

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

3
4 David Szanto, BA¹

5 Michael Wall, MD²

6 Luke X Chong, PhD³

7 Mark J Kupersmith, MD⁴

8
9 ¹Department of Neurology, Icahn School of Medicine at Mount Sinai, New York

10 ²Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa

11 ³School of Medicine, Deakin University, Geelong, Australia

12 ⁴Departments of Neurology, Ophthalmology and Neurosurgery, CNIIC, Icahn School of
13 Medicine at Mount Sinai, NY.

14
15 Corresponding Author:

16 Mark Kupersmith, MD

17 New York, NY 10029

18 (212) 636-3200

19 mark.kupersmith@mountsinai.org

20
21 Funding: The New York Eye and Ear Infirmary Foundation, New York, N.Y.; NEI

22 EY032522; Research to Prevent Blindness, Inc., New York, NY unrestricted grant to the

23 Department of Ophthalmology; Shulman Family NAION Fund at Icahn School of

24 Medicine at Mount Sinai

25
26 Word count: 2108

27
28 Disclosures: D. Szanto, None; M. Wall, None; L.X. Chong, Carl Zeiss Meditec (C, F);

29 M.J. Kupersmith, None

30

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

37

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

43

44 **Results:** The ideal censoring conversion threshold was 21 dB for stimIII and 24 dB for
45 stimV. The difference between stimV and stimIII censored (0.0 ± 1.9 dB) and
46 uncensored (0.4 ± 2.6 dB) TD pairings strongly correlate with each other ($r^2 = 0.70$, $p <$
47 0.001). The line of best fit from these pairings has a slope of 0.92, which is similar to
48 that of the line of unity ($m = 1$).

49

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

53

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

56

57 **Introduction**

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

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

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

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

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

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

97

98 **Methods**

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

104 **Study Design and Participants**

105 **Glaucoma**

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

119

120 **NAION**

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

131

132 **Data Censoring and Conversion**

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

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

141

142 Data Visualization & Analysis

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

151 Results

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

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

173

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

180 **Discussion**

181 We found that censoring and converting sensitivities to TDs enables direct comparison
182 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

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

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

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

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

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

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

234 **Characteristics of Study Participants**

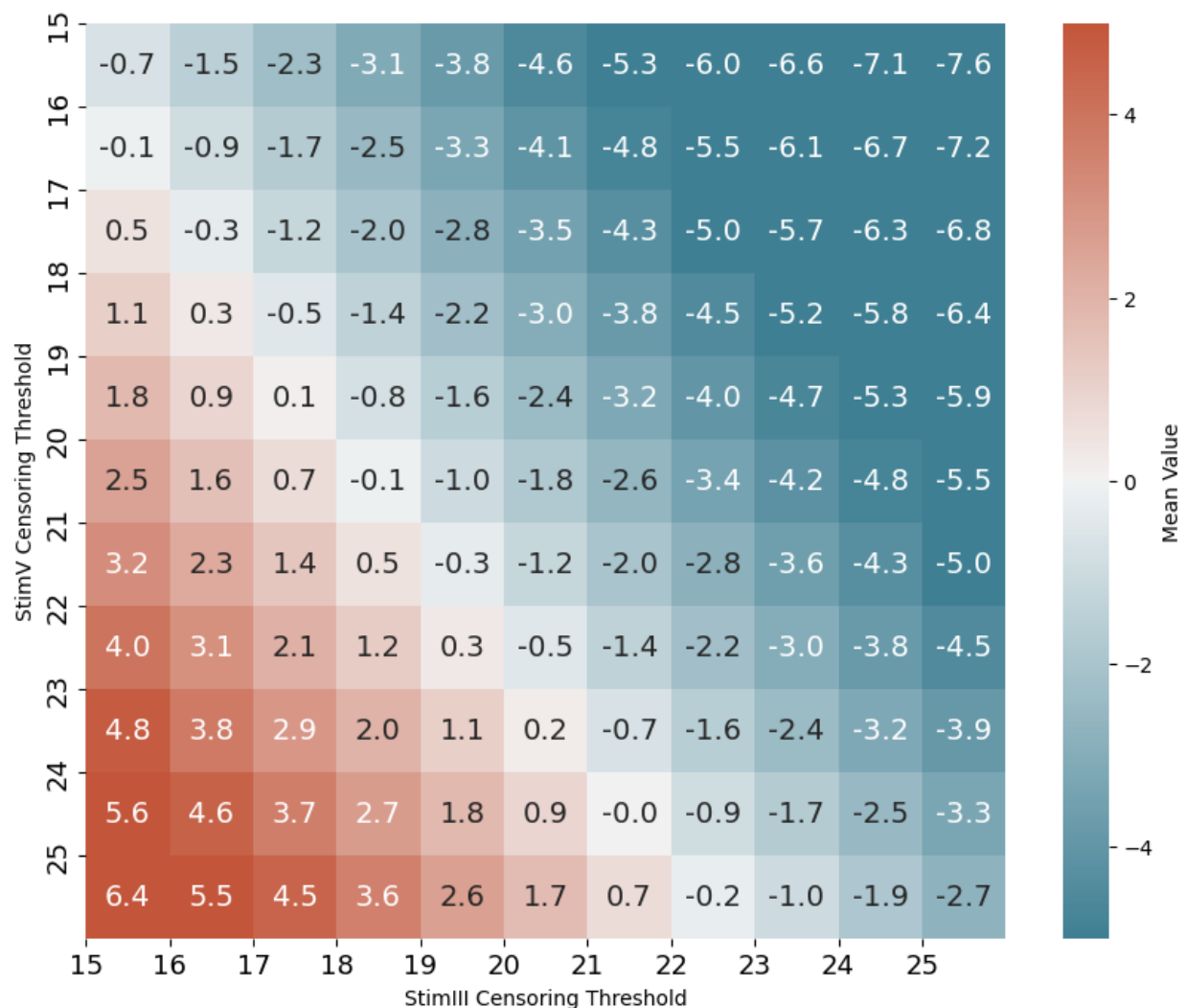
	# VFs	Age (years)	% Male	% Right eye
Glaucoma (n = 120)	1053	67.8 ± 9.3	39%	53%
NAION (n = 586)	939	61.3 ± 7.7	69.3%	51%

