Glaucoma

Applying “Lasso” Regression to Predict Future Visual Field Progression in Glaucoma Patients

Yuri Fujino, Hiroshi Murata, Chihiro 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, 113-8655 Japan, rasaoka-tky@umin.ac.jp.

Submitted: January 13, 2015
Accepted: February 11, 2015

Citation: Fujino Y, Murata H, Mayama C, Asaoka R. Applying “Lasso” regression to predict future visual field progression in glaucoma patients. Invest Ophthalmol Vis Sci. 2015;56:2334–2339. DOI:10.1167/iovs.15-16445

PURPOSE. We evaluated the usefulness of various regression models, including least absolute shrinkage and selection operator (Lasso) regression, to predict future visual field (VF) progression in glaucoma patients.

METHODS. Series of 10 VFs (Humphrey Field Analyzer 24-2 SITA-standard) from each of 513 eyes in 324 open-angle glaucoma patients, obtained in 4.9 ± 1.3 years (mean ± SD), were investigated. For each patient, the mean of all total deviation values (mTD) in the 10th VF was predicted using varying numbers of prior VFs (ranging from the first three VFs to all previous VFs) by applying ordinary least squares linear regression (OLSLR), M-estimator robust regression (M-robust), MM-estimator robust regression (MM-robust), skipped regression (Skipped), deepest regression (Deepest), and Lasso regression. Absolute prediction errors then were compared.

RESULTS. With OLSLR, prediction error varied between 5.7 ± 6.1 (using the first three VFs) and 1.2 ± 1.1 dB (using all nine previous VFs). Prediction accuracy was not significantly improved with M-robust, MM-robust, Skipped, or Deepest regression in almost all VF series; however, a significantly smaller prediction error was obtained with Lasso regression even with a small number of VFs (using first 3 VFs, 2.0 ± 2.2; using all nine previous VFs, 1.2 ± 1.1 dB).

CONCLUSIONS. Prediction errors using OLSLR are large when only a small number of VFs are included in the regression. Lasso regression offers much more accurate predictions, especially in short VF series.

Keywords: glaucoma, visual field, progression, lasso regression, static perimetry, mean deviation

Glaucous visual field (VF) deterioration is irreversible, but may be delayed through appropriate IOP reduction. However, the medical and surgical IOP reduction can be associated with various ocular and general complications. Therefore, it is essential to accurately predict future VF progression when making glaucoma treatment decisions. One popular approach to assess VF deterioration and predict future damage is to apply ordinary least squares linear regression (OLSLR) to global VF indices, such as mean deviation (MD), implemented in the Guided Progression Analysis (GPA) software on the Humphrey Visual Field Analyser (HFA; Carl Zeiss Meditec AG, Dublin, CA, USA). However, VF sensitivity fluctuates in the short-term and long-term, and the reliability of a measured VF is inherently affected by a patient’s concentration. Furthermore, VF measurement noise can be very large even when reliability indices are deemed good.

It has been reported that clinicians should acquire a considerable number of VFs to accurately forecast future progression, such as five or eight VFs. Accumulating this number of VFs can take years in many clinics; hence, various attempts have been made to develop better models to predict VF progression. For instance, Caprioli et al. recently suggested that an exponential method offers more accurate predictions, especially when VF sensitivity approaches the floor level (0 dB). Very recent research has developed dedicated VF regression models, that take into account nonstationary variability and spatial correlations using a Bayesian approach or spatial/temporal patterns of glaucomaous VF progression also using a Bayesian approach. There also has been a renewed interest in applying alternative “ready-made” regression models as an alternative to OLSLR. A number of regression models, such as M-estimator robust regression (M-robust), MM-estimator robust regression (MM-robust), skipped regression (Skipped), and deepest regression (Deepest), have been developed to overcome the sensitivity of OLSLR to outliers. In addition, others have proposed a shrinkage method for OLSLR in which the sum of the absolute values of the regression coefficients is constrained or penalized, known as least absolute shrinkage and selection operator regression (Lasso). The most important merit of using Lasso is the optimum penalty can be decided using the actual clinical information of other patients; how the regression model should be shrunk to accurately predict the future VE. On the contrary, M-robust, MM-robust, Skipped, and Deepest obtain the robustness using only an individual’s data. Lasso regression has been used in many different fields, including the analysis of human perception and genetic analysis. Indeed, we recently applied the method to predict the MD values in the 10-2 HFA VFs from 24-2 HFA VFs. In this study, different robust regression models as well as Lasso regression were applied to the mean of total deviation values (mTD) in patients’ VF series, and the performance of the different methods for predicting future progression was compared.

