How Many Visual Fields Are Required to Precisely Predict Future Test Results in Glaucoma Patients When Using Different Trend Analyses?

Yukako Tacketani, Hiroshi Murata, Yuri Fujino, Chiihiro Mayama, and Ryo Asaoka

Department of Ophthalmology, The University of Tokyo Graduate School of Medicine, Tokyo, Japan

Correspondence: Ryo Asaoka, Department of Ophthalmology, The University of Tokyo Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 1138655 Japan; rasaoka-tky@umin.ac.jp.

Submitted: December 25, 2014
Accepted: April 17, 2015

Citation: Tacketani Y, Murata H, Fujino Y, Mayama C, Asaoka R. How many visual fields are required to precisely predict future test results in glaucoma patients when using different trend analyses? Invest Ophthalmol Vis Sci. 2015;56:4076–4082. DOI:10.1167/iovs.14-16341

PURPOSE. To evaluate the minimum number of visual field (VF) tests required to precisely predict future VF results using ordinary least squares linear regression (OLSLR), quadratic regression, exponential regression, logistic regression, and M-estimator robust regression model.

METHODS. Series of 15 VFs (Humphrey Field Analyzer 24-2 SITA standard) were analyzed from 247 eyes of 155 open-angle glaucoma patients. Future point-wise (PW) VF results and mean VF sensitivities were predicted with varying numbers of VFs in each regression method.

RESULTS. In PW-OLSLR, as expected, the minimum absolute prediction error was obtained using the maximum number of VFs in the regression (14 VFs); mean absolute prediction error was equal to $2.4 \pm 0.9$ dB. Ten VFs were required to reach the 95% confidence interval (CI) of the minimum absolute prediction error. Prediction errors associated with the exponential and quadratic regression models were significantly larger than those from PW-OLSLR, whereas errors from logistic regression were not significantly smaller than those from PW-OLSLR; however, the absolute prediction error from the M-estimator robust regression model was significantly smaller than those associated with PW-OLSLR ($P < 0.01$, paired Wilcoxon test). Like PW-OLSLR, 10 VFs were needed to obtain the minimum absolute prediction error of mean VF sensitivity, but there were no significant differences in errors using the different regression methods.

CONCLUSIONS. Approximately 10 VFs, are needed to achieve an accurate prediction of PW VF sensitivity and mean sensitivity. Prediction error of PW VF sensitivity can be significantly minimized using the M-estimator robust regression model compared with conventional OLSLR.

Keywords: visual field, glaucoma, prediction, regression model

Glaucoma

Glaucoma is the second leading cause of blindness in the world.\(^1\) As glaucoma can lead to irreversible blindness, it is essential to accurately measure visual field (VF) progression so that physicians can treat patients with glaucoma accordingly, especially because medical and surgical IOP reduction interventions can be accompanied with a number of ocular and general complications.\(^2\)–\(^6\) It is therefore imperative to accurately predict future VF progression when making glaucoma treatment decisions. Visual field trend analyses can be used to measure the progression of global indices, such as mean deviation (MD) and the visual field index (VFI), or point-wise (PW) sensitivity using ordinary least square linear regression (OLSLR). This type of analysis is widely used in glaucoma clinics using clinical support tools such as PROGRESSOR (Medisoft, Ltd., London, UK).\(^7\)

Detection of VF progression is hampered by VF variability. Visual field sensitivity fluctuates in the short-term\(^8\) and long-term.\(^9\) The reliability of VF measurements is inherently affected by a patient’s concentration, but previous reports have suggested that measurement noise is considerable even with good reliability indices.\(^10\),\(^11\) The ability of VF trend analyses, in particular PW linear regression (PLR), to detect progression will be significantly impeded by VF variability.\(^12\) Consequently, the minimum number of VFs required to obtain reliable PLR results has been widely discussed in previous studies.\(^13\),\(^14\)

