

# Assessing Trends in Functional and Structural Characteristics: A Survey of Statistical Methods With an Example From Ophthalmology

Johannes Ledolter<sup>1,2</sup> and Randy H. Kardon<sup>2,3</sup>

<sup>1</sup> Department of Management Sciences at the University of Iowa, Iowa City, IA, USA

<sup>2</sup> Iowa City VA Medical Center, Iowa City, IA, USA

<sup>3</sup> Department of Ophthalmology and Visual Sciences at the University of Iowa Hospital, Iowa City, IA, USA

**Correspondence:** Johannes Ledolter, Tippie College of Business, S352 Pappajohn Business Building, University of Iowa, Iowa City, IA 52242, USA. e-mail: johannes-ledolter@uiowa.edu

**Received:** 30 March 2018

**Accepted:** 8 September 2018

**Published:** 31 October 2018

**Keywords:** classification; OCT measurements; patient monitoring; personalized medicine; random effects trend model; subject variability; trend change detection

**Citation:** Ledolter J, Kardon RH. Assessing trends in functional and structural characteristics: a survey of statistical methods with an example from ophthalmology. *Trans Vis Sci Tech.* 2018;7(5):34, <https://doi.org/10.1167/tvst.7.5.34>  
Copyright 2018 The Authors

**Purpose:** Clinical decisions on treatment are usually based on short-term records of consecutive measurements of structure and function. Useful models for analyzing average trends and a description of statistical methods for classifying individual subjects on the basis of subject-specific trend progressions are presented.

**Methods:** Random effects trend models allow intercepts and slopes of the trend regression to vary across subjects around group-specific mean intercepts and mean slopes. Model results assess whether average intercepts and slopes and subject variability in intercepts and slopes are the same across groups. Fisher's discriminant functions are used for classification.

**Results:** Methods are presented and illustrated on structural visual data from a multiyear perimetry study. Average thickness of the ganglion cell layer from the optical coherence tomography macula scan and of the retinal nerve fiber layer from the optic disc scan for both glaucoma patients on optimal treatment and normal subjects are analyzed. The random effects trend model shows that average intercepts of glaucoma patients and normal subjects are quite different, but that average slopes are the same, and that the subject variability in both intercepts and slopes is larger for the glaucoma group. These findings explain why the subject-specific trend progression is not a good classifier; it is the level of the measurement (intercept or baseline value) that carries useful information in this particular cohort example.

**Translational Relevance:** Clinicians base decisions on short-term records of consecutive measurements and need simple statistical tools to analyze the information. This paper discusses useful methods for analyzing short time series data. Model results assess whether there exist significant trends and whether average trends are different across groups. The paper discusses whether clinical measures classify patients reliably into disease groups, given their variability. With ever more available data, classification plays a central role of personalized medicine.

## Introduction

Clinicians study the progression (i.e., the changes) of functional and structural characteristics over time and need to know how to analyze the resulting data with the appropriate statistical methods. This paper has two objectives: (1) Discuss useful statistical approaches for analyzing groups of short time series and for classifying individual subjects on the basis of subject-specific trend progressions. (2) Illustrate the methodology by analyzing

structural optical coherence tomography (OCT) data collected during a multiyear perimetry study conducted at the University of Iowa.

## Methods

### The Linear Trend Model for Observations on an Individual Subject

For each patient,  $n$  consecutive measurements  $Y_t$  are obtained at times  $Time_1 < Time_2 < \dots < Time_n$ .

Times are typically expressed in years, and there is no requirement that times are spaced equally. Times are often expressed relative to the time of the initial visit. For example,  $Time_1 = 0$ ,  $Time_2 = 0.5$ ,  $Time_3 = 1.0$ ,  $Time_4 = 1.25$ ,  $\dots$ ,  $Time_8 = 4.14$  expresses that eight observations are obtained—at the start of the study, after 6 months, after 12 months, after 15 months,  $\dots$ , and after 4 years and  $365(0.14) = 51$  days since the start of the study.

The following trend regression model is considered for the progression of the measurements

$$Y_t = \alpha + \beta Time_t + \varepsilon_t. \quad (1)$$

The parameter  $\alpha$  represents the expected baseline measurement at the time a subject enters the study. The parameter  $\beta$  reflects the expected change in the measurement for a unit change in time (which, in our illustration, is 1 year). The model in equation (1) assumes linearity; the mean change over the first year is the same as the mean change in the second year, and so on. The measurement errors  $\varepsilon_t$  are independent, homoscedastic (that is, with constant variance; if the variation increases with the level, a log transformation of the response will stabilize the variance), and normally distributed. Under these assumptions one can obtain the best (least squares) estimates of the parameters, their SEs, confidence intervals, and tests of hypotheses. Least squares estimates of the slope and the baseline are given by

$$b = \frac{\sum_{t=1}^n (Time_t - \overline{Time}) Y_t}{\sum_{t=1}^n (Time_t - \overline{Time})^2} \text{ and } a = \bar{Y} - b \overline{Time} \text{ where } \bar{Y} = (1/n) \sum_{t=1}^n Y_t \text{ and } \overline{Time} = (1/n) \sum_{t=1}^n Time_t.$$

The estimated SEs are given by  $se(b) = \frac{s}{\sqrt{\sum_{t=1}^n (Time_t - \overline{Time})^2}}$

$$\text{and } se(a) = s \sqrt{\frac{1}{n} + \frac{(\overline{Time})^2}{\sum_{t=1}^n (Time_t - \overline{Time})^2}}, \text{ where } s =$$

$\sqrt{\frac{\sum_{t=1}^n [Y_t - (a + b Time_t)]^2}{n-2}}$  is the unbiased estimate of the measurement variability.

