The Effective Dynamic Ranges of Standard Automated Perimetry Sizes III and V and Motion and Matrix Perimetry

Michael Wall, MD; Kimberly R. Woodward, MS; Carrie K. Doyle, MD; Gideon Zamba, PhD

Objectives: To establish the associations between threshold estimates of 4 perimetric tests and to define and compare the tests' effective dynamic ranges.

Methods: We examined 152 patients with glaucoma and 80 controls using standard automated perimetry (SAP) with stimulus size III, SAP with size V, and motion and matrix perimetry. We explored the intertest associations using principal-components analysis. We defined the effective dynamic range bottom using the frequency of 0-dB trials on retest. We defined the upper effective dynamic range as a value above which fewer than 0.5% of the values fall in the controls. We also calculated the number of discriminable steps from normal to the floor value of the perimeter.

Results: The association between SAP III and V was approximately linear up to a sensitivity of about 20 dB on both tests and with motion and matrix perimetry up to about 25 dB from 0 dB. While the upper bounds were similar among the tests, size V SAP had a lower floor and more discriminable steps.

Conclusions: The effective dynamic range of SAP III is substantially less than its physically tested limits. Size V stimuli have a greater effective dynamic range than size III by about 1 log unit and have about twice as many discriminable steps.


The dynamic range in perimetry is defined as the range of the smallest and largest values of the visual stimuli presented by a perimetry device. It is set by the instrument developer usually either arbitrarily or because of technical limitations. However, some of the values in this range may not be clinically meaningful. For example, standard automated perimetry (SAP) is capable of displaying dim increments of light of 60 dB and higher that are beyond the limits of perception of differential light sensitivity perimetry.

When describing a perimetric instrument, it is useful to characterize the effective dynamic range (EDR). We define this as the clinically useful dynamic range. Developing an operational definition of the EDR, while important, is problematic. Defining the EDR is useful, as testing outside of this range can be both time consuming and misleading. The EDR is important not only for individual instruments but also for comparing instruments.

When perimetric sensitivity using SAP falls to the point where retest variability is so great that the retested values are scattered across several log units (therefore having limited clinical value), one could argue that this represents the lowest value of the EDR. For example, Heijl and colleagues\textsuperscript{1,2} investigated the variability in 51 eyes of 51 experienced glaucoma subjects (they underwent automated perimetry before) representing all stages of optic nerve damage. The patients, all clinically stable, were tested 4 times in a 4-week period with size III full-threshold testing with the Humphrey field analyzer. For test locations, initially measured with 8 to 18 dB of loss, the 95% prediction interval nearly covered the full measurement range of the instrument. Retesting of values in a clinical setting after this much loss would appear to have limited value. One could argue that it should not be included in the EDR.

Our aim in this investigation is to define the EDRs of 4 perimetric instruments, establish the associations between threshold estimates of these tests, and to compare the tests’ EDRs. To explore these associations, we will compare the tests’ EDRs by (1) comparing individual thresholds of subjects tested the same day at the same test location once a week for 5 weeks across instruments; (2) defining the lower bound of the EDR in participants tested twice by the highest sensitivity at which 5%
of the retest values is the floor (0 dB) and the value at which 0 dB is the most frequent retest value; and (3) defining the upper bound of the EDR as the value above which less than 0.5% of the values fall in normal subjects. In addition, we will use a scale-independent measure: the number of discriminable steps to compare the EDRs of the 4 perimetry methods.

METHODS

SUBJECTS

University of Iowa institutional review board approved the visual testing protocol. Thirty-two patients with glaucoma and 20 healthy participants were tested once a week for 5 weeks. Also, 120 different patients with glaucoma and 60 other healthy participants were tested at baseline and again at a separate sitting within 1 to 8 weeks. All 4 cohorts were consecutive series. All subjects gave informed consent to participate in the study. The tenets of the Declaration of Helsinki were followed. The normal participants were volunteers who answered advertisements inviting them to participate in research, paid in accordance with the institutional review board.

