The Usefulness of Gaze Tracking as an Index of Visual Field Reliability in Glaucoma Patients

Yukako Ishiyama, Hiroshi Murata, and Ryo Asaoka

Department of Ophthalmology, University of Tokyo, Tokyo, Japan

PURPOSE. We evaluated the usefulness of gaze tracking (GT) results as an index of visual field reliability in glaucoma.

METHODS. The study population consisted of 631 eyes of 400 patients with open angle glaucoma in an institutional practice, with 10 visual fields (VFs). For the observational procedure, visual fixation was assessed using the gaze fixation chart at the bottom of the VF (Humphrey Field Analyzer, 30-2 SITA standard printout. Average frequency of eye movement between 1° and 2°, 3° and 5°, and greater than or equal to 6° were calculated. In addition, average tracking failure frequency (TFF) and average blinking frequency (BF) were calculated. The relationship between mean deviation (MD), fixation losses (FLs), false-positives (FPs), false-negatives (FNs), and pattern standard deviation (PSD) were evaluated using linear modeling. Main outcome measures included parameters related to over- or underestimation of MD values.

RESULTS. Patients’ mean MD progression rate was –0.23 dB/y. The best model to predict MD values included FL rate, FP rate, move1–2, move3–5, TFF, BF, and PSD as dependent variables with coefficients of 0.90, 9.2, −0.57, −0.52, −2.2, −1.1, and −0.56, respectively (P < 0.001).

CONCLUSIONS. High FL and FP rates tend to raise MD values. By contrast, high values of move1–2, move3–5, TFF, BF, and PSD tend to lower MD values. Thus, GT parameters can be used as new indices of VF reliability through the prediction of over- or underestimation of VF results.

Keywords: gaze tracking, glaucoma, visual field

Assessing the reliability of visual field (VF) results is very important at clinical settings, because the time it takes to detect progression is largely influenced by the variability of VFs, which impedes clinicians when making medical and surgical treatment decisions. In the Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Dublin, CA, USA), several methods have been used to estimate the reliability of VF tests. Fixation loss (FL) is recorded when a stimulus projected onto the area of the blind spot is perceived, and it indicates test reliability and vision fixation. Elevated FLs can mask the presence of early scotoma. False-positive (FP) rate is estimated by the number of positive answers that occur during a “listen time,” which starts shortly after the end of the response window and ends 180 ms after the onset of the next stimulus. False-negatives (FNs) mainly occur when a patient fails to respond to a stimulus that is more intense than that to which the patient had responded previously. A high rate of FP answers is thought to indicate “trigger-happy” patients and a high rate of FN responses is thought to represent inattention during an examination.

While some past studies have reported on the usefulness of these indices, more recent studies have pointed out their limitations; FLs also can result from the mislocalization of the blind spot and fixational instability can be found even in well trained observers. A high FN rate is reported to be associated with the amount of field loss as well as threshold reproducibility. Gaze tracking (GT) is a record of eye movement monitored during the actual sensitivity measurement. Its use in clinical practice has been somewhat limited, since results are merely represented as a printed line diagram at the bottom of the VF printout, and, as a result, can only be evaluated subjectively by clinicians. Nonetheless, it has been reported that GT is useful for evaluating the quality of fixation, particularly when VF defects surround the blind spot, and indeed, we have recently reported the usefulness of GT parameters for VF reliability as measured by test–retest reproducibility. In the study, GT results were evaluated objectively and quantitatively, and GT parameters were closely related to test–retest reproducibility; the FN rate also was significantly related to test–retest reproducibility, but the FP rate and FL rate were not. Nevertheless, the results do not deny the usefulness of the FP and FL indices, because they may be related to over- or underestimation of VF sensitivity. Indeed, Junoy Montolio et al. investigated the residuals from a mean deviation (MD) trend analysis and reported that high FP rates increase MD values. Thus, the objective of the current study was to investigate the usefulness of GT parameters, in addition to classic reliability indices, in the over- and underestimations of VF results.

