

Improvements in Test–Retest Variability of Static Automated Perimetry by Censoring Results With Low Sensitivity in Retinitis Pigmentosa

Ava K. Bittner^{1–3}, Anushka Mistry², Leon Nehmad², Rakin Khan², and Gislin Dagnelie³

¹ Department of Ophthalmology, Stein Eye Institute, University of California, Los Angeles, Los Angeles, CA, USA

² College of Optometry, Nova Southeastern University, Fort Lauderdale, FL, USA

³ Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA

Correspondence: Ava K. Bittner, Department of Ophthalmology, Stein Eye Institute, University of California, Los Angeles, 200 Stein Plaza Driveway, Los Angeles, CA 90095, USA. e-mail: abittner@mednet.ucla.edu

Received: July 15, 2020

Accepted: October 12, 2020

Published: November 20, 2020

Keywords: perimetry; retinitis pigmentosa; variability; visual field; reliability

Citation: Bittner AK, Mistry A, Nehmad L, Khan R, Dagnelie G. Improvements in test–retest variability of static automated perimetry by censoring results with low sensitivity in retinitis pigmentosa. *Trans Vis Sci Tech.* 2020;9(12):26. <https://doi.org/10.1167/tvst.9.12.26>

Purpose: We evaluated whether omitting (censoring) points in more severely damaged visual field areas can reduce test–retest variability of static automated perimetry (SAP) in retinitis pigmentosa (RP), as variability creates a significant challenge when monitoring for changes.

Methods: Cohort 1 included 27 eyes in 16 RP subjects with visual acuity (VA) ranging from 20/20 to 20/70 who completed Humphrey 10-2 size III SAP, once per visit at three visits. Cohort 2 included 15 eyes in nine RP subjects with VA \leq 20/60 who completed Humphrey 30-2 size V SAP, twice per visit at three visits. Variability was assessed using 95% coefficient of repeatability (CR) calculations for uncensored (all threshold values and data points included) and censored data.

Results: In cohort 1, the uncensored between-visit 95% CR was 11.6 decibels (dB); censoring locations with threshold values of <8 to 20 dB resulted in 31% to 53% reductions in the 95% CR. For cohort 2, uncensored 95% CRs were 8.7 and 8.0 dB for within- and between-visit variability, respectively; censoring <8 to 17 dB resulted in 15% to 41% and 15% to 43% reductions in within-visit and between-visit 95% CRs, respectively. For both cohorts, censoring at higher values yielded slightly less variability, at the expense of discarding data from a greater number of eyes and test locations.

Conclusions: For 20/20 to 20/70 VA tested with size III stimuli, censoring lower sensitivity values results in substantially lower test–retest variability, which may help detect true changes for locations without severe baseline loss.

Translational Relevance: A rule of thumb for clinical practices using SAP to monitor RP is that longitudinal losses of >9 dB for individual test locations with initial values ≥ 9 dB are likely to be real and meaningful, as they exceed typical variability.

Introduction

Visual field testing with static automated perimetry (SAP) can identify one of the hallmarks of visual loss in retinitis pigmentosa (RP)—the loss of peripheral vision. Such testing can lead to the diagnosis of RP and aid in monitoring the progression of the disease. However, SAP is plagued by greater test–retest variability in RP than in normally sighted individuals,¹ which is a significant challenge when monitoring for subtle visual changes, such as loss due to progression or possi-

ble improvements from experimental treatments. Previous studies of SAP in RP patients have recognized that, due to the high test–retest variability, only large deviations in sensitivity can be detected and classified as true, meaningful changes,^{1,2} resulting in the need for more frequent testing to confirm changes. The variability of SAP results may be related to both patient-related factors and test-related parameters.³ It has been suggested that the variability of SAP should be documented for each individual with RP, as some cases are more variable than others, and factors such as visual field extent, mean SAP threshold, or patient

age have not been shown to predict SAP test variability in RP.¹

Clinical trials tend to utilize kinetic perimetry to document visual field area in RP, but some recent treatment trials have used 30-2 or 10-2 SAP protocols.⁴⁻⁶ Additionally, because many clinical practices may not have kinetic perimetry (e.g., Goldmann or Octopus 900) or appropriately trained personnel to administer kinetic perimetry, they will continue to rely on SAP to document changes in peripheral vision. A previous study demonstrated that it is feasible to document progression in RP using the 10-2 SAP protocol.⁷ Therefore, it remains important to evaluate novel approaches to attempt to improve the reliability of SAP.

