Effect of a Variability-Adjusted Algorithm on the Efficiency of Perimetric Testing

Stuart K. Gardiner

Discoveries in Sight Laboratories, Devers Eye Institute, Legacy Health, Portland, Oregon, United States

Correspondence: Stuart K. Gardiner, Devers Eye Institute, Legacy Research Institute, 1225 NE 2nd Avenue, Portland, OR 97232, USA; sgardiner@deverseye.org.

Submitted: February 7, 2014

Accepted: March 29, 2014

Citation: Gardiner SK. Effect of a variability-adjusted algorithm on the efficiency of perimetric testing. *Invest Ophthalmol Vis Sci.* 2014;55:2983–2992. DOI:10.1167/iovs.14-14120

PURPOSE. Variability in perimetry increases with the amount of damage, making it difficult for testing algorithms to efficiently converge to the true sensitivity. This study describes a variability-adjusted algorithm (VAA), in which step size increases with variability.

METHODS. Contrasts were transformed to a new scale wherein the SD of frequency-of-seeing curves remains 1 unit for any sensitivity. A Bayesian thresholding procedure based on the existing Zippy Estimation by Sequential Testing (ZEST) algorithm was simulated on this new scale, and results converted back to decibels. The root-mean-squared (RMS) error from true sensitivity based on these simulations was compared against that achieved by ZEST using the same number of presentations. The procedure was repeated after limiting sensitivities to 15 dB or higher, the lower limit of reliable sensitivities using standard white-on-white perimetry in glaucoma, for both algorithms.

RESULTS. When the true sensitivity was 35 dB, with starting estimate also 35 dB, RMS errors of the algorithms were similar, ranging from 1.39 dB to 1.60 dB. When true sensitivity was instead 20 dB, with starting estimate 35 dB, VAA reduced the RMS error from 7.43 dB to 3.66 dB. Limiting sensitivities at 15 dB or higher reduced RMS errors, except when true sensitivity was near 15 dB.

CONCLUSIONS. VAA reduces perimetric variability without increasing test duration in cases in which the starting estimate of sensitivity is too high; for example, due to a small scotoma. Limiting the range of possible sensitivities at 15 dB or higher made algorithms more efficient, unless the true sensitivity was near this limit. This framework provides a new family of test algorithms that may benefit patients.

Keywords: perimetry, testing algorithm, computer simulation

Functional testing for glaucoma is limited by variability. Using a Humphrey Field Analyzer (HFA; Carl-Zeiss Meditec, Inc., Dublin, CA, USA), whose decibel scale is used throughout this report, when sensitivity is near normal between 30 and 35 dB, the 90% confidence limits about the perimetric sensitivity covers approximately 3 dB.¹ At more damaged locations, this variability worsens.^{2,3} For example, when sensitivity is 20 dB, the 90% confidence interval for perimetric sensitivity (i.e., the 5th to the 95th percentile) is approximately 12 dB wide.¹ This makes clinical detection of true functional damage challenging, and necessitates a series of several visual fields to confidently assess the rate of visual field change.⁴ Existing perimetric testing algorithms, such as the Swedish Interactive Testing Algorithm (SITA),⁵ German Adaptive Thresholding Estimation,⁶ and Zippy Estimation by Sequential Testing (ZEST),^{7,8} aim to minimize this variability subject to various constraints, but they are all limited by the need to maintain a short test duration, so that the reliability of subject responses is not compromised by fatigue.^{9,10}

Perimetric variability is not merely caused by subjects making response errors. Experiments using frequency-of-seeing (FOS) curves indicate that the probability of responding to a stimulus a fixed amount higher (lower contrast) than the true psychophysical threshold (which is conventionally defined in perimetry as the contrast responded to on 50% of presentations) is substantially higher in damaged areas than at locations

with normal sensitivity.² For example, if the true sensitivity is 35 dB, the probability of responding to a 39-dB stimulus has been reported as being 5%, whereas if the true sensitivity is 20 dB, the response probability for a 24-dB stimulus is 25%.² If the subject responds to such a stimulus, testing algorithms assume that the sensitivity is probably greater than the contrast of that stimulus. The algorithm will therefore have difficulty converging to the correct sensitivity without requiring an exorbitantly large number of stimulus presentations. This is problematic because in clinical perimetry it is desirable to present only three or four stimuli per location to test the entire central visual field in close to 5 minutes. The “flattening” of the FOS curve that occurs in regions of glaucomatous damage therefore increases the variability about estimates of perimetric sensitivities.