METHODS

The study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at The
indices on over- or underestimation of VF results, VFs with current study was to evaluate the influence of reliability clinically insignificant senile cataract. As the purpose of the disease that could affect the VF, including cataract other than tion, and other anterior and posterior segment of the eye except for cataract extraction and intraocular lens implanta-

tion were equal to or better than 6/12. Eyes with the denoted move

Analyzed Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>FL</th>
<th>FP</th>
<th>FN</th>
<th>Move1–2</th>
<th>Move3–5</th>
<th>Move6</th>
<th>TFF</th>
<th>BF</th>
<th>PSD</th>
</tr>
</thead>
</table>
|**Table 1. Investigated Parameters****|**

Gaze Tracking Measurements

The GT system monitors patients’ gaze position at each stimulus presentation (Fig.).14 An upward bar in the chart indicates fixation disparity and the length of the bar represents the magnitude of disparity, from 1° to a maximum of 10°. A short downward bar represents tracking failure, while a long downward bar indicates eyelid closure. Gaze tracking parameters were calculated as follows: average TFF, average BF, the average frequency of eye movement per stimulus between 1° and 2° (denoted move1–2), 3° and 5° (denoted move3–5), and more than 6° (denoted move6).

We included in the study 631 eyes of 400 open angle glaucoma patients at the glaucoma clinic at The University of Tokyo Hospital. Each patient had at least 10 VFs with the HFA (30-2 Swedish Interactive Threshold Algorithm, SITA, standard program). A patient’s most recent 10 VFs were used in the analysis when more than 10 VFs had been recorded.

All patients enrolled in the study fulfilled the following criteria: (1) glaucoma was the only disease causing VF damage, (2) patients were followed for at least 6 months at The University of Tokyo Hospital and had undergone at least two VF measurements before this study, and (3) all patients had glaucomatous VF defects in at least one eye defined as three or more contiguous total deviation points at P < 0.05, or two or more contiguous points at P < 0.01, or a 10 dB difference across the nasal horizontal midline at two or more adjacent points, or MD worse than –5 dB.7 All visual acuities of the eyes examined were equal to or better than 6/12. Eyes with the following conditions were excluded: previous ocular surgery except for cataract extraction and intraocular lens implantation, and other anterior and posterior segment of the eye disease that could affect the VF; including cataract other than clinically insignificant senile cataract. As the purpose of the current study was to evaluate the influence of reliability indices on over- or underestimation of VF results, VFs with high FL or FP values were not excluded using the HFA criteria, which is FP < 20% or FL < 15%. However VFs with FL, FP, or

PSD, pattern standard deviation.

Sex, male:female 222:178

MD in the initial VF dB, mean ± SD (range) −7.3 ± 5.9 (−27.0 to 2.7)

PSD in the initial VF dB, mean ± SD (range) 8.8 ± 5.0 (1.2 to 19.0)

MD in the last VF dB, mean ± SD (range) −8.7 ± 6.5 (−28.1 to 2.7)

PSD in the last VF dB, mean ± SD (range) 9.9 ± 4.8 (1.2 to 18.6)

FN > 50% were excluded, to avoid the influence of extremely unreliable VF tests.

Gaze Tracking Measurements

The GT system monitors patients’ gaze position at each stimulus presentation (Fig.).14 An upward bar in the chart indicates fixation disparity and the length of the bar represents the magnitude of disparity, from 1° to a maximum of 10°. In the current study, GT data were exported as JPEG images from the Beeline (Tokyo, Japan) data filing system. Then the frequency of the upward and downward bars with each length in the GT records were simply calculated as follows: average frequency of eye movement per stimulus between 1° and 2° (denoted move1–2), 3° and 5° (denoted move3–5), and greater than or equal to 6° (denoted move6–6), average tracking failure frequency (denoted TFF), and average blinking frequency (denoted BF). The three levels of move1–2, move3–5, and move6–6 were chosen following the approach in our recent study.15,16

Table 2. Patient Demographics

<table>
<thead>
<tr>
<th>Demographics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean ± SD</td>
<td>56.5 ± 12.6</td>
</tr>
<tr>
<td>Sex, male:female</td>
<td>222:178</td>
</tr>
<tr>
<td>MD in the initial VF dB, mean ± SD (range)</td>
<td>−7.3 ± 5.9 (−27.0 to 2.7)</td>
</tr>
<tr>
<td>PSD in the initial VF dB, mean ± SD (range)</td>
<td>8.8 ± 5.0 (1.2 to 19.0)</td>
</tr>
<tr>
<td>MD in the last VF dB, mean ± SD (range)</td>
<td>−8.7 ± 6.5 (−28.1 to 2.7)</td>
</tr>
<tr>
<td>PSD in the last VF dB, mean ± SD (range)</td>
<td>9.9 ± 4.8 (1.2 to 18.6)</td>
</tr>
</tbody>
</table>

