ANALYSIS WITH SUPPORT VECTOR MACHINE SHOWS HIV-POSITIVE SUBJECTS WITHOUT INFECTIOUS RETINITIS HAVE mfERG DEFICIENCIES COMPARED TO NORMAL EYES

BY **Michael H. Goldbaum MD MS**,* Irina Falkenstein MD, Igor Kozak MD, Jiucang Hao PhD, Dirk-Uwe Bartsch PhD, Terrance Sejnowski PhD, AND William R. Freeman MD

ABSTRACT

Purpose: To test the following hypotheses: (1) eyes from individuals with human immunodeficiency virus (HIV) have electrophysiologic abnormalities that manifest as multifocal electroretinogram (mfERG) abnormalities; (2) the retinal effects of HIV in immune-competent HIV individuals differ from the effects in immune-incompetent HIV individuals; (3) strong machine learning classifiers (MLCs), like support vector machine (SVM), can learn to use mfERG abnormalities in the second-order kernel (SOK) to distinguish HIV from normal eyes; and (4) the mfERG abnormalities fall into patterns that can be discerned by MLCs. We applied a supervised MLC, SVM, to determine if mfERGs in eyes from patients with HIV differ from mfERGs in HIV-negative controls.

Methods: Ninety-nine HIV-positive patients without visible retinopathy were divided into 2 groups: (1) 59 high-CD4 individuals (H, 104 eyes), 48.5 ± 7.7 years, whose CD4 counts were never observed below 100, and (2) 40 low-CD4 individuals (L, 61 eyes), 46.2 ± 5.6 years, whose CD4 counts were below 100 for at least 6 months. The normal group (N, 82 eyes) had 41 age-matched HIV-negative individuals, 46.8 ± 6.2 years. The amplitude and latency of the first positive curve (P1, hereafter referred to as *a*) and the first negative curve (N1, referred to as *b*) in the SOK of 103 hexagon patterns of the central 28° of the retina were recorded from the eyes in each group. SVM was trained and tested with cross-validation to distinguish H from N and L from N. SOK was chosen as a presumed detector of inner retinal abnormalities. Classifier performance was measured with the area under the receiver operating characteristic (AUROC) curve to permit comparison of MLCs. Improvement in performance and identification of subsets of the most important features were sought with feature selection by backward elimination.

Results: In general, the SOK b-parameters separated L from N and H from N better than a-parameters, and latency separated L from N and H from N better than amplitude. In the HIV groups, on average, amplitude was diminished and latency was extended. The parameter that most consistently separated L from N and H from N was b-latency. With b-latency, SVM learned to distinguish L from N (AUROC = 0.7.30 ± 0.044, $P = .001$ against chance [0.500 ± 0.051]) and H from N (0.732 ± 0.038, $P = .0001$ against chance) equally well. With best-performing subsets (21 out of 103 hexagons) derived by backward elimination, SVM distinguished L from N (0.869 ± 0.030, $P < .00005$ against chance) and H from N (0.859 ± 0.029, $P < .00005$ against chance) better than SVM with the full set of hexagons. Mapping the top 10 hexagon locations for L vs N and H vs N produced no apparent pattern.

Conclusions: This study confirms that mfERG SOK abnormalities develop in the retina of HIV-positive individuals. The new finding of equal severity of b-latency abnormalities in the low- and high-CD4 groups indicates that good immune status under highly active antiretroviral therapy may not protect against retinal damage and, by extension, damage elsewhere. SOKs are difficult for human experts to interpret. Machine learning classifiers, such as SVM, learn from the data without human intervention, reducing the need to rely on human skills to interpret this test.

Trans Am Ophthalmol Soc 2008;106:196-205

INTRODUCTION

Several studies demonstrated retinal damage in the era before highly active antiretroviral therapy (HAART) became available to treat patients with AIDS due to human immunodeficiency virus (HIV). These pre-HAART studies reported both structural[1,2] and functional[3-8] damage even in eyes without infectious retinitis. With the improvement in the health of HIV-positive patients brought about by HAART, it has become important to determine whether retinal damage continues to occur under antiretroviral treatment.

HIV retinopathy is a microangiopathy that presents with cotton-wool spots, intraretinal hemorrhages, and capillary nonperfusion. In the HAART era HIV retinopathy has been seen in higher proportion than HIV-related retinal infections, most likely because of a decreased incidence of the latter. The improvement in the immune system of patients treated with HAART has permitted the grouping of HIV patients into those who have had extended periods with low CD4 counts (low-CD4 group) and those who have never been observed to have low CD4 counts (high-CD4 group). It has been hypothesized that eyes in the high-CD4 group are less likely to be damaged from the HIV infection or secondary infections because the immune system approaches the competency of people without HIV infection, and the eyes are thus protected by the near-normal immune status.[9,10]