In this study, a new testing algorithm is described that increases the step size when estimating sensitivity at damaged visual field locations in direct relation to the variability. This increases the likelihood that the next stimulus will be presented at a contrast that gives a substantially different response probability. By computer simulation, we determine whether this technique could reduce the variability about perimetric sensitivities without increasing the test duration.

A second major contributory factor for the increased variability in regions of glaucomatous damage is that retinal ganglion cell responses saturate when presented with high-

contrast stimuli.¹¹ This means that the response probability asymptotes at some fixed contrast, and further increasing the stimulus contrast will not increase the response probability, hampering the ability of test algorithms to converge to the true sensitivity. We have recently shown that this causes perimetric sensitivities within the central visual field to be unreliable below 15 to 19 dB, with little relation to the true sensitivity as measured using FOS curves.¹² Testing algorithms can therefore be shortened by stopping testing once this contrast has been reached, rather than continuing testing with stimuli of 10 dB or 5 dB, for example, which might not provide further useful information about the true sensitivity. In this study, we assess the potential benefits on variability of using test algorithms that are designed to terminate once the sensitivity has been determined to be below 15 dB, allowing more accurate assessments to be made at locations with higher sensitivities within the same average test duration.

To assess the effect of these two proposed changes to perimetric testing algorithms, a computer simulation model is used, such that the true FOS curve is known and can be used to give the probability that the “subject” will respond to any given stimulus contrast. A solitary location is simulated in each case. The simulation approach allows very large numbers of test sequences to be constructed with known characteristics, such that the results are influenced solely by the changes made to the algorithm and not by external factors (e.g., fatigue). This technique, therefore, reveals promising test algorithms that later can be tested on patients to confirm the findings.

METHODS

Simulated FOS Curves

Given a sensitivity *Sens*, expressed on the decibel scale as in standard automated perimetry using the HFA, the probability that the observer will respond to a stimulus of contrast *Stim* dB can be modeled as follows:

$$\psi(\text{Stim}, \text{Sens}) = FP + (1 - FP - FN)\Theta((\text{Sens} - \text{Stim})/SD)$$

In this equation, $\Theta(X)$ represents a cumulative Gaussian function, such that $\Theta(-\infty) = 0$, $\Theta(0) = 0.5$, and $\Theta(\infty) = 1$. *FP* and *FN* represent the probability of a false-positive response (responding to a stimulus that was not detected) and a false-negative response (failing to respond to a stimulus that was detected), respectively. The primary results assumed *FP* = 5% false-positive responses and *FN* = 5% false-negative responses. The variability is defined by the SD of the FOS curve *SD*, which is taken from the formula of Henson et al.²: $SD = \exp(-0.081 \times \text{Sens} + 3.27)$.

When the simulation “presents” a stimulus, the response is generated using a random number generator with uniform distribution in the range (0, 1). If that random number is greater than $\psi(\text{Stim}, \text{Sens})$, the simulation continues as if there was a response to the stimulus. If the random number generated is less than or equal to $\psi(\text{Stim}, \text{Sens})$, the simulation continues as if there was no response. All simulations were performed using the statistical programming language R.¹³

