Relating Standardized Automated Perimetry Performed with Stimulus Sizes III and V in Eyes With Field Loss due to Glaucoma and NAION

David Szanto, BA[1]
Michael Wall, MD[2]
Luke X Chong, PhD[3]
Mark J Kupersmith, MD[4]

[1]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York
[2]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa
[3]School of Medicine, Deakin University, Geelong, Australia
[4]Departments of Neurology, Ophthalmology and Neurosurgery, CNIIC, Icahn School of Medicine at Mount Sinai, NY.

Corresponding Author:
Mark Kupersmith, MD
New York, NY 10029
(212) 636-3200
mark.kupersmith@mountsinai.org

Funding: The New York Eye and Ear Infirmary Foundation, New York, N.Y.; NEI EY032522; Research to Prevent Blindness, Inc., New York, NY unrestricted grant to the Department of Ophthalmology; Shulman Family NAION Fund at Icahn School of Medicine at Mount Sinai

Word count: 2108

Disclosures: D. Szanto, None; M. Wall, None; L.X. Chong, Carl Zeiss Meditec (C, F); M.J. Kupersmith, None

Objective: Standard automated perimetry (SAP) visual field (VF) results are more repeatable using Goldmann stimulus size V (stimV) in eyes with moderate/severe deficits due to glaucoma. There are few reports relating VFs using stimulus size V and III, typically used in the clinic for glaucoma, and none for non-arteritic anterior ischemic optic neuropathy (NAION). We hypothesized that we could compare and relate the VFs with both stimuli for glaucoma and NAION.

Methods: We utilized 1992 same-day pairs of stimIII and stimV SAP VFs using the 24-2 strategy for eyes with glaucoma or NAION. We explored the optimal threshold to censor the raw sensitivities, prior to calculating age-standardized total deviations (TD). We determined the mean and standard deviation of the differences among all TD pairs. We computed a line of best fit to determine closeness to the line of unity.

Results: The ideal censoring conversion threshold was 21 dB for stimIII and 24 dB for stimV. The difference between stimV and stimIII censored (0.0 ± 1.9 dB) and uncensored (0.4 ± 2.6 dB) TD pairings strongly correlate with each other ($r^2 = 0.70$, $p < 0.001$). The line of best fit from these pairings has a slope of 0.92, which is similar to that of the line of unity ($m = 1$).

Conclusion: Censoring plus age correction is a valid method of comparison between stimIII and stimV SAP VFs with moderate to severe VF loss due to optic nerve disorders.

Translational Relevance: StimIII and stimV TDs are interchangeable in clinical practice.

Introduction

Standard Automated Perimetry (SAP) is the most common method used for VF testing to measure visual sensitivities at multiple points to detect deficits in the central 30 degrees of vision. The Goldmann stimulus size III (stimIII) is widely used but subject to useful dynamic range limitation in eyes that have moderate to severe deficits.[1] The stimulus size V (stimV), which uses a larger stimulus, has a wider useful dynamic range and is more reliable for testing eyes with more advanced glaucoma VF loss and also has better test-retest repeatability.[2]

A key challenge is comparing or converting visual field (VF) results obtained with different stimulus sizes, particularly for longitudinal data or inter-patient comparisons. Converting VFs obtained with stimV to one that would have been obtained with stimIII can standardize data, making it easier to track disease progression and compare results across studies and clinical settings. However, this conversion process has not been standardized due to the differences in sensitivity and response characteristics between the two stimulus sizes.

Censoring sensitivities below a certain threshold is necessary because data points below about 20 dB levels are dominated by noise, making them unreliable and unrepeatable.[3] Within this range, there is high retest variability which distorts statistical measures and obscures meaningful patterns. Censoring involves adjusting all sensitivity

values below a specific threshold to that threshold value. While the exact cutoff for the useful dynamic range is debated (somewhere below between 17 and 25 dB), setting these low sensitivity values in both stimuli to a predefined threshold should minimize variability with more consistent and interpretable data.[4]

Glaucoma and non-arteritic anterior ischemic optic neuropathy (NAION) are both leading causes of vision impairment in adults.[5,6] Both cause irreversible vision loss ranging from mild visual field defects to blindness, and can significantly impact the quality of life; they are monitored using VF testing.[5,6] Since many patients with both diseases have severely depressed VFs, stimV could be used to better monitor the disease progression. This would be particularly helpful in clinical trials that investigate therapies for eyes with moderate to severe VF deficits. Recently, new methods to analyze patterns of VF loss have complemented more global measures of VF loss. Quantifying the specific losses in regions of interest should lead to more precise monitoring and gauging the effects of therapy. Relating these patterns for stimV and stimIII in the same patient or participant is needed. However, we currently cannot directly compare stimV VFs to stimIII VFs, particularly as stimIII VFs are more likely used earlier in the disease course for glaucoma when it is mild.

This study explored whether we could determine a reasonable threshold censoring level for each point in the 24-2 VF and conversion factor of stimV to stim III VFs in individuals with moderate to severe VF loss due to glaucoma and NAION. We used two existing datasets that contained both stimIII and stimV data for the same individuals.

Methods

This study was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai and required no additional consent as the data used were de-identified and derived from participants who had consented for use of their data at multiple study institutions. The study was conducted according to the tenets of the Declaration of Helsinki.

Study Design and Participants

Glaucoma

Glaucoma stimIII and stimV data were received from a trial investigating differences in variability between differently sized perimetric stimuli and their ability to discriminate between healthy and damaged visual fields in glaucoma patients. The study compared abnormal test locations across different stimuli sizes to compare findings and extend the analysis to size modulation perimetry. This study involved data (previously reported) from 120 participants with glaucoma with moderate to severe VF loss who underwent same-day VF testing using both 24-2 SITA-Standard size III and full threshold size V at the University of Iowa Department of Ophthalmology and Visual Sciences.[7] Participants were included if they had glaucomatous optic disc changes with abnormal conventional automated perimetry and were diagnosed with primary, secondary, or normal-tension glaucoma with no other vision-affecting diseases. Exclusion criteria included cataract causing visual acuity worse than 20/30, pupil size less than 2.5 mm, age under 19, or being pregnant at the time of study entry.

