Article

The Effect of Stimulus Size on the Reliable Stimulus Range of Perimetry

Stuart K. Gardiner[1], Shaban Demirel[1], Deborah Goren[1], Steven L. Mansberger[1], and William H. Swanson[2]

[1] Devers Eye Institute, Legacy Research Institute, Portland, OR, USA
[2] Optometry and Vision Science, Indiana University, Bloomington, IN, USA

Purpose: Automated perimetry uses a 3.5 log unit (35dB) range of stimulus contrasts to assess function within the visual field. Using 'Size III' stimuli (0.43°), presenting stimuli within the highest 15dB of available contrast may not increase the response probability at locations damaged by glaucoma, due to retinal ganglion cell response saturation. This experiment examines the effect of instead using 'Size V' (1.72°) stimuli.

Methods: Luminance increment thresholds for circular spot stimuli of each stimulus size were measured in 35 participants (mean deviation −20.9 to −3.4 dB, ages 52–87) using the method of constant stimuli, at four locations per participant. Frequency-of-seeing curves were fit at each size and location, with three free parameters: mean, standard deviation, and asymptotic maximum response probability. These were used to estimate the contrasts to which each participant would respond on 25% of presentations ($c25$).

Results: Using segmented orthogonal regression, the maximum observed response probabilities for size III stimuli began to decline at $c25 = 25.2$ dB (95% confidence interval 23.3–29.0 dB from bootstrap resampling). This decline started at similar contrast for the size V stimulus: $c25 = 25.0$dB (22.0–26.8 dB). Among locations at which the sensitivity was above these split-points for both stimulus sizes, $c25$ averaged 5.6 dB higher for size V than size III stimuli.

Conclusions: The lower limit of the reliable stimulus range did not differ significantly between stimulus sizes. However, more locations remained within the reliable stimulus range when using the size V stimulus.

Translational Relevance: Size V stimuli enable reliable clinical testing later into the glaucomatous disease process.

Introduction

Many studies have demonstrated that test–retest variability in standard automated perimetry (SAP) increases substantially with glaucomatous damage.[1–9] This makes it difficult for clinicians to assess the extent and magnitude of damage, and the rate of disease progression. However, the causes of this variability are not fully understood. This study aims to develop an improved understanding.

One factor that limits the useful dynamic range of perimetry (i.e., the range of disease severities over which useful measurements can be obtained) is noise in the neurovisual system. The responses of retinal ganglion cells (RGCs) increase approximately linearly with stimulus contrast at low contrasts, and then at high contrasts they show response saturation, with their spike rate no longer increasing beyond an asymptotic maximum.[10,11] When RGC response rate is already high (well over one-half of this maximum response rate) then increasing stimulus contrast results in only a small increase in the RGC firing rate. This in turn means that the probability that the observer will respond to the higher contrast is only minimally increased. In regions of severe glaucomatous damage, the observer's maximum response probability may be well below 100%, and even below 50%.[12] A testing algorithm would then be unable to quickly converge to the contrast to which the observer

responds on 50% of presentations (the conventional clinical definition of sensitivity). Clinical automated perimetry only permits 5 to 10 presentations per location in order to maintain an acceptably short test duration below 10 minutes per eye. This would tend to make the sensitivity that is reported by the perimetric testing algorithm unreliable at such locations.

In a recent paper, we used a method of constant stimuli approach and fitting of frequency-of-seeing (FOS) curves to assess the effects of response saturation, and found that it occurs at contrasts above 400% to 1000%, or 15 to 19 dB for clinical perimetry with the conventional "size III" stimulus, a circular luminance increment with a diameter of 26 minutes of arc (0.43°).[12] For example, the probability that participants responded to a 20,000% (2 dB) contrast in a deep visual field defect was typically only slightly higher than the probability that they would respond to a 2000% (12 dB) contrast at the same location (this small increase may reflect effects of light from the stimulus being scattered toward remaining areas of higher sensitivity). For locations with sensitivity worse than 15 to 19 dB (equivalent to contrasts of 1000%–400%), we found that the relation between sensitivity measures from FOS curves and those obtained from clinical perimetry had R^2 less than 0.1, indicating that the true sensitivity explained less than 10% of the observed variance. This implies that it is not possible with current clinical perimetry to reliably distinguish between sensitivities of 2 and 12 dB. We defined the "reliable stimulus range" as being the range of contrasts over which sensitivity estimates from clinical perimetry accurately reflect the sensitivity as measured using FOS curves with R^2 greater than or equal to 0.1 (i.e., above 15–19 dB). We suggested that this phenomenon might partially explain the increase in test–retest variability that is observed in moderate and severe glaucoma.

Perimetric contrast sensitivity is known to increase with stimulus size in both normal and diseased eyes.[13–17] Previously it has been suggested that use of a size V (1.72° diameter) stimulus could also reduce variability, allowing reliable visual field testing to be performed later into the disease process.[18–21] This study examines whether the reduction in variability observed when using size V stimuli can be explained solely by the increase in sensitivity, or whether the reliable stimulus range is also extended to higher contrasts than when using size III stimuli. The results from this study will help clinicians, researchers and manufacturers understand the most reliable testing strategies for glaucoma patients with moderate to severe visual field loss, who have the most variable results and also the highest risk of permanent visual impairment.

Methods

Testing followed the same procedure as in the previous study.[12] Participants with moderate to severe primary open-angle glaucoma were recruited from a tertiary glaucoma clinic at Devers Eye Institute (Portland, OR). Inclusion criteria were a diagnosis of primary open-angle glaucoma as determined by each participant's clinician, and at least two nonadjacent visual field locations with sensitivities between 6 and 18 dB on both of their two most recent clinic visits (24-2 test pattern, size III stimulus, SITA Standard algorithm; Humphrey Field Analyzer [HFA]; Carl-Zeiss Meditec, Dublin, CA). Exclusion criteria were an inability to perform reliable visual field testing, best-corrected visual acuity worse than 20/40 (because this could cause difficulties with maintaining fixation), cataract or media opacities likely to significantly increase light scatter, or other conditions or medications that may affect the visual field. All protocols were approved and monitored by the Legacy Health institutional review board, and adhered to the Health Insurance Portability and Accountability Act of 1996 and the tenets of the Declaration of Helsinki. All participants provided written informed consent once all of the risks and benefits of participation were explained to them.

