Differences in the Relation Between Perimetric Sensitivity and Variability Between Locations Across the Visual Field

Stuart K. Gardiner

Devers Eye Institute, Legacy Health, Portland, Oregon, United States

PURPOSE. Perimetric sensitivities become more variable with glaucomatous functional loss. This study examines the extent to which this relation varies between locations, and whether this can be predicted by eccentricity-related differences in spatial summation.

METHODS. Longitudinal series of visual fields from standard automated perimetry were obtained from participants with suspected or extant glaucoma. For each location in the 24-2 visual field, heterogeneous fixed-effects models were fit to the data, assuming that variability increased exponentially as sensitivity decreased. The predicted variability at each location was calculated when sensitivity was either 30 dB or 25 dB.

RESULTS. Variability significantly increased with damage at all 52 locations. When sensitivity was 30 dB, variability increased with eccentricity, with $P = 0.0003$. The average SD was 1.54 dB at the four most central locations, versus 1.74 dB at the most peripheral locations. When sensitivity was 25 dB, variability did not vary predictably with eccentricity, with $P = 0.340$. The average SD was 2.36 dB at the four central locations, versus 2.24 dB at the most peripheral locations.

CONCLUSIONS. The relation between sensitivity and variability differed by eccentricity. Among healthy locations, variability was lower centrally, where the stimulus size is larger than Ricco's area, than peripherally. Among damaged locations, variability did not systematically vary with eccentricity. This could be because Ricco's area expands in glaucoma, such that stimuli were now smaller than this area at all locations.

Keywords: perimetry, variability, Ricco's area, statistics

Automated perimetry remains an essential part of clinical management of glaucoma, not least because it measures outcomes that correlate directly with a patient's quality of life[1–4]; however, its results are notoriously variable between tests for the same eye,[5–9] delaying detection of damage and disease progression.[10] Most perimetric research can be framed as a series of gradual improvements to the signal-to-noise ratio.[11–13] To reduce this variability in future functional testing paradigms, it is important to first understand the sources of the variability in current testing.

There are several factors that contribute to test-retest variability in perimetry, including the amount of experience the patient has had with the test,[14] the experience level of the technician conducting the test,[15] the time of day,[15] and even the time of year.[15,16] However, the biggest source of variability is short-term fluctuation[8,17] related to the psychometric function. The physiologic response to a stimulus presentation involves an increase in the rate at which retinal ganglion cells produce neural spikes, but this spiking rate is inherently probabilistic.[18,19] Therefore, the observer will not always respond to stimuli of greater contrast than the physiologic detection threshold, and will sometimes respond to stimuli of lesser contrast.[19,20] Indeed, perimetric testing algorithms aim to converge to a contrast to which the observer will respond on a certain percentage of presentations, and not respond to the remaining presentations.

In regions of glaucomatous visual field loss, not only does the sensitivity decrease, but the psychometric function becomes shallower.[21] This causes an increase in intratest variability,[8,22,23] a major component of test-retest variability, because there is a broader range of contrasts at which the response probability will be neither 0% nor 100%. Heijl et al.[22] suggested that this increase in variability may vary by distance from fixation, whereby variability was lower in the central visual field than peripherally in eyes with near-normal sensitivity, but there was no apparent correlation between variability and eccentricity in eyes with moderate glaucomatous loss. That study was performed by testing 51 patients four times within a month. The effect may be confounded by the fact that normal sensitivities are higher centrally than peripherally, which would also be expected to result in lower variability.[21]

Several facets of retinal physiology change with distance from the fovea. Most relevantly for perimetry, the density of retinal ganglion cells decreases with eccentricity, while the size of their receptive fields increases.[24,25] A related psychophysical feature of the normal visual system is that Ricco's area increases with visual field eccentricity.[26] For stimuli smaller than Ricco's area, complete spatial summation occurs, such that stimulus area multiplied by stimulus intensity is constant at the detection threshold (hence, a smaller stimulus must have proportionately higher intensity for it to be equally detectable). For stimuli larger than Ricco's area, only partial spatial summation occurs. Ricco's area is also larger in glaucomatous eyes than in normal eyes.[27] The difference in detection contrast between normal and glaucomatous locations appears to be greater for stimuli smaller than Ricco's area.[27] It has therefore been suggested that

modifying stimulus size so that it remains within Ricco's area may be a way to increase the signal-to-noise ratio of perimetry.[27-29]

The size III stimulus (0.43° diameter) that is most commonly used in clinical perimetry is larger than Ricco's area at central locations in the visual field, and so partial spatial summation occurs. At the most peripheral locations in the visual field,[30] and more centrally in glaucomatous eyes,[27] the size III stimulus is approximately equal to or smaller than Ricco's area, causing complete spatial summation to occur in many eyes. The slope of partial summation (the change in threshold for a given change in stimulus size) becomes steeper with increasing eccentricity,[31] while the relation for complete spatial summation remains unchanged.[32] Therefore, existing perimetric data can be used to reveal important information about the consequences of using stimuli larger or smaller than Ricco's area. In particular, if the relative size of the stimulus and Ricco's area is indeed a key determinant of perimetric variability, then the sensitivity-variability relation should differ with eccentricity.

This study aims to confirm and extend the conclusions of Heijl et al.[22] using a longitudinal dataset, analyzed in a manner so as to remove the effects of possible disease progression. The advantages of this approach are 2-fold: first, it represents a more clinically realistic scenario in particular with regard to learning effects,[33] because the intertest interval is 6 months rather than just a week; and second, it allows a much larger dataset to be used so that more precise pointwise results can be generated. If variability is indeed lower when stimulus size is larger than Ricco's area, then this will provide an essential piece of information concerning the expected variability when deciding on the optimal balance between stimulus size and contrast in testing algorithms.