NAION

NAION data were received from the Quark207 trial, a multinational, prospective, five-armed randomized controlled trial aimed at investigating the safety and efficacy of a biologic in individuals, ages 50–80, diagnosed with NAION within 14 days of vision loss, meeting study entry criteria.[8] The study included 729 participants, who were using the Humphrey Field Analyzer with the 24-2 SITA-Standard size III and full threshold size V, which was added after recruitment began. The two stimulus types were tested on the same day. Raw sensitivity values in decibels (dB) were recorded for each test. VFs were measured at screening, day 1 of enrollment, two months, six months, and up to one year. There were same-day VFs using both stimuli for participants at month 6 (493), month 12 (414), and various unscheduled times (32).

Data Censoring and Conversion

We determined optimal censoring thresholds by comparing the average difference of censored TD values between the results for both stimuli and selecting the thresholds for each stimuli that minimized this difference. We then censored sensitivities by replacing values below the defined threshold with that value. We then converted the sensitivities to age-corrected total deviation (TDs) values using a normative database for both stimuli. We also determined the optimal censoring thresholds for each disease

separately to demonstrate that the data can be effectively combined in a threshold analysis.

Data Visualization & Analysis

We performed all statistical analyses in a Jupyter notebook with Python 3.8.8. All visualizations were done with the open-source python module "matplotlib".[9] We plotted pairs of TDs for all points in all VFs (103,584 pairings) as well as the line of unity (y = x) showing the perfect agreement of points. We also plotted the line of best fit and computed a coefficient of determination. We calculated mean and standard deviations between pairs of TDs, focusing on both uncensored and censored pairs (where at least one pairing is censored) by condition. We performed a paired t-test on the difference of all TD pairings.

Results

We investigated a total of 1992 VFs from 706 participants where 120 had moderate to severe glaucoma and 586 had NAION. In participants with glaucoma, the mean age was 67.8 ± 9.3 years and 39% were male, and for participants with NAION the mean age was 61.3 ± 7.7 years and 69% were male (Table 1). The mean TD after censoring in participants with glaucoma for stimIII and stimV was -4.7 ± 3.7 dB and -4.2 ± 3.9 dB respectively, and for NAION was -7.2 ± 3.4 dB and -7.3 ± 3.8 dB respectively. We discovered that the optimal censoring threshold was 21 for stimIII and 24 for stimV, which provided the lowest mean difference and low standard deviation (Figure 1, Supplemental Figure 1).

Overall, the average difference between stimV and stimIII TD pairings for uncensored pairs was 0.4 ± 2.6 dB, and a censored pair difference of 0.0 ± 1.9 dB (Figure 2). A paired t-test analyzing differences of TDs show a statistically significant nonzero (0.1 dB) difference between the two stimuli ($p < 0.001$). Glaucoma VF differences between stimIII and stimV were remarkably consistent over all subtypes, showing an uncensored pair difference of 0.4 ± 2.6 dB and a censored pair difference of 0.0 ± 1.9 dB. NAION VFs consist of an uncensored pair difference of 0.2 ± 2.9 dB and a censored pair difference of -0.3 ± 1.8 dB. Notably, 63% of pairings in NAION were censored. We plotted a line of best fit along all pairs of points, and we found that it was similar to the line of unity (Figure 3). There was a strong correlation between TD pairings ($r^2 = 0.70$). Opting to not censor TD pairings reveals that, beyond a certain threshold, data points increasingly deviate from the line of unity in an unpredictable way (Figure 4).

In performing threshold analysis on just glaucoma, we found that the optimal censoring threshold was 20 dB for stimIII and 22 dB for stimV. Using these thresholds for glaucoma TD pairings, the resultant line of best fit had a slope of 0.91 with strong correlation ($r^2 = 0.68$). The same analysis on only NAION TD pairings resulted in censoring thresholds of 19 dB for stimIII and 22 dB for stimV, and the resultant line of best fit had a slope of 0.67 with a similar correlation ($r^2 = 0.67$).

Discussion

We found that censoring and converting sensitivities to TDs enables direct comparison between stimIII and stimV. The average difference for censored and uncensored data is

183 marginal, but the retest variability is markedly reduced using censoring. Determining an
184 optimal censoring threshold for VFs for each stimulus improves the comparison
185 between stimuli particularly in VFs with regions of major sensitivity loss.

186 We determined that the optimal censoring thresholds for stimIII and stimV are 21 dB
187 and 24 dB for this data set of two different optic nerve disorders. These thresholds
188 provided the lowest mean differences and low standard deviations, and are both within
189 a reasonable range, allowing for better comparison between eyes with severely
190 depressed VFs. These thresholds are similar to those calculated for each disease
191 separately, and the lines of best fit for the combined and single disease computations
192 have similar slopes and r^2 values. Therefore, our calculated overall thresholds of 21 dB
193 and 24 dB for their respective stimuli may be applicable across a broader range of optic
194 nerve diseases

195

196 A paired t-test revealed a small, but significant, average difference between TD
197 pairings. However, this number is very close to zero, and considering same-day testing
198 variability, this number has minimal clinical relevance. Prior studies have shown that the
199 inter-test variability for specific points range from 1.3 dB - 3.0 dB for stimIII and 1.2 dB -
200 2.0 dB for stimV.[7] Thus, the two stimuli should be largely interchangeable in clinical
201 practice if censoring and age corrected TDs are used. This allows clinicians to switch
202 from a test using one stimulus size to another without clinically meaningful
203 discrepancies affecting clinical decisions. For NAION and glaucoma, the average TD
204 difference and standard deviation were very similar (as well as between censored and

uncensored pairs), suggesting that our methodology may broadly apply for a variety of optic nerve disorders.

