

# Effect of Restricting Perimetry Testing Algorithms to Reliable Sensitivities on Test-Retest Variability

Stuart K. Gardiner and Steven L. Mansberger

Devers Eye Institute, Legacy Research Institute, 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](mailto:sgardiner@deverseye.org).

Submitted: June 2, 2016  
Accepted: September 19, 2016

Citation: Gardiner SK, Mansberger SL. Effect of restricting perimetry testing algorithms to reliable sensitivities on test-retest variability. *Invest Ophthalmol Vis Sci.* 2016;57:5631-5636. DOI:10.1167/iops.16-20053

**PURPOSE.** We have previously shown that sensitivities obtained at severely damaged visual field locations (<15–19 dB) are unreliable and highly variable. This study evaluates a testing algorithm that does not present very high contrast stimuli in damaged locations above approximately 1000% contrast, but instead concentrates on more precise estimation at remaining locations.

**METHODS.** A trained ophthalmic technician tested 36 eyes of 36 participants twice with each of two different testing algorithms: ZEST<sub>0</sub>, which allowed sensitivities within the range 0 to 35 dB, and ZEST<sub>15</sub>, which allowed sensitivities between 15 and 35 dB but was otherwise identical. The difference between the two runs for the same algorithm was used as a measure of test-retest variability. These were compared between algorithms using a random effects model with homoscedastic within-group errors whose variance was allowed to differ between algorithms.

**RESULTS.** The estimated test-retest variance for ZEST<sub>15</sub> was 53.1% of the test-retest variance for ZEST<sub>0</sub>, with 95% confidence interval (50.5%–55.7%). Among locations whose sensitivity was ≥17 dB on all tests, the variability of ZEST<sub>15</sub> was 86.4% of the test-retest variance for ZEST<sub>0</sub>, with 95% confidence interval (79.3%–94.0%).

**CONCLUSIONS.** Restricting the range of possible sensitivity estimates reduced test-retest variability, not only at locations with severe damage but also at locations with higher sensitivity. Future visual field algorithms should avoid high-contrast stimuli in severely damaged locations. Given that low sensitivities cannot be measured reliably enough for most clinical uses, it appears to be more efficient to concentrate on more precise testing of less damaged locations.

**Keywords:** glaucoma, diagnostics, perimetry, visual fields

Clinicians and researchers assess glaucoma by using functional testing with automated perimetry, together with structural imaging techniques. However, results from perimetry are highly variable, especially in regions of more severe damage.<sup>1,2</sup> Simulation studies suggest that reducing variability (defined as the spread of the frequency-of-seeing [FOS] curve) by 20% would enable progression to be detected, on average, one visit sooner,<sup>3</sup> and more than that for many patients.

We have previously used FOS curves to assess the effects of response saturation, whereby increases in stimulus contrast beyond a certain point no longer result in increases in the observer's response. We found evidence that this appears to occur at contrasts above 400% to 1000%, or 15 to 19 "dB" for clinical perimetry on the Humphrey Field Analyzer (Carl Zeiss Meditec, Inc., Dublin, CA, USA) with the Size III stimulus.<sup>4</sup> For example, the probability that participants responded to a 20,000% (2 dB) contrast in a deep visual field defect was typically only marginally higher than the probability that they would respond to a 2000% (12 dB) contrast at the same location, and this small increase may just 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  $R^2 < 0.1$ ,

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 dB and 12 dB. We suggested that this phenomenon might partially explain the increase in test-retest variability that is observed in moderate and severe glaucoma.<sup>4</sup>

In theory, useful information could be obtained at locations with these very low sensitivities. At a location whose asymptotic maximum response probability is 80% (i.e., the patient will respond to 80% of stimulus presentations at that location no matter how much contrast is increased), commercial testing algorithms are likely to produce sensitivity estimates of 10 to 20 dB, because it is very likely that the patient will respond to one or more stimulus presentations. By contrast, the patient is far less likely to respond to presentations at a location at which the maximum response probability is 20%, causing the algorithms to report sensitivity estimates of 0 to 5 dB. However, the high variability associated with such probabilistic responses prevents clinicians from distinguishing between these possibilities without multiple visual field tests, likely over many years, by which time the location may have deteriorated further anyway. A new algorithm could present many 15-dB stimuli at such locations to estimate the maximum response probability, but this would greatly extend the test duration in eyes with severe damage. It may therefore be more prudent to cease

testing locations whose sensitivity is below 15 dB, either reducing the total test duration or allowing time to be spent obtaining more accurate sensitivity estimates elsewhere in the visual field.