For each participant, four test locations from the 24-2 visual field test pattern were chosen based on reviewing their two most recent clinical visual field test results. The "Perimetric Sensitivity" at each location was defined as the mean of the sensitivities measured at that location on those two visits. At least two of the chosen locations had significantly reduced sensitivity that was no lower than 6 dB (to ensure that some function remained at that location), with the remaining locations chosen in different regions of the visual field to promote stable fixation. Testing several locations that were spread around the visual field also ensured spatial uncertainty, which will steepen FOS curves by preventing attention being focused on a single location,[22] and will also make the test conditions more similar to clinical perimetry. Frequency-of-seeing curves were then assessed using the Method of Constant Stimuli (MOCS)[23,24] on an Octopus perimeter, using both size III and size V stimuli in separate runs.

Thirty-five participants were tested (mean age 69.9 years, range, 52–87). The mean deviation of the tested eye on the most recent clinic visit averaged −10.7 dB (range, −20.9 to −3.4 dB). The perimetric sensitivities of the locations tested (i.e., the mean of the pointwise sensitivities at their last two clinic visits) averaged 18.9 dB (range, 8–33 dB). Seventy-seven out of 140 tested locations had perimetric sensitivity less than 19 dB, and 56 were less than or equal to 15 dB.

For size III stimuli, seven contrasts were chosen for testing per location. At the two less damaged locations of the four, these were set at 3-dB intervals centered at the perimetric sensitivity (i.e., so that the range ±9 dB from this value was covered). For two participants, one of these locations had perimetric sensitivity 13 dB, and so the seven contrasts were instead centered on 12.7 dB (i.e., covering the range 3.7–21.7 dB; 3.7 dB on an HFA scale = 1350 cd/m^2 was equivalent to 0 dB on our instrument's native scale; to avoid confusion we use HFA-scale decibels throughout this report). This meant that the greatest contrast to be tested at such locations (lowest on the decibel scale) would be 3.7 dB, which was the highest intensity stimulus available on the Octopus perimeter used. The lowest contrast that could be presented was set at 40 dB.

For the two most damaged locations of the four selected for a given eye, the highest contrast stimulus to be tested was always set to 3.7 dB, and the lowest contrast to be tested was always set to 28.7 dB, so that there were common contrast levels for all participants. The remaining five intermediate contrasts were set at 3-dB intervals centered on the perimetric sensitivity (or centered on 12.7 dB) as before. These unequal intervals between contrasts will not affect parameters taken from the fitted FOS curve.

The same four locations were also tested using size V stimuli. The contrasts to be tested were set 4 dB higher than those for size III (i.e., a lower contrast; this should cover the full range of response probabilities, since increasing the stimulus size is expected to increase sensitivity[15]), except in two situations. Firstly, the highest contrast for the two most damaged locations was kept at 3.7 dB. Secondly, the lowest contrast that could be presented was still set at 40 dB. Among the two less damaged locations per eye, perimetric sensitivity was always greater than 12.7 dB, and so there was no need to enforce a minimum of 3.7 dB.

Stimuli were presented using an Octopus 900 perimeter (Interzeag/Haag Streit, Koeniz, Switzerland), externally controlled using the Open Perimetry Interface.[25] This allows a specified stimulus to be presented, with the perimeter returning information about whether the participant pressed the response button within a designated period of time, which was set as being up to 800 ms after the end of the stimulus. Stimuli were presented for 200 ms, as is standard in clinical static increment automated perimetry when using the HFA perimeter, in order to provide the most direct comparison with each participant's clinical results. Participants were also required to undergo several minutes of adaptation to the perimeter's background, as they would in clinic.

We programmed the perimeter to present seven repetitions at each of the seven chosen contrasts for each of the four chosen locations, plus five blank presentations (i.e., a total of 201 presentations in one run). Five runs were completed per stimulus size, with the order of presentations randomized within each run for both contrast and location. Runs were alternated between size III and size V, in order to prevent any bias due to fatigue and learning effects. Therefore the total number of presentations over the five runs for each stimulus size was 35 per contrast level per location, resulting in 245 presentations per FOS curve. The 50 blank presentations across all 10 runs (i.e., from both stimulus sizes) were used to estimate the false positive rate, which was assumed to be independent of stimulus size. Each run took approximately 6 minutes, designed to be similar to the duration of a SITA Standard visual field test. To reduce fatigue, participants were allowed to take breaks between runs, provided they regained adaptation before recommencing testing. All testing was performed in one visit.

Once testing had been completed, we calculated the proportion of stimuli to which the participant responded for each contrast, averaged across the five runs for a given stimulus size. A cumulative Gaussian curve was fit to each set of FOS data, such that the response probability was given by:

$$fp + (a - fn)\Phi\big((c - cs)/sd\big) \quad (1)$$

fp represents the false positive rate. c represents the contrast of the stimulus in HFA-scale dB. Φ represents a cumulative Gaussian distribution function, such that $\Phi(-\infty) = 0$, $\Phi(0) = 0.5$ and $\Phi(\infty) = 1$. cs represents the contrast sensitivity in dB according to the conventional definition in clinical perimetry (i.e., the contrast that the participant would respond to on 50% of presentations in the absence of false positive or false negative responses). sd represents the stan-

dard deviation of the cumulative Gaussian, such that a higher value of *sd* gives a shallower FOS curve. The values of *cs* and *sd* were fit by constrained maximum likelihood estimation, with *cs* constrained to be greater than −10 dB (to ensure algorithmic convergence) and *sd* constrained to be greater than zero. All analyses were performed using the statistical programming language R (downloaded from http://www. R-project.org, version 2.15.3; R Foundation for Statistical Computing, Vienna, Austria).