METHODS

Data were obtained from participants in the Portland Progression Project (P3), a longitudinal study of glaucomatous progression.[34] Inclusion criteria were a diagnosis of primary open-angle glaucoma, and/or likelihood of developing glaucomatous damage (e.g., ocular hypertension with other risk factors, such as a suspicious-looking optic disc or a family history of glaucoma), as determined by each participant's physician. Exclusion criteria were an inability to perform reliable visual field testing, best-corrected visual acuity worse than 20/40, substantial cataract or media opacities likely to increase light scatter, or other conditions or medications that may significantly affect the visual field. A visual field defect was not a requirement for study entry. Participants provided written informed consent once all of the risks and benefits of participation were explained to them. All protocols were approved and monitored by the Legacy Health Institutional Review Board, and adhered to the Health Insurance Portability and Accountability Act of 1996 and the tenets of the Declaration of Helsinki.

The longitudinal series used in this study date back as far as 1999 for some eyes, at which time participants attended annually; this was changed to twice annual testing in 2009. Visual field testing is performed on both eyes (when eligible), on every visit. Only tests performed using the Swedish Interactive Testing Algorithm (SITA Standard)[35] on the Humphrey Field Analyzer II, with the 24-2 test pattern, were included for analysis. Tests were excluded if >15% false positives or >33% fixation losses were recorded. If the technician considered the test to be unreliable, it was repeated, and only the last test on that day for each eye was included in the analysis. For this study, only series of length of five or more visual fields were included.

Analyses in this study used fixed-effects models. Essentially, this means that every eye is considered to be independent, but with variability estimates and parameterizations of the variance structure being pooled between eyes and individuals. In a fixed-effects model, the fact that both eyes were used for the same individual does not affect parameter estimates or residuals; even though the baseline sensitivity when healthy would be expected to be highly correlated between eyes, an unknown amount of time had passed between disease onset for that eye and study entry, such that sensitivities of the two eyes at the start of their series may be considered as being uncorrelated. Analyses were performed using R (Version 2.15.3; R Core Team, Vienna, Austria, 2013; provided in the public domain, http://www.R-project.org/) with the "nlme" package.[36]

Each of the 52 non-blindspot locations in the 24-2 visual field was analyzed separately. The dataset was first filtered to include only visual fields at which the sensitivity was ≥15 dB at the chosen location, because sensitivities below this value can be considered unreliable[37,38]; see the Discussion for more on this point. Eyes were then excluded if fewer than five visual fields remained in their series, because estimates of their rate of change would be insufficiently accurate.

A fixed-effects model was then fit to the pointwise data, using a generalized least squares formulation, by the following code:

$$Fit < -gls(Sensitivity \sim Eye + Eye{:}TestDate,$$
$$data = LongitudinalData,$$
$$weights = varExp(-0.08, form = \sim Sensitivity))$$

Here, the dataset *LongitudinalData* contains columns for *Sensitivity* (the pointwise sensitivity values at the chosen location), *Eye* (a unique identifier for each eye in the dataset), and *TestDate* (the date on which the visual field was taken, expressed as the number of days since the start of the series). The model assumes that *Sensitivity* changes linearly with *TestDate*, with a different starting point and rate for each eye.

The *weights* argument in the model fitting function means that the error variance is assumed to be related exponentially to *Sensitivity*, as in previous empirical studies.[21] For a given value of *Sensitivity*, the SD of the error terms is assumed to be given by $SD(Sensitivity) = A_0 * e^{B*(Sensitivity - C)}$, which simplifies to $SD(Sensitivity) = A * e^{B*Sensitivity}$. The starting value for the algorithm was set to a value of $B = -0.08$, as this was found to result in successful convergence of the algorithm for all test locations. For heteroscedastic linear models, weighting observations according to the reciprocal of the error variance function has been shown to be more robust to model misspecification than transforming the data to achieve homoscedasticity.[39]

The fitted value of parameter *A* was extracted from the fitted model using the code *sigma(Fit)*. The fitted value of parameter *B* was extracted together with its approximate 95% confidence interval (based on a normal approximation to the distribution of the maximum likelihood estimator), using the code *intervals(Fit)$varStruct*. These were then used to calculate the predicted SD of errors when the sensitivity is 30 dB or 25 dB, to provide examples of the predictions and illustrate the differences in behavior across the visual field (note that these examples are based on the results from the entire dataset, not just on locations with these particular sensitivities).

Two further models were also fit to the pointwise data. The first of these assumes homoscedasticity (i.e., that the error variance is constant across sensitivities):

$$Fit.Homoscedastic < -gls(Sensitivity \sim Eye + Eye{:}TestDate,$$
$$data = LongitudinalData)$$

The models *Fit* and *Fit.Homoscedastic* were compared by Akaike's Information Criterion (AIC) to assess whether incorporating heteroscedasticity improves the fit of the model to the observed data, which can be taken as an indication of whether the error variance is significantly related to sensitivity. The null hypothesis is that the error variance is independent of sensitivity, in which case the homoscedastic and heteroscedastic models will fit the data equally well.