We have previously shown that “censoring” low sensitivities, so that those below 15 dB are set to equal 15 dB, did not reduce the ability to detect glaucomatous progression using pointwise analyses.<sup>5</sup> We have also suggested that it may even slightly improve the ability to detect progression using global indices such as mean deviation (MD) (Pathak M, et al. *IOVS* 2016;57:ARVO E-Abstract 3920). It has previously been shown that ceasing testing of locations with sensitivity below 10 dB does not hinder progression detection using MD, but that censoring at 20 dB resulted in lost information.<sup>6</sup> However, those studies used data that had been collected using existing testing algorithms, which give sensitivity estimates down to 0 dB. In this study, we are interested in the change in variability associated with a new testing algorithm that avoids testing with very high contrast stimuli in areas of poor sensitivity. Clinicians and researchers may use this information to design visual field testing algorithms to more accurately assess visual field sensitivity and detect visual field progression with shorter test time and reduced variability in patients with moderate to advanced glaucoma.

## METHODS

### Participants

Participants were from a tertiary glaucoma clinic at Legacy Devers Eye Institute in Portland, OR. Inclusion criteria were a diagnosis of open-angle glaucoma, and at least six test locations that were outside age-matched normal limits on both of their two most recent visual fields (tested using the Humphrey Field Analyzer [HFA; Carl Zeiss Meditec, Inc., Dublin, CA, USA] and the SITA standard testing algorithm<sup>7</sup>). Additionally, participants were required to have a non-end-stage localized glaucomatous defect, which we defined as having at least two adjacent locations in the same hemifield (not including the blind spot) whose sensitivities differed by  $\geq 6$  dB on both of their most recent two visits. Exclusion criteria were an inability to perform reliable visual field testing, best-corrected visual acuity worse than 20/30 due to nonglaucomatous causes, history of angle closure, or any nonglaucomatous ocular pathology likely to affect the visual field. If both eyes met the inclusion and exclusion criteria, one was chosen at random for testing. 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.

Participants underwent two visits, either on the same day (with a lunch break of more than an hour between so as to reduce fatigue) or on different days as close together as possible. On each visit, they underwent visual field testing using three different test algorithms, in random order; the two algorithms that are relevant to this study are outlined below. Testing was conducted using an Octopus 900 perimeter (Haag-Streit AG, Bern, Switzerland), controlled externally using the Open Perimetry Interface.<sup>8</sup> Although the decibel scales, which are defined relative to the maximal intensity stimulus of the instrument, differ between perimeters, in this study we report all measures using the HFA decibel scale. Software to run the testing, together with all analyses, were written using the R statistical programming language.<sup>9</sup>

## Test Algorithms

Two testing algorithms were defined, both based on the existing Zippy Estimation by Sequential Testing (ZEST) algorithm. This is a variant of the Bayesian maximum likelihood thresholding algorithm QUEST,<sup>10</sup> which has been shown to have good precision and low bias,<sup>11</sup> and has been implemented in some clinically available perimeters.<sup>12</sup> The algorithms will be denoted as ZEST<sub>0</sub> (allowing sensitivities as low as 0 dB), and ZEST<sub>15</sub> (only allowing sensitivities down to 15 dB). Other than the range of sensitivities, all other elements of the algorithms were identical.

The ZEST algorithm works by defining the probability density function (pdf) of the sensitivity at a given location, which is a function that gives the probability  $P(S)$  that the true sensitivity is  $S$  for all values of  $S$ . The pdf before a stimulus presentation is known as the prior distribution, and a stimulus is presented equal to the mean of this prior (rounded to the nearest 0.1 dB due to the available precision of the instrument). According to Bayes' theorem, the posterior probability that the true sensitivity is  $S$  is given by multiplying the prior distribution by the likelihood that you would obtain the observed result (seen or not seen) if the true sensitivity were  $S$ . This likelihood function is based on the FOS curve with sensitivity  $S$ . In this study, this was defined as a cumulative Gaussian, with SD taken from the formula of Henson et al.<sup>1</sup> (capped at an SD of 7.8 dB for sensitivities below 15.0 dB), and assuming 5% false-positive and false-negative responses. The resultant posterior distribution is then used as the prior distribution for the next presentation.