These results can be found in most regression texts; see, for example, Sections 2.2 to 2.5 of Abraham and Ledolter.<sup>1</sup>

### The Random Effects Trend Model: Modeling Linear Trends for Several Subjects From More Than One Group

Typically, there is not just one single subject, but there are several subjects, which are observed at possibly different times and with different sample sizes. And, typically there is not just a single group of

subjects, but several groups. For example, a group of normal/healthy subjects and a group (or groups) of subjects with a certain disease (such as glaucoma in the example in the Results section, or a blast injury when studying traumatic brain injury).

We denote the observation on the  $i$ th subject in group  $j$  ( $j = 1$  for normal,  $j = 2$  for disease) at time  $t$  with  $Y_{it}^{(j)}$  and the time at which the observation is obtained as  $Time_{it}^{(j)}$ , with  $Time_{i1}^{(j)} = 0$  when time is measured from the first visit at the clinic. The model in equation (2) allows for subject variability in the baselines and the slopes. The model

$$Y_{it}^{(j)} = (\alpha^{(j)} + \tilde{\alpha}_i^{(j)}) + (\beta^{(j)} + \tilde{\beta}_i^{(j)}) Time_{it}^{(j)} + \varepsilon_{it}^{(j)} \quad (2)$$

includes group-specific average baselines and slopes (the fixed effects,  $\alpha^{(j)}$  and  $\beta^{(j)}$ ) and subject-specific random effects for the baselines and slopes. The random effects  $\tilde{\alpha}_i^{(j)}$  and  $\tilde{\beta}_i^{(j)}$  are independent across subjects, with zero means and group-specific standard deviations  $\sigma_{\alpha}^{(j)}$  and  $\sigma_{\beta}^{(j)}$ . The model allows for correlation among baseline and slope,  $\rho_{\tilde{\alpha}\tilde{\beta}}^{(j)} = Corr(\tilde{\alpha}_i^{(j)}, \tilde{\beta}_i^{(j)})$ , in case there is reason to believe that a subject's slope changes with the baseline. The measurement errors  $\varepsilon_{it}^{(j)}$  are independent across time and subjects, with group-specific standard deviations  $\sigma_{\varepsilon}^{(j)}$ .

For two groups (in our example, the groups of normal and glaucoma patients), the model in equation (2) can equivalently be written as

$$Y_{it}^{(j)} = [\alpha + \alpha^* IND(Glaucoma)_{it}^{(j)} + \tilde{\alpha}_i^{(j)}] + [\beta + \beta^* IND(Glaucoma)_{it}^{(j)} + \tilde{\beta}_i^{(j)}] Time_{it}^{(j)} + \varepsilon_{it}^{(j)} \quad (3)$$

where  $IND(Glaucoma)_{it}^{(j)}$  is an indicator variable with value 0 when subject  $i$  is from the normal group ( $j = 1$ ) and value 1 when subject  $i$  is from the glaucoma group ( $j = 2$ ). The parameter  $\alpha^*$  expresses the difference in the average baselines of glaucoma and normal subjects. The parameter  $\beta^*$  expresses the difference in the average slopes of glaucoma and normal subjects.

The models in equations (2) and (3) are known as repeated measurements or mixed linear regression models. They are studied in detail by Searle<sup>2</sup> and Searle et al.<sup>3</sup> Estimation of such models can be carried out with the PROC MIXED procedure of the SAS Statistical Software,<sup>4</sup> or with the R Statistical Software<sup>5</sup> through its libraries nlme and lme4. Parameter estimates are usually obtained through restricted maximum likelihood (REML).

The parameter  $\alpha^*$  expresses the difference between the average baselines of the two groups while the parameter  $\beta^*$  assesses whether the average slopes of the two groups are the same. The model is very general, and not all parameters may be needed. Slopes and baselines may be uncorrelated if subject trend changes are unrelated to baselines; that is,  $\rho_{\tilde{\alpha}\tilde{\beta}}^{(1)} = \rho_{\tilde{\alpha}\tilde{\beta}}^{(2)} = 0$ . Subject-variability among baseline intercepts, and also among slopes, may be the same for the two groups; that is,  $\sigma_{\alpha}^{(1)} = \sigma_{\alpha}^{(2)}$  and  $\sigma_{\beta}^{(1)} = \sigma_{\beta}^{(2)}$ . And, the measurement variability for the two groups may be the same; that is,  $\sigma_{\varepsilon}^{(1)} = \sigma_{\varepsilon}^{(2)}$ .

Time series data for  $n_1$  subjects from the normal group and  $n_2$  subjects from the disease/experimental group are available to estimate the model. The model with the general covariance structure and models with restricted covariance structure and equality constraints among the variances can be compared and assessed using either the traditional log-likelihood ratio test or the Bayesian (BIC) or Akaike (AIC) information criteria, which penalize the likelihoods for the number of parameters; see Chapter 3 of Verbeke and Molenberghs.<sup>6</sup>

Random effects models allow us to address important questions. A comparison of the average trend of normal subjects with the average trend of subjects from a comparison group (subjects with a certain disease) is often the main objective of a study, and this objective is addressed by testing the hypothesis whether or not  $\beta^* = 0$ . Another question of interest is whether the variability of slopes in the normal group is the same as the variability of slopes in the disease group, and this question can be answered by testing whether  $\sigma_{\beta}^{(1)} = \sigma_{\beta}^{(2)}$ . Hypotheses about baselines (whether average baselines are the same across groups, and whether subject variability of baselines are the same across groups) can be assessed similarly.

