/m²) for 60 seconds. The subject was then required to detect a faint probe light in a two-alternative, forced-choice procedure, in which the tester randomly pushed one button on the handheld device that did not display the probe light and another that did, and repeated these two choices continuously every 5 to 10 seconds. The SST presents a full-field green LED (peak wavelength = 572 nm) as the stimulus, flashed for half a second. The maximum intensity of the probe light is 0.0045 cd/m², and it can be attenuated over a range of 30 dB (3 log units). On repeated correct detection, the intensity of the light was decreased in 1-dB steps (if necessary, during the rapid initial adaptation phase, 2-dB steps were used). If the subject could not detect the probe light or if an incorrect answer was given, the two-alternative forced choices continued at the same intensity level. The test was continued until the dimmest light sensitivity provided by the sensitivity tester was seen, or for a maximum of 45 minutes if this sensitivity was not attained.

Immediately after dark adaptation was completed, subjects performed dark-adapted static macular perimetry spanning a central 6° square, in the better-seeing eye only. The test was performed on an automated perimeter (model VFA 640, Humphrey Instruments; Carl Zeiss Meditec, Inc., Dublin, CA), customized for dark-adapted testing.23 This test, requiring approximately 10 minutes per wavelength, was performed twice, using size-V stimuli: Once with 500-nm wavelength test stimuli (blue-green), followed by 650-nm wavelength stimuli (red), allowing for determination of rod versus cone thresholds, respectively. Three threshold attenuation levels (maximum 50 dB) were obtained for each of the 16 test spot locations, arranged in a 4 × 4 square with 2° spacing.

After macular perimetry, with the subject still dark adapted, a full-field flash threshold test was completed. The equipment for this test included the ERG Ganzfeld (UTAS-Epic2000; LKC) with a photograph strobe (model PS22; Grass Telefactor, West Warwick, RI) and computer-controlled neutral density filters, spanning a 48-dB range in 2-dB steps. Starting at the dimmest intensity level, a modified two-alternative, forced-choice procedure was used with single flashes presented at random intervals of 1 to 4 seconds. One of the two flashes was always presented at maximum attenuation, whereas the other flash was used to determine threshold. The modification was made to achieve rapid initial convergence to the threshold. The probe flash was presented with increasing intensity, until the subject consistently recognized the correct flash, maintained correct identification through four or more consecutive attenuation steps, and then made an error, prompting a 4-dB increase completing the initial cycle. The second cycle required four consecutive correct responses in two attenuation steps before an error prompting a single-step increase completing the second cycle. This initiated a five-cycle three-up/one-down procedure. Figure 1 shows a log of the test, with the flash intensity, the subject’s response, and reversal numbers according to the modified staircase procedure. The final threshold was estimated by fitting all a subject’s data with a Weibull function, and the final intensity threshold reported was the 82% correct inflection point of this function. In case the subject detected flashes in both intervals at the maximum (+8 dB) attenuation, a pair of welder’s goggles with two types of filters (6; Schott Glas, Mainz, Germany; and 1 ND Kodak Wratten; Eastman Corp., Rochester, NY) was provided to diminish the intensity reaching the retina by a factor of 2500 (34 dB), and 34 dB was added to the final threshold. Although this test was performed on each
eye individually, we are reporting results in the better eye only. The duration of this test procedure was approximately 10 minutes per eye.

Data Analysis

As suggested by Bland and Altman,24 we used a coefficient of repeatability (CR.95), calculated as 1.96 multiplied by the SD of the differences between test and retest data for the dark-adapted Humphrey and dark-adapted flash test results, computed individually for each subject using all test-retest combinations: $n \cdot (N - 1)/2$ values for $n$ test repetitions. CR.95 represents the one-sided test-retest confidence interval and specifies the range around the baseline threshold, outside of which a variable must fall on repeat testing in order for the change to be regarded as significant—that is, not attributable to chance. Exceeding CR.95 represents a high degree of confidence that an individual’s visual function has changed. The CR.95 is an appropriate measure to characterize test–retest reliability across a population, even if there are only one test and retest per subject.

The use of multiple test repetitions per subject in this study allows for establishment of an individual mean and 95% confidence interval (CI.95), defined as 1.96 times the SD of the subject’s test measures. For a large number of samples from a Gaussian distribution one can easily prove that the expected absolute difference between two samples equals $\sqrt{2}$ times the SD; hence CR.95 = $\sqrt{2} \times$ CI.95. We will use CI.95 in particular for analysis of dark-adapted Humphrey macular perimetry data. Note that in this test each threshold measure is repeated three times within a session, and so there also is an opportunity to compare within- and between-session reliabilities.

