

Reducing Variability of Perimetric Global Indices from Eyes with Progressive Glaucoma by Censoring Unreliable Sensitivity Data

Manoj Pathak¹, Shaban Demirel², and Stuart K. Gardiner²

¹ Department of Mathematics and Statistics, Murray State University, Murray, KY, USA

² Devers Eye Institute, Legacy Research Institute, Portland, OR, USA

Correspondence: Manoj Pathak, PhD, Assistant Professor of Statistics Faculty Hall 6C-19 Department of Mathematics and Statistics Murray State University, Murray, KY, USA. e-mail: mpathak@murraystate.edu

Received: 9 January 2017

Accepted: 29 May 2017

Published: 20 July 2017

Keywords: visual field; sensitivity; censored; variability

Citation: Pathak M, Demirel S, Gardiner SK. Reducing Variability of perimetric global indices from eyes with progressive glaucoma by censoring unreliable sensitivity data.

Trans Vis Sci Tech. 2017;6(4):11, doi: 10.1167/tvst.6.4.11

Copyright 2017 The Authors

Purpose: Recent evidence suggests that increasing perimetric contrast all the way to 0 dB may not be clinically useful. This study examines whether raising the floor for point-wise sensitivities affects the ability of global indices to detect change.

Methods: Longitudinal data from eyes with progressive glaucoma were used. Point-wise sensitivities were censored at various cutoffs (12–19 dB). At each cutoff, mean deviations (MD) were recalculated using censored sensitivities, called censored mean deviation (CMD). Both MD and CMD were fitted using a linear model. MD and CMD rate of changes (signal) and the standard deviations (SD) of the residuals (noise) were obtained from the fitted models. The linear signal to noise ratio (LSNR) for MD (LSNR_{MD}) and CMD (LSNR_{CMD}) were compared. Additionally, at each cutoff, the ratios of LSNR_{CMD} to LSNR_{MD} were calculated and tested.

Results: CMD provided significantly ($P < 0.05$) better LSNR than MD when using any point-wise sensitivity cutoff between 15–19 dB for progressing eyes. Moreover, the ratios of LSNR_{CMD} to LSNR_{MD} were significantly ($P < 0.05$) greater than 1 at all cutoffs from 15–19 dB.

Conclusion: This study demonstrates that censoring is an effective tool to reduce variability at low sensitivities for progressing eyes.

Translational Relevance: This study suggests that 15–19 dB could be a more suitable endpoint for perimetric testing algorithms.

Introduction

Standard automated white-on-white perimetry (SAP) has been a benchmark test for assessing visual fields (VF) in glaucoma. Unfortunately, SAP measurements are known to become considerably more variable when VF damage is present.^{1–6} The intrinsic variability associated with this test, especially in areas manifesting damage, masks true glaucomatous VF progression, making it difficult for clinicians to assess true progression rates. Such variability limits the usefulness of functional testing in glaucoma.

Given this limitation, reducing the variability of SAP testing is an important objective. Various approaches have been implemented to address this issue. For example, variability in SAP data can be reduced through improved data acquisition methods.⁷

Alternatively, post-processing techniques, such as spatial filtering of currently available data^{8–11} appear to reduce variability or choosing a more valid statistical model,¹² explains variability in the SAP data to some degree. Furthermore, a recent study¹³ recommended using Goldmann size V stimuli instead of size III to produce more reliable VF data later into the glaucomatous disease process.

In regions damaged by glaucoma, the 95% test-retest confidence intervals around VF sensitivities are wide.¹ A recent paper by Gardiner et al.¹⁴ examined the reliability of low sensitivities in glaucomatous eyes. They concluded that clinical VF testing appeared to be unreliable when locations had estimated sensitivities below approximately 15–19 dB, and recommended restricting analyses to locations with sensitivities ≥ 19 dB. However, their results focused on point-wise sensitivities. No study to date has exam-

ined whether imposing such a cutoff reduces the variability observed in global indices, such as mean deviation (MD), and whether this would result in a loss of essential clinical information. The overall goal of this study is to identify the optimal floor for point-wise VF sensitivity measurements for determining global VF progression rates and significance when extending the previous point-wise results to the global index MD.