The line of best fit for TD pairings closely followed the line of unity. Coupled with its high r^2 value, suggests there is strong agreement between stimIII and stimV across different visual field severities. The variability in our uncensored data scatterplot highlights the necessity of censoring to maintain reliability and consistency.

This study has several clinical implications. Clinicians can choose between stimIII and stimV based on the degree of visual field damage without compromising diagnostic accuracy. We also demonstrated that patients undergoing regular VF testing with a mix of either stimuli will have reliable tracking of disease progression. Future research can leverage both stimIII and stimV within the same dataset, thereby eliminating the need to segregate data by Goldmann stimulus size. Our optimal censoring thresholds also improve the comparability of VFs not only within the same patient but across different patients as well in the setting of severe vision loss, aiding in better disease management.

Our study had limitations typical of retrospective analyses. First, widespread population data sets for VFs performed using stimulus V are lacking and are not included in the current Humphrey perimeters. Differences in testing protocols or equipment settings between the two datasets may have introduced variability that would affect the comparability of stimIII and stimV results, particularly due to the presence of multiple testing sites in both groups. Of course, having NAION VFs performed at 80 sites supports the potential for real world applicability of our method. We also only tested VFs

of participants with NAION and glaucoma with substantial damage so we had relatively fewer data points near a TD of 0 and above.

Our study shows that usage of normative databases for stimIII and stimV, in conjunction with censoring, allows for the direct comparison between VFs using either of the two stimuli. Future directions may include validating these findings across diverse patient populations, extending the comparison to include other stimuli types, or comparing visual field patterns between the stimuli with methods such as archetypal analysis.

Characteristics of Study Participants

	# VFs	Age (years)	% Male	% Right eye
Glaucoma (n = 120)	**1053**	**67.8 ± 9.3**	**39%**	**53%**
NAION (n = 586)	**939**	**61.3 ± 7.7**	**69.3%**	**51%**

Table 1. Demographic information for participants with glaucoma and non-arteritic anterior ischemic optic neuropathy.

Mean TD Differences Between StimIII and StimV Pairs Across Censoring Thresholds

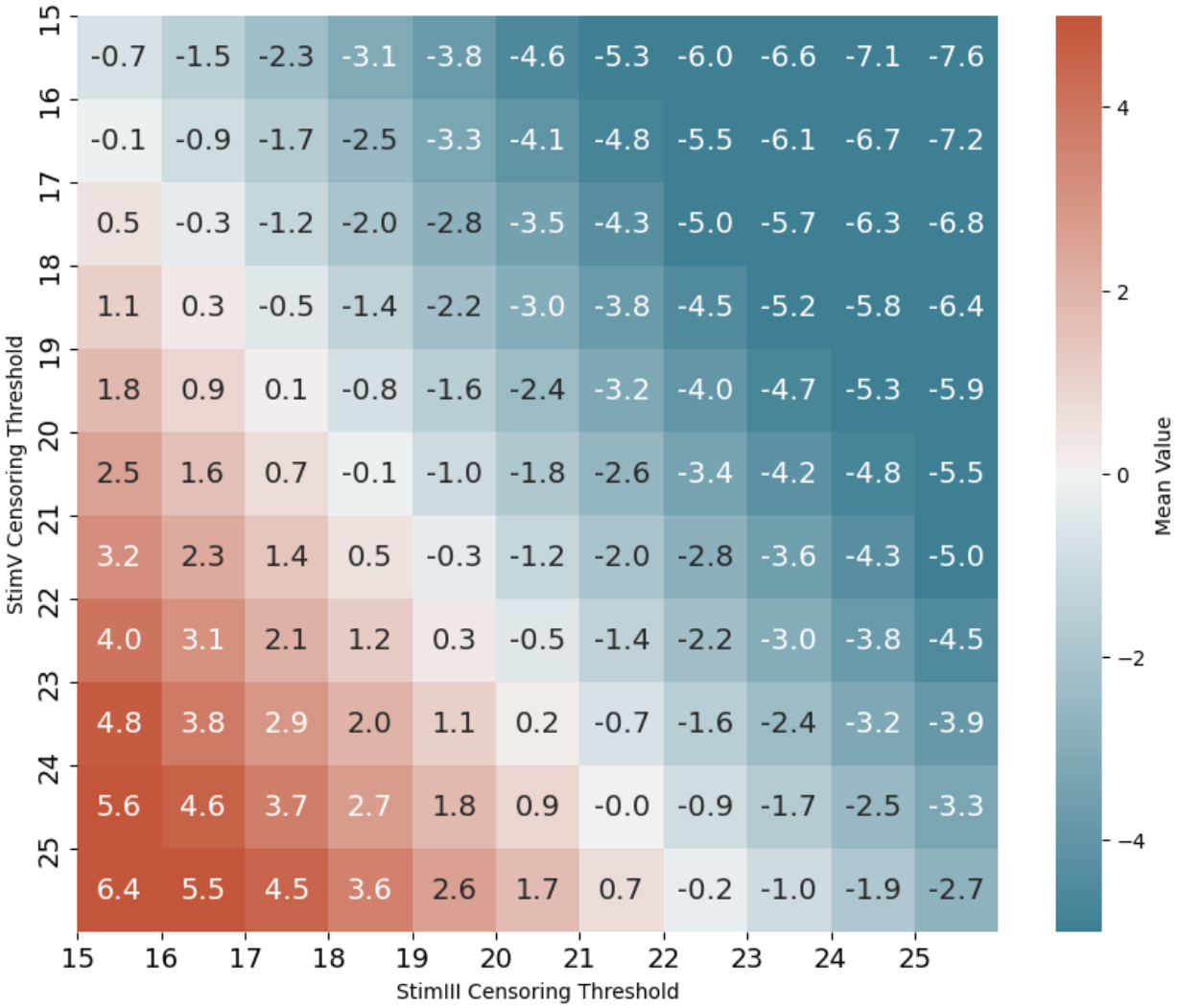
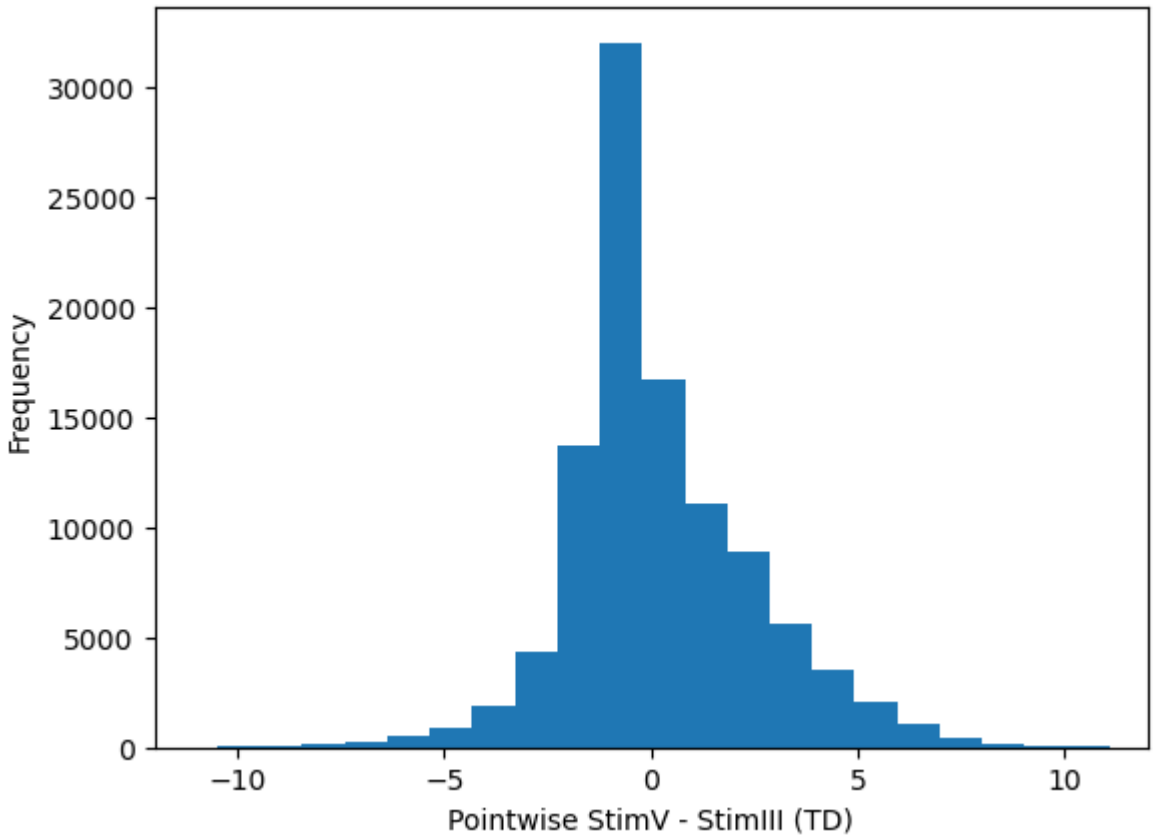