a in Equation 1 represents the asymptotic maximum response probability, that is, the probability that the observer would respond to an arbitrarily high contrast stimulus (in the absence of extraneous light scatter), and was constrained in the fitting process to be between 0 and 1. For a healthy location, *a* should equal *1-fn*, where *fn* represents the proportion of false negative responses. However at more damaged locations, the fitted asymptotic maximum was often well below 1.[12]

In order to assess the fitted asymptotic maxima, confidence intervals were derived using bootstrap analyses. For each location and stimulus size, 500 sets of response data were generated, with the number of responses at each contrast simulated by repeated sampling from a binomial distribution with response probability equal to that observed in the experiment. An FOS curve was fitted to each of these resampled sets of response data in the same manner described above, giving 500 bootstrapped FOS curves and their fitted parameters. Empirical 90% confidence intervals for the fitted asymptotic maximum were obtained using the fifth and 95th percentiles of the distribution of these 500 values.

For each location, two contrasts were extracted based on the fitted FOS curve, *c50* and *c25*, both expressed on a decibel scale. *c50* was defined as the reciprocal of the contrast to which the participant would respond on 50% of stimulus presentations. Similarly, *c25* was defined as the reciprocal of the contrast to which the participant would respond on 25% of stimulus presentations. In the event that the false positive rate *fp* equaled *1-a*, the FOS curve is rotationally symmetrical (i.e., the probability of responding to a stimulus of contrast *cs*+1 equals the probability of failing to respond to a stimulus of contrast *cs*-1). In this case, *c50 = cs*, and so *c50* is equivalent to the conventional contrast sensitivity. However, this was often not the case.

For the first analysis, the value of *c50* was compared against the response probability for the maximal 3.7-dB stimulus for each stimulus size (at

locations where this contrast was tested). The aim was to determine the value of *c50* at which a split-point occurs, such that for locations below this split-point the participant does not always respond, even to the maximal stimulus. This observed response probability was used in preference to the parameter *a* in order to reduce the potential for results being driven by artefacts of the fitting process. At some locations, the asymptotic maximum *a* was below 0.5, indicating that the detection threshold (and hence sensitivity) in its conventional definition is undefined; therefore analyses were repeated using *c25*, to reduce potential biases caused by excluding locations at which *c50* could not be calculated. In a second analysis, *c50* and *c25* were compared directly between stimulus sizes III and V.

It became clear from the results that the response probability for a 3.7-dB stimulus declines as *c50* and *c25* decrease. In order to determine the split-point at which this decline begins, segmented (hockey-stick) models were fit to the data. These models take the form:

$$Above\ s : Probability = m$$

$$Below\ s : Probability = m - (s - c)\lambda \quad (2)$$

The parameters *m* (the maximum response probability at less severely damaged locations), *s* (the split-point, in dB) and λ (the slope of the function below the split-point *s*) were fit using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm[26] to minimize the root-mean-squared perpendicular distance of the data points from the fit. The optimization was performed for both stimulus sizes simultaneously, with a common value for the parameter *m*, on the assumption that in healthy eyes the asymptotic maximum response probability *m* will equal *1-fn*, which is assumed to be independent of stimulus size. Confidence intervals (95%) for *s* were generated empirically from 1000 bootstrapped resamplings of the data, with each random resampling consisting of 140 locations (the same as the number in the complete dataset). The resulting value of *s* indicates the threshold contrast at which the response probability for the maximum stimulus begins to decline.