Normal participants were included if they had (1) no history of eye disease, (2) refractive error within a ±5-diopter sphere and ±2-diopter astigmatism, (3) no history of diabetes mellitus or systemic arterial hypertension, and (4) a normal ophthalmologic examination result, including 20/25 or better Snellen acuity. The subjects either had undergone a complete eye examination within 24 months prior to this study or were examined by an ophthalmologist on the day of testing to ensure normal ocular health. One eye of each participant was randomly chosen as the study eye.

The patients with glaucoma were invited from the glaucoma clinic at the University of Iowa Department of Ophthalmology and Visual Sciences if they met entry criteria. They were enrolled if they had glaucomatous optic disc changes with an abnormal SAP result (glaucomatous visual field defects, ie, ≥3 adjacent test locations that fall outside normal limits in a clinically suspicious area at P < .05 or 2 adjacent locations falling outside normal limits with at least 1 at P < .01; in addition, mean deviation was in the range of 0 to −20 dB on SAP). We included patients with primary, secondary, or normal tension glaucoma. The patients did not have another disease that affected vision and were capable of undergoing SAP and returning for follow-up visits. Patients were excluded if they had cataract that caused visual acuity of worse than 20/30, had a pupil size smaller than 2.5 mm, or were aged younger than 19 years. If both eyes qualified for the study, 1 was chosen at random as the study eye. The mean age of the controls was 57.2 years (SD, 7.9 years), with a range of 41 to 78 years. Thirty-eight of the healthy participants were women and 22 were men. The mean age of the patients with glaucoma was 64.9 years (SD, 9.5 years; range, 38-81 years); their mean deviation was −6.7 dB (SD, 4.4 dB).

VISUAL TESTING

Humphrey Field Analyzer

All subjects underwent automated perimetry using program 24-2 of the Humphrey field analyzer (Carl Zeiss Meditec, Dublin, California). For SAP, the stimuli of Goldmann size III (0.43° diameter, 4 mm²) were used with the standard 24-2 Swedish interactive thresholding algorithm (SITA). Goldmann size V stimuli (1.72° diameter, 64 mm²) were used along with full-threshold testing; there is no SITA program currently available for size V stimuli.

Our pilot data and the work of Artes and coworkers show that the differences between SITA and full threshold are minor. We chose size III SITA so that we would be comparing the test most commonly used in clinical practice.

We followed the manufacturer’s recommendations for using corrective lenses. Care was taken to prevent lens rim artifacts. The subjects had testing done in 1 eye, chosen at random, but the same eye was used for all tests. All visual field examinations met the following reliability criteria: fixation losses of less than 20% or normal gaze tracking, a false-positive rate of less than 10%, and a false-negative rate of less than 33%. The 4 tests were administered in a random order with at least a 5-minute rest between the testing sessions.

Motion Perimetry

Motion perimetry uses random-dot cinematograms as visual stimuli. The dots are randomly displayed on a gray background with a luminance of 31.5 apostilbs using a standard video graphics array video display (640 × 480 pixels). The motion targets were circular random-dot cinematograms within which 50% of the dots moved centrifugally and 50% moved in random directions. This is a commonly used stimulus for motion perception experiments, as the random placement of stimulus dots reduces the effect of positional cues. Subjects were asked to respond by touching the monitor with a light pen in the area in which they had perceived the movement. The circular target itself was stationary; dots moved within the target. Details of the display can be found in previous publications. The stimuli were of 20 sizes with a scaling factor of 1.26. Subjects sat 22 cm from the cathode ray tube monitor. This 17-inch (diagonal [43-cm]) monitor gives a 21° test field (42° × 42° total). The angle subtended by the stimuli ranged from 0.25° to 8.46°. The size of the stimulus window varied from trial to trial, and a 2/1 staircase procedure was used to bracket the threshold. The test, therefore, continued until the smallest circle size seen (size threshold) at each test point was bracketed by the staircase procedure. We tested 44 locations that matched the 24-2 Humphrey perimetry test points except for absence of the top and bottom rows (y = 21° and −21°) and the 2 points along the nasal horizontal axis (x = 27°). Valid responses were defined by (1) a response time between 110 and 1000 milliseconds and (2) a localization error smaller than 10° from the center of where the target was actually presented.