### Classifying Subjects Into One of Two (or More) Groups

Another important objective of medical studies is classification. That is, how to use characteristics (such as the slope in a trend regression) to classify a new subject into either one (the healthy) or the other (the disease) group. There may be strong evidence that the difference between the average slopes of the two groups is statistically significant, especially if the sample sizes in the two groups are large. But with large subject variability among the slopes and with considerable overlap of the slopes from the two

groups, it will be difficult to classify subjects; the misclassification errors (classifying a true healthy subject as diseased or a diseased subject as healthy) will most likely be unacceptable. Establishing statistical significance of a difference of average slopes is, in general, much easier than the classification of subjects into one or the other group.

In order to carry out the classification, we adopt a two-step approach that (1) estimates the baseline measurement and slope for each subject and (2) compares and uses for classification the estimated baseline measurements and slopes of all subjects from the two groups. Least squares estimates of the baseline and of the slope of the trend regression are calculated for each subject. This leads to estimates  $(a_1^{(1)}, b_1^{(1)})$ ,  $(a_2^{(1)}, b_2^{(1)})$ , ...,  $(a_{n_1}^{(1)}, b_{n_1}^{(1)})$  for subjects in group 1 and estimates  $(a_1^{(2)}, b_1^{(2)})$ ,  $(a_2^{(2)}, b_2^{(2)})$ , ...,  $(a_{n_2}^{(2)}, b_{n_2}^{(2)})$  for subjects in group 2. The reliability of the least squares estimates depends on the number and the time-spacing of the observations. One may want to omit estimates from subjects with only few measurements to make the estimates across subjects comparable.

In general, each subject is characterized by  $k$  features, and information about group membership (healthy group and diseased group) is available on each subject (that is, one deals with supervised learning). If the slope of the progression is the only feature to be considered, then  $k = 1$ . If the baseline and the slope of the trend model are used to describe the state of a subject, then  $k = 2$ . The baseline may be a useful feature as it measures the subject's starting level of the measured characteristic. The Fisher<sup>7</sup> linear discriminant analysis classifies a patient with feature vector  $x$  into group 2 (disease) if

$$(\mu_1 - \mu_2)^T \Sigma^{-1} x < (1/2)(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2). \quad (4)$$

The superscript  $T$  stands for the vector (matrix) transpose. Here  $\mu_1$  and  $\mu_2$  are the mean vectors (of dimension  $k$ ) for group 1 (healthy) and group 2 (diseased), and they are estimated with the sample mean vectors of features for subjects in groups 1 and 2. The common covariance matrix of the features  $\Sigma$  (a matrix of dimension  $[k \times k]$ ) is estimated with the pooled covariance matrix  $\frac{(n_1-1)S_1 + (n_2-1)S_2}{(n_1+n_2-2)}$ , where  $S_1$  and  $S_2$  are the covariance matrices for group 1 and group 2, respectively. The linear discriminant function minimizes the expected cost of misclassification if the features have normal distributions with group mean vectors  $\mu_1$  and  $\mu_2$  and identical group covariance matrices, equal



prior probabilities, and equal misclassification costs. See, for example, Chapter 11 of Johnson and Wichern<sup>8</sup> and Chapter 12 of Ledolter.<sup>9</sup> If  $k = 1$ , Fisher's linear discriminant analysis classifies a subject with feature  $x$  into the group with the closest mean.

The optimal discriminant function becomes quadratic if the covariance matrices for groups 1 and 2 are not the same. The Fisher<sup>7</sup> quadratic discriminant function classifies a subject with feature vector  $x$  into group 2 (disease) if

$$-(1/2)x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu_1^T \Sigma_1^{-1} - \mu_2^T \Sigma_2^{-1})x < \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} + \frac{1}{2} (\mu_1^T \Sigma_1^{-1} \mu_1 - \mu_2^T \Sigma_2^{-1} \mu_2), \quad (5)$$

with mean and variances estimated by their respective sample estimates.

Classification uses the known label (that is, the known group association) when constructing the classification rules. Misclassification errors (i.e., the number of times an observation from group 1 is classified into group 2 and the number of times an observation from group 2 is classified into group 1, expressed as a fraction of the number of items classified) can be calculated. Within-sample and out-of-sample misclassification errors can be obtained. For within-sample errors, the data are used "twice"—first for developing the classification rules and then for evaluating the performance. For out-of-sample errors, a subset of the data is used to develop the classification rules while a different subset is used in the evaluation.

Fisher's discriminant analysis method is one of the simplest approaches to classification. It makes the assumption that the  $k$ -variate distributions are normal—an assumption that is often violated. Several nonparametric classification approaches that do not rely on normality are available, among them the Support Vector Machine (SVM) approach. We refer the interested reader to the contributions by Boser et al.<sup>10</sup> and Christianini and Shawe-Taylor,<sup>11</sup> and readily available computer software such as Package e1071 in the R Statistical Software<sup>5</sup> and LIBSVM, an extensive library for SVMs.