Several studies in the HAART era have also investigated the anatomy and function of noninfectious eyes in HIV-positive individuals. They again demonstrated both structural[9,11-13] and functional[10,14,15] changes, especially in patients with a prolonged low immune status evidenced by low CD4 counts.[9,11]

The multifocal electroretinogram (mfERG) was developed for the simultaneous detection of the electrical activity from specific regions of the retina.[16-18] The result is a map of mfERG parameters in the central retina. Diseases that affect the retina may cause focal damage that is diffuse or that may occur in patterns.[19-25] The mfERG has been found to be a sensitive detector of damage in the outer retina and the inner retina in diseases that can affect the retina.[20,21] The first-order kernel (FOK) is a response in a small hexagonal

From the Ophthalmic Informatics Laboratory (Dr Goldbaum), Joan and Irwin Jacobs Retina Center (Dr Falkenstein, Dr Kozak, Dr Bartsch, Dr Freeman), and Institute for Neural Computation (Dr Hao, Dr Sejnowski), University of California at San Diego.

*Presenter.
Bold type indicates AOS member.

region of the retina to a dark-to-light or light-to-dark stimulus. The response in the FOK is from the receptor cells and the "on" or "off" bipolar cells in each hexagon.[16,18] The second-order kernel (SOK) is a nonlinear adaptive response of the retina to a preceding flash.[16,26] Half the stimuli are dark or light single-flash FOKs, and half the stimuli occur after a preflash. Since the FOK primarily represents activity from the outer retina, subtracting the FOK from the preflashed stimulus generates a signal that diminishes the contribution from the outer retina and diffuses the origin of the signal. There is controversy about the source of the SOK,[16] but damage to the inner retina from diseases such as diabetes and glaucoma has been detected with abnormal SOKs.[27,28]

We have previously performed mfERG studies in noninfectious eyes of HIV-positive patients. In a study using the FOK, we found outer retina sparing.[29] In a separate study, we investigated the inner retina with the SOK. This study analyzed the groups with linear classifiers. We found mfERG abnormalities in HIV-positive patients, more prominent in those HIV subjects with prolonged decreased immunity.[15] The SOK changes discovered in this population and using these techniques can be subtle, and the deficits are difficult to detect by human observers, including the experts. Without human expertise or intervention, machine learning classifiers (MLCs) can learn from the data. For that reason we applied support vector machine (SVM), a more powerful method of discriminating data than the previously used linear classifiers.

The human brain is organized to use pattern recognition, yet there is a limit to the number of features humans can consider simultaneously. The simultaneous consideration of the values in 103 mfERG hexagons is an example of an overwhelming pattern recognition problem for humans. Classification is a form of pattern recognition that takes features as input and delivers a decision about class membership (eg, either HIV-affected eyes or normal eyes). Learning classifiers learn from examples how to determine class membership. Machine learning classifiers use computers to accomplish that task. Some data can be adequately separated with a linear classifier, such as multivariate analysis or linear discriminate function. Other data are poorly separated when constrained to the assumptions of linear classification. Support vector machine is a MLC that is designed to adapt to the data.[30,31] With many types of data, it has the potential to separate the data into classes much better (fewer false positives and false negatives) than linear classifiers.

Machine learning classifiers can surpass humans in simultaneously considering a large number of features to make a decision about the class membership of a sample. Machine learning classifiers have been successfully applied to separate normal from affected eyes in glaucoma and HIV disease.[10,15,32-36] For example, MLCs have learned to diagnose glaucoma from standard automated perimetry locations,[36] meridians of peripapillary retinal nerve fiber layer thickness,[37] and parameters of optic nerve topography.[35] Machine learning classifiers have also learned to distinguished HIV eyes from normal eyes with automated perimetry.[10]

This study was designed to use SVM, a strong MLC, to analyze the response of mfERG in eyes from subjects with HIV without observed infectious retinitis, in comparison to normal eyes. The hypotheses tested were as follows: (1) eyes from HIV individuals have electrophysiologic abnormalities that manifest as mfERG abnormalities; (2) the retinal effects of HIV in immune-competent HIV individuals differ from the effect in immune-incompetent HIV individuals; (3) strong MLCs, like SVM, can learn to use subtle mfERG abnormalities to distinguish HIV from normal eyes; and (4) the mfERG abnormalities fall into patterns that can be discerned by MLCs.

METHODS

This study has IRB approval, the participants gave consent for the participation in the study, and the data were managed to fit HIPAA guidelines.

SUBJECTS

There were 165 eyes in 99 HIV-positive subjects and 82 eyes in 41 normal individuals (Table 1). The HIV-positive subjects were part of a cohort of HIV-positive patients seen at the Jacobs Retina Center in the HAART era, from May 5, 2005, to December 31, 2006. The normal individuals volunteered to participate.