The relationship between MD values, and FL, FP, FN, move1–2, move3–5, move6–6, TFF, BF, and pattern standard deviation (PSD) was analyzed using the linear mixed model in which each eye was treated as a random effect, as shown in Table 1. The linear mixed model is equivalent to ordinary linear regression in that the model describes the relationship between the predictor variables and a single outcome variable. However, standard linear regression analysis makes the assumption that all observations are independent of each other. In the current study, measurements are nested within subjects and, thus, dependent of each other. Ignoring this grouping of the measurements will result in the underestimation of standard errors of regression coefficients. The linear mixed model adjusts for the hierarchical structure of the data, modeling in a way in which measurements are grouped within subjects. In the model selection, PSD was included as one of the possible parameters because the purpose of the current study was to decide the parameters related to over- or underestimation of MD values among all possible parameters, not to predict MD values from other measurements.

The best linear model was selected among all possible combinations of predictors: 29 patterns based on the second order bias corrected Akaiake Information Criterion (AICc) index. The AIC is a well-known statistical measure used in model selection, and the AICc is a corrected version of the AIC, which provides an accurate estimation even when the sample size is small.18 All analyses were performed using the statistical programming language ‘R’ (R version 2.15.1; The Foundation for Statistical Computing, Vienna, Austria).

Downloaded From: http://tvst.arvojournals.org/ on 10/14/2017
RESULTS
Characteristics of the study subjects are summarized in Table 2. Subjects comprised 222 males and 178 females. The mean age of the patients was $56.5 \pm 12.6$ (mean $\pm$ SD) years. Ten VFs were obtained in $5.8 \pm 1.3$ years. The mean MD value of the initial VFs was $-7.5 \pm 5.9$ dB and in the last VFs it was $-8.7 \pm 6.5$ dB. The mean PSD value of the initial VFs was $8.8 \pm 5.0$ dB and in the last VFs, it was $9.9 \pm 4.8$ dB. The MD progression rate was $-0.23 \pm 0.039$ dB/y on average.

As shown in Table 3, average rates (mean $\pm$ SD [range]) of FL, FP, and FN were $6.9 \pm 8.6$ [0–48]%, $2.9 \pm 3.8$ [0–43]%, and $3.6 \pm 5.2$ [0–46]%, respectively.

The average eccentricity of eye movement throughout the VF test was $1.9 \pm 1.5$ [0–12]° per stimulus (mean $\pm$ SD [range]). As shown in Table 4, average values of move1–2, move3–5, move6, TFF, BF were $0.65 \pm 0.17$ [0.00–0.97] per stimulus, $0.10 \pm 0.12$ [0.00–0.96] per stimulus, $0.070 \pm 0.15$ [0.00–1.00] per stimulus, $0.037 \pm 0.095$ [0.00–0.90], $0.045 \pm 0.086$ [0.00–0.99] per stimulus, respectively.

As shown in Table 5, FL, FP, move3–5, move6, TFF, BF, and PSD were selected as significant predictors of MD. The coefficients of FL, FP, move3–5, move6, TFF, BF, and PSD were $0.90$, $0.92$, $-0.57$, $-0.52$, $-2.2$, $-1.1$, and $-0.56$, respectively (linear mixed model, $P < 0.001$).

DISCUSSION
In the current study GT results from 30-2 HFA VFs were evaluated quantitatively and objectively. The influence of GT results, FLs, FPs, FNs, and PSD on MD values was then investigated. It was suggested that high FL and FP rates tend to raise MD values. Misfixation of more than 3° during the sensitivity measurement, as represented by move3–5 and move6, was significantly related to low MD values. Misfixation of less than 3° did not have a significant effect on MD values.

With regard to standard reliability indices, FL measures visual fixation during VF tests and is recorded when a patient traces the target stimulus rather than the fixation target. In addition, pseudo-FL can occur when the blind spot is not in the expected position due to change in head tilt and eye rotation.7 Furthermore, in our previous study, it was suggested that FL was not related to test–retest reproducibility of 24-2 and 10-2 VFs.16 Thus, the usefulness of FL as a reliability indicator is somewhat limited, especially when compared to GT parameters, which are derived from real-time eye movements during the sensitivity measurement.