Furthermore, OLSLR is highly sensitive to outliers. Attempts have been made to overcome this problem; for example, Gardiner and Crabb\(^14\) reported that omitting VFs helps to identify progression, whereas Hirasawa et al.\(^15\) suggested that the accuracy of PW-OLSLR can be improved by applying it to the average sensitivities of small sectors. Alternatively, robust linear regression models can be applied to mitigate the effects of outliers and improve the detection of progression. In the robust model, the weight of each data point in the regression is dependent on the size of its residual; thus, the model is more “robust” to the influence of outliers.\(^16\)

The first objective of the current study was to investigate the minimum number of VFs required to accurately predict VF progression using OLSLR. The second purpose of this study was to estimate the usefulness of the M-estimator robust regression model (M-robust) for VF progression prediction as well as the exponential, quadratic, and logistic models, compared with OLSLR.

METHODS

This study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at The University of Tokyo. Written consent was given by patients for their information to be stored in the hospital database and used for research. This study was performed according to the tenets of the Declaration of Helsinki.
Subjects
This retrospective study included 247 eyes of 155 patients with primary open-angle glaucoma (mean ± SD age: 55.7 ± 11.9 years) from the glaucoma clinic at the University of Tokyo Hospital since 2002. All patients had a minimum of 16 reliable VF tests. An unreliable VF was defined as having more than 20% fixation losses or more than 15% false-positive errors. Only a patient’s initial 16 VFs were analyzed when a patient had more than 16 VF test results. The first VF was excluded from the analysis to reduce the learning effect, and the remaining 15 VFs were obtained over 7.5 years (mean ± SD years) of follow-up. Inclusion criteria for this study were patients with a visual acuity better than or equal to 6/12, refraction less than 5 diopters ametropia, no previous ocular surgery (except for cataract extraction), and no other posterior segment eye diseases. Patients with other ocular diseases that could affect VF sensitivity, such as diabetes mellitus retinopathy, corneal opacity, and AMD were excluded. Patients with cataract other than clinically insignificant senile cataract were excluded. All VFs were recorded using the Humphrey Field Analyzer (HFA; Carl-Zeiss Meditec, CA, USA), with the 24-2 or 30-2 test pattern and the STIA standard strategy with a Goldmann size III target. When the VF was measured using the 30-2 test pattern, only the 52 test points overlapping with the 24-2 test pattern were used for the analysis.

Statistical Analysis

Using OLSLR, the PW VF sensitivities of the first (in approximately 6 months), second (in approximately 1 year) and third future (in approximately 1 and a half years) VFs: sixth VF (VF6), seventh VF (VF7), and eighth VF (VF8) were predicted from the first five VFs (VF1–5). Absolute prediction accuracy was calculated as the absolute value of the difference between the predicted and the observed PW sensitivities. In addition, the absolute prediction errors using the exponential, quadratic, logistic, and M-robust methods were also calculated. The process was iterated to predict the PW sensitivities of the seventh, eighth and ninth VFs (first, second, third future VF: VF7, VF8, and VF9) using VF1–6, the PW sensitivities of VF6, VF7, and VF8 using VF1–5, and so on, up to prediction of the PW sensitivities of VF13, VF14, and VF15 using VF1–12. The minimum prediction error was calculated for each regression method and the minimum number of VFs to reach the 95% confidence interval (CI) was identified. These analyses were also performed using the mean sensitivity of the entire VF. As a subanalysis, predictive accuracy was compared in eyes with early-stage glaucoma (initial mTD [mean of total deviation of all 52 test points] larger than 6 dB) and eyes with moderate to advanced-stage glaucoma (initial mTD [mean of total deviation of all 52 test points] larger than 6 dB) and eyes with moderate to advanced-stage glaucoma (initial mTD [mean of total deviation of all 52 test points] larger than 6 dB) and eyes with moderate to advanced-stage glaucoma (initial mTD [mean of total deviation of all 52 test points] larger than 6 dB) and eyes with moderate to advanced-stage glaucoma (initial mTD [mean of total deviation of all 52 test points] larger than 6 dB). In addition, accuracy was evaluated in eyes with an mTD progression rate smaller than −0.25 dB per year and the remaining eyes, which is commonly considered to be the deterioration by pressure-independent damaging factors.