Figure 1. Heatmap depicting the average total deviation difference between pairs of stimulus size III and stimulus size V points across varying censoring thresholds where at least one stimulus was censored. The optimal censoring thresholds were identified at 21 dB for stimIII and 24 dB for stimV.

246 **Pointwise TD Differences Between StimV and StimIII**

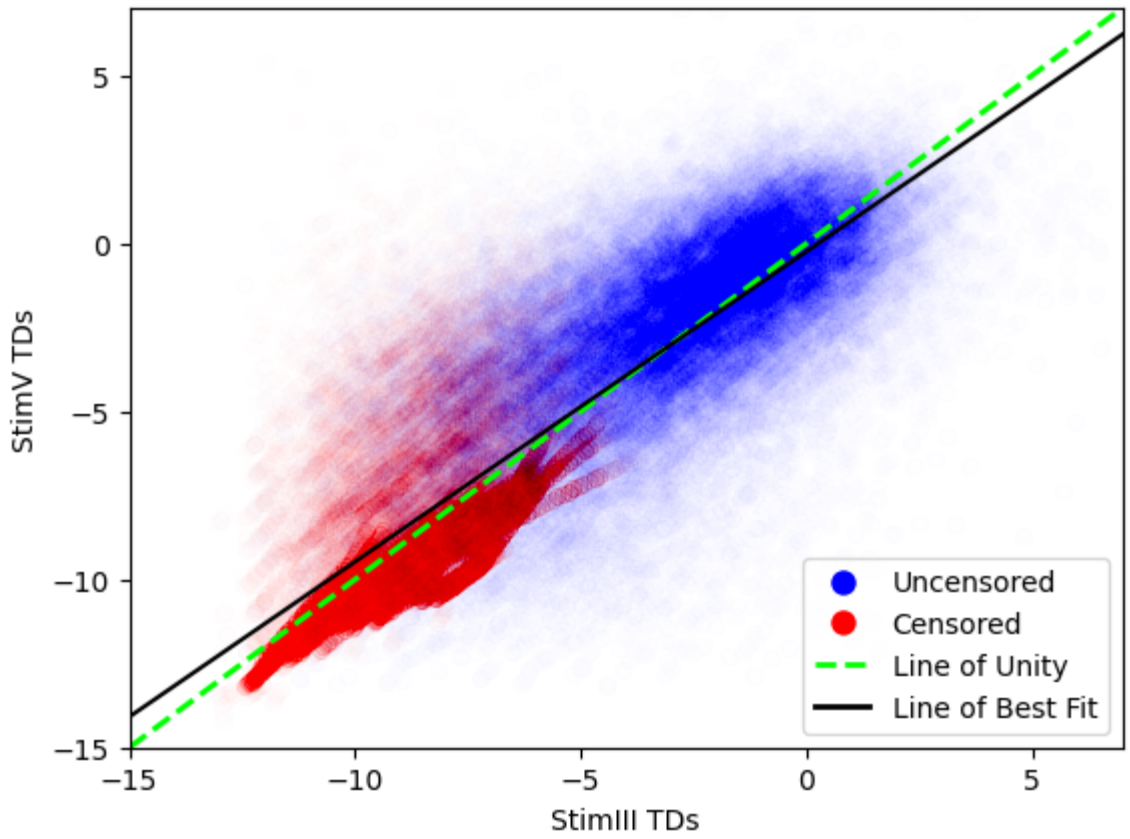


247

248 **Figure 2.** Histogram illustrating the distribution of pointwise total deviation differences
249 between stimulus size V and stimulus size III visual fields. The histogram is centered
250 around 0, indicating comparable differences between the two stimuli across the tested
251 points.

252