The final model allows the error terms to be temporally correlated:

$$Fit.CAR < -gls(Sensitivity \sim Eye + Eye{:}TestDate,$$
$$data = LongitudinalData,$$
$$weights = varExp(fixed = B_{Loc},$$
$$form = \sim Sensitivity),$$
$$correlation = corCAR1$$
$$(0.1, form = \sim Test.Date|Eye))$$

Here, the correlation between the residuals decreases as the time between the measurements increases, according to a continuous autoregressive (CAR) correlation structure.[40] A recent study used an autoregressive moving average correlation structure for similar analyses[41]; however, that approach discretized time using "visit number" as its covariate, and a CAR model would be expected to be more realistic in a situation in which the intervisit time interval can vary substantially (for example, if data from one test visit was excluded due to unreliability of the sensitivity value, resulting in an intervisit interval of 1 year instead of 6 months). When fitting the *Fit.CAR* model, the value of the variance function coefficient was fixed at the value of *B* found in the primary model for this location. Principally, this is to ensure that direct comparisons between *Fit* and *Fit.CAR* are affected only by the presence of an autocorrelation term, and not by differing values of the variance function coefficient; a secondary advantage is that fixing this coefficient greatly aids with achieving algorithmic convergence. The comparison provided by *anova(Fit, Fit.CAR)* can then be used to assess whether significant autocorrelation is present. Note that *Fit* is nested within *Fit.CAR*, hence they can be formally compared using an ANOVA, whereas *Fit* and *Fit.Homoscedastic* are not nested (due to the different weightings of observations) and so have to be compared using AIC.

RESULTS

Data were available from 504 eyes of 255 participants who had series of at least five visits. Their clinical characteristics are outlined in the Table. The mean deviation on their most recent visit was −1.6 dB (range, −26.6 to +2.7 dB). A total of 210 of those eyes (42%) had Pattern Standard Deviation outside normal limits on their most recent visit, and 214 had Glaucoma Hemifield Test results of "outside normal limits" or "borderline," indications of significant localized functional loss. It should be noted, however, that the exact sample size varied slightly according to which location was being tested, because visual fields at which the sensitivity at that particular location was <15 dB were excluded.

Figure 1 shows, for each location, the fitted values of the variance function coefficient *B* (top left); the fitted value of *A*, equivalent to the SD of the error terms that would be predicted when *Sensitivity = 0 dB* (top right); and the fitted SDs of the error terms when *Sensitivity = 30 dB* (bottom left) and when *Sensitivity = 25 dB* (bottom right) as examples illustrating the

TABLE. Characteristics of the Cohort

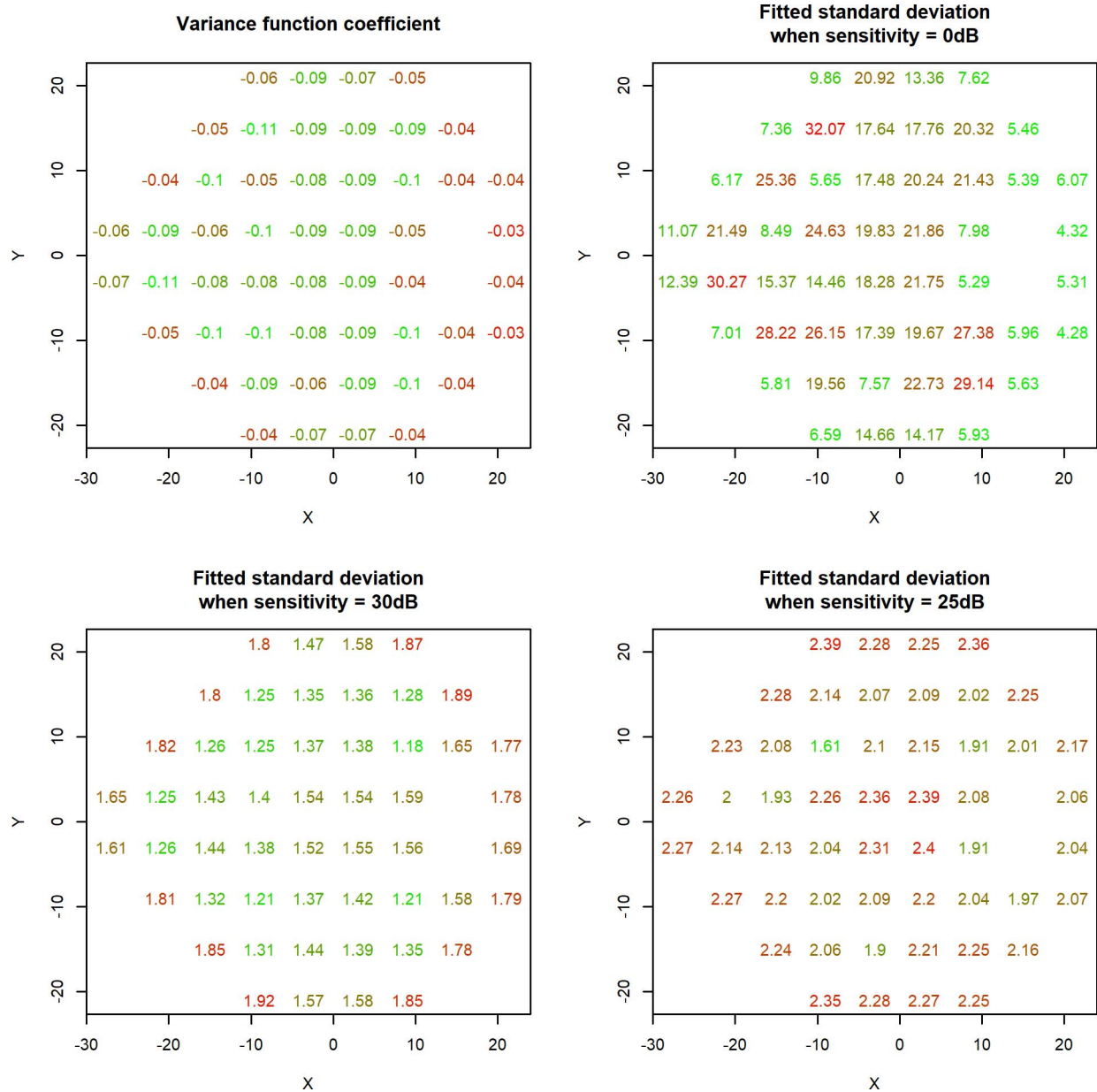|	Median	Interquartile Range
Series length, no. visits per eye	15	10–19
Series length, y	9.4	6.1–13.8
Age at last visit, y	70.9	64–79
Mean deviation at first visit, dB	+0.40	−0.8 to +1.3
Mean deviation at last visit, dB	−0.35	−2.7 to +0.8
Pattern SD at last visit, dB	1.80	1.5–3.1
Rate of change of mean deviation across series, dB/y	−0.10	−0.24 to +0.01

