

The Effective Dynamic Ranges for Glaucomatous Visual Field Progression With Standard Automated Perimetry and Stimulus Sizes III and V

Michael Wall,^{1,2} Gideon K. D. Zamba,³ and Paul H. Artes⁴

¹Department of Ophthalmology, University of Iowa, College of Medicine, Veterans Administration Hospital, Iowa City, Iowa, United States

²Department of Neurology, University of Iowa, College of Medicine, Veterans Administration Hospital, Iowa City, Iowa, United States

³Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, Iowa, United States

⁴School of Health Professions, Peninsula Allied Health Centre, Plymouth University, Plymouth, Devon, United Kingdom

Correspondence: Michael Wall, Department of Neurology, University of Iowa, College of Medicine, 200 Hawkins Drive, No. 2007 RCP, Iowa City, IA 52242-1053, USA; michael-wall@uiowa.edu.

Submitted: June 8, 2017
Accepted: December 18, 2017

Citation: Wall M, Zamba GKD, Artes PH. The effective dynamic ranges for glaucomatous visual field progression with standard automated perimetry and stimulus sizes III and V. *Invest Ophthalmol Vis Sci.* 2018;59:439-445. <https://doi.org/10.1167/iovs.17-22390>

PURPOSE. It has been shown that threshold estimates below approximately 20 dB have little effect on the ability to detect visual field progression in glaucoma. We aimed to compare stimulus size V to stimulus size III, in areas of visual damage, to confirm these findings by using (1) a different dataset, (2) different techniques of progression analysis, and (3) an analysis to evaluate the effect of censoring on mean deviation (MD).

METHODS. In the Iowa Variability in Perimetry Study, 120 glaucoma subjects were tested every 6 months for 4 years with size III SITA Standard and size V Full Threshold. Progression was determined with three complementary techniques: pointwise linear regression (PLR), permutation of PLR, and linear regression of the MD index. All analyses were repeated on “censored” datasets in which threshold estimates below a given criterion value were set to equal the criterion value.

RESULTS. Our analyses confirmed previous observations that threshold estimates below 20 dB contribute much less to visual field progression than estimates above this range. These findings were broadly similar with stimulus sizes III and V.

CONCLUSIONS. Censoring of threshold values < 20 dB has relatively little impact on the rates of visual field progression in patients with mild to moderate glaucoma. Size V, which has lower retest variability, performs at least as well as size III for longitudinal glaucoma progression analysis and appears to have a larger useful dynamic range owing to the upper sensitivity limit being higher.

Keywords: automated perimetry, visual field, glaucoma

The dynamic range in perimetry is the range of the smallest and largest values of the visual stimulus that the device is capable of displaying. It is determined by the physical capabilities of the device. For example, the Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, San Leandro, CA, USA) and many other automated perimeters use a maximum stimulus intensity (brightness) of 3200 cd/m², on a background of 10 cd/m². We and others¹⁻⁴ have shown that stimuli of such high contrasts may not be clinically meaningful. Therefore, it may be more appropriate to characterize an “effective dynamic range” (EDR) for perimetry, defined as that part of the physical range that is physiologically meaningful and clinically useful.

By definition, the limit of the physical dynamic range of standard automated perimetry (SAP) extends to 0 dB (the maximum intensity stimulus). But, owing to the poor precision (large retest variability) of intense stimuli, the threshold estimates are scattered over several log units of range. For example, Heijl and colleagues^{5,6} have found that for test locations initially measured with 8- to 18-dB loss, the 95% prediction interval of future measurements nearly covers the full measurement range of the instrument.

One could argue that measurements of very low precision have limited clinical value, and therefore the high retest variability of threshold estimates below 20 dB may represent a natural limit on the effective dynamic range.^{1,3,7} Retesting of values in a clinical setting after this much loss would appear to have limited value and one could argue that they not be included in the EDR.¹⁻⁴ Supporting this proposal, Gardiner and colleagues² have demonstrated that censoring pointwise perimetric sensitivities below approximately 20 dB has little effect on the ability of pointwise linear regression (PLR) to detect glaucomatous change over time. They have used PLR to find progressing locations, based on the slope criterion of ≤ -1 dB/y and *P* value criterion of *P* < 0.01. Gardiner's group² explored two sets of criteria: an eye was labeled as progressing (1) if three or more locations were progressing by PLR criterion and (2) if three or more neighboring locations were progressing by the PLR criterion. They concluded that readings below 20 dB add little if anything to progression detection. Along the same lines, Junoy Montolio et al.⁸ also have suggested that perimetric sensitivities below approximately 20 dB contribute little to identify glaucomatous progression.