Results

Figure 1 shows the fitted FOS curves at four locations from different participants in the study. In
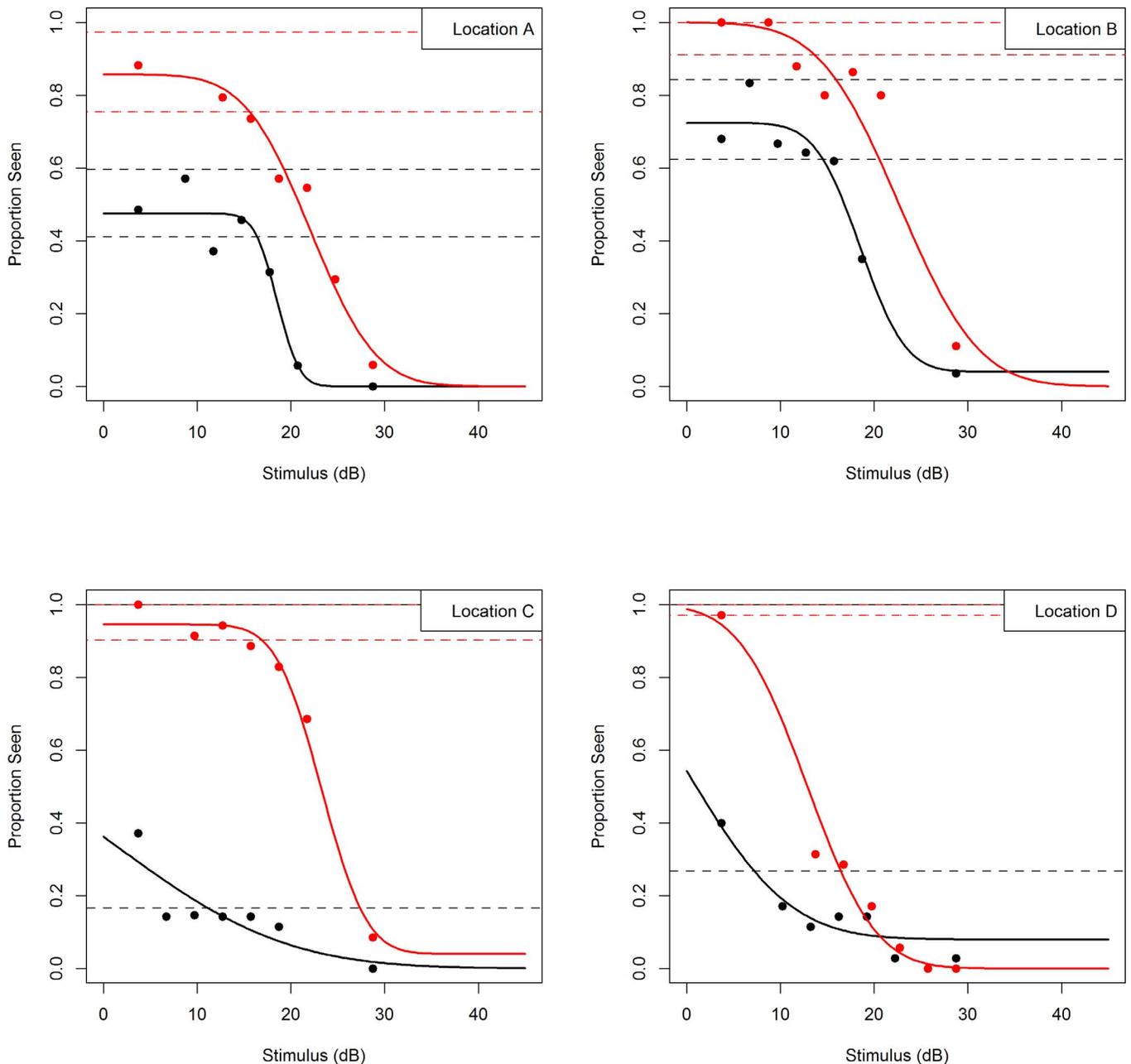
FIGURE 1. Examples of FOS curves obtained at four locations. In each case, the *black symbols* represent the response probability at each stimulus contrast for a Size III target; the *solid black line* shows the best fit curve through those data; and the *horizontal dashed black lines* show the 90% confidence interval for the fitted asymptotic maximum response probability, obtained empirically through bootstrapping. The *red points* and *lines* show the equivalent for a Size V target.

each case, the black symbols show the response probability for Size III stimuli, and the red symbols for Size V stimuli. The solid curve in each case shows the best fit FOS curve. The horizontal lines show the 90% confidence interval for the asymptotic maximum, derived empirically from bootstrapping. In the first two cases (Locations A and B), the response probability appears to reach its maximum by approx-

imately 15 to 19 dB. The confidence intervals for this maximum are quite tight, and do not extend above 90%. In the first example (Location A), the true asymptotic maximum is likely between 40% and 60% for a Size III stimulus, and 75% and 95% for a Size V stimulus. In the second example (Location B), the true asymptotic maximum is slightly higher, and may be as high as 85%; yet this is highly unlikely to be due to
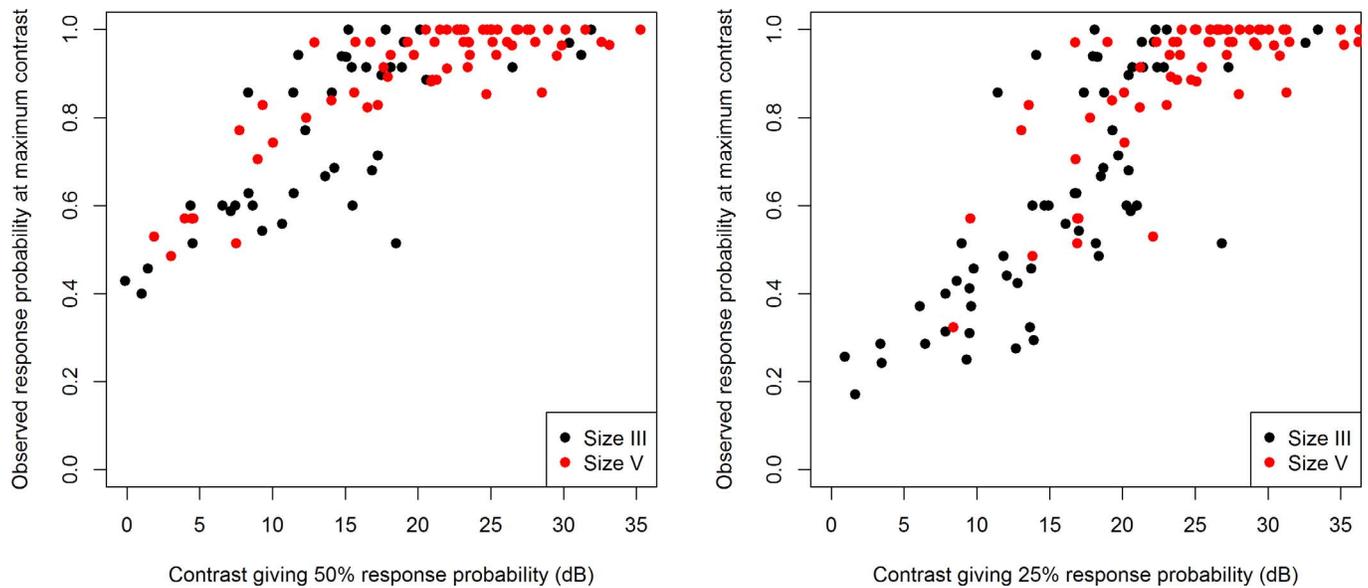
FIGURE 2. The observed response probability for a maximal 3.7-dB stimulus at each location, out of 35 trials, plotted against the contrast (dB) that would result in 50% (*left*) or 25% (*right*) response probability according to the fitted frequency-of-seeing curve. *Black points* are from testing with a size III stimulus, and *red points* are from testing with a size V stimulus.

false negative responses, since the response probability for Size V stimuli reached 100%. In the second two cases (Locations C and D), the fits to the data are worse. In both cases, we would contend that the response probability has reached its asymptotic maximum by 15 to 19 dB, with the further increase at the highest contrast being caused by light scatter to other, more sensitive, parts of the retina. However the confidence intervals for the asymptotic maximum are wide, and it is possible that the response probability has not yet reached its maximum. We note that the perimetric sensitivities at these locations were 12 and 16.5 dB respectively, but the observed response probability was certainly well below 50% at those contrasts. Therefore, even if the asymptotic maximum were in fact 100%, the sensitivity measures obtained from perimetry should still be considered unreliable.