Testing was done in a darkened room using a computer with software that we developed. The patients’ appropriate near correction was again used. Care was taken to prevent lens rim artifacts by asking if the subject could see each corner of the video display while looking at the fixation target.

To facilitate the comparison among the 4 methods, the threshold values of motion perimetry were transformed such that their numerical ranges were similar to those of the other tests. This does not imply that we were able to standardize the tests’ dynamic ranges. Motion perimetry measures a size threshold (18 steps) with the smallest size being 1 dB. We used the simplest way to transform the data so that the range of motion perimetry test results were similar to the range of the other tests by using the following equation: (18−observed threshold) × 2.

Matrix Perimetry

Humphrey matrix frequency doubling perimetry was performed either before or after conventional perimetry testing or motion perimetry with at least a 5- to 10-minute rest between examinations to reduce the effect of fatigue. Testing was performed in a dim room using the Humphrey matrix device. Patients were asked to press a response button whenever they saw
a small patch of alternating light and dark-gray bars at any location within the field of view. Each test lasted about 5 to 6 minutes per eye. For this test, patients wore their own prescription glasses and did not use an eye patch to cover the fellow eye. Breaks were allowed when requested. Details of the testing can be found in publications by Anderson et al.8,9

Figure 1. Patients with glaucoma retested 5 times and mean thresholds plotted 1 test against another using a principal-components analysis and Loess fit. A, Data graphed with the 0-dB trials included. B, Data with the floor effect of 0 dB removed (if a 0-dB trial is included in the average of the 5 retests, the result is not graphed). Note the linear associations above about 20 to 25 dB and the poor fits below this range for most comparisons. Matrix and motion perimetry appear to have similar effective dynamic ranges and standard automated perimetry (SAP) size V has a greater dynamic range than SAP size III by about 1 log unit.
STATISTICAL ANALYSIS

We explored the association between pointwise threshold estimates obtained with the 4 techniques using a principal-components graphic method. This graphical representation describes a smooth curve that passes through the middle of the data. The curves, on each pairwise association, are a nonparametric generalization of a linear principal component. They provide a parsimonious description of the covariance structure among pairs of measurements (on these 4 techniques) and by extension provide a graphical display of associations between each pair of techniques. The curves all use the Loess choice of smoother. The Loess smoother allows one to visualize nonlinear patterns by considering localized fits and by building an approximate function that can describe the deterministic part of variation in the data (Figure 1 and Figure 2).

We defined the upper bound of the EDR as the value above which less than 0.5% of the values fall in normal subjects. To estimate the lower boundary of the EDR, we examined retest variability of the 4 methods. We established the distributions of threshold estimates at retest, conditional on their value at the first test. The EDR lower boundary was then defined in 2 ways: (1) the value at which 5% of the retest values are the same as the floor (0 dB) and (2) the value at which 0 dB is the most frequent retest value.

Lastly, we used a scale-independent method: counting the number of discriminable steps. To do this, we first computed the empirical fifth and 95th percentiles of the distribution of the retest values for the 180 subjects tested twice and used Loess smoothing to create the graph. We then plotted a line of unity, that is, the 45° line that runs from the intersection of the x- and y-axes to the upper right-hand corner of the graph. We defined the first discriminable step as the vertical distance from where the fifth percentile boundary crossed the line of unity to where the line intersects the 95th percentile. The next structure of the staircase is a horizontal line drawn from the 95th percentile to the line of unity. This process starts in the top end of the decibel range and continues until 0 dB is reached and the number of discriminable steps for each perimetry method is then counted.