The aims of our investigation were first, to confirm the findings of Gardiner et al.² in an independent dataset, and with a more comprehensive set of techniques of analyzing progression; second, to investigate how the EDR for visual field progression varies between stimulus size III and V; and third, to analyze the effect of censoring on mean deviation (MD). Our approach was similar to that of Gardiner et al.²; however, we included an additional detection method—permutation of PLR (PoPLR)—and we added an analysis of MD. Like Gardiner et al.,² we analyzed progression in series of visual field examinations from which the lower threshold estimates (corresponding to stimuli of higher intensity) have been systematically “censored,” that is, replaced by a given “floor” value. For example, with a floor value of 20 dB, progression is evaluated in a visual field series in which all threshold estimates < 20 dB are censored to become 20 dB.

SUBJECTS AND METHODS

Participants

We enrolled and tested 120 glaucoma patients at baseline and again at a separate sitting within 1 to 8 weeks. We then tested the subjects every 6 months for 4 years.

The glaucoma patients were consecutive patients invited from the glaucoma clinic at the University of Iowa Department of Ophthalmology and Visual Sciences if they met entry criteria. They were enrolled if they had glaucomatous optic disc changes with abnormal SAP (glaucomatous visual field defects, i.e., three or more adjacent abnormal test locations in a clinically suspicious area at the $P < 0.05$ level or two adjacent locations abnormal with at least one at the $P < 0.01$ level). In addition, MD was in the range of 0 to -20 dB on SAP. We included patients with primary, secondary, or normal tension glaucoma. The patients did not have other disease affecting vision and were capable of performing SAP and returning for follow-up visits. Patients were excluded if they had cataract causing visual acuity of worse than 20/30, pupil size less than 2.5 mm, were younger than 19 years, or were pregnant. If both eyes qualified for the study, one eye was chosen at random as the study eye. The average age of the glaucoma patients was 64.9 years (SD, 9.5 years; range, 38–81 years); their mean deviation was -6.7 dB (SD, 4.4 dB).

Participants were excluded if they had five or fewer time points available for analysis or if they had undergone cataract extraction during the 4-year study period. The study followed the tenets of the Declaration of Helsinki, and the protocol was approved by the University of Iowa Institutional Review Board. All participants gave written informed consent.

Visual Testing

All subjects underwent automated perimetry using program 24-2 of the HFA (Carl Zeiss Meditec, Dublin, CA, USA). For SAP, the stimuli of Goldmann size III (0.43° diameter, 4 mm^2) were used with the SITA Standard 24-2 algorithm. Goldmann size V stimuli (1.72° diameter, 64 mm^2) were used along with Full Threshold (FT) testing; there is no SITA program currently available for size V stimuli. Our pilot data and the work of Artes and colleagues⁹ have shown that the differences between SITA and FT are minor. We chose size III SITA so that we would be comparing the test most commonly used in clinical practice.

We followed the manufacturer's recommendations for using corrective lenses. Care was taken to prevent lens rim artifacts. The subjects had testing in one eye, chosen at random, but the same eye was used for all tests. All visual field examinations met the following reliability criteria: fixation losses < 20% or

normal gaze tracking, false-positive rate < 10%, and false-negative rate < 33%. The four tests were administered in a random order with at least a 5-minute rest break between consecutive tests.

Analyses of Visual Field Progression

We performed three complementary analyses of visual field progression. PLR is a widely used point-by-point analysis of change at individual visual field locations.^{10–12} PoPLR, an extension of this technique, provides an overall statistical significance (P) value for the null hypothesis that there is no statistically significant deterioration anywhere in the visual field, customized to the individual patients' data. Regression of the MD index provides an estimate of the rate of change (dB/y) of an individual's overall visual field damage, along with a P value for the statistical significance of this change. The visual field data were censored in steps of 1 dB over the observed range for all analyses.

Pointwise Linear Regression. PLR is based on least-squares linear regression of sensitivity (decibel measurements) over time. We evaluated the total number of test locations that changed by more than -1.0 dB/y at a $P < 0.01$, across the group of glaucoma patients and healthy participants.