For dark-adaptometry timing, the coefficient of variation (CoV)—that is, the ratio of the SD to the mean of detection time across multiple visits expressed as a percentage—was calculated at each intensity level within each subject.

RESULTS

Dark Adapтомetry

The dark-adaptation curve was measurable in all CTL, MD-I, RP-I, DR, ON, and OR subjects. The resultant curves can be seen in Figure 2. Cone and rod branches cannot be easily distinguished in these curves, due to insufficient bleaching by the adapting light, but both were present in the curves of the MD-I, MD-II, ON, and OR group members. In comparison with the curves of CTL and ON subjects, most RP, MD, DR, and OR subjects exhibited a delay of the entire curve. This was most evident in the MD-II, advanced DR, and RP subjects. None of the subjects in the RP-IV group could see the maximum flash intensity, whereas just 4 of 11 RP-II, and 4 of 13 RP-III eyes tested had measurable dark-adaptation levels with the SST. The RP group members all had mild to severe loss of cone sensitivity, with apparently little to no functional rod vision. Most of the curves for RP subjects show delayed or slowed responses between 30 to 22 dB. Exceptions to this were the thresholds for one RP-III subject, one RP-II subject, three RP-I subjects, and the subject with choroideremia (assigned to the RP-I group). These subjects demonstrated some functional, but delayed and limited, adaptation to levels lower than 20 dB, in addition to a slowed initial response. None of the RP subjects tested was able to detect levels below 10 dB after 45 minutes of testing.

Figure 3A shows the within-subject CoV across visits for dark-adaptometry timing at each intensity level, averaged by group. Generally, the mean CoV in low vision groups fell below 25% at all dB levels, whereas control subjects exhibited slightly higher values during the early part of the test. Some subjects exhibited good reliability throughout, and others showed fluctuation at various intensity levels between visits. Most subject groups averaged — 20% CoV throughout the test across visits. CTL subjects had the highest overall variation for cone-based dark adaptation, along with decreasing variability (down to 11% CoV) noted over time with decreasing scotopic intensity levels. For the MD, ON, DR, and OR groups, the CoV showed less dependence on test intensity. The OR group members had the least timing variability at the lowest intensities, as did those subjects with RP who were able to see test flashes below 22 dB.

Dark-Adapted Humphrey Perimetry

Figure 4 shows how many eyes were able to complete the dark-adapted Humphrey macular perimetry test. Those subjects who were unable to perform this test could not adequately see the red fixation LED. Among the subjects with RP who could perform the test, the sensitivity and/or extent of the central visual field was limited, so that zero values were obtained at certain test target locations. We excluded the results from some of the RP-II and RP-III subjects, and all the RP-IV subjects, who exhibited results that were not consistent within tests and across visits, because they appeared likely to be affected by inherent floor effects, and returned many zero-threshold values. The histogram in Figure 4 shows mean sensitivity levels per group, across test locations, visits, and subjects, for 500 and 650 nm and the log sensitivity ratio (which equals the difference in 500- and 650-nm dB levels.) The findings for the RP-IV, DR, and MD-II groups are not reported in Figure 4, as only two subjects in each of these groups were
able to perform the test, a number considered too low to allow
generalization of the results.

As indicated under Methods, each of the 16 points in the
dark-adapted Humphrey test grid yields three thresholds per
session. Therefore, the initial variability analysis for this test
was based on means and standard deviations. Two measures of
between-visit variability, expressed as CI, are depicted by
the asymmetric error bars in Figure 4: The first assessment
computed the variability at each individual point tested, aver-
aging those measures across points and subjects to give the
upward error bars. The second assessment calculated the vari-
ability of the average sensitivity across all 16 test locations; this
variability was then averaged across subjects to give the down-
ward error bars. As expected, the downward error bars are
smaller than the upward ones for all subject groups at 500 and
650 nm and for the difference scores. CTL subjects showed less
variability than other groups, with the possible exception of
the RP-I group. Variability after averaging across points is 0.5 to
0.75 times the variability by point, whereas a theoretical im-
provement to 0.25 times might be expected (n = 16, improve-
ment as square root). Thus, much of the test-retest variability
is correlated among test points.