Methods

Data

We used data from the ongoing Portland Progression Project conducted at Devers Eye Institute in Portland, OR (see Pathak et al.¹² for a detailed description). Participants were tested every 6 months with SAP using the Humphrey Field Analyzer II (Carl Zeiss Meditec Inc., Dublin, CA), 24–2 test pattern and standard testing protocols, using the SITA standard algorithm. Only reliable VF tests ($\leq 15\%$ false-positives, $\leq 33\%$ false-negatives, and fixation losses) were included. The study adhered to the tenets of the Declaration of Helsinki, and local institutional review boards approved the protocol.

To address objectives of this study, we censored point-wise VF sensitivities at various cutoff levels from 12–19 dB. These cutoffs were chosen in order to cover the range of cutoffs that Gardiner et al.¹⁴ suggested in their paper. The censoring was achieved by replacing VF sensitivities below the specified cutoff by the given cutoff value. For example, to censor sensitivities at 19 dB, all sensitivities below 19 dB were replaced by 19 dB. After censoring the point-wise sensitivities at the given cutoff, total deviations were calculated at each test location. The censored MDs (CMDs) were derived by taking the mean of all total deviations. We then used longitudinal series of MD and CMD to calculate a linear signal-to-noise ratio (LSNR). The LSNR was defined using the same approach described by Gardiner et al.¹⁴; signal is defined as the rate of change of MD (or CMD) from linear regression and noise is defined as the standard deviation (SD) of residuals from this fitted model. A more negative LSNR indicates that it is easier to distinguish true deterioration from measurement variability. Hence, a more negative LSNR of CMD data compared with MD would provide evidence that censoring reduces the variability and improves the ability to monitor change.

Statistical Analysis

All statistical analyses were performed using freely available statistical software R (R Foundation, Vienna, Austria). Initially at each cutoff, only eyes with eight or more VFs in their longitudinal sequence and having at least one point-wise sensitivity below the given cutoff at any of the 52 nonblindspot 24–2 test locations were included in the analyses. Both MD and CMD data from individual eyes were then fitted using the following linear model implementing generalized estimation equation with autoregressive AR(1) type within eye error structure.¹²

$$MD_i(\text{or CMD}_i) = \alpha + \beta \times t + \varepsilon_t$$

$$i = 1, 2, \dots, N; t = 1, 2, \dots, T_i; \varepsilon_t \sim N(0, \Sigma_{T_i \times T_i})$$

where i represents number of eyes included in the analyses, and t is an indicator representing the number of test time points per eye. The model parameters α , β are the intercept and slope respectively, and ε_t are the errors associated with the fitted line. The model's errors are assumed to be temporally correlated according to a continuous autoregressive (CAR1) model, wherein the correlation between two residuals derived from the same eye decreases with the amount of time between them.

The MD and CMD rates of change (signal), the corresponding SDs of the residuals (noise) and the associated P value of the signal were obtained for each eye from the fitted model. In the final analysis, we selected only those eyes that displayed significant deterioration over time that is only eyes with a significant ($P < 0.05$) negative MD and CMD rate of change were selected for further analyses. Furthermore, at each cutoff, the ratio of $LSNR_{CMD}$ to $LSNR_{MD}$ was also calculated. If censoring VF data was effective at reducing the variability then: (i) $LSNR_{CMD}$ would be significantly lower than $LSNR_{MD}$ at the given cutoff and (ii) the $LSNR_{CMD}$ to $LSNR_{MD}$ ratio would be significantly greater than 1. We tested the above two hypotheses using the Wilcoxon matched pairs test.