Statistical Models

The following five regression models were used to make predictions, where y represents PW or mean VF sensitivity and x represents the time from the initial VF:

#1 OLSLR: \( y = ax + b \)
#2 exponential regression: \( y = e^{ax+b} \)
#3 quadratic regression: \( y = ax^2 + bx + c \)
#4 M-estimator robust linear regression:

\[ y_i = x + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k + \epsilon = x\beta + \epsilon \]

for the \( i \)th of \( n \) observations, the general M-estimator minimizes the objective function:

\[ \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - x^T\beta) \]

where the function \( \rho \) gives the contribution of each residual to the objective function.\(^{24} \)

#5 Logistic regression:

\[ Y = \frac{\zeta}{1 + e^{\beta x}}, \]

where \( \alpha, \beta, \) and \( \zeta \) are model parameters to be estimated.

All models were fit using the statistical programming language R (R version 2.15.1; The Foundation for Statistical Computing, Vienna, Austria). The robust regression model was fit using the R package MASS. To compare the prediction error of each model, a Wilcoxon rank test was carried out. Statistical significance was set as less than 0.05, where Holm’s method\(^{25,26} \) was used to correct \( P \) values for the problem of multiple testing.

RESULTS

Patient Data

Patients’ demographic information is shown in Table 1. The mean ± SD MD in VF1 was −6.6 ± 5.5 (range, 1.1 to −25.8) dB and −8.7 ± 6.3 (range, 1.5 to −27.0) dB in VF15. The mean ± SD interval between the first, second, and third future VFs was 5.9 ± 0.2 months, 11.9 ± 0.2 months, and 17.8 ± 0.2 months, respectively.

Prediction Error

When predicting the first, second, and third future VFs using PW-OLSLR, the minimum absolute prediction error was obtained using VF1–14 (2.4 ± 0.9 [range, 0.6–5.9] dB: see Fig. 1A), VF1–13 (2.6 ± 1.0 [range, 0.6–6.5] dB: Fig. 1B), and VF1–12 (2.8 ± 1.2 [range, 0.6–9.1] dB: Fig. 1C), respectively. The absolute prediction errors associated with the exponential and quadratic regressions were significantly larger than OLSLR in all comparisons (paired Wilcoxon test, \( P < 0.05 \) after correction of \( P \) values for multiple testing using Holm’s method\(^{25,26} \)). No significant improvement was observed in the prediction error by applying logistic regression compared with OLSLR (paired Wilcoxon test, \( P > 0.05 \) after correction of \( P \) values for multiple testing using Holm’s method\(^{25,26} \)). The absolute prediction errors (mean ± SD) associated with the M-estimator robust regression were 2.3 ± 0.9 (range, 0.6–5.8) dB using VF1–14, 2.5 ± 0.9 (range, 0.6–6.2) dB using VF1–13, and

<table>
<thead>
<tr>
<th>Table 1. Subject Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>Follow-up duration, ( y ), mean ± SD</td>
</tr>
<tr>
<td>Age, ( y ), mean ± SD</td>
</tr>
<tr>
<td>Sex, male:female</td>
</tr>
<tr>
<td>MD of 1st VFs, dB, mean ± SD</td>
</tr>
<tr>
<td>MD of 15th VFs, dB, mean ± SD</td>
</tr>
<tr>
<td>Interval against 1st future VFs, mo, mean ± SD</td>
</tr>
<tr>
<td>Interval against 2nd future VFs, mo, mean ± SD</td>
</tr>
<tr>
<td>Interval against 3rd future VFs, mo, mean ± SD</td>
</tr>
</tbody>
</table>