## Iowa Ophthalmology Patients: Normal Subjects and Glaucoma Patients

Data on Iowa 105 glaucoma patients and 55 normal subjects are analyzed. Subjects were participants in a multiyear Variability in Perimetry (VIP) study directed by Michael Wall, MD, and funded by the Veterans Administration Rehabilitation Research and Development Division. Adult patients with glaucoma were

recruited from the glaucoma clinic at the University of Iowa Department of Ophthalmology and Visual Sciences. The research was approved by the University of Iowa and Veterans Affairs institutional review boards, and informed consent was obtained from all patients after an explanation of the study. Inclusion criteria for cases were as follows: (1) the clinical diagnosis of glaucoma determined by the presence of glaucomatous optic disc changes confirmed through the examination of fundus images and (2) the presence of visual field defects (mean deviation [MD] between  $-2$  and  $-20$  dB on standard automated perimetry, as well as either three adjacent locations in a clinically suspicious area falling outside the deviation limits compared with normative data at  $P < 0.05$  [fifth percentile] or two adjacent locations at  $P < 0.01$  [first percentile]). Cases were not required to have elevated intraocular pressure, but they were excluded if there were cataracts causing visual acuity worse than 20/30, they were younger than 19 years, or they had a pupil size of less than 2.5 mm. If both eyes qualified, one eye was randomly selected for inclusion in the study.

Healthy adults (controls) were recruited using advertisements inviting participation in a research study. Inclusion criteria for controls were as follows: (1) no history of eye disease, (2) refractive error within a  $\pm 5$ -diopter sphere and  $\pm 2$ -diopter cylinder, (3) no history of diabetes mellitus or systemic arterial hypertension, and (4) a normal ophthalmologic examination result, including 20/30 or better visual acuity. An examination by an ophthalmologist on the day of testing or the results of a complete eye examination within 2 years of the testing date was used to confirm normal ocular health. One eye was randomly selected for the study. If a control subject developed a pattern of visual loss leading to an ophthalmologic diagnosis other than refractive error, the individual was not included in the analysis. This research adhered to the tenets of the Declaration of Helsinki.

Visual field testing was performed with automated threshold perimetry (Humphrey Visual Field 24-2 SITA Standard test with size III, Zeiss-Meditec, Dublin, CA). Field test results were excluded from the analysis if they were unreliable as defined by excessive false-positive ( $>10\%$ ), false-negative ( $>33\%$ ), or fixation losses ( $>33\%$ ).

All glaucoma patients were being adequately treated and monitored by the glaucoma service at the University of Iowa Department of Ophthalmology and Visual Sciences during the course of the study. In addition to measurement of visual field sensitivity in one eye with various perimetry testing

**Table 1.** Estimation Results

	Intercept			Slope			Measurement Error
	$\alpha$	$\alpha^*$	$(\sigma_{\alpha}^{(j)})^2$	$\beta$	$\beta^*$	$(\sigma_{\beta}^{(j)})^2$	$(\sigma_{\varepsilon}^{(j)})^2$
<b>GCL</b>							
M1	78.16 (N)	<b>-15.85 (G)</b> SE = 1.23 PV < 0.001	36.15 (N) 88.19 (G)	<b>-0.291 (N)</b> SE = 0.041 PV < 0.001	-0.027 (G) SE = 0.083 PV = 0.749	0.054 (N) 0.451 (G)	0.659 (N) 0.953 (G)
M2	78.68 (N)	<b>-15.86 (G)</b> SE = 1.23 PV < 0.001	36.15 (N) 88.19 (G)	<b>-0.297</b> SE = 0.035 PV < 0.001		0.054 (N) 0.447 (G)	0.659 (N) 0.953 (G)
<b>RNFL</b>							
M1	90.14 (N)	<b>-25.87 (G)</b> SE = 1.53 PV < 0.001	71.46 (N) 106.56 (G)	-0.145 (N) SE = 0.089 PV = 0.107	+0.089 (G) SE = 0.141 PV = 0.527	0.207 (N) 0.894 (G)	4.390 (N) 4.078 (G)
M2	90.10 (N)	<b>-25.81 (G)</b> SE = 1.53 PV < 0.001	71.44 (N) 106.56 (G)	-0.109 SE = 0.069 PV = 0.116		0.205 (N) 0.892 (G)	4.391 (N) 4.077 (G)

Estimation results for the general random effects model in Equation (3),  $Y_{it}^{(j)} = [\alpha + \alpha^*IND(Glaucoma)_{it}^{(j)} + \tilde{\alpha}_i^{(j)}] + [\beta + \beta^*IND(Glaucoma)_{it}^{(j)} + \tilde{\beta}_i^{(j)}]Time_{it}^{(j)} + \varepsilon_{it}^{(j)}$ , are listed under model M1. G, glaucoma; N, normal. The expressions SE and PV denote the standard error and the probability value of the estimate. Significant results, with probability value smaller than 0.05, are shown in bold. The correlation between the random intercept and slope is small, and  $\rho_{\tilde{\alpha}\tilde{\beta}}^{(1)}$  and  $\rho_{\tilde{\alpha}\tilde{\beta}}^{(2)}$  are set to zero. Estimation results for model M2 assume that the average slopes of the glaucoma and normal groups are the same. This restriction is appropriate because the estimate of  $\beta^*$  in model M1 is insignificant ( $P$  values 0.749 and 0.527, respectively).

methods, all subjects underwent spectral domain OCT testing (Cirrus, Zeiss-Meditec) on the same day. A  $200 \times 200$  volume scan of the macula and disc were obtained at each test date and were included if a signal-to-noise ratio was at least 7:10. The thickness of the average ganglion cell layer (GCL) complex derived from the macula scan and the average retinal nerve fiber layer (RNFL) derived from the optic disc scan were measured and analyzed over time. We omitted subjects with three or fewer observations. In addition, we omitted eight glaucoma patients with unreliable segmentation of the GCL. The median number of observations per subject was 10 (mean = 9.01; minimum = 4; maximum = 11) for the glaucoma group and 11 (mean = 10.38; minimum = 5; maximum = 12) for the normal group. The median time span from the first to the last visit (range) per subject was 3.99 years (mean = 3.65; SD = 0.90) for the glaucoma group and 4.12 years (mean = 4.00; SD = 0.80) for the normal group. The median time between consecutive visits was 0.50 years (mean = 0.46, SD = 0.20) for glaucoma patients and 0.48 years (mean = 0.43, SD = 0.23) for normal subjects.