Results

Data from 270 eyes with series of at least eight visits were available. The number of eyes included in analyses, however, varied greatly depending on the cutoff used, because eyes were required to have a significant worsening of both MD and CMD when using that cutoff. At cutoff 19 dB, a total of 133 eyes

Table 1. Characteristics of the Study Population Where MD and CMD Represent Uncensored Mean Deviations and Censored Mean Deviations, Respectively

	Mean (\pm SD)	Range
Series length (visits)	16 (\pm 4.02)	(8, 22)
Follow-up duration (y)	11.31 (\pm 2.15)	(4.95, 14.43)
Age at last visit (y)	74.60 (\pm 9.98)	(42, 91)
MD at first visit (dB)	-0.29 (\pm 1.96)	(-9.94, 3.21)
CMD at first visit (dB)	-0.13 (\pm 1.54)	(-4.38, 3.46)
MD at last visit (dB)	-3.90 (\pm 3.87)	(-24.50, 1.26)
CMD at last visit (dB)	-2.74 (\pm 1.97)	(-8.58, 1.36)

with progressive glaucoma (defined as a significant rate of worsening of MD and CMD) were available. Table 1 presents the characteristics of the study population when the cutoff was set at 19 dB, restricted to the eligible series. The mean age was 69.1 (\pm 10.97) years. The mean follow-up duration was 11.31 years (\pm 2.15).

Figure 1 shows histograms of signal (rate of change) for MD (left) and CMD (right) respectively. The mean MD rate of change was -0.28 dB/year (\pm 0.15). Likewise, the mean CMD rate of change was -0.26 dB/year (\pm 0.09). This indicates that the mean rate of progression appeared to be faster for uncensored data (MD) than censored data (CMD).

Figure 2 shows histograms of noise for MD (left) and CMD (right) data, respectively. The mean noise for MD rate of change was 0.73 dB (\pm 1.17). Likewise, the mean noise for CMD rate of change was 0.55 dB (\pm 0.42). The mean noise for CMD was smaller than the mean noise for MD.

To identify the ideal cutoff for reducing variability, and hence improving LSNR, we compared $LSNR_{CMD}$ with $LSNR_{MD}$ at various cutoffs. Table 2 presents the means of $LSNR_{MD}$ and $LSNR_{CMD}$ and

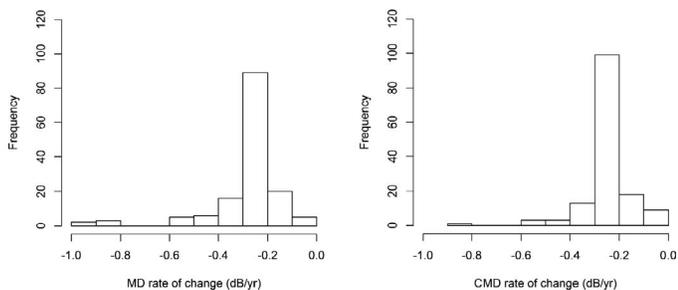


Figure 1. Histograms of signal for MD (left) and CMD (right) at cutoff 19 dB, respectively, where signal is defined as the rate of VF decay in decibel per year.

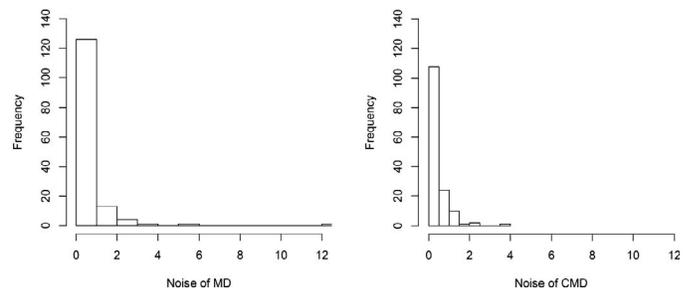


Figure 2. Histogram of noise for MD (left) and CMD (right) at cutoff 19 dB, respectively, where noise is defined as the SD of residuals from the fitted model.