\( \text{dB, decibel; POAG, primary open-angle glaucoma; MD, mean deviation; SD, standard deviation; VE visual field.} \)
2.7 ± 1.1 (range, 0.6–9.0) dB using VF1–12. These values were significantly smaller than those in other models, in all comparisons for first future VF prediction (paired Wilcoxon test, \( P < 0.001 \) except for VF1–5; \( P = 0.010 \), after correction of \( P \) values for multiple testing using Holm’s method\(^{25,26} \)), except for VF1–5 for second future VF prediction (\( P = 0.001 \), except for VF1–6; \( P = 0.022 \), VF1–12; \( P = 0.0047 \)), and except for VF1–5 for third future VF prediction (\( P = 0.001 \) except for VF1–11; \( P = 0.018 \)). In PW-OLSLR, the absolute prediction error for predicting the first, second, and third future VFs reached 95% CI of the minimum prediction error using VF1–11, VF1–10, and VF1–10, respectively.

A total of 126 eyes were suffering from early-stage glaucoma (mTD \( > -6 \) dB, Table 2). Figure 2 shows the comparison between the absolute prediction errors associated with OLSLR and M-estimator robust regression. Significantly smaller prediction errors were observed with M-estimator robust regression when predicting the first future VF using VF1–10, VF1–11, and VF1–11 (\( P = 0.0071, 0.025, \) and 0.010, paired Wilcoxon test, Fig. 2A), and when predicting the second future VF using VF1–9, VF1–10, and VF1–11 (\( P = 0.015, 0.044, \) and 0.044, paired Wilcoxon test, Fig. 2B). No significant difference was observed in predicting the third future VF using VF1–5 to VF1–12 (\( P > 0.05 \), paired Wilcoxon test, Fig. 2C). In eyes with more advanced glaucoma (mTD \( \leq -6 \) dB; 121 eyes, Table 2), significantly smaller prediction errors were observed when predicting the first future VF using VF1–5 to VF1–14 (VF1–5: 0.025, VF1–6: \( P = 0.0010 \), VF1–7: \( P = 0.0014 \), and VF1–8 to VF1–14: \( P < 0.001 \), paired Wilcoxon test, Fig. 3A), predicting the second future VF using VF1–6 to VF1–15 (VF1–6: 0.001, VF1–7 to VF1–15: \( P < 0.001 \), paired Wilcoxon test, Fig. 3B), but not for predicting the third future VF using VF1–5 to VF1–12 (\( P > 0.05 \), paired Wilcoxon test, Fig. 3C).

A total of 117 eyes had an mTD progression rate faster than \( -0.25 \) dB per year (Table 2). Figure 4 shows the comparison between absolute prediction errors associated with OLSLR and M-estimator robust regression. Significantly smaller prediction errors were observed with M-estimator robust regression when predicting the first future VF using VF1–14 and VF1–10 to VF1–12 (\( P = 0.028, 0.0080, 0.029, < 0.001, < 0.001, 0.0020, \) paired Wilcoxon test, Fig. 4A), the second future VF using VF1–7 to VF1–11 and VF1–13 (\( P = 0.0015 \) and VF1–13: \( P = 0.0068 \), paired Wilcoxon test, paired Wilcoxon test, Fig. 4B), and predicting the third future VF using from VF1–7 to VF1–12 (\( P = 0.029, 0.044, 0.031, 0.0085, 0.049, 0.013, \) paired Wilcoxon test, Fig. 4C). In the remaining eyes with mTD progression rate equal to or slower than \( -0.25 \) dB per year (130 eyes, Table 2), significantly smaller prediction errors were observed when predicting the first future VF using VF1–11 to VF1–14 (VF1–7: \( P = 0.0015 \), VF1–10: 0.013, VF1–10: 0.014, VF1–7 to VF1–11: \( P < 0.001 \), paired Wilcoxon test, Fig. 5A), predicting the second future VF using VF1–9 to VF1–13 (\( P < 0.001, 0.0024, 0.0025, 0.0015, < 0.001 \), respectively, paired Wilcoxon test, Fig. 5B). No significant difference was observed in predicting the third future VF using VF1–5 to VF1–12 (\( P > 0.05 \), paired Wilcoxon test, Fig. 5C).