ZEST Testing Algorithm

Although SITA is a commonly used testing algorithm in clinical perimetry,⁵ not all of the details of its implementation are publically available. Therefore, the current simulation is built on the ZEST algorithm,⁷ which has similar variability for a given sensitivity,¹⁴ and may be less variable when a reasonably accurate initial guess of the true sensitivity is made (e.g., based

on information from other locations in the visual field).¹⁴ ZEST is a Bayesian thresholding algorithm and requires an initial prior probability density function (pdf). In this simulation, the prior pdf was set as a uniform distribution spanning the range (−3 dB, 40 dB). The lower end of this range is below the nominal maximum contrast stimulus (0 dB) to aid in algorithmic convergence if the true sensitivity happens to be at or near 0 dB. At each step, a stimulus presentation is simulated as above, at the mean of the current distribution (i.e., the first stimulus is at 18.5 dB, which is the mean of the prior pdf). The posterior probability that the sensitivity is $\hat{s}$ is then calculated by multiplying the prior probability that the sensitivity is $\hat{s}$ by either $\psi(\text{Stim}, \hat{s})$ if the simulated observer “responded” to the stimulus, or $1 - \psi(\text{Stim}, \hat{s})$ if the simulated observer failed to “respond” to the stimulus. The mean of the resulting posterior distribution is used as the new estimate of sensitivity at that location, and also gives the contrast of the next stimulus to be displayed in the sequence.

The accuracy and speed of the ZEST algorithm can be enhanced when an initial guess of the sensitivity is provided, based on information such as sensitivities that have already been measured at other locations in the visual field, the normal hill of vision, test results from previous visits, and so forth. In a clinical situation where the entire field is being tested rather than just one location, this is an essential tool to reduce the overall test duration. Therefore, the simulation experiment was repeated with an initial prior pdf weighted toward an “initial guess,” rather than using a uniform prior pdf. This weighted prior pdf was calculated using $0.1 + \Phi([S - \text{Guess}]/5)/\Phi(0)$, where $\Phi(x)$ represents the pdf of a Gaussian distribution with mean 0 and SD 1; and then normalized such that the integral of the prior pdf over the available range (−3 dB, 40 dB) equaled one. Therefore the weighted prior pdf had a constant, nonzero probability of any sensitivity within the range, plus an increased probability around the initial guess according to a Gaussian distribution with an SD of 5 dB. Other values of this SD were tried in initial simulations (results not shown); 5 dB represented a compromise between being small enough to increase speed and accuracy when the initial guess was close to the true sensitivity, while being large enough that the deleterious effect of an inaccurate initial guess was acceptably small. Note that due to the baseline nonzero probability for any sensitivity within the range, the mean of this weighted prior (which will be the first stimulus contrast presented) will typically not equal the initial guess, but will be between this initial guess and the midpoint of the range. This reduces the effect that an incorrect initial guess will have on the final estimated sensitivity.

For the first analysis, a “true sensitivity” was chosen, and its associated FOS curve was generated. Series of simulated stimuli were then “presented.” After each stimulus presentation, the variance of the posterior pdf was calculated as the numerical integral of $P(s) \times (s - \text{Mean})^2$, where $P(s)$ represents the posterior probability that the true sensitivity is *s*, and *Mean* represents the mean of the posterior pdf. The SD of the posterior pdf was calculated as the square root of this value. The algorithm was said to terminate once the SD of the posterior pdf was less than a criterion value SD_{Crit} , and the estimate of sensitivity was given by the mean of the posterior pdf at that point. After 1000 simulations for each possible value of SD_{Crit} , the mean number of stimuli that had been presented in the sequence to achieve that criterion and the remaining root-mean-squared error (RMS error; the true sensitivity minus the estimate from the posterior pdf) were calculated. The RMS error was plotted against the mean number of presentations taken to reach this criterion, to show the variability that should be expected for sequences with a given number of stimulus presentations.

FIGURE 1. Examples showing how sensitivities are transformed from the decibel scale onto a VAS for use in the new testing algorithm. When two contrasts are 1 SD apart on the decibel scale, they are 1 unit apart on the VAS. The transform is based on the SD of an FOS curve at a location with that sensitivity, taken from the equation of Henson et al.¹⁵

To give an overall idea of the performance of the algorithm, the above procedure was repeated for different “true sensitivities,” at 0.5-dB intervals in the range (15 dB, 35 dB); 1000 runs were simulated for each true sensitivity. The RMS error after three or four presentations was recorded, and plotted against true sensitivity. This was carried out using a flat initial prior pdf, and using various values for the initial guess of the sensitivity.