To classify an eye as progressive we used criteria as defined by Gardiner et al.² Criterion 1 explores progression based on at least three progressing locations of the visual field; without any contiguity restrictions—a more relaxed classification rule compared to criterion 2. In conjunction with criterion 2, we define neighboring locations as those locations within 6° , that is, horizontally, vertically, and diagonally, for a maximum of eight neighbors for each interior location of the visual field, except for points within immediate vicinity of the blind spots and boundary points. Criterion 3 identifies progression by summing all progressing locations in the qualifying dataset. We define progression rate as the proportion of glaucoma patients flagged by each criterion, out of the total sample of patients.

Permutation of Pointwise Linear Regression. In contrast to PLR (where a trend is evaluated separately at each test location), PoPLR¹³ is a technique designed to derive a single P value for the statistical significance of visual field progression across the entire visual field. Because the result of PoPLR is a single P value, it is much more straightforward to make decisions on progression or stability in a visual field series than with PLR where some arbitrary combination of number of test locations, pointwise slope, and P value is required as the criterion for a “real” change. PoPLR has been described in detail previously,¹³ but in summary the technique works as follows: (1) at each test location, change is analyzed by ordinary least-squares linear regression, and a one-sided P value is calculated to isolate deterioration and ignore improvement; (2) the evidence for change is pooled into a single summary statistic across the entire visual field, through a statistical combination of the pointwise P values; (3) the visual field examinations from each patient are randomly reordered (permuted) many times, in order to estimate the null distribution of the summary statistic; and (4) the summary statistic from the actual series is compared to the null distribution, and its rank order within this distribution provides the overall P value. When there is statistically significant visual field progression, the summary statistic of the actually observed visual field series will be much smaller than when the visual fields are randomly rearranged in time.

Because the P value of PoPLR is based on each individual patient's visual field series, it implicitly “corrects” for patients with higher or lower variability than average, and it also is entirely independent of the underlying measurement (decibel)