In the first phase, four “seed points” were tested, located at ( $\pm 9^\circ$ ,  $\pm 9^\circ$ ) in the visual field. The algorithm started with a flat prior pdf, defined as  $P(S) = 1/35$  extending from  $S = 0$  to 35 dB for ZEST<sub>0</sub>, and  $P(S) = 1/20$  from  $S = 15$  to 35 dB for ZEST<sub>15</sub>. Five presentations are made at each of these locations. The possible series of stimuli for these seed points are illustrated in Figure 1. In subsequent phases, an “initial guess” of the sensitivity is made, equaling the mean of the sensitivities at those neighboring locations that have already been tested. The prior pdf at these locations was given by  $P(S) = C * (0.1 + k * \varphi(S))$ . Here,  $\varphi(S)$  describes a normal distribution with mean equal to the initial guess, and SD 5 dB.  $k$  is a constant such that when  $S$  equals the initial guess,  $k * \varphi(S) = 1$ .  $C$  is a constant defined such that the integral of the prior pdf over the defined range equals 1, which is a necessary condition for a well-defined pdf. In these subsequent phases, four presentations are made per location, with the set of tested locations expanding with each phase. Within each phase, the location at which the next stimulus would be presented was chosen randomly among those locations that had not yet reached their designated number of presentations.

### Analysis: Repeatability

Each eye underwent two test runs with each algorithm, and the difference at each test location between the two runs was plotted against the mean of the two runs. The SD of test-retest differences was calculated for each algorithm.

To formally test whether test-retest variability differed between the algorithms, a random effects model was formed to predict the intertest difference (sensitivity in run 1 minus sensitivity in run 2). One random effect was fit for each eye. The model assumed homoscedastic within-group errors, implicitly making the assumption that even though sensitivities are correlated between locations, the intertest differences represent noise and so are uncorrelated. However, the variance of these error terms was allowed to differ between the two algorithms. Thus, the pointwise intertest difference for ZEST<sub>0</sub>