Generally, the variability of SAP increases as sensitivity decreases.⁸ Censoring or omitting test points in areas of severe visual field loss in glaucoma patients has been shown to reduce SAP test-retest variability in previous studies.^{9,10} This approach can be performed after the standard SAP protocol test has been completed by excluding test locations with low sensitivity. A previous study of SAP in RP patients found that the greatest variability tended to occur in regions with low sensitivities, around 8 to 12 decibels (dB), for both stimulus sizes V and III,² a pattern of variability similar to that previously documented for glaucoma.¹¹ Therefore, we hypothesized that censoring points with low sensitivity may have similar effects in RP as in glaucoma to help improve the SAP test reliability. In the current study, we applied this approach to SAP results in two previously collected datasets from RP subjects to explore whether it was possible to reduce the level of between- and within-visit variability.

Methods

The study used de-identified data from adult RP study participants in two different study cohorts who completed multiple SAP tests as part of baseline (pretreatment) testing for two different clinical trials conducted from 2001 to 2002 (cohort 1) and from 2004 to 2005 (cohort 2).^{12,13} Eligibility criteria for the clinical trials have been published previously^{12,13} and included a previously confirmed diagnosis of RP in both eyes, constricted visual fields (VFs) of $<30^\circ$ radius in all directions, and visual acuity (VA) of 20/100 or better for cohort 1 or 20/60 or worse for cohort 2. Both clinical trials received approval from institutional review boards and followed the tenets of the Declaration of Helsinki. Written informed consent was obtained from all subjects before data collection.

The participants' demographics and characteristics of the testing completed for the two study cohorts are shown in Table 1. Both study cohorts completed the Humphrey SAP visual field test during three visits, but different test protocols and number of tests per visit were performed. Cohort 2 completed two SAP tests per eye per visit, whereas cohort 1 completed only one test per eye per visit. Between-visit variability was determined for each cohort by using the results from three tests over three visits—that is, all data collected for cohort 1, as there was only one test per visit, and data from the first test completed at each visit for cohort 2, as the participants completed two tests per visit. Within-visit variability was determinable only for cohort 2. The time between visits was relatively similar between study cohorts, about 1 to 2 months. All VF data were included in the analyses, with none being excluded on the basis of the number of fixation losses, false negatives, or false positives.

Data Analysis

We analyzed SAP threshold values determined at each of the test points but excluded any test points with values < 0 dB at any test session; as such, we included only test locations with sensitivities ≥ 0 dB across all three or six tests for the uncensored dataset. In order to include a test point location in the between-visit or within-visit analyses, all three or six values, respectively, across the three visits had to exceed the censoring value. For cohort 2, when the Humphrey FASTPAC threshold program (Carl Zeiss Meditec, Jena, Germany) provided two values for a single test location, we used the mean of the two values. The 95% coefficient of repeatability (CR) values were calculated in Microsoft Excel ($1.96 \times \text{SD}$ of the differences; Microsoft Corporation, Redmond, WA) for each eye for all threshold values (uncensored) and for those points that included only thresholds above various decibel values, ranging from 8 to 20 dB, for the censored analyses. For between-session variability in both cohorts, we calculated the 95% CR for each subject using all three possible differences across visits for each individual test location: (1) visit 2 – visit 1, (2) visit 3 – visit 2, and (3) visit 3 – visit 1, as well as across all test locations within-subject. For within-session variability in cohort 2, we calculated the 95% CR for each subject using all three possible differences for the two tests performed at each of three visits for each individual test location: (1) test 1 – test 2 at visit 1, (2) test 1 – test 2 at visit 2, and (3) test 1 – test 2 at visit 3, as well as across all test locations within-subject. In order to allow comparison of 95% CR values across censoring levels, the