To obtain a single measure of test-retest variability, we
calculated CR from differences in the mean threshold across
all points at each visit. The results for each group at 500 and
650 nm, as well as for the 500/650 sensitivity ratio, are de-
picted by the upward bars in Figure 5. The values are expected
to be \sqrt{2} times the downward error bars in Figure 4, and
the actual ratio scatters around this value. Note that CR can vary
between wavelengths, even within groups. The low 500-nm
variations for the RP-II and RP-III groups may be related to
low sensitivity, as shown in Figure 4, but this explanation does
not hold for the ON group.

Three threshold attenuation levels (maximum 50 dB) were
obtained for each of the 16 test spot locations, allowing us to
determine within-session reliability, shown by the downward
bars in Figure 5. CR within sessions ranged between 1 to 3.5
dB for 500 and 650 nm across all subject groups, and was most
repeatable in subjects with good visual acuity, namely CTL,
RP-I, and RP-II. In general, within-session variability was lower
than between-session variability, as would be expected, and
particularly so in groups with high between-session variability
(e.g., MD-I).

Dark-Adapted Full-Field Flash Test

Test measures were obtainable for the dark-adapted full-field
flash test for all but two of our 77 subjects, and vision was not
the limiting factor in these two subjects, who were also unable to understand the dark-adaptometry test directions because of confusion and an obvious language barrier. The mean flash threshold for each subject group is shown in Figure 6. Although advanced RP subjects maintain retinal functioning in only a very small central island, the mean threshold levels that were attainable for the RP-II/III/IV groups are substantial, meaningful scores. The error bars in Figure 6 indicate the repeatability for the full-field flash test, and all groups (mean CR.95/H11005/H11006 14.7 dB) exhibited similar or lower variability than CTL subjects (CR.95/H11005/H11006 12.5 dB). The ON, MD-I, MD-II, RP-I, RP-II, and RP-III groups demonstrated the most reliable results, with mean CR.95/H11005/H11006 7.5 dB. Meaningful and reliable (CR.95/H11005/H11006 12.3 dB) results were also obtained for all RP-IV group members, most of whom had limited or no responses to other tests. The mean threshold for the DR subject group was the most variable.

Consistency among Dark-Adapted Measures

One would expect to find a close relationship between dark-adapted thresholds obtained with the dark-adapted flash test, minimum Humphrey threshold (500 nm for most subjects, but 650 nm for some patients with RP) and dark-adaptometry end values (except for those subjects whose dark adaptation exceeds the range of the SST). If all three measures truly detect absolute threshold sensitivity, then the regression lines in pairwise scatterplots should have unit slope. To verify this, scatterplots in Figures 7 and 8 were created mapping the dark-adapted flash test thresholds, against dark-adaptometry end values and dark-adapted minimum perimetric mean thresholds, respectively.
The mutual regression slope for the dark-adapted, full-field flash thresholds versus dark-adaptometry end values is 2.60. Similarly, the slope of the mutual regression line for dark-adapted full-field flash thresholds versus dark-adapted Humphrey thresholds is $1.42$. Note that neither of these lines comes close to unit slope. The slope in Figure 8 is not significantly different from 1 ($P = 0.13$). In contrast, the slope in Figure 7 differs significantly from 1 ($P = 0.001$) and may confirm that the limited response range of the SST tends to cause ceiling effects, which limits the thresholds compared with those of the full-field flash test. When the slope for the regression line in Figure 8 is forced to 1, the $y$-intercept is $-35.6$, indicating that the flash tests have a 3.5-log unit brightness margin over the perimeter test, allowing testing of subjects with substantially lower remaining light perception than even the most severely impaired subjects in this study. This has been previously put to good use in a pilot study of retinal cell transplantation. $^{25}$