the ratios of $LSNR_{CMD}$ to $LSNR_{MD}$ at various cutoffs, 12 to 19 dB. Censoring appeared to be effective for any cutoff between 15 and 19 dB. For example, at cutoff 19 dB, the $LSNR_{CMD}$ was -0.68 (\pm 0.62), which was significantly ($P < 0.001$) lower (better) than the corresponding $LSNR_{MD}$, -0.57 (\pm 0.52). Furthermore, the ratios $LSNR_{CMD}/LSNR_{MD}$ was significantly greater than 1 at the 19 dB ($P < 0.001$). At all cutoffs below 15 dB, $LSNR_{CMD}$ s and $LSNR_{MD}$ s were statistically equivalent ($P > 0.05$). At the 12 dB cutoff, the ratio $LSNR_{CMD}/LSNR_{MD}$ was not significantly greater than 1 ($P = 0.062$).

Discussion

Several studies have reported that test-retest variability is substantially increased in SAP when glaucomatous VF damage exists.^{1-6,15-20} Various efforts have been made to control perimetric variability, either using better data acquisition methods or by applying filtering or other post-processing techniques to existing data. Our study applied data censoring as an alternative means of reducing variability, extracting a more reliable signal, and attempted to do so without losing the ability to detect and monitor change.

This study demonstrates that censoring leads to a significant increase in the longitudinal signal-to-noise ratio in progressive eyes. For example, $LSNR_{CMD}$ was significantly better than $LSNR_{MD}$ when using cutoffs ranging from 15–19 dB. It may be more efficient to stop estimating perimetric sensitivity when it falls below 15 dB. Notably, this is within the 15–19 dB range for the lower limit of reliable sensitivities that was suggested by Gardiner et al.,¹⁵ using a completely different method that compared the correlation between perimetric sensitivities and those measured using frequency-of-seeing curves. Intuitively, censoring would be expected to reduce variability

Table 2. Comparisons of the Longitudinal Signal to Noise Ratios for Mean Deviations (LSNR_{MD}) and Censored Mean Deviations (LSNR_{CMD})

Cutoff	# of Eyes	LSNR _{MD} (\pm SD)	LSNR _{CMD} (\pm SD)	<i>P</i> Value*	Ratio	<i>P</i> Value**
20	146	-0.56 (0.50)	-0.66 (0.61)	<0.001	1.24	<0.001
19	133	-0.57 (0.52)	-0.68 (0.62)	<0.001	1.23	<0.001
18	124	-0.58 (0.53)	-0.68 (0.63)	<0.001	1.22	<0.001
17	122	-0.58 (0.53)	-0.68 (0.63)	0.002	1.19	<0.001
16	119	-0.58 (0.53)	-0.67 (0.63)	0.005	1.17	0.001
15	114	-0.59 (0.53)	-0.67 (0.64)	0.010	1.15	0.003
14	111	-0.59 (0.54)	-0.66 (0.64)	0.069	1.12	0.024
13	105	-0.60 (0.54)	-0.66 (0.64)	0.093	1.11	0.028
12	105	-0.58 (0.55)	-0.64 (0.64)	0.147	1.10	0.062

The LSNR_{MD} and LSNR_{CMD} were compared using Wilcoxon matched pairs test. The ratio of LSNR_{CMD} to LSNR_{MD} was compared against 1 using Wilcoxon matched pairs tests.

* *P* value for LSNR_{CMD} > LSNR_{MD}.

** *P* value for LSNR_{CMD}/LSNR_{MD} > 1.

at locations below the chosen cutoff. However, if threshold values in this region were reliable and documenting change then the signal would also likely be reduced. Our results show that the LSNR improves; that is, that the reduction in variability outweighs any reduction in signal.

If low sensitivities are not providing useful information about glaucomatous VF progression, then there is no benefit in trying to measure sensitivities that fall this low on the decibel scale. Instead, testing algorithms could be adjusted such that they stop testing below 15 dB, instead of continuing testing down to 0 dB. This would shorten test durations, reducing fatigue for patient, and hence potentially improve the reliability of results from other locations in the VF. Furthermore, it would allow test duration to be more consistent across patients; currently testing takes substantially longer in eyes with severe defects than in relatively healthy eyes. Even though it is useful to know whether sensitivity at a VF location is below 15 dB, using current automated perimetry to determine whether that sensitivity is in fact 5 or 12 dB does not appear to be a worthwhile use of time, because thresholds this low are not sufficiently reliable to be useful.