Table 1. Demographics and Characteristics of the Study Cohorts

	Cohort 1	Cohort 2
Test type	Humphrey 10-2 SITA Standard	Humphrey 30-2 FASTPAC
Stimulus size	III	V
Number of test points	68	76
Visual acuity	20/20–20/70	20/60 or worse
Population size	16 subjects (27 eyes) for censoring <8 to <20 dB; 10 subjects (14 eyes) with first test values in the ranges of 2–5, 6–9, 10–13, 14–17, 18–21, and 22–25 dB	Nine subjects (15 eyes) for censoring <8 to <20 dB
Age range (y)	23–58	32–69
Number of visits	3	3
Number of VF tests per visit	1	2
Time between visits	1–2 mo apart	All within 1 mo

analyses for both cohorts were limited to eyes that contributed values across all censoring levels between <8 and <20 dB; that is, we did not include any eyes that were lost to censoring as the censoring level increased. The total numbers of included eyes and subjects for the two cohorts are listed in Table 1. We are not reporting any censoring analyses for the cut-off levels of 18 to 20 dB in cohort 2, as there was a significant loss of subjects and eyes at those higher censoring levels compared to censoring at 8 to 17 dB. If we had included censoring levels of 18 to 20 dB for cohort 2, then our sample size would have been reduced to only seven eyes in five subjects, which we deemed to be an insufficient number from which to draw conclusions.

We completed an additional subgroup analysis of data from cohort 1 that included 10 eyes in 14 subjects who had baseline (i.e., first test) values within each of the categories of 2 to 5.5 dB, 6 to 9.5 dB, 10 to 13.5 dB, 14 to 17.5 dB, 18 to 21.5 dB, and 22 to 25.5 dB to determine sensitivity-specific mean variability (95% CR) across three visits for thresholds within a limited range, in order to evaluate how variability was related to VF sensitivity in RP and to determine sensitivity-specific criteria for variability for values in the low- to mid-sensitivity range. Subsequent results at the second or third visit were allowed to fall outside of each category range for each test location. This subgroup analysis was not completed in cohort 2, which had worse vision, as only one subject had results across each of the six defined threshold categories, resulting in what was deemed to be an insufficient sample to contribute to a meaningful analysis.

Results

Figure 1 and Table 2 display the between-visit 95% CR values and average number of included test locations for each censoring level ranging from <10 to <20 dB for the SAP 10-2 protocol with size III stimulus completed by subjects in cohort 1 with visual acuity 20/20 to 20/70 or 0 to 0.54 logMAR. For these subjects, censoring points at the lowest (<8 dB) and highest (<20 dB) levels that we explored resulted in 31% and 53% reductions in variability for the between-visit 95% CRs when compared to the uncensored analysis including all data points, as well as mean losses of 8 and 22 test locations (i.e., 24% and 65%, for censoring at <8 dB and <20 dB, respectively) versus no censoring. For cohort 1, censoring between <10 and <13 dB revealed relatively similar 95% CR values, whereas censoring at levels of <16 dB or <20 dB yielded slightly less variability but at the expense of fewer included data points. The standard deviation of individual 95% CR values across all subjects in cohort 1 was 7.9 dB for the uncensored data and ranged from 2.4 to 3.8 dB for the censoring levels between <8 and <20 dB.

Also shown in Figure 1 and Table 2 are the within-visit and between-visit 95% CR values and average number of included test locations for no censoring (all data included) and each censoring level ranging from <8 to <17 dB for the SAP 30-2 protocol with size V stimulus completed by cohort 2 with visual acuity 20/60 or worse. Comparisons between the within- and between-visit variability tended to be very similar at each censoring level. For cohort 2, censoring points