The linear correlation coefficient for the scatterplot in Figure 7 is low ($r = 0.37$). In Figure 8 it appears that the MD-I subjects are outliers, and we therefore calculated correlation coefficients ($r$) that did ($r = -0.60$) and did not ($r = -0.84$; $m = -1.31$) include the MD-I subjects. This analysis confirms our hypothesis that a strong linear relationship exists between the dark-adapted flash test and dark-adapted Humphrey perimeter thresholds, whereas the relationship with dark adaptometry end values is limited by the range of the SST. In contrast, as can be seen from the MD-I subjects in Figure 8, a more complex relationship exists for tests with localized targets: dark-adapted Humphrey perimetry requires good fixation stability, and reduced macular sensitivities can be caused by a combination of sensitivity loss and increased variability caused by fixation instability. In all three tests, variability appears to depend on subjects' eye disease. However, different subject groups also have different sensitivities, so one could hypothesize that $\text{CR}_{0.95}$ increases with decreasing sensitivity (i.e., increasing threshold). To investigate this, the scatterplots in Figures 9 and 10 show the relationship of $\text{CR}_{0.95}$ versus mean dark-adapted, full-field flash and Humphrey thresholds, respectively. The 0.04 slope in Figure 10 is not significantly different from 0 ($P = 0.59$). The slope of $-0.04$ for the regression line in Figure 9, on the other hand, is significantly different from 0 ($P = 0.002$), suggesting
that the subjects with the highest variability are those with the highest sensitivity. This may be attributable to intermittent light leakage around the goggles worn by subjects with high dark-adapted sensitivity. We removed any data where leakage from the mask was indisputably obvious, but were conservative in removing outliers. The goggles attenuate 2500×, so any leakage is likely to be far in excess of the light transmitted by the goggles. Therefore, such leakage was immediately obvious from the data for subjects whose thresholds were in the 48- to 72-dB range, due to the excessive changes noted during the test session. Only for subjects whose thresholds are in excess of 72 dB would fluctuations be 10 dB or less, and thus go unnoticed until they were suspected in the analysis. Note that none of those highly sensitive subjects is included in Figures 7 to 9. It is important to note that both regression lines contradict the commonly held belief that reduced sensitivity in psychophysical tests is associated with greater between-sessions variability.

We calculated the ratios of the CI.95 and CR.95 for dark-adapted Humphrey perimetry by computing each subject’s own ratio, and calculating the mean ratio of all study participants within a group and for all subjects, shown in Table 2. These calculations were intended to minimize the effect of inhomogeneity of various disease subtypes, due to different remaining levels of retinal function within each group. The mean ratios at 500 and 650 nm for all subjects are larger than √2, predicted by theory, whereas more than half of the ratios computed by subject group fall below this value. These variations are presumably due to a non-Gaussian distribution of the data.

**Reliability of Individual Eyes**

The eyes of a low-vision subject may have different levels of visual function, dependent on the type and course of the ocular disease. For this study, the mean values for the better eye versus the worse eye for the full-field flash test (r = 0.96), in addition to visual acuity (ETDRS; r = 0.83) and contrast sensitivity (Pelli-Robson; r = 0.88), correlated highly. For normally sighted individuals, it is generally thought that the eyes are not independent with respect to reliability, and that for this reason data from only one eye should be evaluated. No careful analysis of this issue in subjects with severe vision loss due to retinal degeneration can be found in the literature. To compare the reliability between eyes of visually impaired individuals, each subject’s mean test–retest SD of the better eye versus that of the worse eye was plotted. Low correlation coefficients were obtained for the full-field flash test (r = 0.28), as well as for visual acuity (r = 0.013) and contrast sensitivity (r = 0.42).26

Similar to the findings in Figures 9 and 10, there was no significant difference between CR.95 for better and worse eyes. The ON and RP-IV groups had the largest differences of approximately 2 dB between the mean CR.95 for the better and worse eyes, with the worse eyes performing more reliably, indicating that less vision does not necessarily mean that the test results will be more variable. All our data suggest, therefore, that the level of variability (noise) is independent of the retinal function level.

A second set of results for the dark-adapted flash test was obtained from six subjects among the original 77, who completed the study requirements and then volunteered to repeat the test battery for all test sessions 1 to 2 years later. We created a scatterplot of the CR.95 for the first series versus those for the second series (performed 1 to 2 years later) for each eye from the six subjects who had more than one test–retest (n = 10) and computed the linear regression of this scatterplot. The CR.95 on test and retest correlated highly (r = 0.82). The slope of the regression line (m = 1.11) indicated that the variability was similar between the sessions performed several years apart. Thus, while the reliability between a subject’s eyes did not correlate well, the reliability within eyes appeared to be stable over the course of a few years.