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

237

238 **Mean TD Differences Between StimIII and StimV Pairs Across Censoring**

239 **Thresholds**

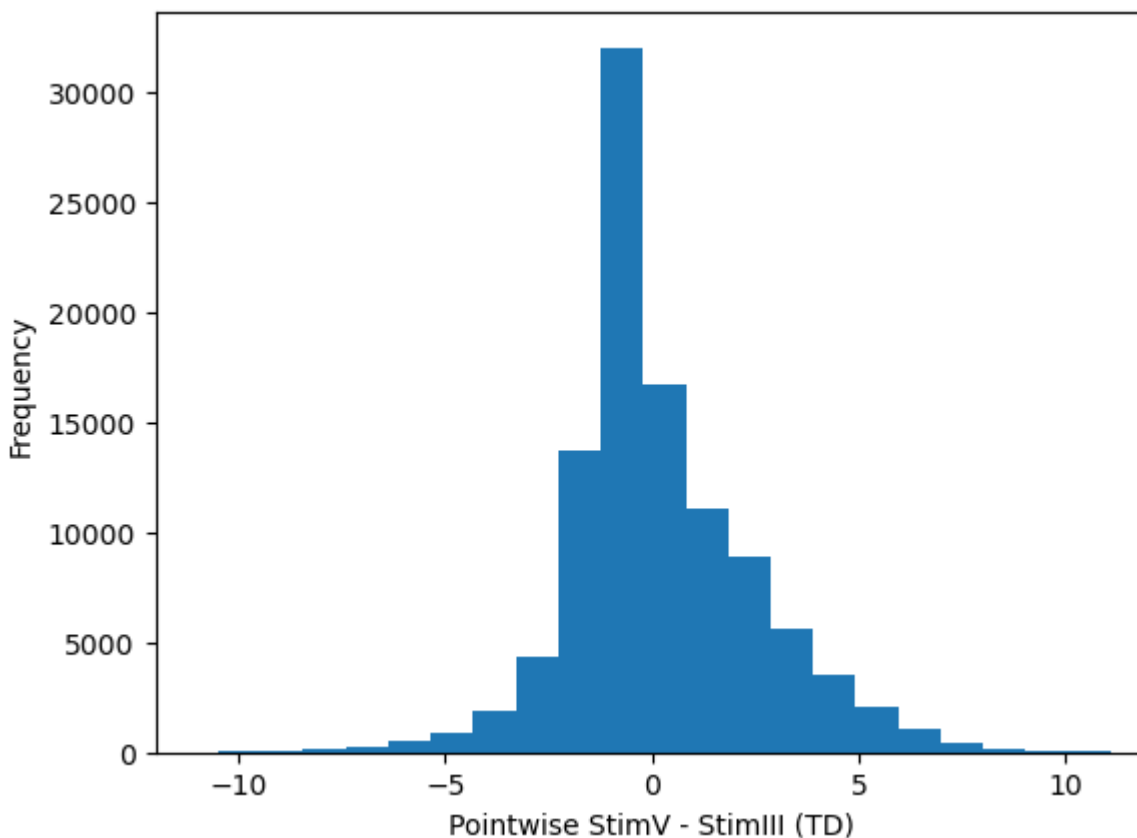


240

241 **Figure 1.** Heatmap depicting the average total deviation difference between pairs of
242 stimulus size III and stimulus size V points across varying censoring thresholds where