RESULTS

After averaging values from the 5 retests and graphing 1 test against another using a principal-components technique, the association between SAP size III and V was approximately linear to a sensitivity of about 20 dB (Figure 1). When matrix or motion perimetry was plotted against SAP size III or V, the results were linear to about 25 dB. Matrix and motion perimetry had similar EDRs, but the associations were complex. Despite averaging 5 tests, the associations were noisy, especially below the 20 to 25 dB range.

The floor and ceiling criteria results are found in the Table and come from test-retest data on 120 patients with glaucoma and 60 healthy ocular participants. The ceiling results are similar for the 4 tests. With regard to the EDR floor, when comparing the total number of 0-dB trials, SAP size III had the greatest floor effect in terms of 0-dB trials (3 times as many as SAP size V), suggesting that it has the smallest dynamic range. Matrix and motion perimetry were intermediate and SAP size V had the fewest number of 0-dB trials. Standard automated perimetry size V appeared to have a greater EDR than SAP size III by about 1 log unit. Standard automated perimetry size III and matrix and motion perimetry appear to have similar EDRs with poor fits below the 20 to 25 dB range and a similar EDR by the criteria in the Table (the highest sensitivity at which 5% of the trials are 0 dB and the value at which 0 dB becomes the most frequent retest result).

Whether the floor of the EDR was defined by the point where 5% of the retest trials were 0 dB (Figure 2) or the sensitivity where 0 dB becomes the most frequent value...
Figure 3. Floor of effective dynamic range as defined by results of histograms of the proportion of retest estimates by sensitivity where 0 dB becomes the most frequent value on retest. Standard automated perimetry (SAP) size III, SAP size V, and matrix and motion perimetry. Again, the floor for size V is about 1 log unit lower than size III, and below about 20 dB the histogram is dispersed.
on retest (Figure 3), SAP V was again about 1 log unit of EDR more than SAP III. The lowest values of the EDR for matrix and motion perimetry are similar to SAP V with these analyses. As shown in Figure 3, the retest histograms are dispersed below about 20 dB and 0 dB becomes the most frequent value on retest while retesting below this level.

When comparing the number of discriminable steps, SAP size V again appears to have the greatest dynamic range with 8 steps. Other tests had about half as many discriminable steps (Figure 4 and Table).

Defining the EDR has not been previously attempted, so there is no available agreed upon classification. Therefore, we defined and described the EDR in multiple ways. Because 5% levels of significance are standard in medicine (and the lower the sensitivity, the higher the number of 0-dB results on retest), we thought a reasonable definition would be to use the sensitivity at which 0 dB reaches 5% of the total retest values. Our reasoning was that clinically, if 0 dB was more likely than the original value, then there was a limited predictive value to that sensitivity value and an EDR had been reached. Along these lines, we also computed the sensitivity value at which 0 dB was the most frequent value on retest. The upper limit of the EDR is related to the limits of perception. To account for occasional false-positive responses, we defined the upper limit of the EDR as the value above which less than 0.5% of the values fell in normal subjects.

Possibly the clearest, most clinically relevant, and most intuitive method to compare EDRs of these 4 perimetry tests is to count the number of discriminable steps that could be estimated for each test’s decibel range. This number of steps from normal to blind is determined by the retest variability and the stimulus brightness or stimulus strength range. Unlike the other measures, this one has a major benefit in being scale-independent and allows an unbiased comparison of the EDR for the different tests. In Figure 4, for example, SAP III has 4 discriminable steps for progression. The final step goes from about 20 dB to 0 dB...
23 dB to about 0 dB; further worsening cannot be measured. Using this technique, SAP size V had by far the greatest EDR, having 8 discriminable steps for progression. This was about twice as many steps as the other tests.