Figure 2 shows the probability that the participant responded to a 3.7-dB stimulus for both stimulus sizes, plotted against either *c50* (left-hand plot) or *c25* (right-hand plot) from the fitted FOS curve. The 3.7-dB stimulus was tested at 72 locations using the size III stimulus, per the protocol above, either because the location was one of the worst two for that participant (70 locations) or because the perimetric sensitivity was 13 dB (two locations). At 26 of these locations the fitted asymptotic maximum response probability for the size III stimulus was less than 50%, and so *c50* (the contrast resulting in 50% response) is not defined, and the location does not appear on the

left-hand plot. At 11 locations this asymptotic maximum was less than 25% and so the location also does not appear on the right-hand plot. For the size V stimulus, the 3.7-dB contrast stimulus was tested at 70 locations (the worst two locations per participant), of which only one had an asymptotic maximum less than 50% (the fitted maximum at this location was 14% so it appears on neither plot).

It is clear from Figure 2 that the plots are similar for size III (black) and size V (red) stimuli. When a segmented hockey-stick model was fit to the data on each plot, the split-points when using *c50* (left-hand plot) were 23.1 dB for size III (95% confidence interval 20.0–27.0 dB, from bootstrapped resampling) and 20.0 dB for size V (15.8–23.6 dB), with a fitted response probability of 97.0% for both stimulus sizes above the respective split-points. The true difference may be smaller than this, since the segmented fit for size III stimuli is likely biased toward a flatter slope below the split-point and hence a higher estimate for the split-point, due to the exclusion of the majority (26 out of 33) of locations whose observed response probability at 3.7 dB was below 50% (due to their fitted asymptotic maximum *a* also being below 50%, causing an inability to calculate *c50*). Using *c25* (right-hand plot) reduces this potential for bias since only 11 locations were excluded on this basis. For this plot, the split-points were 25.2 dB for size III (95% confidence interval 23.3–29.0 dB) and 25.0 dB for size
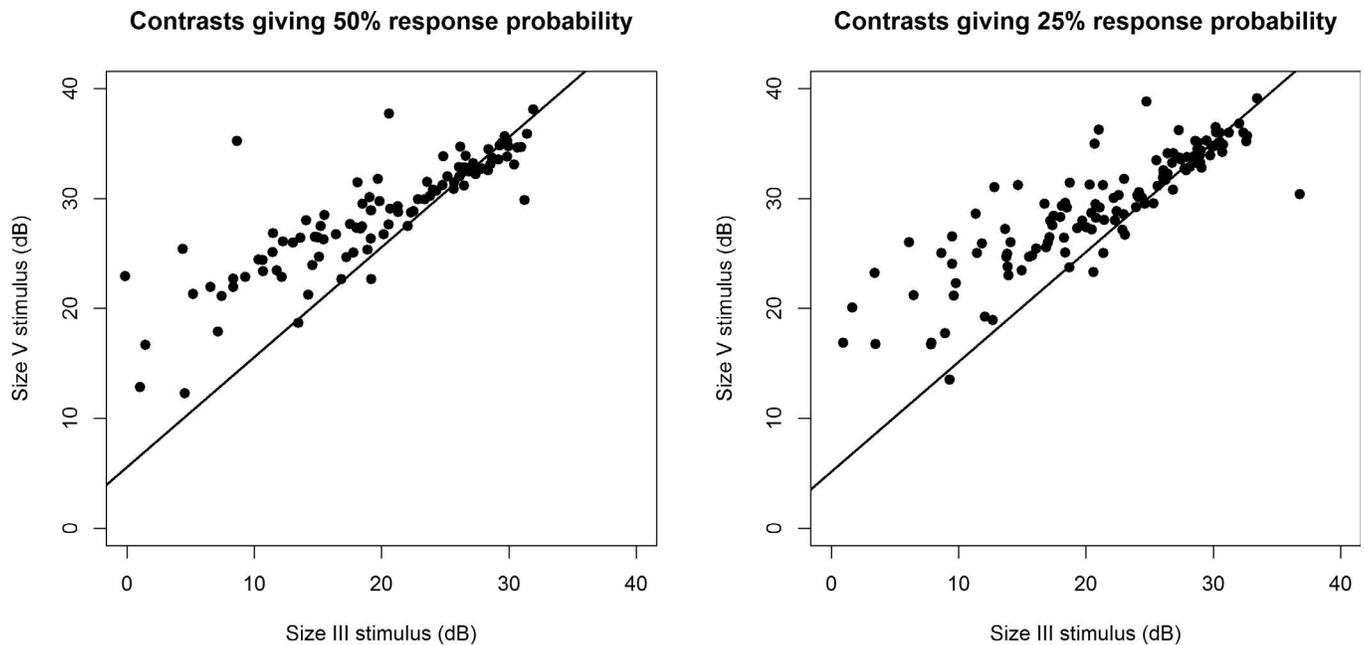
FIGURE 3. The contrasts (dB) that would give 50% (*left*) or 25% (*right*) response probability for each stimulus size, based on the fitted frequency-of-seeing curves, at each location tested. The *solid lines* give the difference between the stimulus sizes, based only on those locations with sensitivity above the split-points derived from Figure 2, indicating that a maximal contrast stimulus would be reliably detected. Therefore the *line* is based on locations with *c50* greater than or equal to 23.1 dB and *c25* greater than or equal to 25.2 dB, respectively, for size III stimuli. Locations for which the asymptotic maximum response probability was below 50% (*left*) or 25% (*right*) for either stimulus size are omitted.