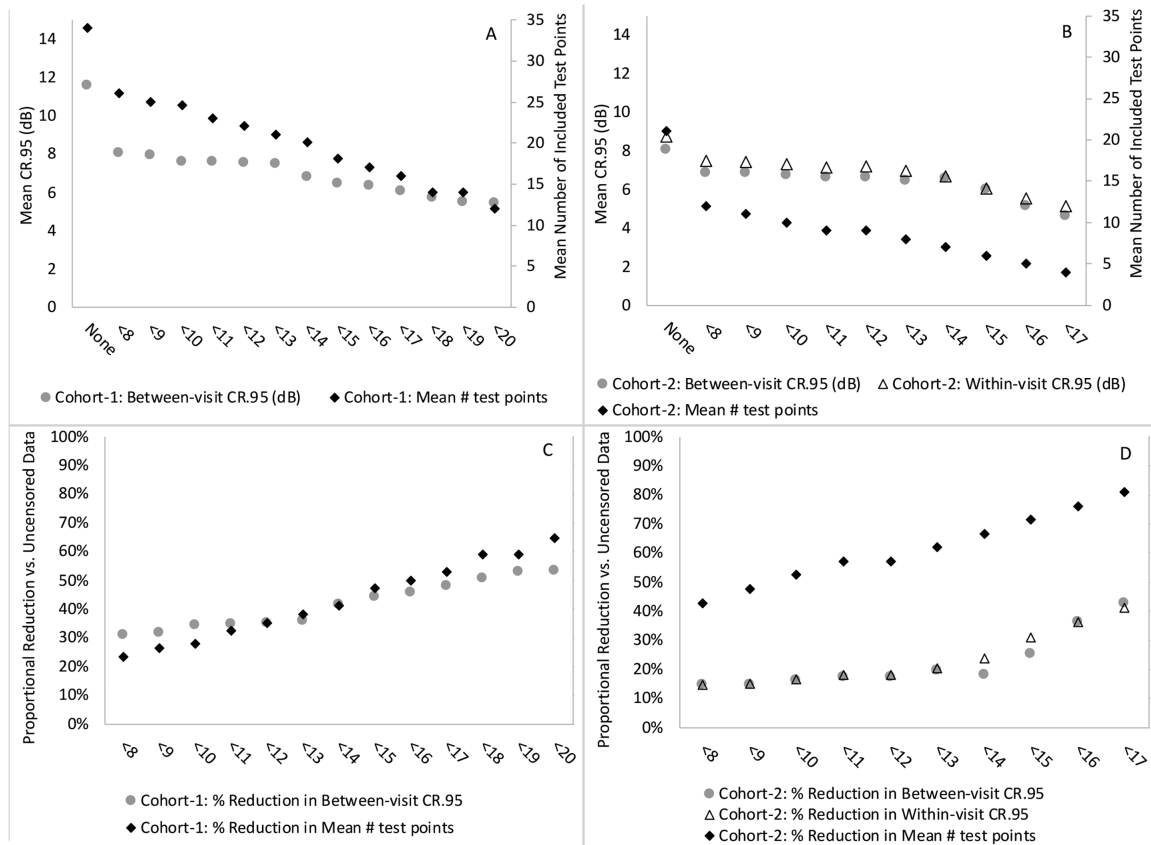


Figure 1. (A, B) Graphs show the mean number of included test points and mean 95% CR values for each censoring level and uncensored data for between-visit variability for cohort 1 (A) and cohort 2 (B), as well as within-visit variability for cohort 2 (B). (C, D) Graphs display the proportional reduction in included test points and mean 95% CR values when comparing each censoring level to uncensored data for between-visit variability for cohort 1 (C) and cohort 2 (D), as well as within-visit variability for cohort 2 (D).

at the lowest level (<8 dB) and highest level (<17 dB) that we explored resulted in 15% and 41% reductions in within-visit variability compared to no censoring and 15% and 43% reductions in between-visit variability for mean losses of 43% and 81% of test locations, respectively. For cohort 2, between-visit variability remained relatively constant for censoring levels of <10 to 14 dB, then decreased for higher levels of censoring at <15 to 17 dB. The standard deviations for the within-visit and between-visit 95% CR values across all subjects in cohort 2 ranged from 1.8 to 3.0 dB for uncensored data and censoring levels between <8 and <17 dB.