We also plotted the sensitivities of each test on another map after taking an average of 5 tests and adjusting the test’s physical dynamic range to be numerically similar. Despite this, the associations among the tests were noisy when sensitivity fell below about 20 to 25 dB. We found the EDR of SAP III to be considerably less than its physically tested limits. Also, motion and matrix perimetry appeared to have similar EDRs, but these comparisons are problematic because different visual functions are being tested and the tests have some fundamental differences.

To define the lower ends or floors of EDRs, we used test-retest data. This more approximates the clinical situation in which one is comparing a current test with the test of the previous visit. With this paradigm, whether one uses the criteria of when 0 dB becomes the most frequent retest value or when 5% of values on retest are 0 dB, it is clear that once sensitivity falls below about 20 dB, the data have a substantial amount of noise and the repeatability of retest at that test location has limited clinical value (retest histogram is dispersed, Figures 2-4). With this poor usefulness of the initial value to predict the retest value, it begs the question of whether test locations that likely have sensitivities in this range of little practical benefit should be retested. In other words, should we stop the thresholding procedure when we reach the floor of the EDR? This issue is more complicated than it appears. For example, if a cutoff value of about 20 dB is used to truncate testing and the true sensitivity is 22 dB, with retest some values will be less than 20 dB. We are not implying that 20 dB is the defined cutoff. Whether 15, 20, or 25 dB or an intermediate value is used, the EDR of SAP III appears to extend for about 2 log units, and then retests become highly variable, with 0 dB becoming the most frequent value on retest below 15 dB. This information is important, as clinicians should be very careful when interpreting visual field change when the initial value is below about 20 dB.

If the thresholding algorithm were to be interrupted when the EDR is exceeded, a Bayesian procedure, such as the zippy estimation by sequential testing, might be used. When there is sufficient certainty that the test location’s sensitivity is less than about 20 dB (or whatever cutoff is calculated), the procedure would be terminated. This would substantially reduce the retest variability of SAP and might allow a more accurate and less noisy cutoff determination of visual field change. So, rather than trying to use test locations with clinically unacceptable variability, only the test locations with retest variability in an acceptable range would be used. Something similar is already done with the Humphrey-guided progression analysis.

The values we found at the lower end of the EDR with SAP are interesting. Using optical coherence tomography data, if one graphs retinal nerve fiber layer thickness against sensitivity, an asymptote of about 50 μm is reached at about 20 dB. In other words, at about 20 dB, there is no measurable nerve fiber layer by optical coherence tomography. Clearly, some receptive fields remain, but their topography is likely a sparse irregular array. This might be the reason that 0 dB becomes the most frequent value on retest with SAP for values below 15 dB. That is, with the sparse irregular receptive field array, a stimulus may cover a sufficient number of receptive fields on the first test and then “fall in a hole” of the remaining receptive fields on retest and be measured as 0 dB.

This idea is supported by the results from size V testing. Here, we gain about a 1 dB unit in the EDR and 0 dB, as the most frequent value on retest falls from 15 dB to 4 dB. Possibly this is because it is more difficult for the much larger stimulus to fall in the remaining receptive field “holes” and be missed (ie, measured as 0 dB).

In conclusion, the effective dynamic range of SAP size III is substantially less than its physically tested limits. Standard automated perimetry size III, motion perimetry, and matrix perimetry have similar EDRs, but their associations are complex. Standard automated perimetry V stimuli have a greater EDR than SAP III by about 1 log unit and have about twice as many discriminable steps. Size V SAP stimuli may therefore be useful in testing glaucoma patients with moderate to severe visual field damage and may be superior in detecting true progression in patients with moderate to severe visual field loss.

Submitted for Publication: September 15, 2008; final revision received April 3, 2009; accepted April 12, 2009.

Correspondence: Michael Wall, MD, Department of Neurology, College of Medicine, University of Iowa, 200 Hawkins Dr, #207 Roy Carver Pavilion, Iowa City, IA 52242-1053 (michael-wall@uiowa.edu).

Financial Disclosure: None reported.

Funding/Support: This study was supported by a Veterans Affairs Merit Review Grant and an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, New York, New York.

References