## Results

### The Random Effects Trend Model

The model in equation (3) is estimated for RNFL and GCL, and the results are shown in Table 1. The assumption of uncorrelated errors is appropriate for this particular data set as the autocorrelations of the errors in the trend regression are all small (average lag 1 autocorrelation  $-0.19$  for RNFL and  $-0.15$  for GCL). The estimation results (model M1) for RNFL show that the measurement variability is about the same in each group, but that the subject-variability among the baselines and also the slopes is larger in the glaucoma group than in the normal group. The average baseline (RNFL thickness, intercept) in the glaucoma group is significantly thinner (by 25.87 micron units) than the average baseline in the normal group. There is evidence for a negative average slope in the normal group (0.145 micron unit reduction per year, with SE = 0.089;  $P = 0.107$ ). The average rate of change for the glaucoma group is actually smaller ( $0.145 - 0.089 = 0.056$  micron unit reduction per year),

but the difference is not statistically significant. The second model M2 with a common average slope implies a reduction of 0.109 units per year ( $SE = 0.069$ ;  $P = 0.11$ ). In summary, we find that the average RNFL baseline of the glaucoma group is considerably thinner than the average RNFL baseline of the normal group (as expected), some evidence for an overall average reduction over time (probability value around 0.10), no statistically significant difference between the average slopes in the normal and the glaucoma groups, and larger subject variability among the slopes in the glaucoma group.

The estimation results (model M1) for GCL show that the measurement variability is roughly the same in each group, but that the subject-variability among the baselines and also the slopes is larger in the glaucoma group than in the normal group. The average baseline of GCL thickness in the glaucoma group is significantly thinner (by 15.85 micron units) than the one in the normal group. We find statistically significant evidence for a negative average slope in the normal group (0.291 micron unit reduction per year, with  $SE = 0.041$ ). The average rate of change for the glaucoma group is slightly larger ( $0.291 + 0.027 = 0.318$  micron unit reduction per year), but the difference is not statistically significant. In summary, we find the average GCL baseline in the glaucoma group very much and significantly smaller than the average GCL baseline of the normal group, evidence for a significant overall average reduction over time (0.297 per year), but no statistically significant difference between the average slopes in the normal and the glaucoma groups. Furthermore, we find larger subject variability among the slopes in the glaucoma group.

The results—no difference in the average slopes across the two groups and larger variability in the glaucoma group—imply that a strategy that uses just slopes will be unsuccessful in classifying subjects into either the normal or the glaucoma groups. The only distinguishing feature that reliably classifies subjects into one of these two groups is the baseline measurement at the start of the study.

### Classification Using Subject-Specific Least Squares Estimates

We use data on 104 glaucoma patients (we omitted one glaucoma patient with missing GCL measurements) and 55 normal subjects that have been analyzed above. Least squares estimates of baselines

and slopes from the regression model in [equation \(1\)](#) are used to obtain the Fisher linear and quadratic discriminant functions (discussed in the Methods section) for classification. Scatter plots of baselines and slopes, for GCL and RNFL, are shown in [Figures 1](#) and [2](#). Normal subjects are shown as red dots and glaucoma subjects are shown as blue dots. Subjects that are misclassified by the linear (quadratic) discriminant function are displayed with an open circle in a color that reflects the incorrect classification. A red dot with a blue circle indicates a normal subject incorrectly classified as coming from the glaucoma group. A blue dot with a red circle indicates a glaucoma patient incorrectly classified as coming from the normal group.

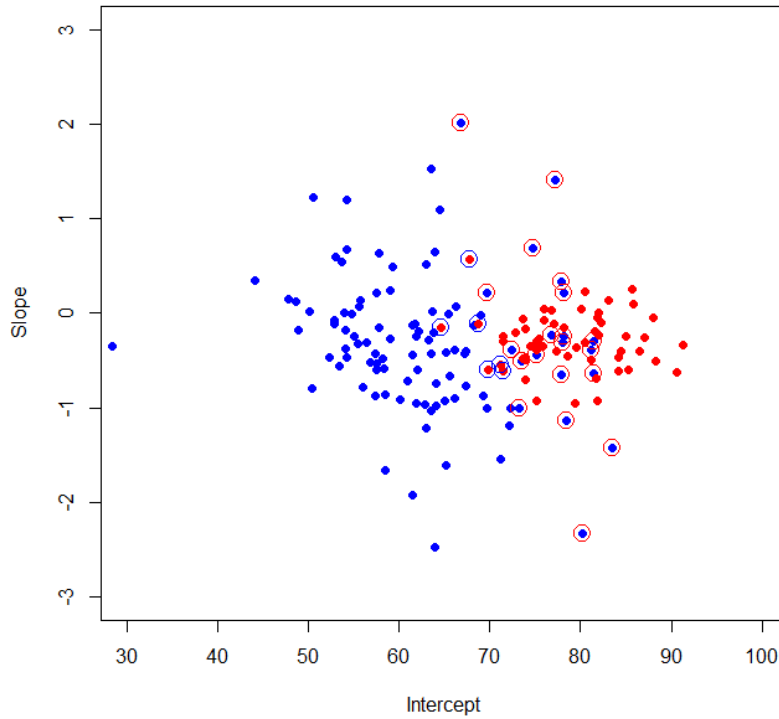
The cross-classification matrix and the misclassification rates of linear and quadratic discriminant functions are summarized in [Table 2](#). Quadratic discriminant functions perform best. The misclassification rates are 10.7% for GCL and 6.9% for RNFL. It is the baseline that helps classify subjects. Quadratic discriminant classification on just the baseline leads to misclassification rates of 9.4% (for both GCL and RNFL), while the misclassification rates when using just slopes are 48.8% (for both GCL and RNFL). For this particular group of treated glaucoma subjects and the group of the age-matched normal subjects, it is the baseline that best classifies the subjects, and not the slope.