In a separate analysis of all included study eyes for cohort 1 and cohort 2, we determined the between-visit variability (95% CR) for test locations at which the initial, first test value was between 0 and 9.5 dB. This analysis included an average of 5.7 and 5.8 test locations per eye for cohorts 1 and 2, respectively. The results revealed that test–retest variability was much greater for cohort 1 (95% CR, 13.7 dB) with

better vision tested with the smaller size III stimulus than for cohort 2 (95% CR, 7.5 dB) with worse vision tested with the larger size V stimulus when considering only the least sensitive initial test values between 0 and 9.5 dB.

In a separate subgroup analysis, we determined the between-visit variability (95% CR) for test locations at which the initial, first test value was within the six threshold range categories defined in the Methods section for cohort 1 eyes with at least one test location that contributed to each of these six categories. Figure 2 displays the between-visit 95% CRs across three tests for each of the baseline threshold categories that included an average of 2.6 to 7.9 test locations across eyes. The results indicate that variability was greatest for the least sensitive baseline values, between 2 and 5.5 dB (95% CR, 10.8 dB) and then is gradually reduced with increasing sensitivity to the lowest variability for baseline values between 22 and 25.5 dB (95% CR, 7.1 dB).

Table 2. Mean Number of Included Test Points and Mean 95% CR Values for Each Censoring Level and Uncensored Data for Between-Visit Variability for Cohorts 1 and 2 and Within-Visit Variability for Cohort 2

Censoring Value (dB)	Cohort 1, VA 20/20–20/70 (27 Eyes)		Cohort 2, VA 20/60 or Worse (15 Eyes)		
	Between-Visit		Between-Visit		Within-Visit
	Average Number of Test Points	95% CR (dB)	Average Number of Test Points	95% CR (dB)	95% CR (dB)
None	34	11.57	21	8.03	8.74
<8	26	8.01	12	6.85	7.47
<9	25	7.92	11	6.86	7.44
<10	25	7.59	10	6.72	7.29
<11	24	7.56	9	6.63	7.16
<12	23	7.50	9	6.64	7.17
<13	22	7.44	8	6.46	6.97
<14	20	6.78	7	6.58	6.66
<15	18	6.45	6	6.01	6.03
<16	18	6.30	5	5.13	5.56
<17	16	6.02	4	4.60	5.15
<18	14	5.69	—	—	—
<19	14	5.44	—	—	—
<20	12	5.42	—	—	—

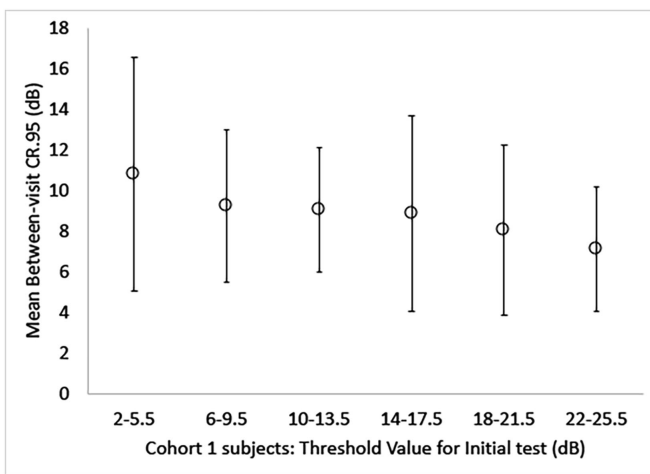


Figure 2. Graph of mean between-visit variability as 95% CR values for six baseline threshold range categories for cohort 1 eyes. The error bars represent the standard deviations across all subjects' 95% CR values.

Discussion

Censoring of low sensitivity values at various levels (<8–20 dB) reduced both within- and between-visit variability in SAP testing for RP patients, with a tendency for a greater reduction in variability when censoring a larger number of thresholds within this range. However, a major consequence of censoring is that substantially fewer data points, eyes, and subjects were included as the censoring levels increased, which was a greater concern for RP patients with vision 20/60 or worse, as these patients tend to have lower sensi-