V (22.0–26.8 dB), with a common response probability of 96.1% above the split-points.

The models above using *c50* suggest that when the sensitivity equals 15 dB (the lowest limit of the reliable stimulus range found in our previous study[12]), the maximum observed response probability would be 79.8% for size III and 85.7% for size V. Even though these asymptotic maxima have started to decline, perimetric testing algorithms may still be able to make reasonable estimates of sensitivity at some of these locations, but the likelihood of a misleading estimate is increased.

For the size III stimulus, 41 of the 140 locations were above the split-point based on *c50*, (i.e., their fitted sensitivity was ≥ 23.1 dB; and 46 locations were above the split-point based on *c25*). For the size V stimulus, 112 of the 140 locations were above the split-point for *c50*, and 107 for *c25*. In neither case were there any locations that were below the split-point for size V but above the split-point for size III. Even though the split-points are not significantly different for the two stimulus sizes, the increase in sensitivity observed using a larger stimulus size means that more locations remained within the reliable stimulus range.

Figure 3 compares *c50* (left) and *c25* (right) between stimulus sizes. Among the 41 locations that were above the split-points derived in the previous paragraph for both of the stimulus sizes, *c50* (which in this case corresponds closely to the perimetric sensitivity) was on average 5.6 dB higher for size V than size III stimuli, as indicated by the diagonal line on the left-hand plot. This difference was significant with P less than 0.001, based on a generalized estimating equation model[27] (predicting the difference between stimulus sizes with an intercept term as the sole predictor, adjusting for multiple tested locations per participant). Equivalently, among 46 locations above the split-points based on 25% response probability (right-hand plot), *c25* was on average 5.1 dB higher for size V than for size III stimuli ($P < 0.001$).

Discussion

Increasing the stimulus size has been suggested as a method to counteract the variability inherent in perimetry.[18–21] A larger stimulus could be used for all testing, or the stimulus size could be increased when testing is being conducted at damaged visual

field locations. Our recent findings showed that for size III stimuli, the lower limit of the reliable stimulus range for clinical perimetry appears to be 15 to 19 dB. Below this value, measures of sensitivity from clinical perimetry were only weakly associated with more accurate measures obtained from FOS curves, with R^2 less than 0.1. This could explain much of the high variability seen below this limit.[12] This study sought to use that new understanding to confirm, quantify and explain the benefits of using larger stimuli for perimetry. We found no evidence that use of a size V stimulus decreased the lower limit of the reliable stimulus range beyond 15 to 19 dB. However, using a size V stimulus resulted in a higher sensitivity at the same location. This higher sensitivity means that a location will not reach the lower limit of reliable testing until later in the disease process, resulting in more reliable and less variable estimates of sensitivity at damaged visual field locations.

Although we report 15 to 19 dB as the lower limit of the reliable stimulus range, that does not mean that sensitivities below this point are entirely uninformative. Firstly, they provide the information that the sensitivity is likely below this cutoff, yet so long as the reported sensitivity is greater than or equal to 0 dB at that location some function still remains. Further, a location whose maximum response probability is, for example, 80% will tend to result in higher estimates of sensitivity than a location whose maximum is only 20%, since there is a greater likelihood of obtaining a response to one or more stimuli. However, the reported perimetric sensitivity does not equate to a true psychophysical detection threshold, and so the perimetric sensitivity should not be considered reliable.

If the lower limit of the reliable stimulus range is approximately the same for size V as for size III, then the benefits of using size V would be entirely attributable to the higher sensitivity observed when using this larger stimulus, which we found to be on average 5.6 dB. This estimated difference is slightly smaller than the 7.6 dB that has been previously reported by Choplin et al.,[15] but within the confidence limits for that study; and within the range reported by Swanson et al.[17] in eyes free of disease. Even though our findings place a limit on the amount of benefit that can be gained from using size V stimuli, this would still represent a considerable advantage. For a rapidly progressing eye, in which the true pointwise sensitivity is worsening at −1 dB/yr, this represents more than 5 years of additional useful and reliable test results. For a treated eye worsening at a less rapid

rate, use of size V stimuli could provide many years of reliable clinical information that would not have been available using size III stimuli.

The idea that the lower limit of the reliable range of perimetry would be relatively independent of the stimulus size in our experiment has a sound logical underpinning, based on our new understanding of the reasons behind this limit. Our hypothesis is that when contrast is very high (i.e., a low dB stimulus), the responses of individual RGCs saturate, with their firing rate limited by factors such as the cell's refractory period.[10,11] The firing rate caused by a 12-dB stimulus will be very nearly the same as the firing rate caused by a 2-dB stimulus. This saturation effect may be responsible for the perimetric response probability reaching an asymptotic maximum below 15 to 19 dB because further increasing the contrast of the stimulus may not affect the spiking frequencies of the RGCs. Increasing the stimulus area has its greatest effect on primate RGC firing rate when the stimulus is smaller than the RGC's receptive field center.[28] Mean RGC receptive field size is smaller than the size III stimulus over much of the central visual field,[29] so RGC response saturation could be expected to occur at similar contrasts for size III and size V stimuli.