243 at least one stimulus was censored. The optimal censoring thresholds were identified at
244 21 dB for stimIII and 24 dB for stimV.

245

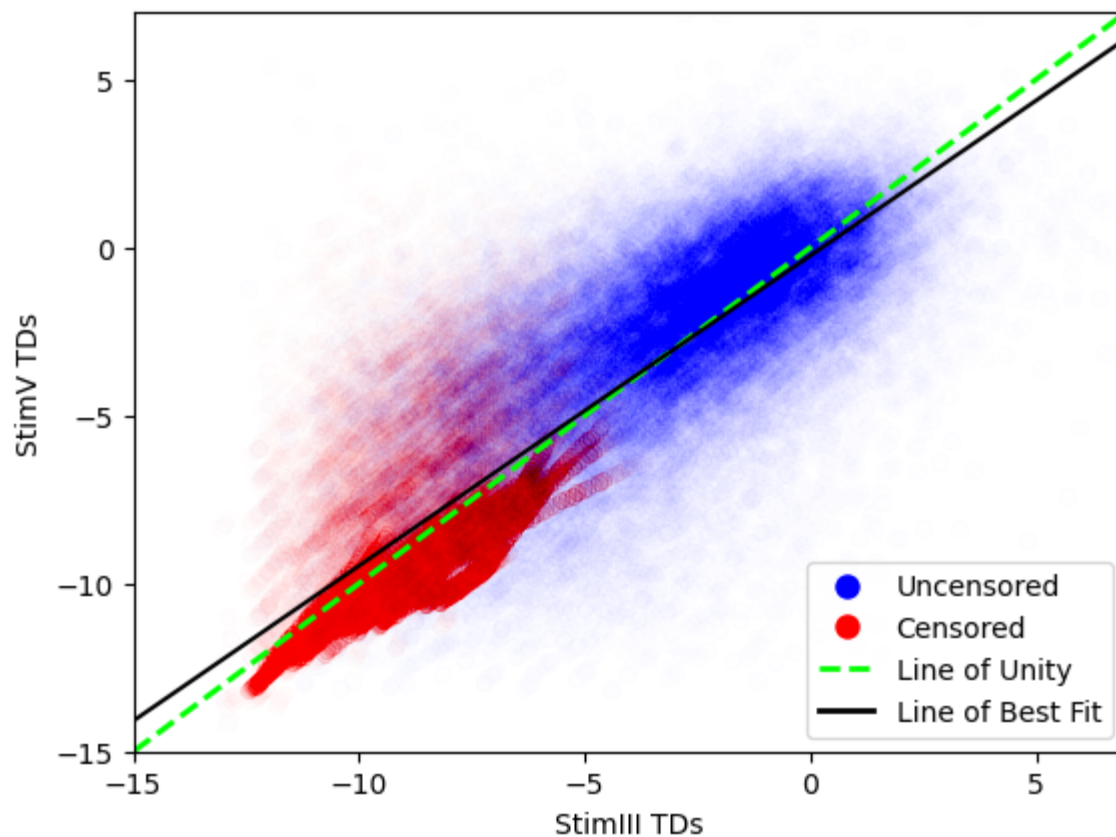
246 **Pointwise TD Differences Between StimV and StimIII**



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.