**DISCUSSION**

In the past, the standard for dark adaptation has long been the Goldmann-Weekers dark adaptometer (Haag-Streit, Kôniz, Switzerland); however, this instrument is no longer manufactured. More recently, lightweight, portable, relatively inexpensive, LED-based dark adaptometers have become commercially available, such as on used in this study (SST-1; LKC Technologies). One study determined the sensitivity of this unit compared with that of the Goldmann-Weekers, for detecting changes in night vision in patients with hereditary retinal degeneration and loss of rod function. Linear regression analysis, discrepancy analysis, and receiver operator characteristic curves for both devices show that the SST-1 quantifies psychophysical rod function nearly as well as the Goldmann-Weekers, though with some limitations.27

The major limitation of the current SST-1 unit is its narrow (3.0 log unit) range of intensities, at its maximum too dim to evaluate some subjects in the RP-II/III groups or any RP-IV, who have severe scotopic sensitivity losses (floor effect), and at its minimum too bright to evaluate thresholds in subjects with good scotopic sensitivity (ceiling effect). The potential to obtain meaningful results with such a compact instrument for all subjects exists, as the SST-1 could be modified to test a wider range of intensities. Another aspect of the SST-1 that could be improved is the initial adapting light that currently does not produce adequate photopigment bleaching. By increasing the intensity or duration of the adapting light, one could distinguish better between the rod and cone branches in the dark-adaptation curve and obtain a rod–cone plateau, as typically seen in well-bleached, normally sighted eyes at approximately 22 dB on the SST-1 scale.

In its present form, dark adaptometry with the SST-1 entails a tedious process that requires a well-trained operator. Although the timing variability during dark adaptometry testing for all subject groups was relatively good, averaging approximately 10% CoV throughout the test, this could probably be reduced by building in an automated test procedure, helping to improve uniformity and repeatability by reducing variability caused by examiner technique or speed. Such a procedure should include automatic recording of the thresholds—instead of the current procedure, which requires manual scoring on paper with the aid of a red penlight—and a continuous staircase procedure for determining running adaptation thresholds and the final dark-adapted, full-field threshold.

To determine whether variations during dark adaptometry were dependent on and proportional to the timing, we plotted the log of the mean CoV against the log of the mean timing values, shown in Figure 3B. Note that a slope of −1 would be expected for log(CoV) against log(time) if the standard deviation of the timing were independent of the mean, whereas a
The mean thresholds that are attainable with the dark-adapted, full-field flash test reflect a summation of function across the entire retina. Threshold values for the ON and MD-I subject groups are only slightly less than those for CTL subjects, as these subject groups typically retain functional rod vision over a substantial portion of the retina. Advanced DR and OR conditions typically affect function throughout large retinal areas, including the macula, posterior pole, and mid- or far-peripheral regions. Therefore, as expected, these subjects tend to have lower full-field flash sensitivities than ON and MD-I subjects, but greater than the RP subject groups. Among the RP subject groups, the RP-I members had the greatest retinal sensitivity during flash testing, and the RP-II, -III, and -IV groups had progressively decreasing mean sensitivity values, as would be expected since rod and cone loss gradually progresses over the course of the disease process.

The low correlation coefficient obtained when comparing each subject’s mean test-retest SD of the better eye versus that of the worse eye for the full-field flash test suggests at most that 10% of the variance can be attributed to a common mechanism between the two eyes (e.g., cortical variability). The lack of correlation indicates that most of the variation is present before combination of the signals from the two eyes, most likely at the earliest processing levels, and can be thought of as detector noise in retinal processing. Therefore, it is likely that a retinal contribution dominates the dark-adapted flash test threshold variability in legally blind subjects.

**Conclusions**

Each of the three dark-adapted psychophysical tests yielded comparable repeatability across most of the subject groups, indicating that these tests can be used to monitor change across many types of ocular diseases. Our results suggest that the most valuable measure of remaining vision in subjects with severe vision loss was the dark-adapted, full-field flash test, as it was capable of producing meaningful and repeatable results across all degrees of vision loss. These psychophysical tests performed under scotopic conditions will be useful for monitoring outcomes in early trials to reverse or halt vision loss. Such functional measures can be used to investigate safety in phase 1 clinical trials and efficacy in subsequent phase 2 and 3 trials, complementing measures obtained through clinical examination. Although our precise instrumentation may not be available to other centers, similar instrumentation and/or software can be developed for this purpose, using a Ganzfeld to present full-field stimuli, with built-in fixation control for the presentation of localized stimuli, and automated psychometric stimulation and recording procedures. Ideally, developing a dedicated instrument to perform all three dark-adapted tests would allow multiple centers to adopt identical outcome measures, making this series of psychophysical visual function tests all the more suitable for use in clinical research at multiple sites.

**Acknowledgments**

The authors thank Nancy Lewis and Jim Stone for their contributions as study coordinators, and Janet Sunness, MD, and Shirin Hassan, PhD, for their insightful comments on an earlier draft of this manuscript.

**References**