Only eyes with at least five visual fields are included.

difference in behavior at different sensitivities. Note that the predicted SD of the test-retest variability would be a factor of $\sqrt{2}$ greater than this, due to there being variability in both the test and the retest observations. Notably, the variance function coefficient was negative at all locations (*B* was significantly less than zero with $P < 0.05$), consistent with previous findings that variability decreases as sensitivity increases.[21] Averaging across all locations gave mean values of *A = 14.8* and *B = −0.070*.

At *Sensitivity = 30 dB*, the predicted variability generally increases with eccentricity, from a mean SD of 1.54 dB at the four central locations to a mean of 1.74 dB at the most peripheral ring of locations. The Pearson correlation between eccentricity and the fitted SD at *Sensitivity = 30 dB* was 0.483, with $P = 0.0003$. The width of the 95% confidence interval for the variance function coefficient *B* varied by location, from ±0.0037 dB to ±0.0067 dB. Treating the values of *A* as fixed, this results in 95% intervals for the fitted SD at *Sensitivity = 30 dB* with widths ranging from ±0.14 dB to ±0.31 dB away from the values shown. This indicates that the coefficients at locations around the edge of the 24-2 field were significantly higher than those at central locations with $P < 0.05$.

By contrast, at *Sensitivity = 25 dB*, there is no obvious trend of variability against eccentricity. The Pearson correlation between eccentricity and the fitted SD at *Sensitivity = 25 dB* was 0.135, with $P = 0.340$. Now, the mean predicted SD was 2.36 dB at the four central locations, versus a mean of 2.24 dB at the most peripheral locations. At this sensitivity, the 95% confidence intervals for the fitted SDs had widths ranging from ±0.15 dB to ±0.39 dB. It is therefore possible that central locations may be even more variable than many of the peripheral locations once this level of functional loss has been sustained. However, such conclusions must be viewed with caution, because the proportion of eyes that had sensitivity between 15 dB and 25 dB was below 1.6% at all four of the most central locations, compared with more than 20% at the four most superior locations.

The model incorporating heteroscedasticity (designated as *Fit* in the Methods section above) fit the data significantly better than the homoscedastic model (*Fit.Homoscedastic*) at all 52 locations, as evidenced by a lower AIC value. The model incorporating autocorrelation between observations that were close together in time significantly improved the fit of the model at 5 of the 52 locations. However, even at these locations, the fitted value of the correlation between the residuals at two observations 1 year apart was below 0.01, indicating that any autocorrelation was not meaningfully affecting the results.

To provide an example of the eccentricity effect, two locations were chosen at which the variance functions appeared to differ substantially. At location (−15°, +3°) (in right eye orientation), the fitted SD was 1.06 dB when sensitivity was 35 dB, increasing to 1.43 dB when sensitivity

Variance function coefficient

```
 20               -0.06 -0.09 -0.07 -0.05
 10         -0.05 -0.11 -0.09 -0.09 -0.09 -0.04
       -0.04 -0.1 -0.05 -0.08 -0.09 -0.1 -0.04 -0.04
  0 -0.06 -0.09 -0.06 -0.1 -0.09 -0.09 -0.05       -0.03
    -0.07 -0.11 -0.08 -0.08 -0.08 -0.09 -0.04      -0.04
-10        -0.05 -0.1 -0.1 -0.08 -0.09 -0.1 -0.04 -0.03
              -0.04 -0.09 -0.06 -0.09 -0.1 -0.04
-20              -0.04 -0.07 -0.07 -0.04
    -30  -20  -10   0   10   20
              X
```

**Fitted standard deviation
when sensitivity = 0dB**

```
 20            9.86 20.92 13.36 7.62
 10      7.36 32.07 17.64 17.76 20.32 5.46
     6.17 25.36 5.65 17.48 20.24 21.43 5.39 6.07
  0 11.07 21.49 8.49 24.63 19.83 21.86 7.98      4.32
   12.39 30.27 15.37 14.46 18.28 21.75 5.29      5.31
-10     7.01 28.22 26.15 17.39 19.67 27.38 5.96 4.28
          5.81 19.56 7.57 22.73 29.14 5.63
-20          6.59 14.66 14.17 5.93
    -30  -20  -10   0   10   20
              X
```

**Fitted standard deviation
when sensitivity = 30dB**

```
 20            1.8  1.47 1.58 1.87
 10       1.8 1.25 1.35 1.36 1.28 1.89
     1.82 1.26 1.25 1.37 1.38 1.18 1.65 1.77
  0 1.65 1.25 1.43 1.4 1.54 1.54 1.59       1.78
    1.61 1.26 1.44 1.38 1.52 1.55 1.56       1.69
-10      1.81 1.32 1.21 1.37 1.42 1.21 1.58 1.79
          1.85 1.31 1.44 1.39 1.35 1.78
-20          1.92 1.57 1.58 1.85
    -30  -20  -10   0   10   20
              X
```