The estimated slopes cannot distinguish the two groups. This is because the average slopes are not significantly different (see results in [Table 1](#)), and there is no separation between the two distributions. However, the clinician may not be interested in such classification. Instead, he may be interested in identifying known glaucoma subjects with a reduction in either RNFL or GCL (progressive worsening in retinal structure) that is larger than what is expected for the normal group. Results in [Table 1](#) show that the standard deviation of the slopes in the glaucoma group is about twice as large as the standard deviation of the slopes in the normal group. A rule that uses the 10th percentile of the normal slope distribution as the cutoff will certainly identify more than 10% of glaucoma patients with slopes smaller than that cutoff (negative slope, which then can be targeted for treatment).

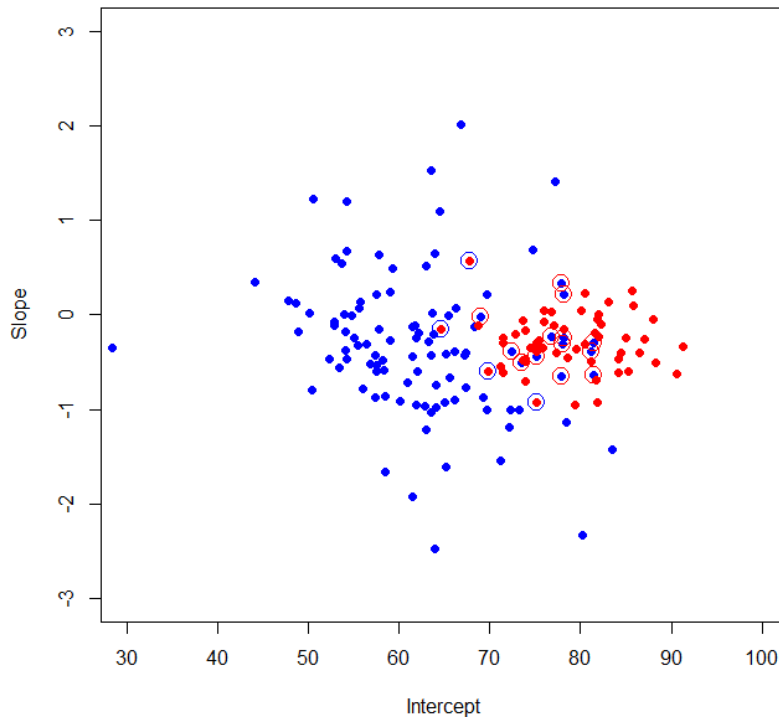
## Discussion

Clinicians base decisions on short records of consecutive measurements. This paper discusses

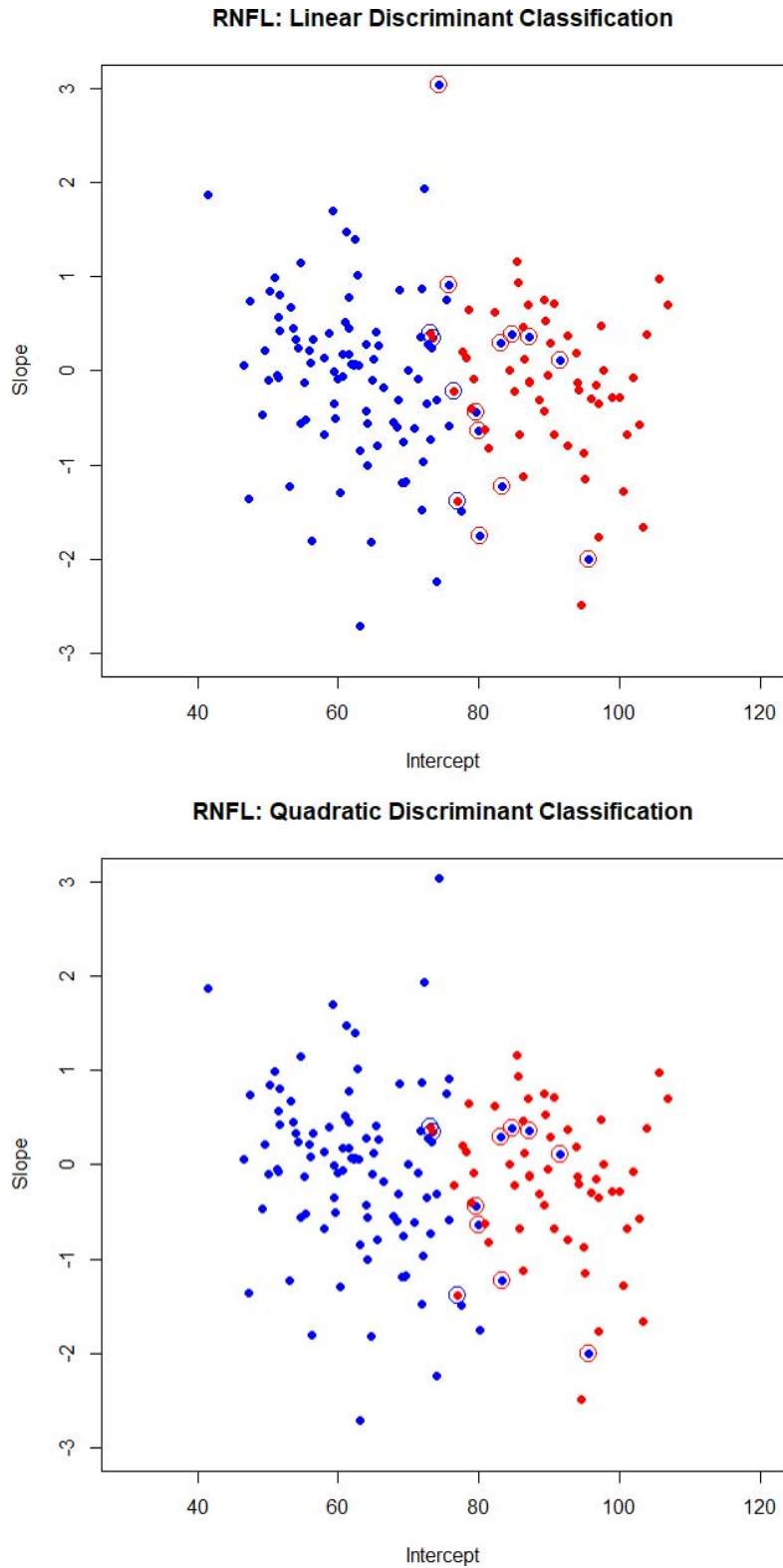
**GCL: Linear Discriminant Classification**