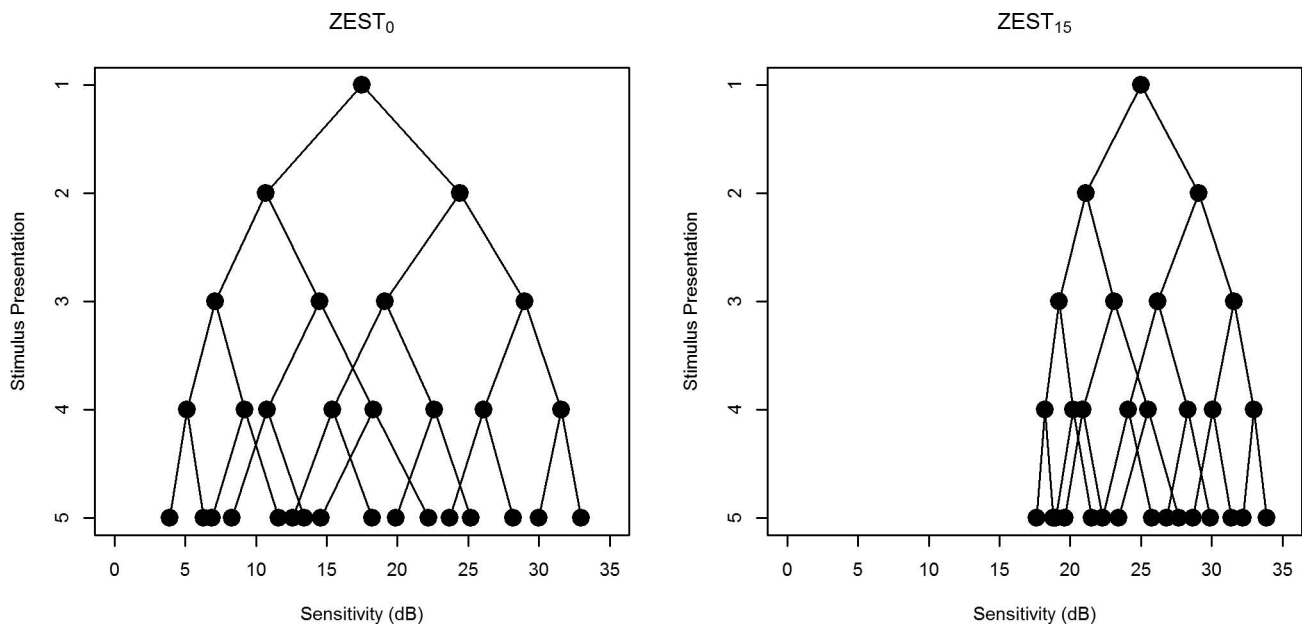


FIGURE 1. The possible series of stimulus presentations at the first four locations tested in the ZEST<sub>0</sub> and ZEST<sub>15</sub> algorithms. The first stimulus at each of those locations always equals the mean of a flat prior pdf, which extends over the range 0 to 35 dB for ZEST<sub>0</sub>, giving an initial stimulus of 17.5 dB, and extends over the range 15 to 35 dB for ZEST<sub>15</sub>, giving an initial stimulus of 25 dB. Subsequent stimuli are found by following the line down to the right if the subject responded to the stimulus, and down to the left if the subject did not respond.

was assumed to follow a normal distribution with variance *Var*<sub>z</sub>; and the pointwise intertest difference for ZEST<sub>15</sub> was assumed to follow a normal distribution with variance (*Effect* \* *Var*<sub>z</sub>). Using this random effects formulation allows a 95% confidence interval for *Effect* to be produced. If *Effect* is less than one, then this implies that ZEST<sub>15</sub> is less variable than ZEST<sub>0</sub>.

The analysis used the R code:

```
intervals(lme(Difference~1, random=~1|ID,
weights=varIdent(form=~1|Algorithm), data=Data))
```

Here, “Difference” represents the intertest difference in sensitivity estimates, with an intercept term as the sole predictor (which will be very close to zero because no trend was observed for sensitivities to be consistently higher or lower on the second test than the first). The term “random=~1|ID” introduces a random effect for each eye. The term “weights=varIdent(form=~1|Algorithm)” means that the residuals from the model (which equal the interrun differences when the intercept is zero) have a different variance for each algorithm.

The lowest sensitivity that was possible using the ZEST<sub>15</sub> algorithm was 17 dB. Therefore, the analyses were repeated among the subset of locations whose observed sensitivity was ≥17 dB on both runs using the ZEST<sub>0</sub> algorithm. This allows comparison of the test-retest variability among those locations whose sensitivity has not been “censored” at 17 dB.

**Analysis: Quantification of Damage**

To test whether the ability to detect and stage functional loss was affected by reducing the range of the testing algorithm, the pointwise sensitivities were averaged over the two runs for ZEST<sub>0</sub> and for ZEST<sub>15</sub>, and the averages compared against the sensitivity at the same location from the most recent clinical visual field examination. These visits were the second of the two clinical examinations used to assess eligibility for the study as detailed above, and were performed using the SITA standard testing algorithm.<sup>7</sup> The correlations between sensitivities from the algorithms were assessed.

A random effects model of the same form as above was formed, but with “Difference” now representing the sensitivity from SITA minus the average sensitivity of the two runs for the same ZEST algorithm. Hence this difference, which will be denoted as (SITA - ZEST<sub>0</sub>), was assumed to follow a normal distribution with variance *Var*<sub>s</sub>; and the difference (SITA - ZEST<sub>15</sub>) was assumed to follow a normal distribution with variance (*Effect*<sub>s</sub> \* *Var*<sub>s</sub>). If *Effect*<sub>s</sub> < 1, then it can be concluded that sensitivities from ZEST<sub>15</sub> are more closely correlated with SITA than those from ZEST<sub>0</sub>.