**Fitted standard deviation
when sensitivity = 25dB**

```
 20            2.39 2.28 2.25 2.36
 10      2.28 2.14 2.07 2.09 2.02 2.25
     2.23 2.08 1.61 2.1 2.15 1.91 2.01 2.17
  0 2.26 2 1.93 2.26 2.36 2.39 2.08        2.06
   2.27 2.14 2.13 2.04 2.31 2.4 1.91        2.04
-10     2.27 2.2 2.02 2.09 2.2 2.04 1.97 2.07
          2.24 2.06 1.9 2.21 2.25 2.16
-20          2.35 2.28 2.27 2.25
    -30  -20  -10   0   10   20
              X
```

FIGURE 1. Changes in perimetric variability with sensitivity at each location in the 24-2 visual field. The fitted SD of the variability for a given location is given by $A * e^{B*Sensitivity}$, where A is the value shown in the *top right* panel, and B is the value shown in the *top left* panel. The *bottom* two panels show as examples the resultant predicted SD when sensitivity is 30 dB and 25 dB, respectively. Within each panel, the locations are colored along a continuous scale from *green* (the lowest value) to *red* (the highest value).

was 30 dB and 1.92 dB when sensitivity was 25 dB, as seen in Figure 1. At location (+15°, −15°), the fitted SD was higher in near-normal eyes, being 1.47 dB when sensitivity was 35 dB and 1.78 dB when sensitivity was 30 dB; but more similar in damaged eyes, being 2.16 dB when sensitivity was 25 dB.

To ensure that the results are not merely an artifact of the heteroscedastic model fitting process, the fitted values from the homoscedastic model *Fit.Homoscedastic* at each of the 52 locations were split into 1-dB-wide bins, and the SDs of the residuals from all data points within each bin were calculated. These results are plotted in Figure 2 for the two locations chosen above. It can be clearly seen that when the sensitivity is near-normal, the variance is higher for location (+15°, −15°) than for location (−9°, +3°), but this difference disappears once more substantial damage has occurred. Note that the

apparently high variability at location (−9°, +3°) for sensitivities 34 to 35 dB may not be a reliable result, because only 30 visual fields (0.4%) fell within that bin.

The range of these individual-location SDs, among all 52 locations, can be taken as a measure of spatial heterogeneity, and is shown for each 1-dB-wide bin in the boxplot in Figure 3. It is seen that although variability increases with damage, spatial heterogeneity is lowest when sensitivity is approximately 28–29 dB, indicating that the variability is most homogeneous between locations at this sensitivity level.

DISCUSSION

As expected, variability significantly increased as sensitivity decreased at all locations, as evidenced by the significantly
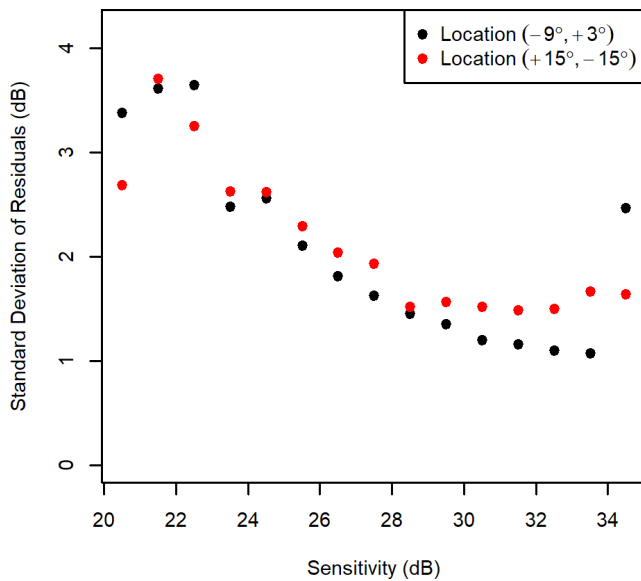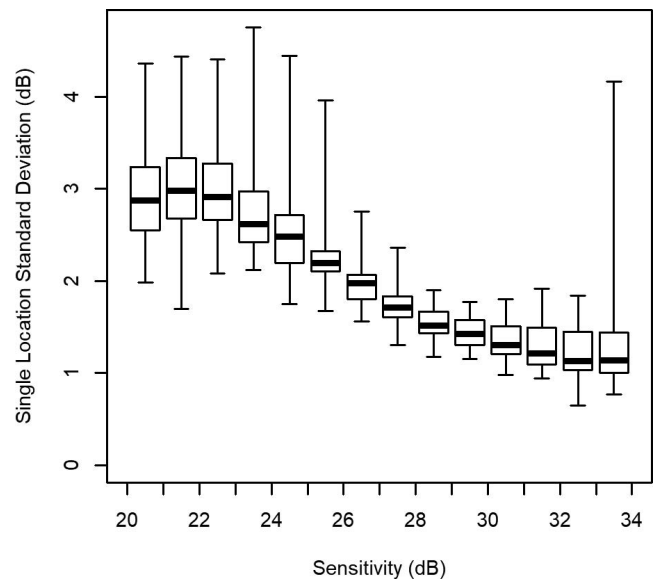
FIGURE 2. The observed variability at two locations in the visual field, when the sensitivity fell within each 1-dB-wide bin. The y-axis shows the SD of residuals using a homoscedastic fixed-effects model, equivalent to the residuals from ordinary least squares linear fits for the longitudinal series at that particular location for each individual eye, with those residuals binned according to the fitted sensitivity on that test date.