tivity values for many of the test points than patients with better vision (20/20–20/70). The data in Figure 1 and Table 2 suggest that censoring at levels higher than 8 to 13 dB can further reduce variability but at the expense of a reduction in the number of test locations contributing to the result. For the cohort 1 data up to 13 dB, the percent reduction in variability compared to the uncensored value exceeded the percent reduction in eyes, but that balance reversed beyond this level, so for this particular sample it would not be advisable to censor beyond 13 dB, and <9 dB may be a more prudent, recommended level given the variability of 9 dB demonstrated in our subgroup analysis. For the data from cohort 2, even censoring at <8 dB came at the expense of a greater percent drop in contributing eyes than the percent reduction in variability. For this particular sample, censoring at values less than <8 dB would not make sense, as any lower values fall within a range of 95% CR from 0 dB and are thus affected by a floor effect; therefore, censoring in cohort 2 would not be beneficial as it was in cohort 1 due to the much greater loss of test locations and much less reduction in variability for censoring <8 dB in cohort 2.

Censoring low sensitivity values with high variability can simplify the analysis and interpretation process, as longitudinal changes can be assessed for global measures, such as the mean sensitivity, for all included points above a threshold value rather than having to consider the sensitivity and variability at individual test locations when censoring is not applied. Another alternative to censoring might be to utilize 95% CRs for different baseline sensitivity ranges, as in our subgroup analysis in Figure 2, in order to determine whether

a given test location exceeds the variability threshold and thus shows significant improvement or worsening. We prefer to advocate for censoring <9 dB values or considering the variability of individual test locations rather than using the average of all (uncensored) test locations, as the profile for the hill of vision (i.e., steep with very few locations with low sensitivity or flat with many areas of low sensitivity) will vary across RP patients, and the average of all test locations may not reflect how many are in the low sensitivity range that contributes to greater test variability.

For both RP cohorts included in the current analysis, the improvement in variability when censoring at higher values was accompanied by a loss of subjects and eyes; thus, it would not generally be advisable to censor at levels greater than 9 to 10 dB for most RP patients. When we censored values <10 dB, only one eye was excluded from cohort 1, and about a third of cohort 2 eyes were excluded when compared to the original uncensored dataset for all of the clinical trial participants. In contrast, when we censored values <18 to 19 dB, a much larger proportion of eyes was excluded, such that a fifth of the subjects in cohort 1 and half of the subjects in cohort 2 were no longer included, as all of their test results were within the censored range. If treatment effects or visual changes occur in the range of 10 to 20 dB, censoring at levels within this range would be problematic because improvements or losses in vision would be missed; thus, we recommend restricting censoring to <10 dB for an ~ 4 -dB improvement in between-visit SAP variability (95% CR) with size III stimuli when compared to no censoring in people with RP who have visual acuity better than 20/70. For those RP patients, a general rule of thumb for clinicians monitoring for longitudinal changes might be to disregard points with sensitivities that drop below 9 dB and to consider any change that exceeds the typical variability of ~ 9 dB for the remaining points to be a real, meaningful change. We recommend a more judicious consideration or use of censoring, which may not be helpful at even <8 dB for RP patients with vision 20/60 or worse who are tested with the large size V stimulus due to exclusion of substantial amounts of data for only a modest improvement in test variability. Clinical trials may consider our recommendations for censoring but should have the flexibility to choose to modify or tailor our suggestions according to their needs (e.g., for therapies with anticipated effects within a specific range of retinal sensitivities) or if the trial participants may vary from our study cohort. Given the relatively small sample size in cohort 2 and the subgroup analysis for cohort 1, there is a risk that our findings may not generalize to other cohorts with RP if our subjects are significantly different than

RP patients seen clinically or who would be enrolled in future clinical trials.