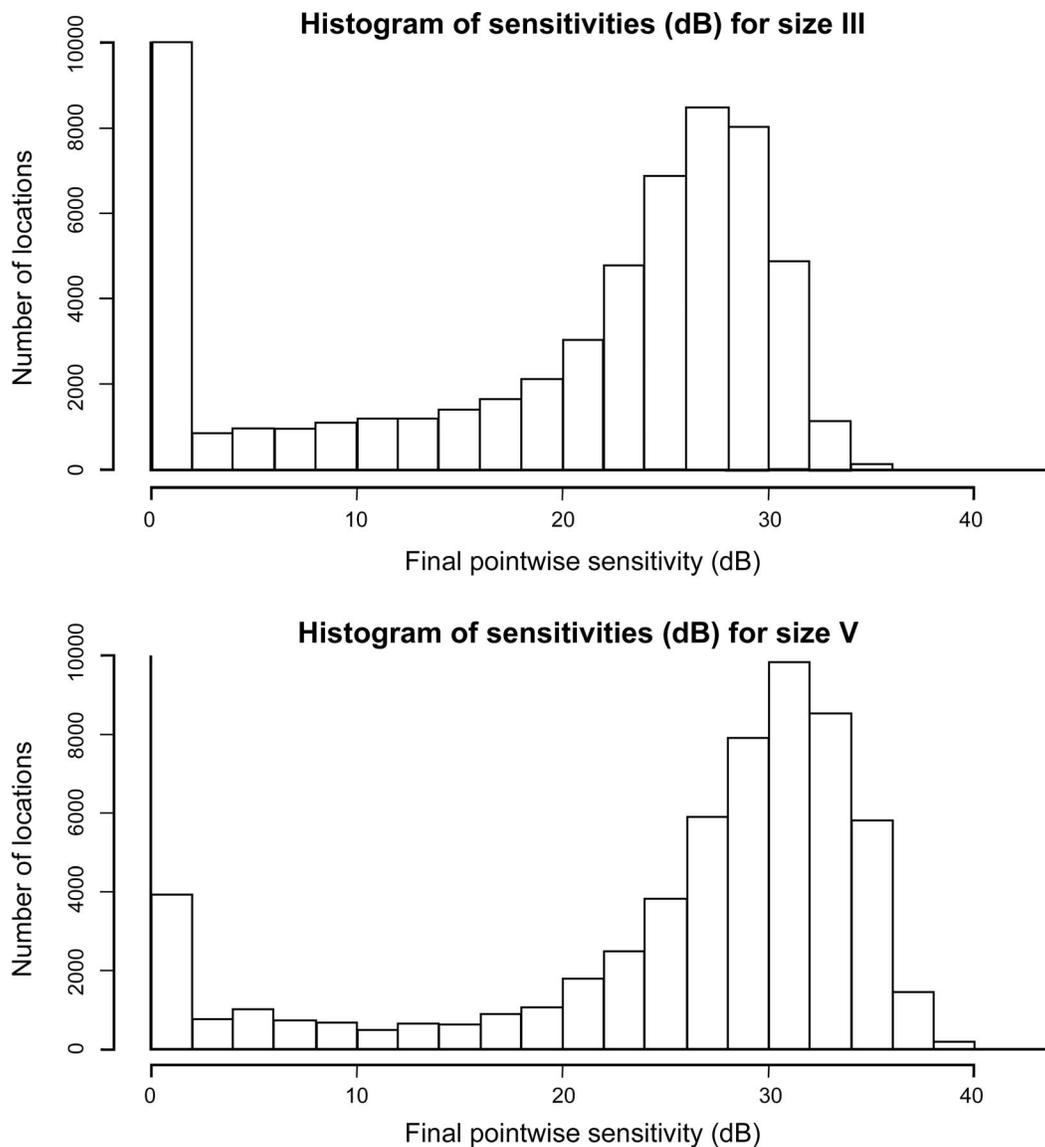


FIGURE 1. Histogram of the absolute frequency of pointwise sensitivities (in decibels) with Goldmann size III stimuli (*top*) and size V stimuli (*bottom*).

scales.¹⁴ These advantages over other, more conventional, analyses of visual progression are particularly important for the investigations reported here, since it may not be easy to equate between point-by-point change-over-time with different stimulus sizes.

Rate-of-Change Analysis of Mean Deviation. Regression of the MD index over time provides a global estimate of rate of change.^{7,15} At each sensitivity level from 0 to 40 dB, we censored by test location and then computed the MD. We then compared the number of eyes classified as changing based on a slope of less than -0.5 dB/y and P value < 0.05 . We used less strict criteria for this analysis because of the known differences in assessing global versus pointwise changes.¹⁶

Lastly, we statistically compared the proportions of eyes flagged as progressing based on censored data with the original proportions (i.e., proportion progressing in uncensored data). This was done for censoring values from 1 to 40 dB, to investigate at what point the censoring leads to a statistically significantly smaller proportion with progression.

RESULTS

Of the 120 glaucoma subjects, 94 met the complete analysis criteria. Twenty-six glaucoma subjects were excluded from the analyses either owing to cataract extraction, owing to dropout, or not having had at least six time points available for a meaningful estimation of regression parameters. The average MD of the SITA tests was 6.6 dB with an SD of 9.5 dB (range, -0.5 to -19.9 dB). The median was -9.2 dB with an interquartile range of -14.8 to -4.8 dB. Figure 1 shows the histogram of the pointwise sensitivities, at the end of the series for Goldmann stimuli size III and size V; and Figure 2 shows the pointwise PLR rates of change. Figures 3 through 5 show the detection rates when PLR criteria 1, 2, and 3 are applied to the glaucoma population. Figure 6 assesses detection rate under PoPLR criterion for global detection by using a P value of 0.01. Figure 7 explores detection rate by using MD regression criterion of slope less than -0.5 dB/y and P value < 0.05 . The results are as follows.

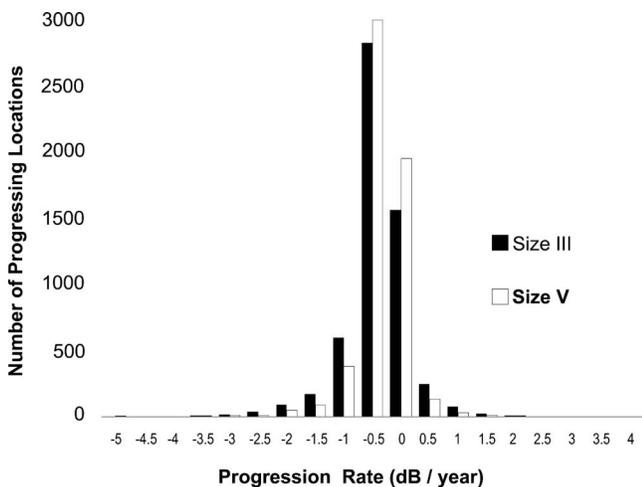


FIGURE 2. Histograms of the pointwise regression slopes as an estimate for rate of change within the population.