**GCL: Quadratic Discriminant Classification**



**Figure 1.** Scatter plots of intercepts and slopes for GCL. Normal subjects are shown as *red dots* and glaucoma subjects are shown as *blue dots*. Subjects that are misclassified by the linear (quadratic) discriminant function are displayed with an *open circle* in a color that reflects the incorrect classification. A *red dot with a blue circle* indicates a normal subject incorrectly classified as coming from the glaucoma group. A *blue dot with a red circle* indicates a glaucoma patient incorrectly classified as coming from the normal group.



**Figure 2.** Scatter plots of intercepts and slopes for RNFL. Normal subjects are shown as *red dots* and glaucoma subjects are shown as *blue dots*. Subjects that are misclassified by the linear (quadratic) discriminant function are displayed with an *open circle* in a color that reflects the incorrect classification. A *red dot* with a *blue circle* indicates a normal subject incorrectly classified as coming from the glaucoma group. A *blue dot* with a *red circle* indicates a glaucoma patient incorrectly classified as coming from the normal group.



**Table 2.** Cross-Classification Matrix and Misclassification Rates of Linear and Quadratic Discriminant Functions

	Classified as Normal	Classified as Glaucoma
<b>GCL</b>		
Linear discriminant function		
Normal	49	6
Glaucoma	20	84
Overall misclassification rate: $26/159 = 0.164$		
Quadratic discriminant function		
Normal	51	4
Glaucoma	13	91
Overall misclassification rate: $17/159 = 0.107$		
<b>RNFL</b>		
Linear discriminant function		
Normal	51	4
Glaucoma	13	91
Overall misclassification rate: $17/159 = 0.107$		
Quadratic discriminant function		
Normal	52	3
Glaucoma	8	96
Overall misclassification rate: $11/159 = 0.069$		

useful methods for analyzing such data. Random effects trend models have been applied previously to clinical data sets where progression of disease is the focus, such as glaucoma.<sup>17–22</sup> Model results assess whether there exist significant trends and whether average trends are different across groups.

The paper also discusses whether clinical measures, such as the baseline and the change of the OCT thickness of the RNFL and the retinal GCL, classify patients reliably into disease groups, given their variability. Interestingly, the glaucoma data set from the Iowa glaucoma study used as an example here did not show a significant change (slope) in structure over time compared to the normal group. This was not unexpected, since the glaucoma patients enrolled were already on maximal treatment and were being carefully monitored in the glaucoma clinic over the study period. An untreated (or not adequately treated sample) would be expected to have very different results, showing more subjects with a significant, negative slope.

In this study, it was also found that more than 10% of the glaucoma patients enrolled showed slopes larger than the 90th percentile of the normal slope distribution (positive slope, or increasing retinal layer thickness over time). This was an unexpected result,

since the thickness of the inner retinal layers is usually stable with glaucoma treatment or becomes less over time in the case of progression and is not traditionally expected to increase in thickness over time in glaucoma. However, this concept may need to be reconsidered. Although measurement variability of retinal layer thickness can occur due to factors that influence segmentation of the retinal layers by software (e.g., signal strength, blood vessels), such factors would not fully explain trend changes resulting in systematic increases in the retinal layer thickness over time (resulting in an increase in slope), which was greater in the glaucoma group compared to age-matched normal subjects. The retinal layers are not static; the retinal GCL complex contains the cell bodies of retinal ganglion cells and dendritic connections. Structural changes associated with neuroplasticity in response to the glaucomatous process may include dendritic sprouting and changes in cytoplasm content and could conceivably cause an increase in thickness.<sup>12</sup> Similarly, the RNFL contains the axons of the retinal ganglion cells, and their axoplasmic contents may be influenced by the characteristics of axoplasmic flow and status of the axon,<sup>12–16</sup> which in turn may cause dynamic changes in the thickness of this layer. The present analysis raises important questions about the increase in variance of the distribution of slopes of the GCL and RNFL in glaucoma patients compared to normal subjects. Specifically, future investigation is warranted to better understand the reason why some subjects with glaucoma show an increase in thickness of the layers of the inner retina over time that the normal group does not.