Variability-Adjusted Algorithm (VAA)

As described above, the SD of FOS curves increases with damage. It has been suggested² that this increase can be described by the equation $SD(Sens) = \exp(-0.081 \times Sens + 3.27)$. The minimum detectable difference, the smallest difference between sensitivities that can confidently be ascribed to pathophysiologic difference rather than variability, is proportional to this SD. The variability-adjusted algorithm

FIGURE 2. The possible sequences of stimulus presentations that could occur when the initial guess of sensitivity from prior information is 35 dB. The *top row* shows results using the standard ZEST algorithm; the *bottom row* shows results using our new VAA. The *left-hand plots* allow sensitivities down to 0 dB, whereas the *right-hand plots* limit sensitivities to being 15 dB or higher.

FIGURE 3. The possible sequences of stimulus presentations that could occur when the initial guess of sensitivity from prior information is 20 dB. The *top row* shows results using the standard ZEST algorithm; the *bottom row* shows results using our new VAA. The *left-hand plots* allow sensitivities down to 0 dB, whereas the *right-hand plots* limit sensitivities to being 15 dB or higher.

seeks to ensure that the step size between consecutive stimulus contrasts increases with this minimum detectable difference. This should ensure that the probability that the subject will respond changes substantially between consecutive stimuli.

Suppose that two locations have true sensitivity 30 dB and 20 dB, and have just responded to stimuli of that contrast (i.e., 30 dB and 20 dB, respectively). In the standard ZEST algorithm, the next stimuli to be presented might be 36 dB and 26 dB (the exact values will depend on the number of stimuli already presented, the initial guess used, and so forth). For the first location, the probability that the subject will respond to this 36-dB stimulus, leading the testing algorithm astray, is very low due to the low variability (the FOS curve will likely have $SD \propto 2.3$ dB). However, the probability that the subject responds to the 26-dB stimulus at the second location is much higher, because the FOS curve will have a higher SD, approximately equal to 5.2 dB. The variability-adjusted algorithm aims to ensure that the step size for this second location would be increased proportionally to the SD of the FOS curve at that location. The response probabilities for the next stimulus presentation are then very nearly the same between these two locations.

First, the algorithm expresses sensitivities on a new variability-adjusted scale, which we shall refer to as VAS, such that the SDs of FOS curves are constant across sensitivities.

Starting (arbitrarily) at 40 dB, each 1-dB decrease in sensitivity is equivalent to a decrease of $1/SD(Sens)$ on the new scale. This means that a 1-unit difference on the VAS scale is always equivalent to a change on the decibel scale that equals the SD of a FOS curve with that sensitivity. Hence $40 \text{ dB} \leftrightarrow 40 \text{ VAS}$ (by definition); and below this point:

$$x\text{dB} \leftrightarrow 40 - \int \left(1/SD(s) \right) ds,$$

where the integral is calculated over the range $(Sens, 40)$

$$x\text{dB} \leftrightarrow 40 - 1/0.81$$

$$\times [\exp(0.081 \times 40 - 3.27) - \exp(0.081 \times Sens - 3.27)]$$

This means that a 1-unit change on the VAS scale corresponds to a larger decibel difference when sensitivity is reduced, and vice versa, as illustrated in Figure 1. For example, a change from 40 VAS to 35 VAS $\leftrightarrow$ a difference of 6.7 dB, whereas a change from 35 VAS to 30 VAS $\leftrightarrow$ a difference of 15.5 dB.

A standard ZEST algorithm is performed but using the new VAS scale, as described above. Note that the uniform prior pdf described above extended from -3 dB to 40 dB. To ensure consistency and allow direct comparisons, in the new algorithm the range of the uniform prior pdf extends from

FIGURE 4. The RMS error (true sensitivity – estimated sensitivity) achieved when the algorithm is set to terminate once the posterior pdf has SD less than some criterion value, plotted against the mean number of stimulus presentations taken to achieve that criterion, based on 1000 simulations per algorithm, when no initial guess of the sensitivity is provided to the testing algorithms, and the “true sensitivity” is 20, 25, 30, or 35 dB.