As shown in Figures 6A through 6C, the minimum absolute prediction error associated with OLSLR of mean VF sensitivity was obtained using VF1–14 (2.4 ± 0.9 dB [0.004–5.3]) for predicting the first future VF, using VF1–13 (2.6 ± 1.0 dB [0.002–5.2]) for predicting the second future VF, and using VF1–12 (2.8 ± 1.2 dB [0.008–9.1]) for predicting the third future VF. There were no significant differences in the absolute error.
errors associated with OLSLR, exponential regression, and M-estimator robust regression models at any time point. The absolute prediction errors with the quadratic regression model were significantly worse than the OLSLR method (paired Wilcoxon test, \( P > 0.05 \) after correction of \( P \) values for multiple testing using Holm’s method\(^{25,26} \)).

**DISCUSSION**

In this study, the relationship between the number of VFs and the prediction accuracy of VF test results was investigated. As expected, the prediction accuracy improved as the number of VFs in the regression increased. A considerable number (approximately 10 on average) of VFs were needed to saturate the prediction accuracy. Importantly, we observed no merit when using the exponential, quadratic, or logistic models, both in the PW and mean sensitivity analyses, compared with OLSLR. The M-robust regression model, however, enabled significantly better predictions than OLSLR for the PW analyses, although this improvement was not observed for mean sensitivity predictions.

Many studies have investigated the prediction performance of linear and nonlinear regression models.\(^{13,27–29} \)

Caprioli et al.\(^{28} \) recently fitted the exponential regression model, the quadratic regression model, and the OLSLR model to series of VFs and suggested that the exponential model was the best-fitting because it obtained the smallest Akaike’s Information Criterion (AIC); the authors argue that the exponential method better models VF sensitivity as it approaches the floor level of VF sensitivity (0 dB).\(^{29,30} \)

Nevertheless, the model with the smallest AIC does not guarantee it is always the best model for prediction accuracy.\(^{31} \)

Indeed, in the current study, the exponential regression model was significantly worse in terms of prediction accuracy, compared with OLSLR in the PW analysis, in agreement with a previous study.\(^{32} \) One possible reason for this apparently contradictory result may be the difference in datasets analyzed. In our study, the average MDs of the initial and last VFs were \( /C0 6.6 \) and \( /C0 8.7 \) dB, respectively, whereas in the Caprioli et al.\(^{28} \) article, the MDs were much worse: \( /C0 10.9 \) and \( /C0 12.9 \) dB, respectively. As a result, many more VFs will tend to reach the floor level in the previous study. It is possible that the “best”
model depends on the length of the observation period and the stage of glaucoma. Nonetheless, the clinical usefulness of applying the exponential regression model and also the logistic regression model instead of OLSLR, may be limited because it requires approximately 7 or 8 years to obtain 15 VFs under an assumption that VF measurements are carried out every 6 months.

Moreover, accurately predicting patients’ VFs in the early to moderate stages of glaucoma is perhaps clinically more important than predictions in an advanced stage.

Attempts have been made to improve the accuracy of OLSLR for VF trend analysis. Gardiner and Crabb recommended applying a “Three-Omitting” rule in which the last VF is omitted and instead two future VFs are used to detect and confirm progression. Over 6 years of follow-up, they showed that the standard PW-OLSLR method was sensitive for detecting VF progression with a sensitivity value equal to 97.1%; however, the specificity of the method was very low (25.4%). The “Three-Omitting” method, on the other hand, achieved a sensitivity of 65.7% with a specificity of 87.4%.

Thus, the proposed method affords a significant improvement in specificity, but nevertheless is accompanied with a notable loss in sensitivity. It should be pointed out, however, that the “Three-Omitting” approach is not a method to improve the regression model itself and, furthermore, future VFs are always needed to diagnose progression; this is in contrast to the robust regression method used in the current study.