The statistical approach presented in this study could also be applied to visual field metrics collected at baseline and monitored over time. The example presented here for classification of individual subjects into the diseased group (glaucoma in this case) or normal group could be applied to visual field data as well as the combination of functional and structural data features for a variety of disorders in addition to glaucoma. A similar approach could also be used to classify patients into a progression versus a non-progression group. With ever more available data, classification plays a central role of personalized medicine. Classification requires more than significant differences among average trends across groups as the subject variability in the trend progression determines whether the classification will be successful.

## Acknowledgments

Michael Wall, MD, kindly allowed us to make use of the OCT imaging data for this manuscript that was collected as part of his Variability in Perimetry (VIP) study, funded by the Department of Veterans Affairs Rehabilitation, Research, and Development Division.

Randy Kardon is supported by grants from the Department of Veterans Affairs Rehabilitation Research & Development (RR&D) Division; Center for the Prevention and Treatment of Visual Loss, C9251-C, Chronic Effects of Neurotrauma Consortium (CENC); DOD, CENC0056P; C1786-R; 1 IO1 RX002101 VA-ORD; 2 IO1 RX000889-05A2; National Eye Institute (NEI).

Disclosure: **J. Ledolter**, None; **R.H. Kardon**, None

## References

1. Abraham B, Ledolter J. *Introduction to Regression Modeling*. Belmont, CA: Duxbury Press; 2006.
2. Searle S. *Linear Models*. New York, NY: John Wiley and Sons; 1971.
3. Searle S, Casella G, McCulloch C. *Variance Components*. New York, NY: John Wiley and Sons; 1992.
4. SAS Statistical Software (Version 9.4). Cary, NC: SAS Institute, Inc.; <https://support.sas.com/en/support-home.html>. Accessed October 14, 2018.
5. The R Project for Statistical Computing. <https://www.r-project.org/>. Accessed October 14, 2018.
6. Verbeke G, Mohlenberghs G. *Linear Mixed Models in Practice: A SAS-Oriented Approach*. New York, NY: Springer; 1997.
7. Fisher RA. The use of multiple measurements in taxonomic problems. *Annals Eugenics*. 1936;7:179–188.
8. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis*. 6th ed. Essex, UK: Pearson; 2012.
9. Ledolter J. *Data Mining and Business Analytics With R*. New York, NY: John Wiley and Sons; 2013.
10. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY: ACM Press. 1992;144–152.
11. Christianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press; 2000.
12. Fry LE, Fahy E, Chrysostomou V, et al. The coma in glaucoma: retinal ganglion cell dysfunction and recovery. *Prog Retin Eye Res*. 2018;65:77–92.
13. Abbott CJ, Choe TE, Lusardi TA, Burgoyne CF, Wang L, Fortune B. Evaluation of retinal nerve fiber layer thickness and axonal transport 1 and 2 weeks after 8 hours of acute intraocular pressure elevation in rats. *Invest Ophthalmol Vis Sci*. 2014;55:674–687.
14. Fortune B, Burgoyne CF, Cull G, Reynaud J, Wang L. Onset and progression of peripapillary retinal nerve fiber layer (RNFL) retardance changes occur earlier than RNFL thickness changes in experimental glaucoma. *Invest Ophthalmol Vis Sci*. 2013;54:5653–5661.
15. O'Leary N, Artes PH, Hutchison DM, Nicolela MT, Chauhan BC. Rates of retinal nerve fibre layer thickness change in glaucoma patients and control subjects. *Eye (Lond)*. 2012;26:1554–1562.
16. Grewal DS, Sehi M, Paauw JD, Greenfield DS. Advanced imaging in glaucoma study group. Detection of progressive retinal nerve fiber layer thickness loss with optical coherence tomography using 4 criteria for functional progression. *J Glaucoma*. 2012;21:214–220.
17. Zhang P, Luo D, Li P, Sharpsten L, Medeiros FA. Log-gamma linear-mixed effects models for multiple outcomes with application to a longitudinal glaucoma study. *Biom J*. 2015;57:766–776.
18. 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.
19. Medeiros FA, Zangwill LM, Alencar LM, et al. Detection of glaucoma progression with stratus OCT retinal nerve fiber layer, optic nerve head, and macular thickness measurements. *Invest Ophthalmol Vis Sci*. 2009;50:5741–5748.
20. Liu T, Tatham AJ, Gracitelli CP, Zangwill LM, Weinreb RN, Medeiros FA. Rates of retinal nerve fiber layer loss in contralateral eyes of glaucoma patients with unilateral progression by conventional methods. *Ophthalmology*. 2015;122:2243–2251.
21. Lee J, Sohn SW, Kee C. Effect of Ginkgo biloba extract on visual field progression in normal tension glaucoma. *J Glaucoma*. 2013;22:780–784.
22. Schuman JS, Pedut-Kloizman T, Pakter H, et al. Optical coherence tomography and histologic measurements of nerve fiber layer thickness in normal and glaucomatous monkey eyes. *Invest Ophthalmol Vis Sci*. 2007;48:3645–3654.