FIGURE 3. Spatial homogeneity at different sensitivity levels. At each location in the visual field, the SD of residuals (using a homoscedastic fixed-effects model) was calculated among eyes whose predicted sensitivity fell within a 1-dB-wide bin on that test date. For each sensitivity level, the median of these individual-location SDs is shown by the thick horizontal line, the box shows the interquartile range, and the whiskers show the full range of values between locations. The width of the box (or alternatively the width of the whiskers) can thus be taken as a measure of spatial heterogeneity in the variability.

negative fitted values of the variance function coefficient B. The SD for an averaged location was given by $SD(Sensitivity) = 14.8 * e^{-0.070*Sensitivity}$. Equivalently, $log(SD) = -0.070*Sensitivity + 2.70$. This is close to, but slightly smaller than, the estimate from Henson et al.[21] of $log(SD) = -0.081*Sensitivity + 3.27$. It is not surprising that these differ, because that article was estimating the slopes of frequency-of-seeing curves, and does not incorporate the effect of the perimetric testing algorithm. However, the similarity in the form of the function and in the magnitude of its parameters is reassuring. Other studies using perimetric data but with otherwise different methodologies have shown very similar effect sizes. In our results, averaging all locations, the predicted SD was 1.10 dB at a sensitivity of 35 dB, and 2.15 dB at a sensitivity of 25 dB. By way of comparison, Artes et al.[42] reported SDs of approximately 0.9 dB and 2.1 dB at those sensitivities, whereas Russell et al.[43] reported SDs of 1.2 dB and 2.2 dB.

This study goes farther than those, also assessing the effect of location on this sensitivity-variability relation. The variability was lower at central locations than peripheral locations when sensitivity was high. This implies that a healthy peripheral location whose sensitivity is (for example) 30 dB would be more variable than a central location in the same eye whose sensitivity has declined to 30 dB due to glaucoma. However, this eccentricity effect did not persist as the amount of damage increased. By the time sensitivity was reduced to 25 dB, the variability was similar at locations across the visual field; the correlation between eccentricity and variability was no longer significant, with $P = 0.340$. This confirms and extends previous findings by Heijl et al.[22]

The magnitude of the intratest variability (and its effect on the test-retest variability) is clinically relevant, because it indicates the magnitude of apparent change in sensitivity that is required before a clinician can conclude that functional status has truly changed. The finding that variability is lower at central than peripheral locations for the same sensitivity, when above approximately 28 dB, implies that even small changes at those central locations are more likely to represent true

change, and so can be considered clinically meaningful. The study also has implications for researchers interested in psychophysics and/or the development of perimetric techniques for future clinical use. The best stimulus size to use for clinical perimetry remains an open question; the reduction in variability when using larger stimuli may be accompanied by a reduction in the ability to detect early defects,[29] and indeed it may be suboptimal to use the same stimulus size at all locations and all levels of damage.[13] This study demonstrates that the optimal stimulus size will not only be affected univariately by damage level and location, but also by the interaction between them.

A possible explanation for this finding follows from spatial summation. Ricco's area increases both with eccentricity[30,31] and with glaucoma,[27] but contrast sensitivity for a stimulus of area equal to Ricco's area has been reported to remain constant.[27,32,44] It seems reasonable to propose then that variability might also remain constant between locations at which the sensitivity has deteriorated (due to aging and/or glaucoma) to the point at which Ricco's area exactly equals the stimulus size. In a healthy eye, Ricco's area equals that of a size III target at locations that are around the edge of, or just outside, the 24-2 visual field grid. At such locations, normal sensitivity for a subject of the average age in our cohort is between 27 and 30 dB. Notably, as seen in Figure 3, variability was most spatially homogeneous when sensitivity was 28 to 29 dB.

Once sensitivity has deteriorated below this point, Ricco's area is now larger than the stimulus size, and so complete spatial summation occurs: the stimulus luminance at the detection threshold is inversely proportional to stimulus area. This relation does not depend on eccentricity or disease status. Similarly, in the results shown here, variability (for a given sensitivity) does not appear to vary with eccentricity once sensitivity is below approximately 28 to 29 dB; however, when sensitivity is above this point, the stimulus is still larger than

Ricco's area, and so partial spatial summation occurs. The slope of partial summation (the change in threshold for a given change in stimulus size) becomes steeper with increasing eccentricity.[31] Similarly here, variability (for a given sensitivity) does appear to vary with eccentricity while sensitivity is above approximately 28 to 29 dB. There may be a discontinuity in the sensitivity-variability relation at the point at which Ricco's area starts to exceed the $0.43°$ diameter of the size III stimulus, with the magnitude of this change in the relation varying with eccentricity.

Apart from spatial summation, another factor that varies with eccentricity is the presence and magnitude of optical aberrations.[45] It is possible that these aberrations are the limiting factor for contrast sensitivity in normal eyes, causing variability to increase with eccentricity; whereas in a glaucomatous eye, undersampling resulting from loss of retinal ganglion cells may become the limiting factor and so the eccentricity effect is diminished. Further evidence would be needed before concluding that the magnitude of such an effect could entirely explain our results.

The correlation between variability and eccentricity was 0.502, but this did not explain all of the difference between locations. The locations surrounding the physiologic blind spot had higher variability when sensitivity was 35 dB than other locations of the same eccentricity. It is possible that this is caused by fixation instability and anatomic differences between eyes causing a proportion of stimuli at these locations to fall within the blind spot. It has previously been reported that fluctuation is increased near the blind spot,[46] even though this effect is not as pronounced as at the edge of a glaucomatous scotoma.[47]

We have previously reported that sensitivities below 15 to 19 dB are unreliable in clinical perimetry, having very low correlation ($R^2 < 0.1$) with sensitivities at the same location measured more accurately using frequency-of-seeing curves.[37] This is consistent with the fitted SD of the error terms increasing exponentially as sensitivity decreases. In this study, locations with sensitivity ≤ 15 dB were excluded from the analysis, so that this unreliability would not affect the main results; however, it should be noted that the heteroscedastic linear model used in this study fits the model by weighting observations according to the reciprocal of the error variance.[39] Therefore sensitivities ≤ 15 dB would have low weighting in the regression model, and so their omission is unlikely to have greatly affected our results or conclusions.