Future clinical trials for RP may need to define criteria for censoring a priori and assess the effects of censoring after baseline testing but prior to intervention to determine if there are excluded individuals with severe vision loss after censoring, in which case other participants may have to be enrolled or investigators may need to consider using other approaches. When monitoring changes in areas with mild to moderate loss of sensitivity, the approach of censoring test points with low sensitivities is ideal; however, censoring would not be appropriate when attempting to detect changes at the edge of a patient's peripheral field of vision due to the typical hill of vision usually noted in RP. In such cases, kinetic perimetry would be more valuable to reliably measure changes in viable peripheral retinal area.¹⁴ If using SAP to evaluate changes in very low sensitivity values, it could be valuable to utilize sensitivity-dependent criteria when monitoring for natural progressive vision loss or improvements during clinical trials when a new treatment may target damaged photoreceptors with sensitivities in the range of 0 to 10 dB. If the intention is to monitor individual SAP test locations with values between 0 and 9.5 dB at the initial baseline test, we recommend using a size V stimulus and a rule of thumb for gauging longitudinal improvements in sensitivity that subsequent longitudinal changes ≥ 8 dB in RP patients are likely to be real and meaningful, as they exceed typical test-retest variability. However, only test locations with initial values of 9 dB can be reliably assessed longitudinally for true loss of sensitivity, as subsequent longitudinal changes greater than or equal to -9 dB could indicate either a change in the extremely low sensitivity values of 0 or 1 dB or a total loss of sensitivity <0 dB. Test locations with initial values of 0 to 8 dB cannot be reliably assessed longitudinally for loss of sensitivity, as the typical test-retest variability exceeds the value itself.

A comparison of the between-visit variability for the two study cohorts revealed that the uncensored test variability was much lower (by ~ 3.5 dB) for cohort 2 subjects who had visual acuity 20/60 or worse and completed the testing with the size V stimulus than the subjects in the other cohort with better vision who completed the testing with the size III stimulus. For censoring between <8 and <17 dB, the between-visit variabilities between cohorts were more similar, with differences between cohorts ranging from 0.3 to 1.5 dB; cohort 2, with worse vision and larger stimulus size, had slightly less variability than cohort 1 across all censoring levels. Our finding for the difference in uncensored, between-visit variability when comparing the cohorts was somewhat unexpected because previous

research in glaucoma patients found that the greatest factor for increased test–retest variability was increasing scotoma depth¹⁵ or decreased visual field sensitivity, which explained over half of the total variability.¹⁶

For our subgroup analysis in cohort 1, test variability was greater for threshold values of 2 to 5.5 dB than for values of 6 to 9.5 dB. That factor related to test variability, however, was not likely responsible for the difference in uncensored variability between cohorts because both cohorts had approximately the same proportion of test locations with values between 2 and 5.5 dB at the first test (i.e., 2.6% for cohort 1 and 2.0% for cohort 2), as well as about the same proportion of test locations with values between 0 and 9.5 dB at the first test (i.e., 7.2% for cohort 1 and 8.7% for cohort 2). Subjects in cohort 1 had much greater variability on average for the test locations with initial values from 0 to 9.5 dB than cohort 2 subjects with initial values from 0 to 9.5 dB. Thus, we hypothesize that the difference in uncensored variability between our RP cohorts was most likely attributable to the difference in the test target size (i.e., retinal areas with low sensitivity, 0–9.5 dB, are more reliably evaluated with the larger size V stimulus), as previous research has demonstrated less test variability with the larger stimulus size V compared to size III in glaucoma patients.^{15,17} The larger stimulus implemented in the study of cohort 2 subjects has another potential advantage of shifting patients' sensitivity to higher levels, which can allow for more test locations and/or eyes to be included during censoring, in addition to increasing the dynamic range to assess change and providing the potential for less variability at greater sensitivity values.¹⁸

In cohort 2, the differences between within-visit and between-visit variability were quite minimal, with differences ranging from 0.02 to 0.62 dB across the censoring levels of <8 to <17 dB, and slightly greater variability within visits than between visits. A previous study in patients with severe vision loss due to ocular diseases such as RP, macular disease, optic nerve disease, or diabetic retinopathy, found that between-visit variability for the Humphrey visual field test in scotopic test conditions was greater than or similar to within-session variability,¹⁹ a finding similar to our current findings for censored and uncensored data.

Longitudinal studies involving improvements or loss of sensitivity on SAP in RP are needed to explore the recommendations proposed here for censoring or using sensitivity-specific criteria for monitoring change. In patients with mild to moderate glaucoma, censoring SAP values <20 dB had relatively little impact on the ability to detect progression rates.²⁰ Future studies using a longitudinal dataset should confirm whether censoring values of <8 dB has

any substantial impact on the ability to detect the progression of RP. Finally, for censoring to be implemented in future clinical trials or in clinical practice, it will be imperative to develop software to automate the calculations for each censoring level and set an optimization criterion to make the process more efficient.