METHOD

The study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo. Informed consent was obtained from all

Copyright 2015 The Association for Research in Vision and Ophthalmology, Inc.
www.iovs.org | ISSN: 1552-5785

2334
predict glaucoma progression with lasso regression

Subjects and VFs

The VF data were retrospectively obtained from a total of 513 eyes in 324 patients with normal tension glaucoma, primary open-angle glaucoma, or exfoliation glaucoma. Patients were followed up in the glaucoma clinic at the University of Tokyo Hospital between 2002 and 2013; all of the patients had their VFs measured at least 12 times. Glaucoma was diagnosed when the following findings were present: (1) presence of typical glaucomatous changes in the optic nerve head, such as a rim notch with a rim width < 0.1 disc diameters or a vertical cup-to-disc ratio of >0.7 and/or a retinal nerve fiber layer defect with its edge at the optic nerve head margin greater than a major retinal vessel, diverging in an arcuate or wedge shape confirmed by a panel of glaucoma specialists (HM and RA) after inspection of stereo-fundus photographs, and (2) a glaucomatousVF defined following the criteria of a cluster of ≥3 points in the pattern deviation plot in a single hemifield (superior/inferior) with P < 0.01, one of which must have been P < 0.01, excluding the outermost test point of Humphrey Field Analyzer 30-2 program; glaucoma hemifield test (GHT) result outside of normal limits; or abnormal pattern standard deviation (PSD) with P < 0.05. The VF measurements were performed using the HFA with either the 30-2 or 24-2 program and the Swedish Interactive Threshold Algorithm Standard. When VFs were obtained with the 30-2 test pattern, only the 52 test locations overlapping with the 24-2 test pattern were used in the analysis and for the calculation of mTD. Patients' first two VFs were excluded from the analysis. Other inclusion criteria in this study were best corrected visual acuity better than 6/12, refraction within ± 6 diopter (D) ametropia, no previous ocular surgery except for cataract extraction, and intraocular lens implantation, and no other anterior and posterior segment of the eye disease that could affect the VF, including cataract other than clinically insignificant senile cataract. Reliability criteria for VFs were applied: fixation losses less than 20% and false-positive responses less than 15%; the false-negative rate was not applied as a reliability criterion based on a previous report. The VF of a left eye was mirror-imaged to that of a right eye for statistical analyses.

Prediction Accuracy

Prediction accuracy was compared between conventional OLSLR, M-robust, MM-robust, Skipped, Deepest, and Lasso regression. For each method, regression was done using mTD values from the first to the third VFs (VF1–3) of each patient, and the mTD values of the 10th VF (VF10) were predicted. The same procedure was carried out using the TD values in different series: VF1–4, VF1–5, VF1–6, VF1–7, VF1–8, and VF1–9, and the mTD values of VF10 were predicted every time. The predictive accuracy of each method was compared using absolute errors. As a subanalysis, predictive accuracy also was compared in eyes with an mTD progression rate <−0.25 dB/y (based on a patient's first 10 VFs), which commonly is considered to be deterioration by pressure-independent damaging factors.

Statistical Models

The following regression models were performed. For the OLSLR method:

\[ y = ax + b \]

For the M-robust regression method:

\[ y^* = x + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon_1 \]

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

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

where the function \( \rho \) gives the contribution of each residual to the objective function. In short, M-robust regression, due to the formula above, is much less affected by outliers than OLSLR. In the current study, M-statistics were calculated using Huber’s method.

The MM-robust regression method obtains further robustness to outliers by combining M estimation and high breakdown value estimation, as introduced by Yohai in 1987. The Skipped regression method is identical to the Theil–Sen estimator. This method also gains robustness to outliers by choosing the median slope among all lines constructed by all pairs of sample points.

The Deepest regression method gains robustness to outliers as well, by identifying the regression line with largest “depth” (as opposed to weighting residuals), which is defined as how well the data are balanced by the model.