When the fitted asymptotic maximum is below 100%, there are in fact two possible causes. We would contend that the response probability really does asymptote to some maximum below 100%, making the perimetric sensitivity unreliable. However, it has been shown that estimates of the asymptotic maximum can be unstable, hence the wide confidence intervals observed at some locations.[30,31] It is therefore possible that the response probability actually would have gone up to 100% (other than possibly a few false negatives), but that the seven contrasts tested were not well placed on the transition zone of the observer's frequency of seeing curve, causing a poor fit of the FOS curve to the data. The contrasts tested were centered on the perimetric sensitivity, and so the fact that even a range of ±9 dB either side of this does not cover the true 50% point on the frequency of seeing curve, despite where the algorithm is designed to converge, means that the perimetric sensitivity is unreliable. Hence, whichever of these two explanations holds true for a given location, the perimetric sensitivity remains unreliable in either case. Although it is possible that the asymptotic maximum is being underestimated by our methodology, the main conclusions still hold; namely that low perimetric sensitivities do not reliably

represent the true psychophysical contrast sensitivity, and that use of a Size V stimulus extends the range of disease severities over which reliable measures can be obtained.

Firm conclusions about RGC saturation for size III and size V stimuli cannot be drawn from the physiological literature. Certainly, the true situation is more complex than presented above. At any given eccentricity there are a range of RGC receptive field sizes, with the mean size increasing with eccentricity. Few studies of primate RGCs have used stimuli similar to those used in clinical perimetry. There is also some debate about the role of inhibitory surrounds for circular stimuli.[32] In this study, we did not find evidence that the contrast beyond which the asymptotic maximum detection probability reduces was affected by either stimulus size or the location within the visual field. However this may be because we did not have sufficient power to detect such an effect, since it was not a primary aim of the study. The mean eccentricity of the locations tested in this study was 15.7°, with a range from 4.2° (at locations ± 3°, ±3°) to 22.9° (at locations ± 21°, ±9° and ±9°, ±21°). Therefore, further testing would be needed to determine whether our conclusions hold in the peripheral visual field.

By a similar logic, our conclusions cannot be extrapolated to smaller stimuli such as size I, since they may not cover even the center of an RGC receptive field. Given these factors, it cannot be concluded that saturation occurs at exactly the same contrast for each stimulus size. However, our results support the idea that any difference between saturation effects for stimulus sizes III and V is small, at least at the eccentricities tested.

The above interpretation also assumes negligible effects of light scatter. In reality, as the stimulus contrast and/or size increase, there will be a greater amount of light scattered to other locations within the visual field, potentially causing a small increase in the response probability. However, such an increase due to scattered light would not be informative of the true sensitivity at the location being tested. Indeed, this represents another argument in favor of limiting testing to lower contrasts (higher decibel values) than 15 dB. Instead of continuing to increase the stimulus contrast as disease progresses, the stimulus size could be increased, reducing the effect of light scatter (because a lower contrast is required for detection) and producing more reliable sensitivity estimates.

The effective retinal contrast at the RGC receptive fields will be lower than that presented by the perimeter, particularly in the presence of substantial media opacity. It would therefore be expected that in eyes with media opacity, the stimulus contrast at which RGC responses saturate would be altered accordingly, and reliable sensitivities may be measurable even below 15 dB.[33] Such eyes were not tested in this study. However, it would be expected that this would alter the lower limit of the reliable stimulus range for all stimulus sizes, so it is unlikely to have a major impact on comparisons between size III and size V stimuli.

It can be seen in Figure 3 that both $c50$ and $c25$ decline more rapidly due to disease with size III than with size V when the location is within the reliable stimulus range for both sizes. This is consistent with prior findings that the "size effect" (the difference in perimetric sensitivities for size V vs. size III) is greater in glaucomatous defects.[14,34,35] A similar finding has been reported in patents with retinitis pigmentosa.[16,17] There have been several efforts to define the physiological changes responsible for this finding,[36,37] but these are beyond the scope of this manuscript. The difference between stimulus sizes in both $c50$ and $c25$ increases still further at more damaged locations. This may be because, due to our emphasis on locations with perimetric sensitivity between 6 and 18 dB using size III stimuli, the majority of those locations were outside the reliable stimulus range for size III but remained within that range for size V.

While there appear to be benefits associated with using larger stimuli in terms of extending the effective dynamic range later into the disease process, it is possible that a larger stimulus would hamper detection of small, early defects, by stimulating not just the developing scotoma but also surrounding healthier regions of the visual field. Wall et al.[38] recently reported that size V stimuli appeared to be able to detect early visual field loss just as well as size III stimuli, but that this ability decreased when further increasing the stimulus size to size VI (3.44° diameter). Increasing the stimulus size helps but does not provide a panacea to the problems of variability, since it cannot be increased indefinitely without causing other problems. Additionally, other strategies may be better for leveraging information from the visual field than altering the stimulus size, for example individualizing test locations. It has been suggested that it may be useful to place extra test locations near the fovea,[39] or in regions where the sensitivity

gradient is greater,[40,41] to produce more accurate measures of the depth and extent of scotomata.

Conclusions

Our results support the idea that use of a size V stimulus instead of size III could improve the use of perimetry later in the course of glaucoma. As expected, the larger stimulus increases sensitivity, especially at greater eccentricities.[13] However, the lower limit of the reliable stimulus range, that is, the range over which the perimetric estimate of sensitivity relates closely to the psychophysical detection threshold, appears to be very similar for size III and size V stimuli, and so this limit is not reached until later in the disease process when using the larger stimulus. Therefore, reliable sensitivities may be obtained in severe glaucoma by increasing stimulus size when they cannot be obtained by increasing stimulus contrast. It remains to be seen whether it is preferable to increase the stimulus size when sensitivity declines,[38] or to use a larger stimulus at all disease severities.