28.4 VAS to 40 VAS, the equivalent values expressed on the VAS scale. This means that the only difference between the standard ZEST algorithm and the new VAA is the transformation of the scaling used to calculate pdfs; all other aspects of the algorithm are the same. On the VAS, the SD of a simulated FOS curve is always 1 unit, regardless of its sensitivity. Once the testing algorithm has been simulated, the final estimate of sensitivity is transformed from the VAS scale back to the decibel scale, before calculation of the RMS errors, and so on, is performed. Results from 1000 simulated runs per true sensitivity can then be directly compared against those from the standard ZEST algorithm described above.

Censored Algorithms

We have recently shown that once stimulus contrast is below 15 to 19 dB, the probability that a subject responds to a stimulus asymptotes, and does not increase appreciably when the contrast is further increased, other than through the effects of light scatter (which are uninformative of the true sensitivity at the location being tested).¹² Therefore, there is little benefit to continuing testing with stimulus contrasts stronger than 15 dB, because the response probability will be the same as if the 15-dB stimuli were presented again. Responses to stimuli lower than 15 dB may be helpful in determining the likelihood that the true sensitivity is lower than 15 dB, but the same

information could be gained by presenting a 15-dB stimuli again, which also would be less likely to have caveats due to light scatter. Algorithms can therefore be made more efficient if they stop testing at 15 dB, so that the time saved can then be spent on deriving more accurate measures of sensitivity at this and other locations in the visual field.

This change will clearly reduce the number of stimulus presentations at locations with sensitivity below 15 dB. To assess the effects of the change at locations with sensitivity of 15 dB or higher, simulations using both the standard ZEST and the VAA were repeated using this smaller range of values. The uniform prior distribution used in the standard ZEST was set to cover the range (12 dB, 40 dB) instead of (–3 dB, 40 dB) (in both cases the lower limit is 3 dB below the maximum available stimulus contrast). For the VAA, the lower limit of the prior pdf was transformed to the VAS scale as outlined above (i.e., 12 dB $\leftrightarrow$ 29.3 VAS).

RESULTS

Figures 2 and 3 show the possible presentation sequences that can be obtained from each of the four algorithms, when provided with an initial guess of 35 dB or 20 dB, respectively. The first stimulus presented (the mean of the weighted prior pdf) is shown on the top row. At each stage, if the observer

FIGURE 5. The RMS error achieved using either three or four stimulus presentations, with each of the four algorithms tested. Results are based on 1000 simulations per true sensitivity, when the initial guess of the sensitivity input to each testing algorithm is 35 dB (*top row*), 30 dB (*second row*), 25 dB (*third row*), or 20 dB (*bottom row*).

responds to a stimulus, the resultant estimate of the sensitivity (which will also be the next stimulus to be presented in the sequence) is found by following the line down one level to the right (i.e., resulting in a higher dB than the previous stimulus). If the observer does not respond to the stimulus, the resultant estimate of sensitivity is found by following the line down one level to the left (lower dB). Therefore the circles on the bottom row represent the possible sensitivity estimates that could be obtained after four stimulus presentations, which are also the possible contrasts for the fifth stimulus presentation if the series were to be continued. Looking at the bottom-left plot in each case, it can be seen that the VAA results in a denser concentration of possible sensitivity estimates toward the healthier end of the range (higher decibels), with larger gaps toward the damaged end of the range where variability is greater and so the minimum detectable difference is increased. In comparison, the ZEST algorithm in the top-left plot has a more even spread of possible sensitivity estimates. For both the ZEST and VAAs, looking at the right-hand plots reveals a denser concentration of possible sensitivity estimates. This indicates that limiting the range such that contrasts beyond 15 dB are not tested allows a more accurate estimate of sensitivity, with finer precision, to be made using the same number of stimulus presentations.

Figure 4 shows the RMS error (the difference between the final sensitivity estimate and the “true” sensitivity input to the simulations) that can be achieved for different algorithm termination criteria, plotted against the mean numbers of presentations required to achieve those criteria, when no initial guess of the sensitivity is provided (uniform prior pdf). When the criterion demands a smaller SD of the posterior pdf before terminating the algorithm, the RMS error goes down, at the cost of requiring a greater number of presentations. No loss of efficiency is apparent when using the VAA based on the new scale, or when limiting the pdfs such that only stimuli of 15 dB or higher are presented.