The SITA algorithm reports “<0 dB” at any locations where the subject did not respond to the most intense stimulus available (i.e., 0 dB), and cannot estimate the actual sensitivity at such locations. We treated these locations as having sensitivity equal to -1 dB for this analysis; but also repeated the analysis excluding them from the correlation analysis.

**RESULTS**

Thirty-six eyes of 36 participants (12 male, 24 female) were tested. Demographic information is presented in Table. Thirty-four eyes had MD outside normal limits, and 35 had pattern standard deviation (PSD) outside normal limits.

**Repeatability**

The SD of test-retest differences was 4.28 dB for ZEST<sub>0</sub>, and 2.36 dB for ZEST<sub>15</sub>. In the random effects model, the estimated test-retest variance for ZEST<sub>15</sub> was 54.5% of the test-retest variance for ZEST<sub>0</sub>, with 95% confidence interval 52.0% to

TABLE. Demographics of the Study Population

Measure	Median	Interquartile Range	Range
Age, y	71	68 to 73	48 to 82
MD, dB	-7.1	-8.9 to -4.5	-26.4 to +0.3
PSD, dB	7.9	4.9 to 10.8	1.6 to 16.1

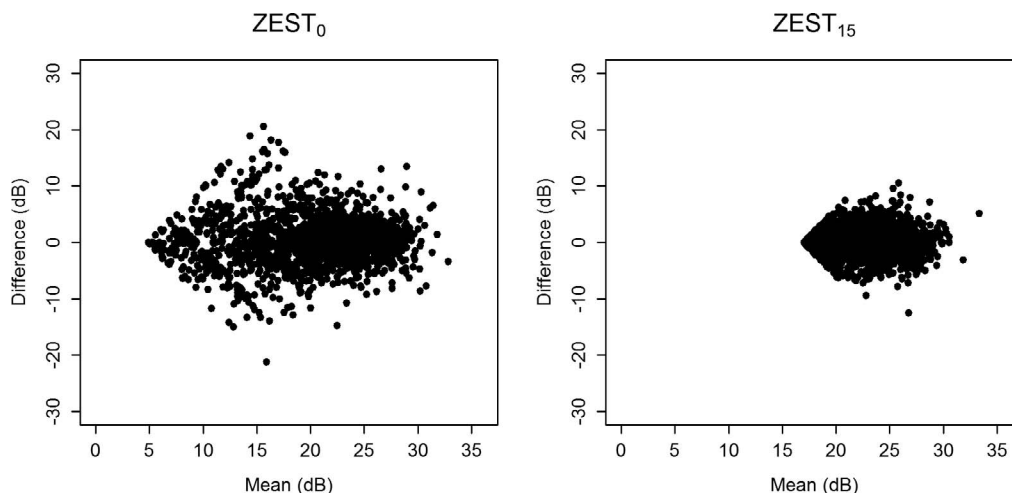


FIGURE 2. The test-retest difference between two runs plotted against the mean, for all test locations, for two different testing algorithms.

57.1%. The test-retest differences are plotted against the mean of the two runs in Figure 2, showing a decreased spread of data and variability in ZEST<sub>15</sub> when compared with ZEST<sub>0</sub>.

Because ZEST<sub>15</sub> could not result in sensitivity estimates below 17 dB, the analysis was repeated on just those locations that also had sensitivity  $\geq 17$  dB on both runs for ZEST<sub>0</sub>. Among these locations, the SD of test-retest differences was 2.97 dB for ZEST<sub>0</sub>, and 2.59 dB for ZEST<sub>15</sub>. In the random effects model, the estimated test-retest variance for ZEST<sub>15</sub> was 86.5% of the test-retest variance for ZEST<sub>0</sub>, with 95% confidence interval 81.6% to 91.7%. This subset of the test-retest differences is plotted against the mean of the two runs in Figure 3, for each algorithm. It is clear that not only is the spread of data visibly narrower (indicating lower variability) for ZEST<sub>15</sub>, but this remains true even when the sensitivity is above 30 dB. Figure 4 presents a Bland-Altman plot comparing the mean of the two sensitivities from ZEST<sub>15</sub> against the mean of those from ZEST<sub>0</sub>, and shows no systematic bias between the algorithms.