As with any research, the best statistical analysis to use depends on the details of the question being asked. We have previously shown that an accelerating exponential model (whereby the rate of change worsens over time) may fit longitudinal perimetric data better than a linear model (whereby the rate of change is constant over time) when conducting mixed effects model analyses,[40,48] as would be performed when analyzing possible risk factors for rapid progression. Intuitively, such an exponential model is more consistent with longitudinal series in which the eye is functionally normal for an extended period of time before visual field loss develops. Certainly, a reduced baseline sensitivity is predictive of a worse rate of change.[49] However, an accelerating exponential model whose variability increases exponentially as sensitivity decreases cannot be fit to the data, due to lack of identifiability between the exponential parameters. Just as importantly, this study used data from patients' entire series of visual fields, even if a treatment change occurred partway. If an eye were seen to be rapidly progressing, then the patient's clinician would alter the management strategy accordingly, hopefully slowing the rate of progression. As a result, accelerating and decelerating series were approximately equally common in our dataset. When performing quadratic fits for individual eyes, 47.4% had a negative coefficient for Time squared, indicating that the series was decelerating (the rate of change was slowing over time); 52.6% had a positive coefficient, indicating that the series was accelerating (the rate of change was becoming faster). Overall, these coefficients were not significantly different from zero ($P = 0.206$, t-test). This causes a linear model to fit the data better than an accelerating exponential model in this particular study. Any such treatment changes would not occur at a consistent time in the series, or at a consistent sensitivity for an individual location, and so they should not influence the distribution of residuals.

All analyses in this study used fixed-effects models. Essentially, every eye is considered to be independent, but with variability estimates (and their structure) pooled between eyes and individuals. There are two free parameters for each eye: the intercept and the rate of change. This contrasts with random effects models, whereby the parameters for each eye are considered as being drawn from a (usually Gaussian) random distribution, greatly reducing the number of free parameters to be estimated. Due to this lower number of free parameters, random effects models tend to perform better for identifying and testing risk factors for progression. By contrast, fixed-effects models fit the observed data better, and so are better for analyzing the structure and distribution of residuals, which is the aim of this study.

One remaining caveat with this study is that it did not include any truly "normal" eyes. Patients were referred to the study on the basis of having already developed, or being at significant risk of developing, glaucoma in at least one eye. There were many eyes in which no significant visual field loss had yet been detected, and both eyes were included even if only one would be considered suspicious for glaucoma. However, it is possible that some of those eyes had early damage, even if it had not yet exceeded normative limits, and such eyes would likely have higher variability than in a perfectly healthy patient. A further caveat is that all of these participants had substantial experience of undergoing perimetric testing, and variability would be expected to be higher in inexperienced patients.[14]

A further confounding factor is that the sensitivities were measured using the SITA Standard testing algorithm.[35] After all stimulus presentations have been made, the algorithm applies a proprietary spatial smoothing algorithm, aiming to reduce variability without obscuring true defects. It seems unlikely that this would cause a consistent eccentricity-related difference in the sensitivity-variability relation, but because the details are not public knowledge, it cannot be ruled out that this may have affected the results.

The increase in variability as sensitivity decreases is often seen as representing a key flaw with current clinical perimetry. It has been shown that this increase in variability is lessened by using alternative stimuli, such as in frequency doubling perimetry,[23,50] or size modulation.[13] In this study, it is apparent that the increase in variability with damage is also far less pronounced at peripheral than central locations when using static size III stimuli. It is therefore important to consider visual field location (and in particular eccentricity) as an additional relevant factor in such studies, because this represents a considerable confound when attributing such findings to stimulus type alone.

In conclusion, variability increases as sensitivity decreases in automated perimetry, but the magnitude of this increase differs between locations. In damaged visual fields, the variability observed at locations with the same sensitivity does not vary consistently with eccentricity; however, when sensitivity is near-normal, variability is lower in the central visual field than peripherally (for the same sensitivity). This

suggests that using stimuli larger than Ricco's area of complete spatial summation may lower test-retest variability, although the effect on the signal-to-noise ratio remains to be determined.

References

1. Gutierrez P, Wilson M, Johnson C, et al. Influence of glaucomatous visual field loss on health-related quality of life. *Arch Ophthalmol*. 1997;115:777–784.

2. Nelson P, Aspinall P, Papasouliotis O, Worton B, O'Brien C. Quality of life in glaucoma and its relationship with visual function. *J Glaucoma*. 2003;12:139–150.

3. Mckean-Cowdin R, Wang Y, Wu J, Azen SP, Varma R; Los Angeles Latino Eye Study Group. Impact of visual field loss on health-related quality of life in glaucoma: The Los Angeles Latino Eye Study. *Ophthalmology*. 2008;115:941–948.e1.

4. Alqudah A, Mansberger SL, Gardiner SK, Demirel S. Vision-related quality of life in glaucoma suspect or early glaucoma patients. *J Glaucoma*. 2016;25:629–633.

5. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol*. 1987;105:1544–1549.

6. Chauhan BC, House PH. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology*. 1991;98:79–83.

7. Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci*. 1999;40:648–656.

8. Spry PG, Johnson CA, Mckendrick AM, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci*. 2001;42:1404–1410.

9. Gardiner SK, Demirel S, Johnson CA. Modeling the sensitivity to variability relationship in perimetry. *Vision Res*. 2006;46:1732–1745.

10. Turpin A, Mckendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci*. 2011;52:3237–3245.

11. Artes PH, Chauhan BC. Signal/noise analysis to compare tests for measuring visual field loss and its progression. *Invest Ophthalmol Vis Sci*. 2009;50:4700–4708.

12. Gardiner SK, Fortune B, Demirel S. Signal-to-noise ratios for structural and functional tests in glaucoma. *Trans Vis Sci Tech*. 2013;2(6):3.