The Lasso regression is a shrinkage method for OLSLR in which the sum of the absolute values of the regression coefficients is constrained or penalized. More precisely, let \( x_i \) denote the variables and let \( y \) denote the response (please note \( x_i \) are normalized and \( y \) has mean zero). The Lasso algorithm solves the following problem:

\[ \min_{\| \beta \|_0 \leq q} \left[ \frac{1}{2N} \sum_{i=1}^{N} (y_i - (\beta_0 - x_i^T \beta))^2 + [\lambda P_\delta(\| \beta \|_l)] \right] \]

\[ P_\delta(\| \beta \|_l) = (1 - z) \frac{1}{2} \| \beta \|_l^2 + z \| \beta \|_l \]

In Equation 1, \( \min_{\| \beta \|_0 \leq q} [(1/2N) \sum_{i=1}^{N} (y_i - (\beta_0 - x_i^T \beta))^2] \) is identical to OLSLR and \( \lambda P_\delta(\| \beta \|_l) \) is the penalty term for the shrinkage. Thus the \( \lambda \) (Lambda) value represents the degree of penalty in Lasso. Equation 1 is Lasso regression when \( \| \beta \|_l = 1 \) and Ridge regression when \( \| \beta \|_l = 0 \); however, this discrimination is not applicable to the current study, because there is only one variable (mTD).

Statistical Analysis

Absolute prediction accuracy was calculated as the absolute value of the difference between the model-predicted and the observed mTD value. For Lasso regression, the prediction error was calculated using leave-one-out cross validation. In this validation method, the data from a single patient (one or two eyes) were used as a testing dataset and all other data were used as training data; this procedure was repeated so that each patient was used only once as the testing dataset. In other words, for each individual, only the data from all other subjects (\( n = 323 \) in 324) were used to produce a diagnosis. An optimum \( \lambda \) value was identified for each iteration (patient), and the prediction error was calculated.

All statistical analyses were done using the statistical programming language R (ver. 2.15.0; The R Foundation for Statistical Computing, Vienna, Austria). The M-robust, MM-robust, Skipped, Deepest, and Lasso regressions were calcu-
lated using the R packages “WRS” and “glmnet.” Absolute prediction errors were compared using the repeated measures ANOVA. Benjamini’s method was used to correct $P$ values for the problem of multiple testing. A linear mixed model was used to analyze the relationship between two values, whereby patients were treated as a “random effect.”

**RESULTS**

Characteristics of the study population are summarized in the Table. The mTD at baseline was $-6.9 \pm 6.2$ dB (mean $\pm$ SD) and initial patient age was $54.2 \pm 12.3$ years. The progression rate of mTD was $-0.24 \pm 0.63$ dB/yr (Fig. 1).

Figure 2 shows the absolute prediction errors associated with OLSLR, M-robust, MM-robust, Skipped, Deepest, and Lasso regression. The prediction error for MM-robust regression with VF1,4 could not be calculated because a leverage point could not be calculated. Absolute prediction errors became smaller as the number of VF tests included in the regression increased. There was no significant improvement in error by applying M-robust, MM-robust, Skipped, and Deepest, compared to using the OLSLR at any time point ($P > 0.05$, repeated ANOVA with Benjamini’s correction for multiple testing), except for M-robust with VF1,4 ($P = 0.028$, repeated ANOVA with Benjamini’s correction for multiple testing). The absolute prediction errors with the Lasso model were significantly better than OLSLR when VF1,3 to VF1,8 were used for prediction ($P < 0.0001$, repeated ANOVA with Benjamini’s correction for multiple testing). Among the 513 eyes, 234 eyes showed progression faster than $-0.25$ dB/yr. A significant improvement was observed when applying Lasso, compared to OLSLR, when the initial one or two VFs were used to predict ($P = 0.007$, 0.035, repeated ANOVA with Benjamini’s correction for multiple testing).

Figure 4 shows the optimum $\lambda$ value derived in relationship to the number of VFs used for prediction. The $\lambda$ value decreased as the number of VFs used in the prediction increased.

As shown in Figure 5, there was no significant relationship between the optimum $\lambda$ value derived for VF1,3, VF1,4, and VF1,5, and the mTD value of VF1 ($P = 0.18$, 0.12, and 0.31, respectively, linear mixed model). As shown in Figure 6, there was no significant relationship between the optimum $\lambda$ value derived for VF1,3, VF1,4, and VF1,5, and the difference between mTDs in VF1 and VF10 ($P = 0.18$, 0.33, and 0.94, respectively, linear mixed model). Figures 5 and 6 are smoothed scatter plots (plotted using the R package “graphics”), which better differentiate dense regions of points.