**Quantification of Damage**

Based on the random effects model, the variance of (SITA - ZEST<sub>15</sub>) was 89.2% of the variance of (SITA - ZEST<sub>0</sub>), with a 95% confidence interval 85.2% to 93.4%. This implies that the

correlation with sensitivities from the SITA standard algorithm was significantly stronger for ZEST<sub>15</sub> than for ZEST<sub>0</sub>. When excluding locations whose sensitivity (from SITA) was  $< 0$  dB, the variance of (SITA - ZEST<sub>15</sub>) was 85.8% of the variance of (SITA - ZEST<sub>0</sub>), with confidence interval 81.8% to 90.0%. The Pearson correlations with these SITA sensitivities were 0.161 for ZEST<sub>0</sub> (95% confidence interval 0.115-0.207) and 0.171 for ZEST<sub>15</sub> (0.125-0.217).

Among locations whose sensitivity using the SITA standard algorithm was  $\geq 17$  dB, the correlations were 0.143 for ZEST<sub>0</sub> (0.093-0.192) and 0.212 for ZEST<sub>15</sub> (0.163-0.259). In this case, the variance of (SITA - ZEST<sub>15</sub>) was 73.5% of the variance of (SITA - ZEST<sub>0</sub>), with 95% confidence interval 69.7% to 77.4%.

**DISCUSSION**

It has been known for many years that the test-retest variability of perimetry increases with glaucomatous damage. A severely damaged visual field location might have to deteriorate by more than 15 dB from its baseline value before a clinician would confidently state that localized progression had occurred. Even then the decision might not be made until multiple locations have changed by that amount, and/or there

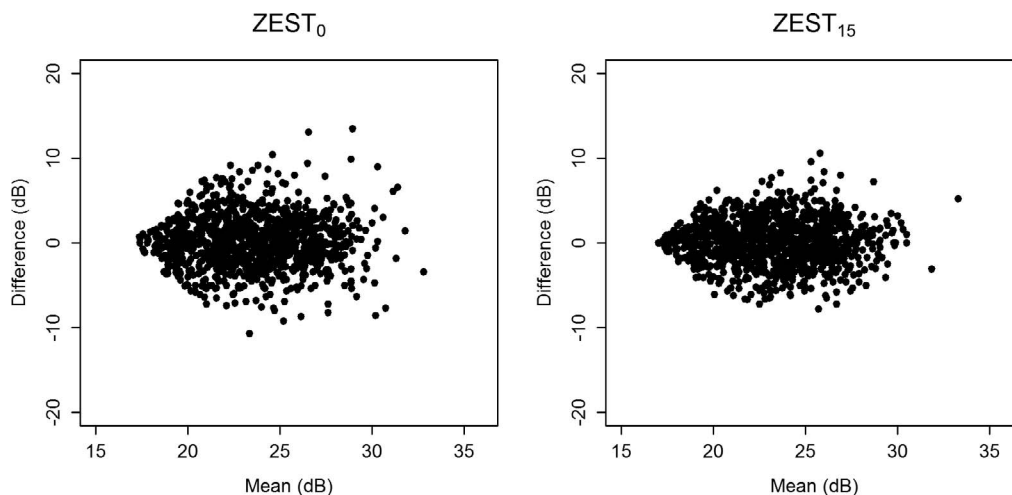
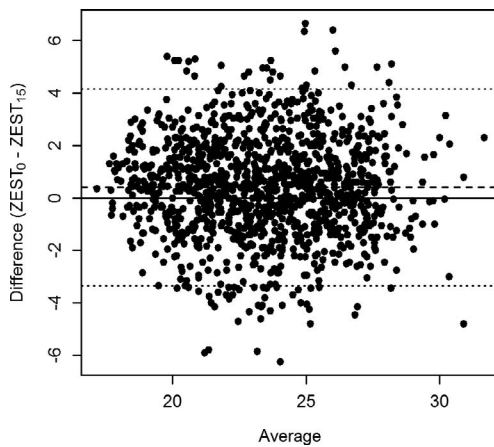


FIGURE 3. The test-retest difference between two runs plotted against the mean, for all test locations that had sensitivities  $\geq 17$  dB on each run, for two different testing algorithms.