Pointwise Linear Regression

PLR criteria 1 and 2 show a sharp change in detection rates with censoring around 20 dB. This result holds for both size III and size V stimuli, although size V appears to offer a larger effective dynamic range present mostly at the upper end of the range. It is not surprising that PLR criterion 2 yields a lower detection rate, since this criterion requires localized progression. Regardless of the restrictiveness of the criterion used, the underlying message is that testing below 20 dB may add little, if any, useful clinical information regarding progression detection within this population (Figs. 3–5).

Permutation of PLR

The results from PoPLR suggest a sharp change in detection rates with censoring around 25 dB (Fig. 6). This suggests that with the more global approach of PoPLR, measuring sensitivity below 25 dB may add little if anything to progression detection.

Mean Deviation

The MD analysis also showed the same pattern although not as pronounced as those of PLR and PoPLR (Fig. 7).

DISCUSSION

By investigating longitudinal data from glaucoma patients with mild to moderate glaucomatous damage, we confirmed that censoring threshold estimates below 20 dB results in only a minor reduction in the number of eyes classified as having changed. This finding holds across a large spectrum of progression criteria.

We have observed that the EDR for retest variability for the size V stimulus is greater by approximately a log unit than for the size III stimulus.¹ We have also observed a larger EDR for glaucomatous progression with the size V stimulus. Much of the enhanced range of size V testing appears to derive from sensitivities at the upper end of the useful range. This finding has important implications suggesting that using the size V stimulus may give clinicians a wider effective range of progression detection in glaucoma than using the size III stimulus.

Our findings are in concert with those of Gardiner et al.² These investigators, using two datasets (one with less damage than our subjects and one with more damage) and the same PLR criteria for progressing test locations that we used, have found that the progression rate for PLR does not materially change when test locations with low sensitivity (those below 15–19 dB) are censored from the analysis. We extended their findings to include the comparison with size V stimulus results, PoPLR analysis, and the MD regression censoring results.

The findings from Gardiner and colleagues² and our study are important for several reasons. If little is gained from testing below 20 dB for glaucomatous progression, testing could be truncated at this sensitivity level with resultant time-savings. This saved time could be used to better map the defects present or add test locations in the periphery or macular area. Second, with the poor repeatability of SAP testing below 20 dB, clinical decisions are being made on very noisy data.¹ We have shown previously that with size III stimuli, when the initial test values are between 0 and 15 dB, the most frequent value on

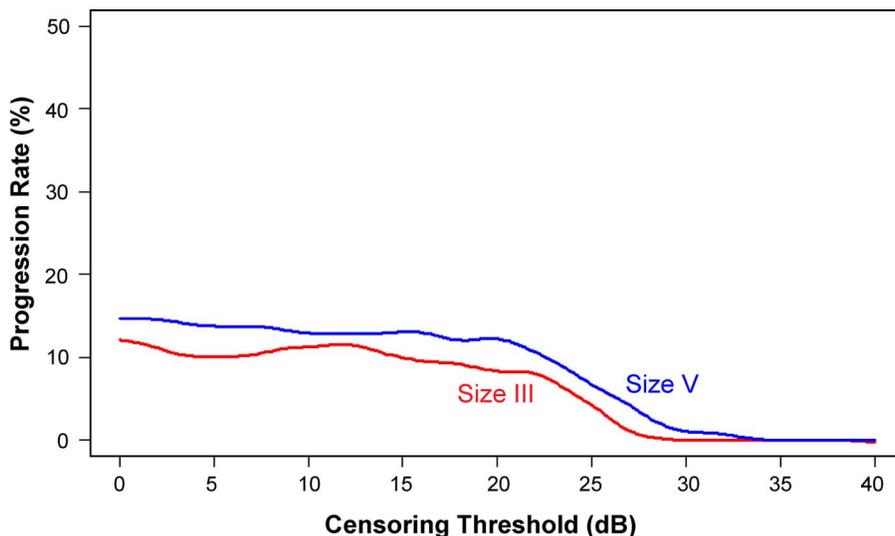


FIGURE 3. Results of progression rate using PLR criterion 1 in which three or more locations in the visual field changed by -1 dB/y at a P value of 0.01; in percentage of subjects. The reduction in progression rate (compared to the uncensored data) became statistically significant ($P < 0.05$, paired test of proportions) at censoring thresholds of 24 dB (size III) and 25 dB (size V).