In this study, only eyes progressing significantly by MD (and CMD) were included, so we may not have included data from eyes in which only one or two locations were deteriorating. This may make our results less applicable in eyes with very early and/or suspected glaucomatous VF damage that have nonsignificant rates of MD and CMD change, or eyes with only a few locations that are changing over

time. However, other analyses have shown that censoring does not harm the ability to distinguish point-wise change from variability.¹⁸ Moreover, even though the potential correlation between repeated measurements was accounted for, linear regression was used to estimate MD and CMD rate of change, which may not be optimal,¹² especially when sensitivities reach the imposed cut off. This happens more for CMD than MD, so our approach may actually underestimate the potential benefits of censoring.

In summary, this study demonstrated that in eyes with progressive glaucomatous VFs changes, censoring point-wise VF sensitivities below 15 dB does not reduce the ability to detect and monitor change using global indices such as MD. Threshold values below 15 dB should be treated cautiously for clinical use and in research studies. Furthermore, this study suggested that 15–19 dB could be a more suitable endpoint for perimetric testing algorithms than continuing testing down to 0 dB.

Acknowledgments

Disclosure: **M. Pathak**, None; **S. Demirel**, None; **S.K. Gardiner**, None

21

References

1. Artes PH, Hutchison DM, Nicolela MT, et al. Threshold and variability properties of matrix frequency-doubling technology and standard

- automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci.* 2005;46:2451–2457.
2. Blumenthal EZ, Sample PA, Berry CC, et al. Evaluating several sources of variability for standard and SWAP visual fields in glaucoma patients, suspects, and normals. *Ophthalmology.* 2000;110:1895–1902.
 3. Chauhan BC, House PH. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology.* 1991;98:79–83.
 4. Heijl A, Lindgren A, Lindgren G. Test retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1989;108:130–135.
 5. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix and motion perimetry. *Invest Ophthalmol Vis Sci.* 2008;50:974–979.
 6. Henson D, Chaudry S, Artes P, Faragher E, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417–421.
 7. Bengtsson, B Olsson, J Heijl, A Rootzen. H A new generation of algorithms for computerized threshold perimetry SITA. *Acta Ophthalmol Scand.* 1997;75:2654–2659.
 8. Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vision Res.* 2004;44:839–848.
 9. Crabb DP, Edgar DF, Fitzke FW, McNaught AI, Wynn HP. New approach to estimating variability in visual field data using an image processing technique. *Br J Ophthalmol.* 1995;79:213–217.
 10. Crabb DP, Fitzke FW, McNaught AI, Edgar DF, Hitchings RA. Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology.* 1997;104:517–524.
 11. Deng L, Demirel S, Gardiner SK. Reducing variability in visual field assessment for glaucoma through filtering that combines structural and functional information. *Invest Ophthalmol Vis Sci.* 2014;55:4593–4602.
 12. Pathak M, Demirel S, Gardiner SK. Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma. *Invest Ophthalmol Vis Sci.* 2013;54:5505–5513.
 13. Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. the effect of stimulus size on the reliable stimulus range of perimetry. *Transl Vis Sci Technol.* 2015;4(2):10.
 14. 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.
 15. 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.
 16. Werner E, Petrig B, Krupin T, Bishop K. Variability of automated visual fields in clinically stable glaucoma patients. *Invest Ophthalmol Vis Sci.* 1989;30:1083–1089.
 17. Piltz J, Starita R. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1990;109:109–110.
 18. Chauhan BC, House PH. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology.* 1991;98:79–83.
 19. Spry P, Johnson C, McKendrick A, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci.* 2001;42:1404–1410.
 20. 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.