253 **Scatterplot of TD Pairs Between StimIII and StimV**



254

255 **Figure 3.** Pairs of total deviation values from stimulus size III and stimulus size V visual
256 fields. Data points are colored based on whether at least one of the pair is censored.
257 The dotted green line represents the line of unity (y = x). The black line of best fit (y = -
258 0.24 + 0.92x) demonstrates a strong linear relationship between the two stimuli (r^2 =
259 0.70). Data points are made transparent for improved visibility and to illustrate relative
260 density.

261

Scatterplot of Uncensored TD Pairs Between StimIII and StimV

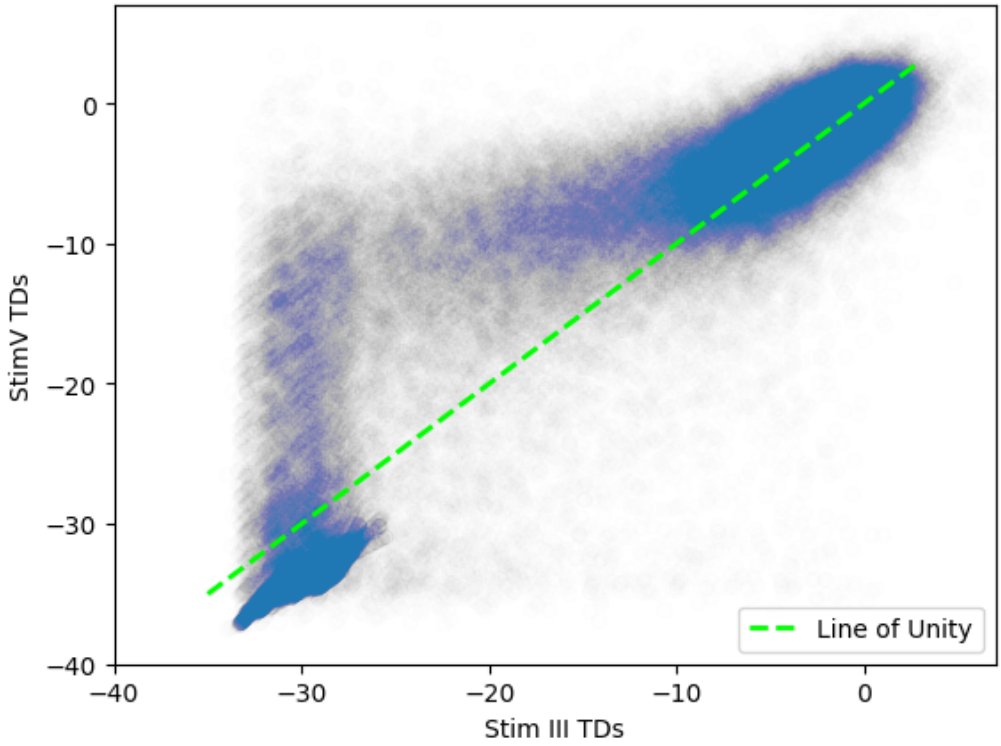
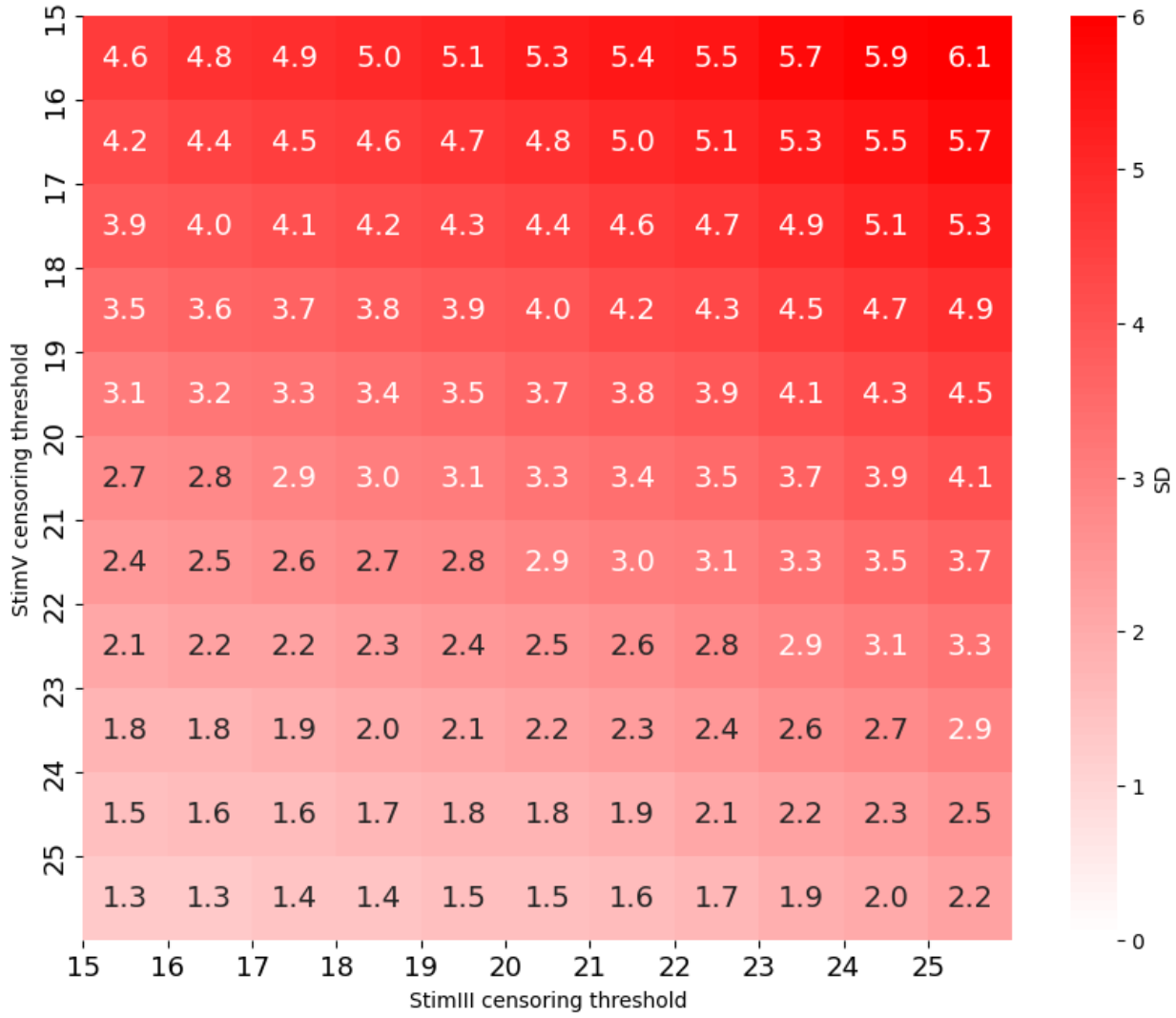