Bengtsson et al. investigated the usefulness of the initial five VFs for predicting future progression, using 11 VFs captured over 8.2 years on average. They showed that linear extrapolation of the five initial VF results offered a reliable predictor of future field loss in most patients. Until now, however, it has not been investigated how many VFs are needed to saturate the prediction accuracy. In the current study, longer series of VFs were investigated and it was shown that a considerable number (approximately 10 VFs taken over 5.1 ± 1.6 years) are required for this purpose.

**Figure 4.** Absolute prediction error associated with OLSLR and M-estimator robust regression in PW sensitivity trend analysis in eyes with mTD progression rate faster than −0.25 dB per year (n = 117 eyes). (A) The first future VF (from the 6th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 14th VFs). (B) The second future VF (7th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 13th VFs). (C) Third future VF (8th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 12th VFs). **P < 0.01 and *P < 0.05 between OLSLR and M-estimator robust regression (paired Wilcoxon test). Gray line, 95% CI range of minimum prediction with OLSLR.

**Figure 5.** Absolute prediction error associated with OLSLR and M-estimator robust regression in PW sensitivity trend analysis of eyes group with mTD progression rate equal to or slower than −0.25 dB per year (n = 130 eyes). (A) The first future VF (from the 6th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 14th VFs). (B) The second future VF (7th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 13th VFs). (C) Third future VF (8th VF to the 15th VF) was predicted using prior VFs (1st to 5th VFs up to 1st to 12th VFs). **P < 0.01 and *P < 0.05 between OLSLR and M-estimator robust regression (paired Wilcoxon test). Gray line, 95% CI range of minimum prediction with OLSLR.
The absolute prediction error was significantly smaller using PW-OLSLR versus the exponential and quadratic models. There was no significant improvement of the prediction accuracy by using the logistic regression model, instead of OLSLR. Moreover, M-estimator robust regression obtained significantly smaller prediction errors than PW-OLSLR. Interestingly, this tendency was more profound in eyes with moderate to advanced-stage glaucoma (Figs. 2, 3). This is probably because test-retest reproducibility of measured VF sensitivity is poor in these eyes\textsuperscript{35} and so the merit of reducing the effect of outliers in the M-estimator robust regression is greater. On the other hand, reducing the effect of outliers could be argued to affect the ability of robust regression to identify fast-progressing eyes. The current results, however, suggest that the prediction accuracy of M-estimator robust regression tended to be significantly better than those with OLSLR in such eyes (Figs. 4, 5). When predicting mean VF sensitivity, no significant differences were observed in prediction errors among the OLSLR, exponential, quadratic, and robust regression models. This is probably because outliers are much less frequent in mean sensitivity measurement compared with PW sensitivity. Indeed, there was no significant difference between the prediction errors associated with M-estimator robust regression and OLSLR when predicting mean VF sensitivity (Fig. 6).

Despite the smaller prediction errors of M-robust regression, the same numbers of VFs were required as OLSLR to reach the 95% CI of minimum absolute error. Both regression models assume that the distribution of VF errors is normally distributed; however, the distribution is actually somewhat bimodal (due to truncation of VF sensitivity) and skewed.\textsuperscript{36} This may have contributed to improve prediction accuracy in the robust regression method. Nonetheless, the improvement of the prediction accuracy is relatively small and it might have only marginal benefit at the clinical settings. There are recent articles that succeeded to improve the prediction accuracy by implementing clinical knowledge into the prediction model, such as clustering eyes with the progression patterns,\textsuperscript{37–40} setting a penalty term in the regression model,\textsuperscript{41} and dividing VF into small sectors.\textsuperscript{15} A future study should be carried out to further improve the diagnostic accuracy by combining the merits of these approaches.

In conclusion, a considerable number of VFs (approximately 10), are required to obtain accurate predictions of future VF test results; moreover, prediction error can be significantly improved by using the robust regression model when PW sensitivity is analyzed.

Acknowledgments

Supported in part by the Japan Science and Technology Agency (JST)-CREST and Grant 26462679 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Disclosure: Y. Taketani, None; H. Murata, None; Y. Fujino, None; C. Mayama, None; R. Asaoka, None

References