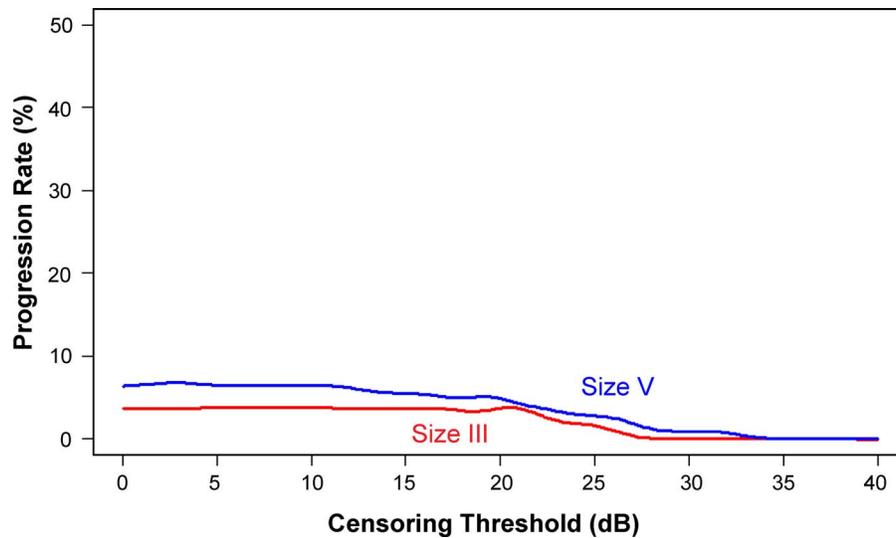


FIGURE 4. Results of progression rate using the more restrictive PLR criterion 2, according to which three or more neighboring locations in the visual field changed by -1 dB/y at a P value of 0.01; in percentage of subjects. The reduction in progression rate (compared to the uncensored data) became statistically significant ($P < 0.05$, paired test of proportions) at censoring thresholds of 27 dB (size III, V).

retest is 0 dB. It may improve clinical care to simply omit this highly variable data from clinical decision-making.

We found, as did Gardiner et al.,² that the progression rate for glaucoma subjects is slightly higher when the censoring threshold is 0 dB (i.e., no censoring) than when it is 20 dB. Although test locations with sensitivity below 20 dB may not drive progression detection, they are not entirely without utility. However, the time cost of a modest addition of information may not justify this additional testing. In other words, the time may be better spent at measuring those locations that are within the effective dynamic range of the technique rather than outside it. Gardiner and Mansberger,¹⁷ for example, have evaluated a perimetric algorithm that does not present stimuli below 15 dB and compared it to an identical testing algorithm that uses a standard decibel range. The difference between the two runs for the same algorithm

has been used as a measure of test-retest variability. They have found a lower retest variability for the test that does not use stimuli brighter than 15 dB. These findings, coupled with ours, suggest that equipment manufacturers may need to reconsider the maximum stimulus luminance of their devices.

A valid concern is that censoring, or any other change in the measurement range, will systematically bias the test results. However, this potential bias is no different in principle from that observed with any current definition of the 0-dB stimulus. Let us consider a case where the physical dynamic range of the instruments are extended (rather than reduced, as we propose here) by increasing the maximum stimulus luminance by a factor of 10. This would shift the decibel scale upwards (e.g., 30 dB becomes 40 dB), and the shift would apply to any threshold estimates as well as to the normative values. In turn, those total deviation values corresponding to absolute scotoma