Acknowledgments

Supported by Optobionics Corporation (2004–2007) and a National Center for Complementary and Alternative Medicine award (R21AT000292, GD). The work of AKB is supported by an unrestricted grant from Research to Prevent Blindness, Inc., to the Department of Ophthalmology at the University of California, Los Angeles.

Disclosure: **A.K. Bittner**, None; **A. Mistry**, None; **L. Nehmad**, None; **R. Khan**, None; **G. Dagnelie**, None

References

1. Seiple W, Clemens CJ, Greenstein VC, Carr RE, Holopigian K. Test-retest reliability of the multifocal electroretinogram and Humphrey visual fields in patients with retinitis pigmentosa. *Doc Ophthalmol*. 2004;109(3):255–272.
2. Kim LS, McAnany JJ, Alexander KR, Fishman GA. Intersession repeatability of Humphrey perimetry measurements in patients with retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2007;48(10):4720–4724.
3. Rozanski C, Haythornthwaite JA, Dagnelie G, Bittner AK. Applying theories and interventions from behavioral medicine to understand and reduce visual field variability in patients with vision loss. *Med Hypotheses*. 2014;83(2):190–195.
4. Hoffman DR, Hughbanks-Wheaton DK, Spencer R, et al. Docosahexaenoic acid slows visual field progression in X-linked retinitis pigmentosa: ancillary outcomes of the DHAX Trial. *Invest Ophthalmol Vis Sci*. 2015;56(11):6646–6653.
5. Birch DG, Bennett LD, Duncan JL, Weleber RG, Pennesi ME. Long-term follow-up of patients with retinitis pigmentosa receiving intraocular ciliary neurotrophic factor implants. *Am J Ophthalmol*. 2016;170:10–14.
6. Nakazawa M, Suzuki Y, Ito T, Metoki T, Kudo T, Ohguro H. Long-term effects of nilvadipine against progression of the central visual field defect

- in retinitis pigmentosa: an extended study. *Biomed Res Int*. 2013;2013:585729.
7. Sayo A, Ueno S, Kominami T. Longitudinal study of visual field changes determined by Humphrey Field Analyzer 10-2 in patients with retinitis pigmentosa. *Sci Rep*. 2017;7(1):16383.
 8. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci*. 2012;53(10):5985–5990.
 9. Pathak M, Demirel S, Gardiner SK. Reducing variability of perimetric global indices from eyes with progressive glaucoma by censoring unreliable sensitivity data. *Transl Vis Sci Technol*. 2017;6(4):11.
 10. 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(1):288–294.
 11. Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Res*. 2007;47(7):925–936.
 12. Bahrami H, Melia M, Dagnelie G. Lutein supplementation in retinitis pigmentosa: PC-based vision assessment in a randomized double-masked placebo-controlled clinical trial. *BMC Ophthalmol*. 2006;6:23.
 13. Chow AY, Bittner AK, Pardue MT. The artificial silicon retina in retinitis pigmentosa patients (an American Ophthalmological Association thesis). *Trans Am Ophthalmol Soc*. 2010;108:120–154.
 14. Bittner AK, Iftikhar MH, Dagnelie G. Test-retest, within-visit variability of Goldmann visual fields in retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2011;52(11):8042–8046.
 15. Maddess T. The influence of sampling errors on test-retest variability in perimetry. *Invest Ophthalmol Vis Sci*. 2011;52(2):1014–1022.
 16. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41(2):417–421.
 17. Wall M, Kutzko KE, Chauhan BC. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Invest Ophthalmol Vis Sci*. 1997;38(2):426–435.
 18. 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.
 19. Kiser AK, Mladenovich D, Eshraghi F, Bourdeau D, Dagnelie G. Reliability and consistency of dark-adapted psychophysical measures in advanced eye disease. *Invest Ophthalmol Vis Sci*. 2006;47(1):444–452.
 20. 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*. 2017;59(1):439–445.