Figure 4. Pairs of uncensored total deviation values from stimulus size III and stimulus size V visual fields. The dotted green line represents the line of unity (y = x). Beyond a certain threshold, points increasingly deviate from the line of unity, highlighting point variability in uncensored measurements. Data points are made transparent for improved visibility and to illustrate their relative density.

269 SD of Differences Between StimIII and StimV Pairs Across Censoring Thresholds



270

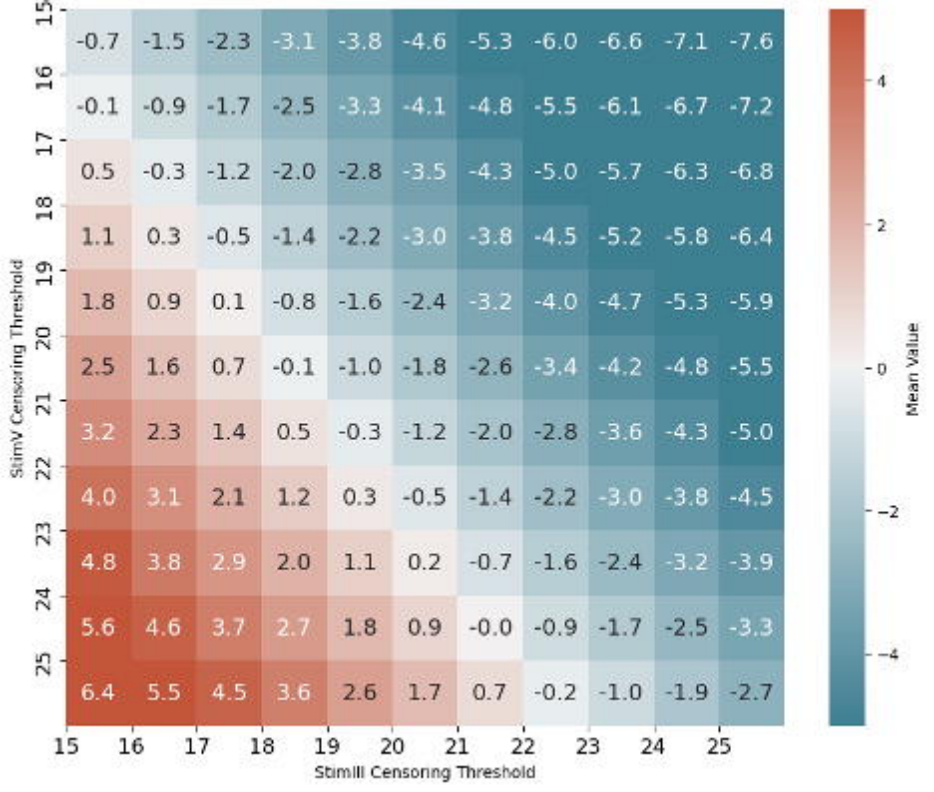
271 **Supplemental Figure 1.** Heatmap depicting the standard deviation of total deviation
272 differences between pairs of stimulus size III and stimulus size V points across varying
273 censoring thresholds where at least one stimulus was censored. For the optimal
274 censoring threshold of 21 dB for stimulus size III and 24 for stimulus size V, the
275 standard deviation is 1.9 dB.