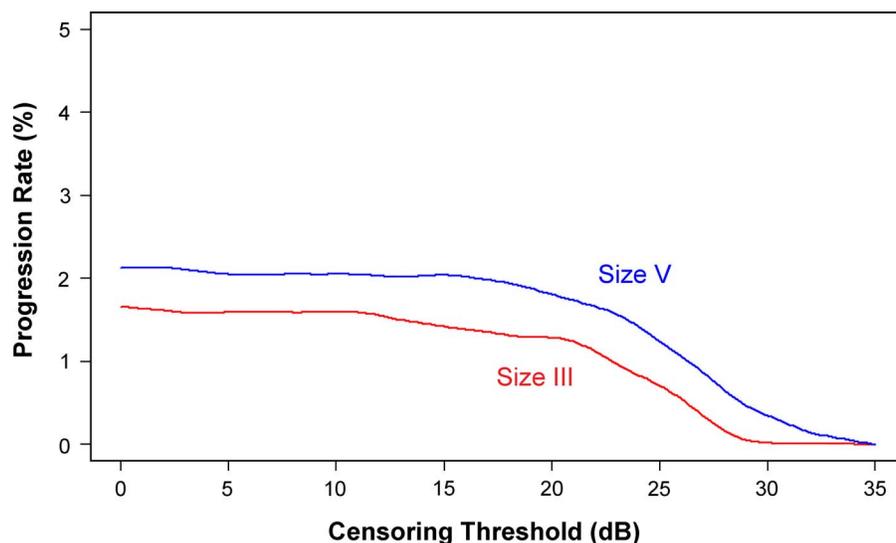


FIGURE 5. Effect of censoring on size III and V thresholds on total number of progressing test locations in subjects with glaucoma with mild to moderate visual loss. Note that below approximately 20 dB there is little increase in progression rate (criterion 3); in percentage of test locations; slope worse than -1 dB/y, $P=0.01$. The reduction in progression rate (compared to the uncensored data) became statistically significant ($P < 0.05$, paired test of proportions) at censoring thresholds of 24 dB (size III) and 25 dB (size V).

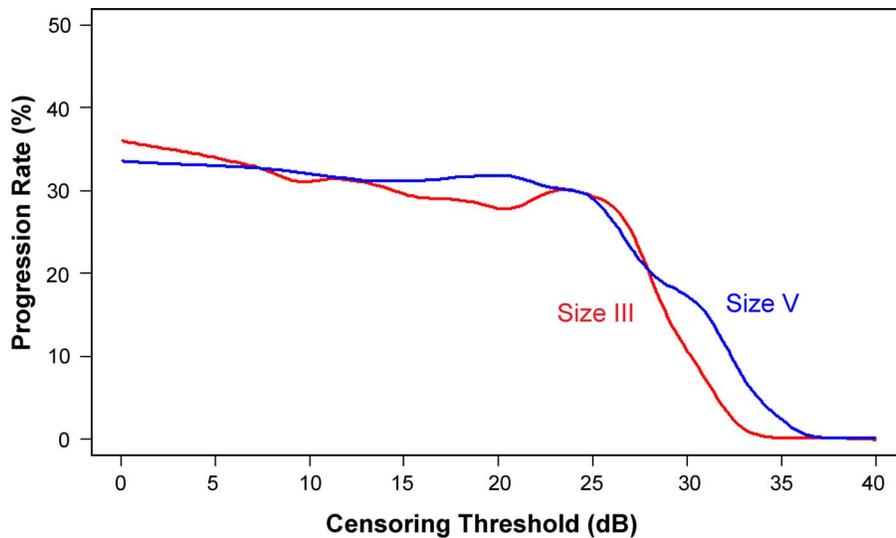


FIGURE 6. PoPLR results are similar to PLR results, although sensitivities below approximately 25 dB have little impact on progression rate; in percentage of subjects with *P* value set at 0.01. The reduction in progression rate (compared to the uncensored data) became statistically significant ($P < 0.05$, paired test of proportions) at censoring thresholds of 26 dB (size III) and 27 dB (size V).

(sensitivity, < 0 dB) would become more negative. In other words, threshold estimates, normative data, and global indices such as MD are all dependent on the specific choice of the measurement range, and test results can only be compared with each other when they are obtained under the same definition. Our case is that a large part of the current measurement range does not seem to contribute much value to the measurement of progression. But, this is best addressed through a redesign of visual field tests, and we do not argue that censoring should be applied with current tests to shorten the test time (e.g., by avoiding to retest locations with sensitivity below a certain value). Simple and ad hoc approaches would likely make it difficult to compare patients' data to previous results.

One weakness of this study was that our cohort was confined to mild to moderate glaucoma. Our findings need to

be corroborated for more severely damaged and end-stage glaucomatous visual field examinations. Another weakness was that the testing algorithms of the two stimulus sizes are different with the Bayesian-based SITA Standard used for Goldmann size III stimuli and the FT strategy for size V. Artes and colleagues⁹ have found retest variability for SITA Standard slightly better than FT above 25 dB and slightly worse below 25 dB. We doubt that this issue has much influence on our findings but realize a better comparison would have been to use FT testing for size III. However, this would not have allowed us to validate the findings of Gardiner et al.,² which was one of our main aims.