**FIGURE 4.** Bland-Altman plot comparing the two test algorithms, ZEST<sub>0</sub> versus ZEST<sub>15</sub>. For each location whose sensitivity was  $\geq 17$  dB on all tests, the mean of the two runs per algorithm was taken. The difference between the algorithms (in decibels) is then plotted against the average of the two algorithms (also in decibels). The *solid horizontal line* is at zero difference. The *dashed line* represents the mean difference, +0.40 dB. The *dotted lines* represent the 95% limits of agreement, defined as  $(\text{mean} \pm 1.96 * \text{SD})$ , which were (-3.35 dB, +4.16 dB).

is convincing evidence of corresponding structural change. With this in mind, we have recently questioned whether it is worthwhile performing perimetry using very high contrast stimuli, because they provide so little reliable information.<sup>4</sup> We have previously shown that “censoring” sensitivities below 15 dB (approximately 1000% contrast) and setting them equal to 15 dB does not harm, and may possibly improve, the ability to detect progression.<sup>5</sup> It has also been shown that censoring sensitivities below 10 dB did not harm the ability to detect progression using MD.<sup>6</sup> The next step is to assess actually altering the testing algorithm so that it does not test beyond 15 to 19 dB. In this study, we show that this would reduce test-retest variability, not only by removing the variability at these very low values, but also by allowing smaller step sizes and hence more accurate threshold determination at less damaged locations. In the eyes tested here, test-retest variability was reduced by 13.5% among locations that were  $\geq 17$  dB, and by an even greater degree among more severely damaged locations, without harming the ability to quantify functional loss. This suggests that visual field testing algorithms could be designed to detect visual field progression sooner without increasing test duration.

The reason why variability among severely damaged locations is reduced is intuitively clear. If the “true” sensitivity were, for example, 15 dB, then assuming the variability equation of Henson et al.,<sup>1</sup> the observer would have a 14% probability of failing to respond to a 5-dB stimulus. This would cause the sensitivity estimate when using the ZEST<sub>0</sub> algorithm to be below 5 dB on some test dates but above 15 dB on other test dates, giving high test-retest variability. We have previously provided evidence suggesting that the response probability does not substantially increase below 15 to 19 dB,<sup>4</sup> in which case the observer would have a nearly 50% probability of failing to respond to a 5-dB stimulus, causing even higher test-retest variability. However, with the ZEST<sub>15</sub> algorithm, the estimated sensitivity at such locations would be no lower than and likely very close to 17 dB every time, giving lower test-retest variability.

Perhaps less intuitively, the test-retest variability was also reduced among less damaged locations with sensitivity above 17 dB. This is because the probability density function

describing sensitivity estimates is narrower, extending from 15 to 35 dB instead of 0 to 35 dB. Therefore, step sizes are smaller between successive contrasts presented at a location. This allows more precise estimation of sensitivity, and hence lower test-retest variability.

Reducing test-retest variability is only one desirable aspect of a testing algorithm. Because variability is reduced even at near-normal locations, it is reasonable to assume that detectability of functional damage would not be impaired, and may be improved, by switching to an algorithm that omits very high contrast stimuli. In this study, sensitivities measured using ZEST<sub>15</sub> were more highly correlated with SITA than those measured using ZEST<sub>0</sub>, albeit with both correlations being weak (below 0.2). This suggests that the staging and quantification of functional defects may actually be improved by not using very high contrast stimuli in testing algorithms. Sensitivities from SITA are also imperfect due to necessary constraints on test duration, and it is impossible to know for certain at present whether an algorithm accurately reflects true functional status, but it is reassuring to see that narrowing the range of stimuli does not harm performance in comparison with the current clinical standard. It is also necessary to demonstrate that reducing the stimulus range would not adversely affect the structure-function relation, and patient testing is under way to examine this issue.