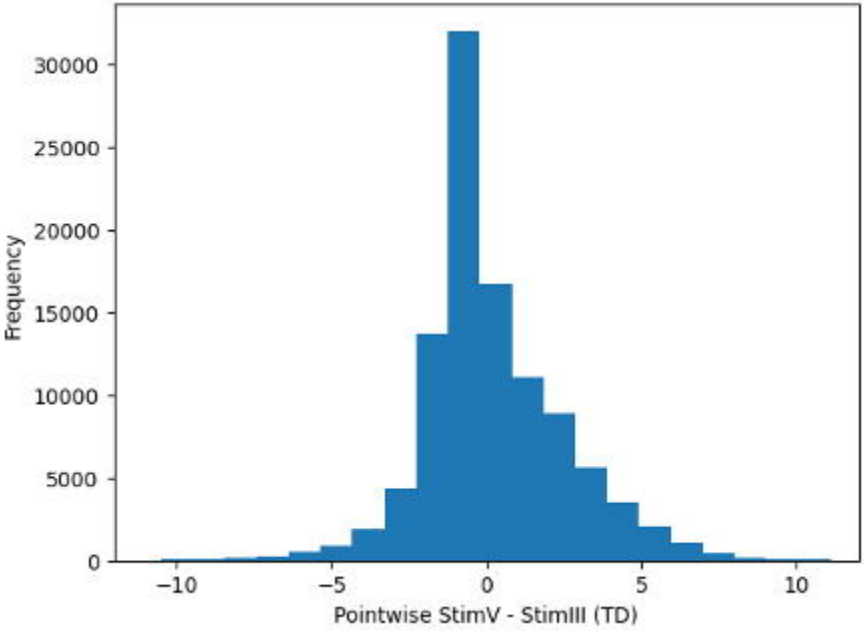
276

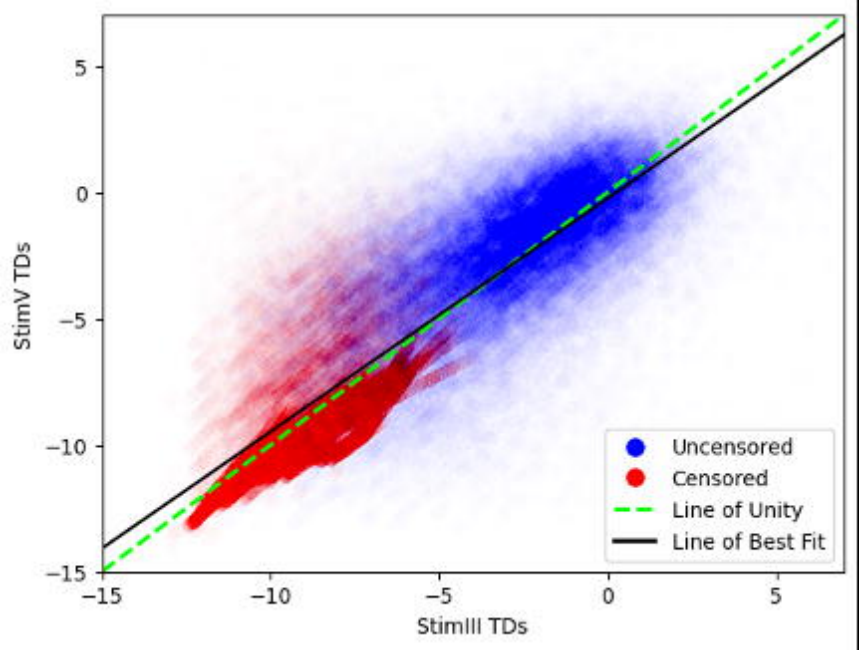
277 References

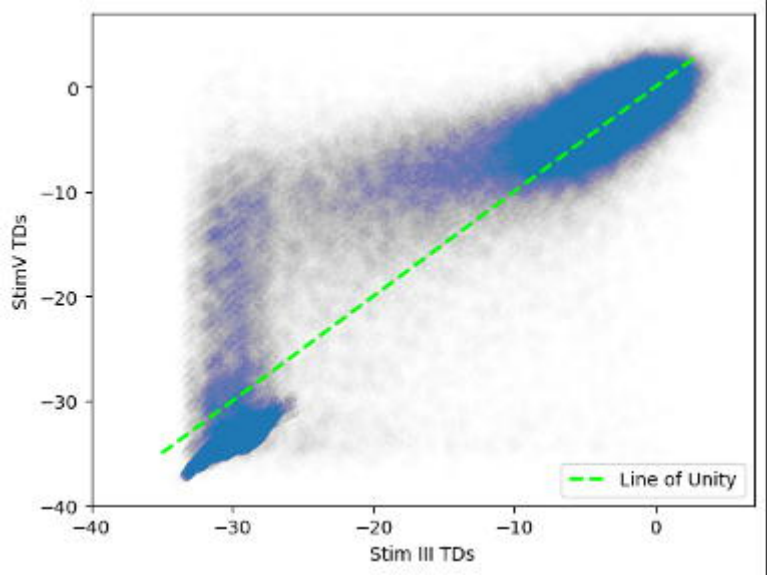
1. Wall M, Brito CF, Woodward KR, Doyle CK, et al. Total deviation probability plots for stimulus size v perimetry: a comparison with size III stimuli. *Arch Ophthalmol*. 2008;126(4):473-479.
2. Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. The effect of stimulus size on the reliable stimulus range of perimetry. *Transl Vis Sci Technol*. 2015;4(2):10-10.
3. 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(7):1359-1369.
4. Wall M, Zamba GK, Artes PH. The effective dynamic ranges for glaucomatous visual field progression with standard automated perimetry and stimulus sizes III and V. *Invest Ophth Vis Sci*. 2018;59(1):439-445.
5. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ*. 2004;82:887-888.
6. Hattenhauer MG, Leavitt JA, Hodge DO, Grill R, Gray DT. Incidence of nonarteritic anterior ischemic optic neuropathy. *Am J Ophthalmol*. 1997; 123(1):103-107.
7. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophth Vis Sci*. 2013;54(2):1345-1351.
8. Kupersmith MJ, Fraser CL, Morgenstern R, et al. Ophthalmic and Systemic Factors of Acute Nonarteritic Anterior Ischemic Optic Neuropathy in the Quark207 Treatment Trial. *Ophthalmology*. 2024;131(7):790-802.