In conclusion, censoring sensitivity estimates < 20 dB leads to only a minor reduction in progression rate, for a variety of criteria including PLR, PoPLR, and MD. SAP therefore may be improved by limiting testing to the range of sensitivities with

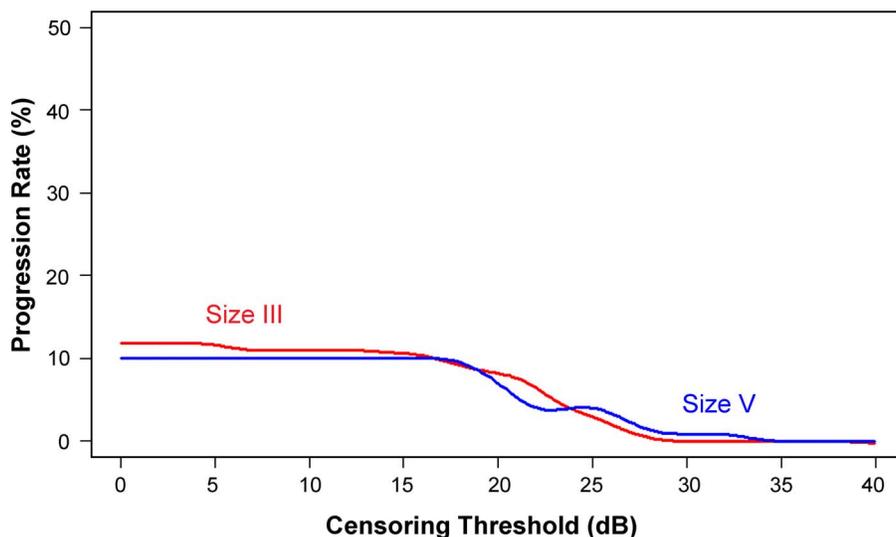


FIGURE 7. The effect of censoring on MD on progression rate for stimulus size III and stimulus size V, with slope less than -0.5 dB/y and *P* value < 0.05 , again shows that values below approximately 20 dB add little to detection; in percentage of subjects. The reduction in progression rate (compared to the uncensored data) became statistically significant ($P < 0.05$, paired test of proportions) at censoring thresholds of 19 dB (size III) and 23 dB (size V).

high repeatability. This would shorten the test in subjects with moderate to severe glaucomatous damage and allow for this saved time to be used for assessment of other more clinically useful test locations.

Acknowledgments

Supported by a VA Merit Review grant, and by an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, New York, New York, United States.

Disclosure: **M. Wall**, None; **G.K.D. Zamba**, None; **P.H. Artes**, None

References

1. Wall M, Woodward KR, Doyle CK, et al. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010;128:570-576.
2. 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.
3. Gardiner SK, Demirel S, Goren D, et al. The effect of stimulus size on the reliable stimulus range of perimetry. *Trans Vis Sci Tech*. 2015;4(2):10.
4. Swanson WH, Horner DG, Dul MW, et al. Choice of stimulus range and size can reduce test-retest variability in glaucomatous visual field defects. *Trans Vis Sci Tech*. 2014;3(5):6.
5. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130-135.
6. Heijl A, Lindgren A, Lindgren G, et al. Inter-test threshold variability in glaucoma: importance of censored observations and general field estimate. In: Mills RP, Heijl A, eds. *Perimetry Update 1988/1989. Proceedings of the VIIIth International Perimetric Society Meeting*. Amsterdam: Kugler; 1989:313-324.
7. Wall M, Doyle CK, Zamba KD, et al. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2013;54:1345-1351.
8. 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.
9. Artes PH, Iwase A, Ohno Y, et al. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654-2659.
10. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci*. 2002;43:1400-1407.
11. Kummet CM, Zamba KD, Doyle CK, et al. Refinement of pointwise linear regression criteria for determining glaucoma progression. *Invest Ophthalmol Vis Sci*. 2013;54:6234-6241.
12. Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci*. 2003;44:3873-3879.
13. O'Leary N, Chauhan BC, Artes PH. Visual field progression in glaucoma: estimating the overall significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Invest Ophthalmol Vis Sci*. 2012;53:6776-6784.
14. Redmond T, O'Leary N, Hutchison DM, et al. Visual field progression with frequency-doubling matrix perimetry and standard automated perimetry in patients with glaucoma and in healthy controls. *JAMA Ophthalmol*. 2013;131:1565-1572.
15. Funkhouser AT, Fankhauser F. A comparison of the mean defect and mean deviation indices within the central 28 degrees of the glaucomatous visual field. *Jpn J Ophthalmol*. 1990;34:414-420.
16. Artes PH, Nicolela MT, LeBlanc RP, et al. Visual field progression in glaucoma: total versus pattern deviation analyses. *Invest Ophthalmol Vis Sci*. 2005;46:4600-4606.
17. 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.