Figure 5 show the RMS errors achieved using either three or four stimulus presentations per simulated location, for different true sensitivities between 15 dB and 35 dB, when the initial guess used is 20, 25, 30, or 35 dB, respectively. The RMS errors when true sensitivity is 20 dB or 35 dB are shown in the Table for these same four initial guesses. It is apparent that when the initial guess is reasonably accurate, the algorithms perform similarly. In some cases, the ZEST algorithm performs slightly better, but this does not occur consistently,

and the difference is small. However, when the initial guess is inaccurate, the VAA performs better than the ZEST algorithm, particularly when the algorithm is limited to sensitivities of 15 dB or higher. This is particularly noticeable in the top two rows of Figure 4 when the true sensitivity is low and so the variability is high. In these cases, there are a wide range of contrasts that would have response probability between 25% and 75%. The VAA uses larger step sizes at these high contrasts, giving a greater change in the response probability, and so the efficiency of the algorithm is improved. These situations are important because they represent the case where the algorithm is trying to detect a small, localized defect, which is often the first sign of functional damage in glaucoma.

The simulations that were provided an initial guess of 30 dB were repeated to test the robustness of the algorithms to subject errors. The top row of Figure 6 shows that the VAA is just as robust as ZEST to the introduction of 20% false-negative responses and 0% false-positive responses (for comparison, the main simulations reported in Fig. 5 assumed 5% false-positive responses and 5% false-negative responses). The bottom row of Figure 6 shows the equivalent results introducing 20% false-positive responses and 0% false-negative responses. It appears that the VAA may be slightly more robust than ZEST to false-negative responses, because these would bias the algorithm toward lower sensitivities, at which the new algorithm reduces variability compared with ZEST. However, the VAA may be slightly less robust in the face of elevated false-positive responses. Nonetheless, in both cases, the differences made by introducing these extra response errors are small.

DISCUSSION

Automated perimetry remains the principal means of assessing visual function in patients with ocular hypertension or glaucoma. Despite advances in imaging technologies, the poor correlation between structure and function,^{16,17} especially cross-sectionally, as is important when assessing a new patient,^{18,19} means that perimetry will remain an essential tool in glaucoma management for the foreseeable future. However, its limitations are substantial. The test-retest variability can be high, owing to the necessity of using short sequences of stimulus presentation per tested location driven by the need to maintain acceptably short test durations. This makes detection and quantification of both damage and progression challenging, requiring repeated testing over a

TABLE. The RMS Error Between the True Sensitivity and That Estimated by Each Algorithm After Four Stimulus Presentations, Based on 1000 Simulated Runs in Each Case

True Sensitivity, dB	Initial Guess, dB	RMS Error Using ZEST	RMS Error Using ZEST Limited at ≥ 15 dB	RMS Error Using VAA	RMS Error Using VAA Limited at ≥ 15 dB
35	35	1.53	1.39	1.39	1.60
35	30	1.47	1.52	1.57	1.66
35	25	2.87	2.92	1.99	1.66
35	20	4.84	4.60	2.23	1.94
20	35	7.43	6.20	3.66	3.71
20	30	5.24	5.00	3.99	2.19
20	25	3.61	3.37	4.35	2.21
20	20	2.88	2.61	4.23	2.38

The four testing algorithms simulated were ZEST and VAA, as described in this article, either based on the range of possible sensitivities from 0 to 40 dB or 15 to 40 dB.

FIGURE 6. The RMS error achieved from the four tested algorithms using either three (*left column*) or four (*right column*) stimulus presentations, based on 1000 simulations per true sensitivity, when the initial guess of the sensitivity input to each testing algorithm is 30 dB, after introducing 20% false-negative responses (*top row*) or 20% false-positive responses (*bottom row*) into the simulation.

prolonged period and hence delaying assignment of appropriate management strategies. Any strategies that can reduce this variability without compromising either test duration or the ability to detect localized defects could be useful in both clinical and research settings. This paper provides proof-of-principle that a relatively simple alteration to existing test algorithms, namely performing the Bayesian calculations on a variability-adjusted scale instead of a decibel scale, could reduce variability in certain circumstances, with comparatively minor deleterious effects in other circumstances.