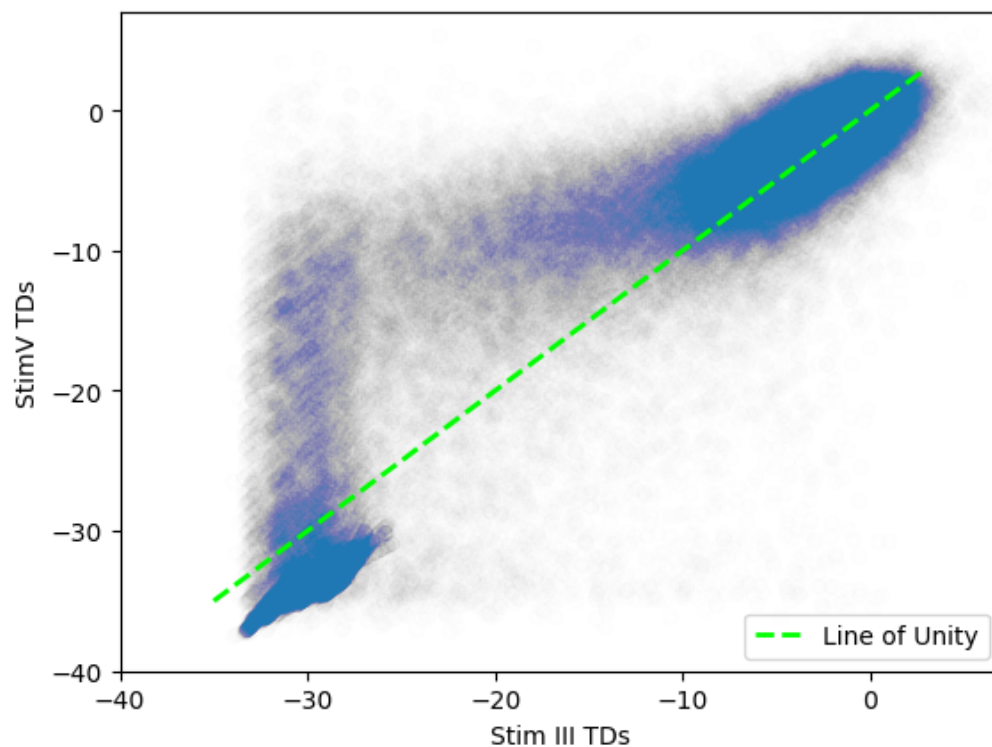
252

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



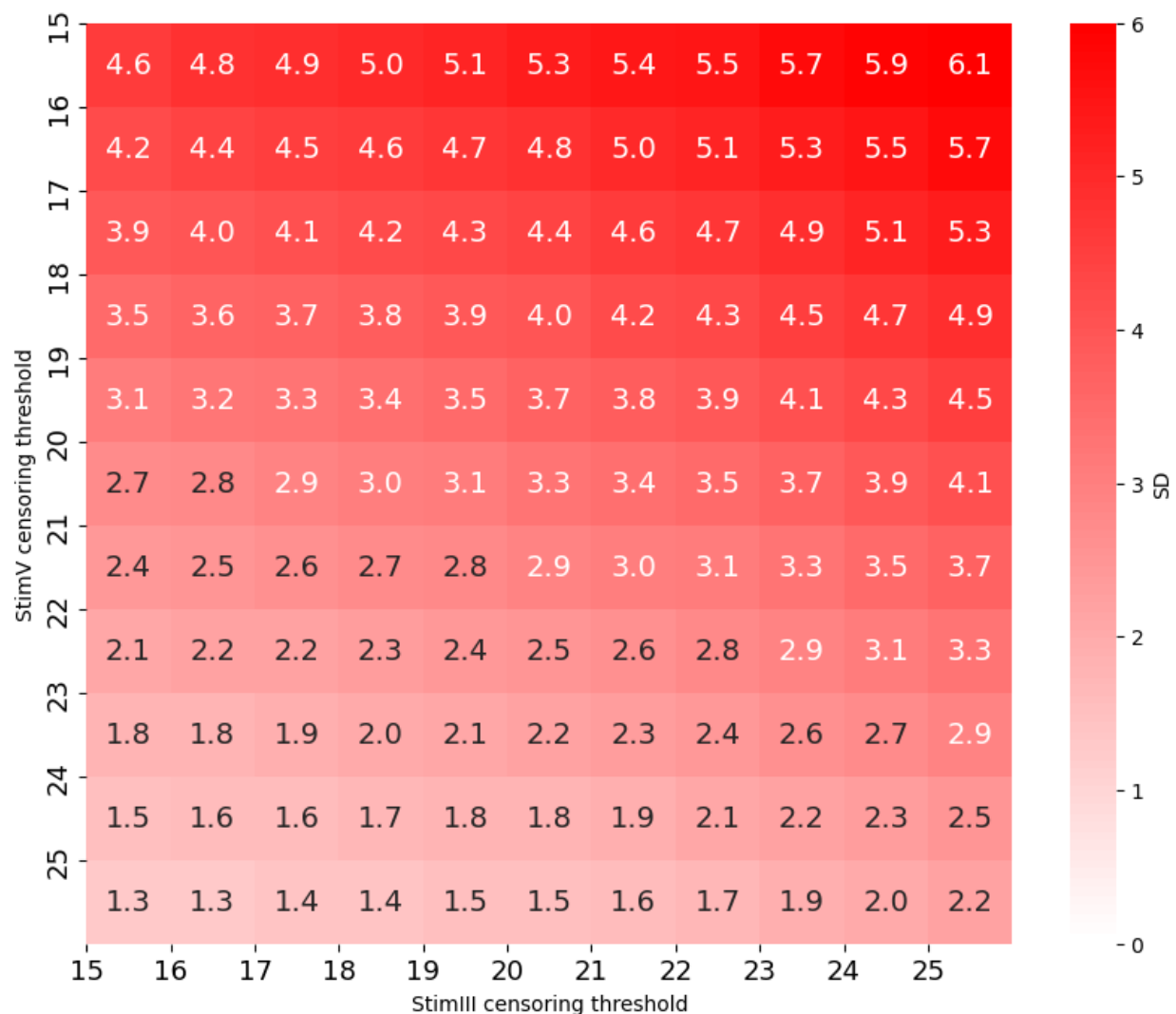
254
255 **Figure 3.** Pairs of total deviation values from stimulus size III and stimulus size V visual
256 fields. Data points are colored based on whether at least one of the pair is censored.
257 The dotted green line represents the line of unity ($y = x$). The black line of best fit ($y = -$
258 $0.24 + 0.92x$) 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

262 **Scatterplot of Uncensored TD Pairs Between StimIII and StimV**



263
264 **Figure 4.** Pairs of uncensored total deviation values from stimulus size III and stimulus
265 size V visual fields. The dotted green line represents the line of unity ($y = x$). Beyond a
266 certain threshold, points increasingly deviate from the line of unity, highlighting point
267 variability in uncensored measurements. Data points are made transparent for improved
268 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

- 278 1. Wall M, Brito CF, Woodward KR, Doyle CK, et al. Total deviation probability plots
279 for stimulus size v perimetry: a comparison with size III stimuli. *Arch Ophthalmol.*
280 2008;126(4):473-479.
- 281 2. Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. The effect of