The ZEST algorithm with a fixed number of presentations per location was chosen for this study because it allows just one element to be altered (the range of sensitivities), making the effect of this change clear. To further this simplicity, the algorithm was not fully optimized. For example, if the probability density function extends from 0 to 35 dB, then it is impossible for its mean to ever equal 0 dB even if the observer does not respond to any stimuli; consequently, there were no sensitivity estimates below 4.8 dB for ZEST<sub>0</sub>, or below 17.0 dB for ZEST<sub>15</sub>. Clinically, more complex algorithms such as SITA standard,<sup>7</sup> SITA Fast,<sup>13</sup> or GATE<sup>14</sup> are often used, which aim to increase efficiency and accuracy, and may involve postprocessing to filter out some of the variability. The magnitude of the reduction in variability obtained by stopping testing at 15 dB will vary between these algorithms. It remains to be seen whether this would reach the 20% reduction in variability that has been reported to be needed to reduce the average time taken to detect progression by one visit.<sup>3</sup> However, our study demonstrates the principle that reducing the technical range of perimetry would reduce variability not just in highly damaged regions of the visual field, but also in less damaged areas by allowing more precise thresholding within the same test duration.

Using Size V (1.72° diameter) stimuli instead of the more conventional Size III (0.43° diameter) stimuli reduces variability at any given location,<sup>15</sup> but this appears to be driven solely by the increase in sensitivity caused by use of larger stimuli,<sup>16</sup> implying that it postpones but does not prevent the point at which variability is too high to be able to distinguish true change from noise in a clinic situation. It may be more efficient to present larger stimuli if the test subject does not respond to a 15-dB Size III stimulus, as is done by the Heidelberg Edge Perimeter (Heidelberg Engineering, Heidelberg, Germany). Another possibility would be to use larger stimuli throughout the range to reduce variability<sup>15</sup> and extend the range of severities over which reliable measurements could be obtained.<sup>16,17</sup> Size threshold perimetry, whereby stimulus size is altered rather than contrast, may also extend the dynamic range without increasing variability,<sup>18</sup> although this would require more extensive modifications to clinical perimeters so as to meet the need for a sufficient number of different stimulus sizes.

In summary, adjusting perimetric testing algorithms so that they do not present stimuli at very high contrasts (above approximately 1000% contrast, or equivalently below approximately 15 dB) reduced the test-retest variability of sensitivity estimates without harming the ability to quantify functional loss. The sensitivity information obtainable from such locations is very limited due to their inherent unreliability and high variability. Therefore, it is more efficient for visual field algorithms to instead concentrate on more precise and accurate estimation of sensitivities at remaining locations with sensitivity  $\geq 15$  dB.

### Acknowledgments

Supported by National Institutes of Health R01-EY020922 (SKG). The authors alone are responsible for the content and writing of the paper.

Disclosure: **S.K. Gardiner**, Haag-Streit, Inc. (C, R); **S.L. Mansberger**, None

### References

- 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.
- Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci.* 2002;43:2654-2659.
- Turpin A, McKendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci.* 2011; 52:3237-3245.
- 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.
- Gardiner SK, Swanson WH, Demirel S. The effect of limiting the range of perimetric sensitivities on pointwise assessment of visual field progression in glaucoma. *Invest Ophthalmol Vis Sci.* 2016;57:288-294.
- Junoy Montolio FG, Wesselink C, Jansonius NM. Persistence, spatial distribution and implications for progression detection of blind parts of the visual field in glaucoma: a clinical cohort study. *PLoS One.* 2012;7:e41211.
- Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand.* 1997;75:368-375.
- Turpin A, Artes PH, McKendrick AM. The open perimetry interface: an enabling tool for clinical visual psychophysics. *J Vis.* 2012;12(11):22.
- R Development Core Team. R: A language and environment for statistical computing. Vienna Austria: R Foundation for Statistical Computing; 2013.
- Watson A, Pelli D. QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys.* 1983;33:113-120.
- King-Smith P, Grigsby S, Vingrys A, Benes S, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations experimental evaluation and practical implementation. *Vision Res.* 1994;34:885-912.
- 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.
- Bengtsson B, Heijl A. SITA. Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand.* 1998;76:431-437.
- 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.
- 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.
- Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. The effect of stimulus size on the reliable stimulus range of perimetry. *Trans Vis Sci Tech.* 2015;4(2):10.
- 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.
- 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.