In particular, the VAA presented here showed markedly lower RMS errors than ZEST when provided with an inaccurate initial guess of sensitivity. Use of such an initial guess based on sensitivities already measured at other locations in the visual field is vital in the interests of reducing test duration. The Full Threshold²⁰ and SITA⁵ algorithms implemented in the HFA perimeter both start by measuring the sensitivity at seed points, and use those measurements to derive the first contrast to be tested at neighboring locations. However, if this initial sensitivity guess is higher than the actual sensitivity at a location, because there is a localized glaucomatous defect at the location in question that does not extend to the seed points used, then the algorithm will tend to overestimate the sensitivity and underestimate the depth of that localized

defect. The VAA gives more accurate sensitivity measurements (lower RMS error) when the initial guess is 30 dB or above and the true sensitivity is below 20 to 25 dB. As the discrepancy between the initial guess and the true sensitivity increases, so does the benefit of the new algorithm.

We have recently shown that below 15 to 19 dB, the probability of responding to a stimulus asymptotes to some maximum probability, rather than continuing to increase with contrast.¹² Presenting stimuli of contrast 5 dB, 2 dB, and so on, does not help assess the sensitivity at the location being tested, because any change in response is likely due to light scattering to remaining healthier locations elsewhere in the visual field and so is uninformative of the functional status at the location being tested. Therefore, in this study, the algorithms also were tested after limiting the range of available stimuli to 15 dB or higher, and limiting all pdfs to 12 dB or higher. For the ZEST algorithm, this gave a small but consistent improvement to the variability (as assessed by the RMS errors for a given input true sensitivity). For the VAA, the benefits were larger but less consistent. When the true sensitivity is low, allowing the pdfs to extend down to 28.4 VAS (equivalent to -3 dB) improved efficiency, by allowing the algorithm to converge toward these low sensitivities more rapidly. However, in the rest of the range, which constitutes most pointwise sensitivities in a

clinical situation, limiting the pdf at 29.3 VAS (equivalent to 12 dB) improved variability. In the peripheral visual field, a size III stimulus does not cover the entire receptive field of a retinal ganglion cell, and so response saturation will not occur until a greater contrast. Therefore the lower limits for the censored algorithm should be lowered at peripheral locations. However, both the actual sensitivity and the initial guess also would be lower, so the benefits of the VAA should be similar.

When deciding on a new testing algorithm, there are many factors that can be altered that will affect variability. These include the range of available stimulus contrasts, the termination criterion/number of stimulus presentations to be made, the shape and range of the prior pdf (this study used a flat prior, whereas the SITA uses a prior pdf weighted toward the more common sensitivities³), the formula used to weight the prior pdf based on the initial guess, and assumptions made about the likelihood of false-positive and false-negative responses. Adjusting any of these is likely to cause the performance of an algorithm to improve in some circumstances and worsen in others. For example, introducing a higher probability of false-positive errors worsens the variability achieved when the true sensitivity is low for any of the current clinical testing algorithms, and the magnitude of that worsening may vary between algorithms. In this study, all of these other factors were kept constant among the four algorithms, allowing a fair assessment of the impact of the two main changes being considered, namely variability-adjustment and limiting the range of available contrast levels. As such, this study should be viewed as a “proof of principle” rather than prescribing an optimal algorithm.

It is also important to note that the algorithms and simulations described here are specific to glaucoma. In particular, the relation between sensitivity and variability may be different depending on the disease present, and could be changed, for example, by the presence of cataract. The SITA also uses assumptions about the distribution of sensitivities that are based on normal and glaucomatous eyes,⁵ and so it may not be optimal for patients with other pathologies, even though it is commonly used in such situations.