282 stimulus size on the reliable stimulus range of perimetry. *Transl Vis Sci Technol.*
283 2015;4(2):10-10.
- 284 3. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of
285 the reliability of standard automated perimetry in regions of glaucomatous
286 damage. *Ophthalmology.* 2014;121(7):1359-1369.
- 287 4. Wall M, Zamba GK, Artes PH. The effective dynamic ranges for glaucomatous
288 visual field progression with standard automated perimetry and stimulus sizes III
289 and V. *Invest Ophth Vis Sci.* 2018;59(1):439-445.
- 290 5. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World*
291 *Health Organ.* 2004;82:887-888.
- 292 6. Hattenhauer MG, Leavitt JA, Hodge DO, Grill R, Gray DT. Incidence of
293 nonarteritic anterior ischemic optic neuropathy. *Am J Ophthalmol.* 1997;
294 123(1):103-107.
- 295 7. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean
296 defect with size III and size V standard automated perimetry. *Invest Ophth Vis*
297 *Sci.* 2013;54(2):1345-1351.
- 298 8. Kupersmith MJ, Fraser CL, Morgenstern R, et al. Ophthalmic and Systemic
299 Factors of Acute Nonarteritic Anterior Ischemic Optic Neuropathy in the
300 Quark207 Treatment Trial. *Ophthalmology.* 2024;131(7):790-802.

- 301 9. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*
302 2007;9(3):90-95.