**DISCUSSION**

In this study, the prediction performance of various robust regression models and Lasso regression were compared against OLSLR. As a result, a considerable number of VFs were needed to obtain an accurate prediction with OLSLR; it also was shown that there was no significant improvement in error by applying robust regression (M-robust, Skipped, and Deepest regression methods). On the contrary, a significantly smaller prediction error was observed with Lasso regression even when the number of VFs included was just three long. This difference should be attributed to the algorithms; in Lasso regression, the model is optimized using other patients’ information (i.e., the degree of shrinkage is decided by optimizing the penalty value), while in M-robust, MM-robust, Skipped, and Deepest regression method only data from each individual eye were used.

The VF variability significantly hampers the usefulness of VF trend analyses, and consequently, the minimum number of VFs required to obtain reliable VF trend analysis results has been widely discussed in previous studies, with research suggesting that at least five or eight VFs, or even higher are required. Indeed, prediction accuracy associated with OLSLR was poor when a small number of VFs were used in the current study. This poor prediction accuracy was not improved by applying a number of different robust regression methods. On the other hand, Lasso regression performed much better due to the fact that the method uses a penalty term, which helps to reduce prediction errors. In other words, the coefficient terms in Lasso regression are adjusted by the optimum penalty ($\lambda$), which is obtained from real data (from other patients). As a result, Lasso mTD trend analysis becomes much more robust to measurement noise and, consequently, prediction accuracy was dramatically improved. This is different to M-robust, MM-robust, Skipped, and Deepest regression, which attempt to improve robustness using only an individual’s data.

In the current study, leave-one-out cross validation was used to evaluate the performance of Lasso regression. As described in the Methods section, the original dataset was divided into a testing dataset (one patient) and training data (all other patients), and the $\lambda$ value was calculated using only the
training dataset; this process was repeated so that each patient was used as a testing dataset once. This is identical to the clinical situation—a new patient can be classified according to the predetermined optimum $k$ value. Furthermore, in clinical practice, a $k$ value could be calculated continuously by adding the data of new patients to an ever-growing database of patient data, which would further improve prediction accuracy. In addition, the Lasso regression performed in this study was built using free statistical software and packages, specifically "R" (ver. 3.1.0; The R Foundation for Statistical Computing).

As shown in Figure 4, a large optimum $\hat{\lambda}$ value was observed when a small number of VFs was used for prediction and it decreased as the number of VFs used for prediction increased. This suggested that any mTD trend analysis should be penalized according to the number of VFs used, which would further improve prediction accuracy. In addition, the Lasso regression performed in this study was built using free statistical software and packages, specifically "R" (ver. 3.1.0; The R Foundation for Statistical Computing).

In the subanalysis, in which only progressing eyes were analyzed, the advantage of using Lasso regression was observed when a small numbers of VFs were used for prediction. However, no advantage was observed when a larger numbers of VFs were used. As shown in Figure 1, more than half of eyes...
in the complete sample did not progress, but the Lasso λ value was decided based on all eyes. Recent studies have improved prediction accuracy by clustering eyes with different progression patterns. Thus, it may be advantageous to optimize Lasso regression (the λ value) based on a patient’s progression pattern; that is, prediction accuracy may be further improved by using a clustering approach in combination with Lasso regression; this should be carried out in a future study.

It has been suggested using a VF progression analysis tool, such as PROGRESSOR, improves clinicians’ decisions regarding VF progression. A possible caveat of the current results is that the prediction with the Lasso regression is not readily useable at the clinical setting. Therefore, it would be clinically beneficial to develop software/support tools to predict VF progression, similar to PROGRESSOR. In particular, only standard data are needed to apply the current methodology in the clinical setting; having the record of MD values with the date of VF measurements of a patient, since the optimum penalty (λ) value can be calculated from other patients.

In conclusion, prediction accuracy of VF progression is poor when OLSLR is used with a small number of VFs. Prediction performance was not improved by applying M-robust, MM-robust, Skipped, and Deepest regressions; however, considerable improvement was observed by applying the Lasso regression technique.

Acknowledgments

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

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

References

Predict Glaucoma Progression With Lasso Regression