Acknowledgments

References

1. Heijl A, Lindgren G, Lindgren A. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol.* 1987;105:1544–1549.
2. Heijl A, Lindgren G, Olsson J. Test retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1989;108:130–135.
3. Werner E, Petrig B, Krupin T, Bishop K. Variability of automated visual fields in clinically stable glaucoma patients. *Invest Ophthalmol Vis Sci.* 1989;30:1083–1089.
4. Piltz J, Starita R. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1990;109:109–110.
5. Chauhan BC, House PH. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology.* 1991;98:79–83.
6. Henson D, Chaudry S, Artes P, Faragher E, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417–421.
7. Spry P, Johnson C, McKendrick A, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci.* 2001;42:1404–1410.
8. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer simulation. *Invest Ophthalmol Vis Sci.* 2003;44:4787–4795.
9. Artes PH, Hutchison DM, Nicolela MT, LeBlanc RP, Chauhan BC. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci.* 2005;46:2451–2457.
10. Kaplan E, Shapley R. The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc Natl Acad Sci U S A.* 1986;83:2755–2757.
11. Swanson WH, Sun H, Lee BB, Cao D. Responses of primate retinal ganglion cells to perimetric stimuli. *Invest Ophthalmol Vis Sci.* 2011;52:764–771.
12. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology* 2014;121:1359–1369.
13. Sloan LL. Area and luminance of test object as variables in examination of the visual field by projection perimetry. *Vision Res.* 1961;1:121–122.
14. Wilensky J, Mermelstein J, Siegel H. The use of different-sized stimuli in automated perimetry. *Am J Ophthalmol.* 1986;101:710–713.
15. Choplin NT, Sherwood MB, Spaeth GL. The effect of stimulus size on the measured threshold values in automated perimetry. *Ophthalmology.* 1990;97:371–374.
16. Kim LS, McAnany JJ, Alexander KR, Fishman GA. Intersession repeatability of Humphrey perimetry measurements in patients with retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 2007;48:4720–4724.
17. Swanson WH, Felius J, Birch DG. Effect of stimulus size on static visual fields in patients with retinitis pigmentosa. *Ophthalmology.* 2000;107:1950–1954.

18. Gilpin L, Stewart W, Hunt H, Broom C. Threshold variability using different Goldmann stimulus sizes. *Acta Ophthalmol*. 1990;68:674–676.

19. Wall M, Kutzko K, Chauhan B. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Invest Ophthalmol Vis Sci*. 1997;38:426–435.

20. Wall M, Woodward K, Doyle C, Artes P. Repeatability of Automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix and motion perimetry. *Invest Ophthalmol Vis Sci*. 2008;50:974–979.

21. Wall M, Woodward KR, Doyle CK, Zamba GJ. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010;128:570–576.

22. Pelli D. Uncertainty explains many aspects of visual contrast detection and discrimination. *J Opt Soc Am A* 1985;2:1508–1531.

23. Fechner GT. *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel; 1860.

24. Laming D, Laming JF. Hegelmaier: on memory for the length of a line. *Psychol Res*.1992;54:233–239.

25. Turpin A, Artes PH, McKendrick AM. The Open Perimetry Interface: an enabling tool for clinical visual psychophysics. *J Vis*. 2012;12(11):pii: 22.

26. Nocedal J, Wright S. *Numerical Optimization. 2 ed*. New York: Springer; 2006.

27. Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.

28. Crook JM, Lange-Malecki B, Lee BB, Valberg A. Visual resolution of macaque retinal ganglion cells. *J Physiol*. 1988;396:205–224.

29. Swanson W, Felius J, Pan F. Perimetric defects and ganglion cell damage: interpreting linear relations using a two-stage neural model. *Invest Ophthalmol Vis Sci*. 2004;45:466–472.

30. Swanson WH, Birch EE. Extracting thresholds from noisy psychophysical data. *Percept Psychophys*. 1992;51:409–422.

31. Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys*. 2001;63:1293–1313.

32. Cleland BG, Enroth-cugell C. Quantitative aspects of sensitivity and summation in the cat retina. *J Physiol*. 1968;198:17–38.

33. Heuer DK, Anderson DR, Knighton RW, Feuer WJ, Gressel MG. The influence of simulated light scattering on automated perimetric threshold measurements. *Arch Ophthalmol*. 1988;106:1247–1251.

34. Fellman R, Lynn J, Starita R, Swanson W. Clinical importance of spatial summation in glaucoma. In: Heijl A, ed. *Perimetry Update 1988/89*. Amsterdam, The Netherlands: Kugler & Ghedini; 1989:313–324.

35. Felius J, Swanson WH, Fellman RL, Lynn JR, Starita RJ. Spatial summation for selected ganglion cell mosaics in patients with glaucoma. In: Wall M, Heijl A, eds. *Perimetry Update 1996/1997*. Amsterdam, The Netherlands: Kugler; 1997:213–221.

36. Pan F, Swanson W, Dul M. Evaluation of a two-stage neural model of glaucomatous defect: an approach to reduce test-retest variability. *Optom Vis Sci* 2006;83:499–511.

37. Redmond T, Garway-Heath DF, Zlatkova M, Anderson RS. Sensitivity loss in early glaucoma can be mapped to an enlargement of the area of complete spatial summation. *Invest Ophthalmol Vis Sci* 2010;51:6540–6548.

38. Wall M, Doyle CK, Eden T, Zamba KD, Johnson CA. Size threshold perimetry performs as well as conventional automated perimetry with stimulus sizes III, V, and VI for glaucomatous loss. *Invest Ophthalmol Vis Sci*. 2013;54:3975–3983.

39. Hood DC, Nguyen M, Ehrlich AC, et al. A test of a model of glaucomatous damage of the macula with high-density perimetry: implications for the locations of visual field test points. *Transl Vis Sci Technol*. 2014;3:5.

40. Aoyama Y, Murata H, Tahara M, et al. A method to measure visual field sensitivity at the edges of glaucomatous scotomata. *Invest Ophthalmol Vis Sci*. 2014;55:2584–2591.

41. Chong LX, McKendrick AM, Ganeshrao SB, Turpin A. Customized, automated stimulus location choice for assessment of visual field defects. *Invest Ophthalmol Vis Sci*. 2014;55:3265–3274.