13. Rountree L, Mulholland PJ, Anderson RS, Garway-Heath DF, Morgan JE, Redmond T. Optimising the glaucoma signal/noise ratio by mapping changes in spatial summation with area-modulated perimetric stimuli. *Sci Rep*. 2018;8:2172.

14. Werner EB, Krupin T, Adelson A, Feitl ME. Effect of patient experience on the results of automated perimetry in glaucoma suspect patients. *Ophthalmology*. 1990;97:44–48.

15. Junoy Montolio FG, Wesselink C, Gordijn M, Jansonius NM. Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. *Invest Ophthalmol Vis Sci*. 2012;53:7010–7017.

16. Gardiner SK, Demirel S, Gordon MO, Kass MA; Ocular Hypertension Treatment Study Group. Seasonal changes in visual field sensitivity and intraocular pressure in the ocular hypertension treatment study. *Ophthalmology*. 2013;120:724–730.

17. Flammer J, Drance S, Zulauf M. Differential light threshold. Short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol*. 1984;102:704–706.

18. Pan F, Swanson WH. A cortical pooling model of spatial summation for perimetric stimuli. *J Vis*. 2006;6:1159–1171.

19. Gardiner SK, Swanson WH, Demirel S, Mckendrick AM, Turpin A, Johnson CA. A two-stage neural spiking model of visual contrast detection in perimetry. *Vision Res*. 2008;48:1859–1869.

20. Bebie H, Fankhauser F, Spahr J. Static perimetry: strategies. *Acta Ophthalmol (Copenh)*. 1976;54:325–338.

21. 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;41:417–421.

22. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.

23. Artes PH, Hutchison DM, Nicolela MT, Leblanc RP, Chauhan BC. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci*. 2005;46:2451–2457.

24. Croner LJ, Kaplan E. Receptive fields of p and m ganglion cells across the primate retina. *Vision Res*. 1995;35:7–24.

25. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol*. 1990;300:5–25.

26. Volbrecht VJ, Shrago EE, Schefrin BE, Werner JS. Spatial summation in human cone mechanisms from 0 degrees to 20 degrees in the superior retina. *J Opt Soc Am A Opt Image Sci Vis*. 2000;17:641–560.

27. Redmond T, Garway-Heath DF, Zlatkova MB, Anderson RS. Sensitivity loss in early glaucoma can be mapped to an enlargement of the area of complete spatial summation. *Invest Ophthalmol Vis Sci*. 2010;51:6540–6548.

28. Wall M, Doyle CK, Eden T, Zamba KD, Johnson CA. Size threshold perimetry performs as well as conventional automated perimetry with stimulus sizes III, V, and VI for glaucomatous loss. *Invest Ophthalmol Vis Sci*. 2013;54:3975–3983.

29. Phu J, Khuu SK, Zangerl B, Kalloniatis M. A comparison of Goldmann III, V and spatially equated test stimuli in visual field testing: the importance of complete and partial spatial summation. *Ophthalmic Physiol Opt*. 2017;37:160–176.

30. Khuu SK, Kalloniatis M. Standard automated perimetry: determining spatial summation and its effect on contrast sensitivity across the visual field. *Invest Ophthalmol Vis Sci*. 2015;56:3565–3576.

31. Choi AYJ, Nivison-Smith L, Khuu SK, Kalloniatis M. Determining spatial summation and its effect on contrast sensitivity across the central 20 degrees of visual field. *PLoS One*. 2016;11:e0158263.

32. Wilson ME. Invariant features of spatial summation with changing locus in the visual field. *J Physiol*. 1970;207:611–622.

33. Gardiner SK, Demirel S, Johnson CA. Is there evidence for continued learning over multiple years in perimetry? *Optom Vis Sci*. 2008;85:1043–1048.

34. Gardiner SK, Johnson CA, Demirel S. Factors predicting the rate of functional progression in early and suspected glaucoma. *Invest Ophthalmol Vis Sci*. 2012;53:3598–3604.

35. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997;75:368–375.

36. Pinheiro J, Bates D, Debroy S, Sarkar D; R Core Team. *Nlme: Linear and Nonlinear Mixed Effects Models*. R Foundation for Statistical Computing, 2017.

37. 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.

38. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010; 128:570–576.

39. Carroll RJ, Ruppert D. A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *J Am Stat Assoc*. 1982;77:878–882.

40. 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.

41. Ledolter J, Kardon RH. Does testing more frequently shorten the time to detect disease progression? *Trans Vis Sci Tech*. 2017;6(3):1.

42. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from full threshold, SITA standard, and SITA fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.

43. 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;53:5985–5990.

44. Kaplan E, Lee B, Shapley R. New views of primate retinal function. *Progress in Retinal Research*. 1990;9:273–336.

45. Navarro R, Moreno E, Dorronsoro C. Monochromatic aberrations and point-spread functions of the human eye across the visual field. *J Opt Soc Am A*. 1998;15:2522–2529.

46. Haefliger IO, Flammer J. Increase of the short-term fluctuation of the differential light threshold around a physiologic scotoma. *Am J Ophthalmol*. 1989;107:417–420.

47. Haefliger IO, Flammer J. Fluctuation of the differential light threshold at the border of absolute scotomas. Comparison between glaucomatous visual field defects and blind spots. *Ophthalmology*. 1991;98:1529–1532.

48. Pathak M, Demirel S, Gardiner SK. Nonlinear trend analysis of longitudinal pointwise visual field sensitivity in suspected and early glaucoma. *Trans Vis Sci Tech*. 2015;4(1):8.

49. Gardiner SK, Demirel S, Johnson CA. Perimetric indices as predictors of future glaucomatous functional change. *Optom Vis Sci*. 2011;88:56–62.

50. 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*. 2009;50: 974–979.