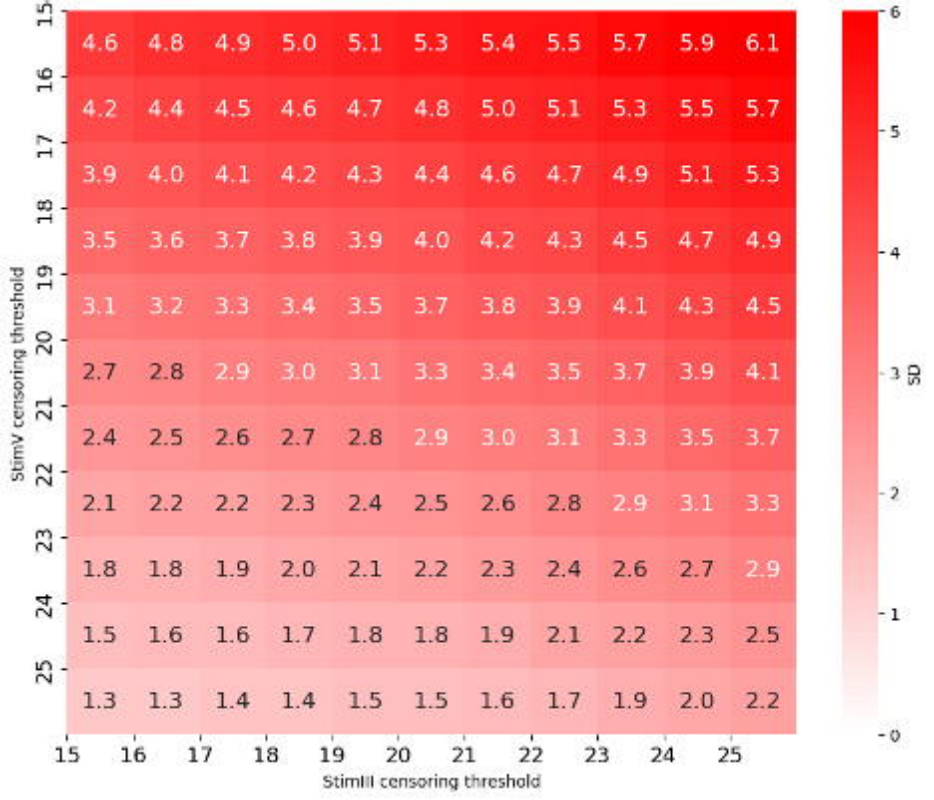
9. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng*. 2007;9(3):90-95.


StimV Censoring Threshold
15 16 17 18 19 20 21 22 23 24 25
-0.7 -1.5 -2.3 -3.1 -3.8 -4.6 -5.3 -6.0 -6.6 -7.1 -7.6
-0.1 -0.9 -1.7 -2.5 -3.3 -4.1 -4.8 -5.5 -6.1 -6.7 -7.2
0.5 -0.3 -1.2 -2.0 -2.8 -3.5 -4.3 -5.0 -5.7 -6.3 -6.8
1.1 0.3 -0.5 -1.4 -2.2 -3.0 -3.8 -4.5 -5.2 -5.8 -6.4
1.8 0.9 0.1 -0.8 -1.6 -2.4 -3.2 -4.0 -4.7 -5.3 -5.9
2.5 1.6 0.7 -0.1 -1.0 -1.8 -2.6 -3.4 -4.2 -4.8 -5.5
3.2 2.3 1.4 0.5 -0.3 -1.2 -2.0 -2.8 -3.6 -4.3 -5.0
4.0 3.1 2.1 1.2 0.3 -0.5 -1.4 -2.2 -3.0 -3.8 -4.5
4.8 3.8 2.9 2.0 1.1 0.2 -0.7 -1.6 -2.4 -3.2 -3.9
5.6 4.6 3.7 2.7 1.8 0.9 -0.0 -0.9 -1.7 -2.5 -3.3
6.4 5.5 4.5 3.6 2.6 1.7 0.7 -0.2 -1.0 -1.9 -2.7
15 16 17 18 19 20 21 22 23 24 25
StimII Censoring Threshold
Mean Value
4
2
0
−2
−4


30000
25000
20000
15000
10000
5000
0
Frequency
−10
−5
0
5
10
Pointwise StimV - StimIII (TD)


StimV TDs
StimIII TDs
5
0
−5
−10
−15
−15
−10
−5
0
5
Uncensored
Censored
Line of Unity
Line of Best Fit


0
−10
−20
−30
−40
StimV TDs
−40
−30
−20
−10
0
Stim III TDs
Line of Unity


StimV censoring threshold
StimII censoring threshold
SD
15 16 17 18 19 20 21 22 23 24 25
0 1 2 3 4 5 6
4.6 4.8 4.9 5.0 5.1 5.3 5.4 5.5 5.7 5.9 6.1
4.2 4.4 4.5 4.6 4.7 4.8 5.0 5.1 5.3 5.5 5.7
3.9 4.0 4.1 4.2 4.3 4.4 4.6 4.7 4.9 5.1 5.3
3.5 3.6 3.7 3.8 3.9 4.0 4.2 4.3 4.5 4.7 4.9
3.1 3.2 3.3 3.4 3.5 3.7 3.8 3.9 4.1 4.3 4.5
2.7 2.8 2.9 3.0 3.1 3.3 3.4 3.5 3.7 3.9 4.1
2.4 2.5 2.6 2.7 2.8 2.9 3.0 3.1 3.3 3.5 3.7
2.1 2.2 2.2 2.3 2.4 2.5 2.6 2.8 2.9 3.1 3.3
1.8 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.6 2.7 2.9
1.5 1.6 1.6 1.7 1.8 1.8 1.9 2.1 2.2 2.3 2.5
1.3 1.3 1.4 1.4 1.5 1.5 1.6 1.7 1.9 2.0 2.2