In the current results, FN rate was not selected as a predictor for MD values. It has been reported that FNs increase with the progression of glaucoma13 and FN rate is no longer used as an official criterion of reliable VFs.14 However, this does not deny the usefulness of FNs to assess test reliability. Indeed, we have shown that the FN rate is useful for estimating test–retest reproducibility16 and, hence, it is not recommended to ignore FN results when interpreting VFs in the clinical setting. Furthermore, Bengtsson19 investigated the relationship between VF reproducibility and FLs, FPs, and FNs, and found that only FNs were associated with reproducibility.

Our previous study suggested that FP rate is not a significant predictor of test–retest reproducibility.16 In the SITA algorithm, FP rates are calculated differently from those in the Full-Threshold test, in which classic catch trials are used. In the SITA algorithm, any response before the minimum response time (approximately 180 ms), which also is adjusted according to the patient’s individual mean response time, is considered an FP error.5 This may suggest that all actual “FP” responses after the minimum response time are ignored in the FP calculation. Still, the current results suggest that a high FP rate raises MD values (the “trigger-happy” patients), which is in agreement with a previous report.17

In contrast with the standard reliability indices, GT parameters directly measure eye position during threshold measurements. Among the GT parameters analyzed in the current study, move3–5, move6, TFF, and BF were significantly associated with low MD values, probably because the patient was not well-fixated and could not see the target stimulus during blinking, as suggested in a previous report.20 On the other hand, move1–2 was not significantly related with MD values. This is unsurprising given the 6° spacing of VF test points in the 30-2 VF. Moreover, a previous study has reported that eye movement of less than 3° are commonly observed in VF tests, even in well-trained healthy observers.4,21

The relationship between PSD and MD can be explained using a quadratic linear regression model,16,22 where PSD decreases in the moderate to advanced stages of the disease.7 PSD was included as a predictor of MD in the best linear model in the current study. This may be because the progression of MD, in our study patients, was slow in general (~0.25 dB/y on average) and so the relationship between PSD and MD was linear in this narrow range of progression.

One caveat of the current investigation is the limited information derived from the GT record. Gaze tracking parameters were merely extracted as the average frequency throughout the VF measurement. A more detailed investigation could be carried out if the “real-time” GT tracking results were available to researchers, thus making it possible to analyze fixation status at each sensitivity measurement. A further caveat is that GT results can be related to dry eye,23 hence, further investigation is needed to shed light on this issue.

<table>
<thead>
<tr>
<th>Table 3. Results of Classic Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classic Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>FL, %, mean ± SD (range, per stimulus)</td>
<td>6.9 ± 8.6 (0–48)</td>
</tr>
<tr>
<td>FP, %, mean ± SD (range, per stimulus)</td>
<td>2.9 ± 3.8 (0–43)</td>
</tr>
<tr>
<td>FN, %, mean ± SD (range, per stimulus)</td>
<td>3.6 ± 5.2 (0–46)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Frequency of Gaze Tracking Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GT Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Move1–2, mean ± SD (range, per stimulus)</td>
<td>0.65 ± 0.17 (0–0.97)</td>
</tr>
<tr>
<td>Move3–5, mean ± SD (range, per stimulus)</td>
<td>0.10 ± 0.12 (0–0.96)</td>
</tr>
<tr>
<td>Move6, mean ± SD (range, per stimulus)</td>
<td>0.070 ± 0.15 (0–1)</td>
</tr>
<tr>
<td>TFF, mean ± SD (range, per stimulus)</td>
<td>0.057 ± 0.95 (0–0.9)</td>
</tr>
<tr>
<td>BF, mean ± SD (range, per stimulus)</td>
<td>0.045 ± 0.086 (0–0.99)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Selected Parameters</th>
<th>FP</th>
<th>FL</th>
<th>PSD</th>
<th>Move3–5</th>
<th>Move6</th>
<th>TFF</th>
<th>BF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>9.2</td>
<td>0.9</td>
<td>-0.56</td>
<td>-0.57</td>
<td>-0.52</td>
<td>-2.2</td>
<td>-1.1</td>
</tr>
</tbody>
</table>
the current study, GT data were exported as JPEG images from the Beeline data filing system and various GT parameters were simply calculated by reading the JPEG image. Thus, GT parameters can be obtained on a personal computer; simple software could be built to give clinicians access to this GT information to estimate the reliability of patients’ VFs at clinical settings.

In conclusion, we analyzed the influence of eye movements derived from the GT record on HFA VF tests. Gaze tracking parameters are significantly related to the underestimation of sensitivity in 30-2 VF tests.

Acknowledgments
Supported in part by Grant 26462679 (RA) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and Supported in part by Grant 26462679 (RA) from the Ministry of

References