Although the VAA is derived using the VAS scale specified above, we do not recommend wider use of this VAS scale as an alternative to the decibel scale. The decibel scale has a strong theoretical underpinning, namely that according to Fechner's Law, the human brain perceives any doubling of contrast as being the same difference.^{21–23} Subject to a few simplifying assumptions, a 100-cd/m² stimulus presented on a 10-cd/m² background will appear very similar to a 50-cd/m² stimulus presented on a 5-cd/m² background. Furthermore, the dB scale is anchored such that 0 dB represents the highest contrast that can be presented by the perimeter, and although this varies between perimeters, it makes the results easier for practitioners and patients to understand. Finally, the dB scale is easily converted to a linear scale so that the resulting sensitivity values are linearly related to structural measures,^{18,24,25} whereas values on the VAS scale would have to be converted to decibels first.

The simulations shown in this study suggest that a VAA may have benefits over a decibel-scaled Bayesian algorithm such as ZEST. By reducing the influence of the known increase in variability that occurs at damaged visual field locations, the algorithm becomes more efficient when measuring low sensitivities. This beneficial effect is most noticeable when the initial guess is that the sensitivity is near normal, such as would be the case in the presence of a small localized defect. Use of simulations allows this effect to become apparent by keeping all other factors consistent, but the true benefits on test–retest variability will need to be confirmed using patient testing.

Acknowledgments

The author thanks Shaban Demirel for assistance in the preparation of this manuscript.

Supported in part by National Institutes of Health Grant NEI R01-EY020922. The author alone is responsible for the content and writing of the paper.

Disclosure: **S.K. Gardiner**, None

References

- Artes P, Hutchison D, Nicoleta M, LeBlanc R, Chauhan B. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci*. 2005;46:2451–2457.
- 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.
- Heijl A, Lindgren G, Olsson J. Test retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
- Chauhan B, Garway-Heath D, Goni F, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92:569–573.
- Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol*. 1997;75:368–375.
- Schiefer U, Pascual JP, Edmunds B, et al. Comparison of the new perimetric GATE strategy with conventional full-threshold and SITA standard strategies. *Invest Ophthalmol Vis Sci*. 2009;50:488–494.
- Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Development of efficient threshold strategies for frequency doubling technology perimetry using computer simulation. *Invest Ophthalmol Vis Sci*. 2002;43:322–331.
- Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Performance of efficient test procedures for frequency-doubling technology perimetry in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2002;43:709–715.
- Johnson CA, Adams CW, Lewis RA. Fatigue effects in automated perimetry. *Appl Opt*. 1988;27:1030–1037.
- Hudson C, Wild J, O'Neill E. Fatigue effects during a single session of automated static threshold perimetry. *Invest Ophthalmol Vis Sci*. 1994;35:268–280.
- 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.
- Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage [published online ahead of print March 12, 2014]. *Ophthalmology*. doi:10.1016/j.ophtha.2014.01.020.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- 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.
- 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.
- Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implica-

- tions for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci*. 2012;53:6939-6946.
17. Gardiner S, Johnson C, Demirel S. The effect of test variability on the structure-function relationship in early glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 2012;250:1851-1861.
 18. Hood D, Kardon R. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. 2007;26:688-710.
 19. Nilforushan N, Nassiri N, Moghimi S, et al. Structure-function relationships between spectral-domain OCT and standard achromatic perimetry. *Invest Ophthalmol Vis Sci*. 2012;53:2740-2748.
 20. Werner E. *Manual of Visual Fields*. New York: Churchill Livingstone; 1991.
 21. Fechner GT. *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel; 1860.
 22. Rushton WAH. Visual adaptation. In: Moses RA, ed. *Adler's Physiology of the Eye*. St. Louis, MO: C. V. Mosby; 1981:652-655.
 23. Laming D. Fechner's law: where does the log transform come from? *Seeing Perceiving*. 2010;23:155-171.
 24. Garway-Heath D, Caprioli J, Fitzke F, Hitchings R. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci*. 2000;41:1774-1782.
 25. Harwerth RS, Wheat JL, Fredette MJ, Anderson DR. Linking structure and function in glaucoma. *Prog Retin Eye Res*. 2010;29:249-271.