The HIV-positive subjects were divided into 2 groups based on the CD4 count. The high-CD4 count group (H) members never were observed to have a CD4 count less than 100 cells/mm^3. From this group, 104 eyes from 59 persons were studied. The low-CD4 count group (L) members were observed to have CD4 counts less than 100 cells/mm^3 for a minimum of 6 months. In this group, 61 eyes from 40 patients were studied.

The exclusion criteria included a history of infectious retinitis, such as cytomegalovirus (CMV) retinitis; visible retinal abnormalities on ocular fundus by indirect ophthalmoscopy or slit-lamp ophthalmoscopy with a 90 diopter (D) lens; Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity less than 20/40; intraocular pressure 22 mm Hg or higher; spherical equivalent refractive error below –5 D or above +2.5 D; and concurrent disease that could cause retinal damage, such as diabetes or glaucoma. Of the 165 eyes in HIV-positive subjects, only 4 had noninfectious cotton-wool spots; these eyes were included.

MULTIFOCAL ELECTRORETINOGRAPHY

The pupils were dilated with tropicamide 1% and phenylephrine hydrochloride 2.5%. Bipolar corneal electrodes have been reported as yielding a better signal-to-noise ratio and more consistent waveform. Nevertheless, we selected the Dawson-Trick-Litzkow thread electrode (DTL) (Plus Electrode; Diagnosys LLC, Lowell, Massachusetts) because it was more practical to maintain sterility with these electrodes for testing eyes in HIV-positive patients. Following anesthesia with proparacaine 0.5%, the electrode was placed under the lower eyelid, and the other eye was occluded. After the skin was cleansed with an abrasive gel, a gold cup electrode was attached to the forehead for ground, and another was placed at the temple for reference.

The pattern of stimuli was displayed on a monitor with RETIscan, version 3.20.15 (Roland Consult Electrophysiologische Diagnostik Systeme, Wiesbaden, Germany), with a double-flash model to elicit a pronounced SOK response. The pattern subtended

28° at 28-cm viewing distance and was viewed in a darkened room. Presbyopic correction was used as necessary.

TABLE 1. DEMOGRAPHICS OF STUDY PARTICIPANTS

VARIABLE	LOW CD4	HIGH CD4	HIV TOTAL	NORMAL	TOTAL
No. of participants	40	59	99	41	140
No. of eyes	61	104	165	82	247
Mean age (yr) ± SD	46.2 ± 5.6	48.5 ± 7.7	47.5 ± 7	46.8 ± 6.2	47.2 ± 7.1
Men (%)	35 (87.5)	48 (81.4)	83 (83.8)	32 (78.1)	115 (82.1)
Women (%)	5 (12.5)	11 (18.6)	16 (16.2)	9 (21.9)	25 (17.9)
Hispanics (%)	24 (60)	40 (67.8)	64 (64.7)	19 (46.3)	83 (59.3)
Caucasians (%)	13 (32.5)	12 (20)	25 (25.3)	15 (36.5)	40 (28.6)
African Americans (%)	3 (7.5)	6 (10)	9 (9)	5 (12)	14 (10)
Pacific Asians (%)	0 (0)	1 (2.2)	1 (1)	2 (5.2)	3 (2.1)
VA logMAR ± SD (ETDRS equivalent)	0.2 ± 0.08 (20/32)	0.1 ± 0.06 (20/25)	0.18 ± 0.04 (20/30)	0.01 ± 0.01 (20/20)	0.1 ± 0.1 (20/25)
Mean refraction, Dsph ± SD	–0.85 ± 0.64	–1.08 ± 0.52	–1.10 ± 0.51	–0.68 ± 0.40	0.86 ± 0.55

Dsph, spherical equivalent in diopters; ETDRS, Early Treatment Diabetic Retinopathy Study; HIV, human immunodeficiency virus; VA, visual acuity.

The pattern of stimuli was an array of 103 hexagons, scaled to contain an equivalent number of cones. The display sequence was 511×10 ($H_1DDDH_2H_3DDD$), where *H* was a light frame and *D* was a dark frame.[15] The sample interval was 1.1 ms (901 Hz), and the frame frequency was 70 Hz. The FOK was recorded as a response after a single H_1 flash, with a 66.4-ms interval of 4 dark frames, permitting the retina to recover before a double flash from H_2 and an immediate H_3 flash, which was used to construct the SOK response when the retina did not have sufficient time to recover from the preceding flash. Each hexagon was modulated temporally between black (<2 candela [cd]/m^2) and white (200 cd/m^2), with a contrast of 98%, following a pseudorandom binary m-sequence with a base interval of approximately 16.6 ms. Each step of the m-sequence consisted of 5 frames of 83-ms length. To increase the signal-to-noise ratio for amplitudes and latencies, instead of the more common 10 to 300 Hz, the bandpass filters were set between 10 and 100 Hz, because this filtering was reported to make the mfERG more sensitive to inner retinal responses in diabetes.[38,39] Responses were amplified 100,000 times. Occasional artifacts, such a blinks during the recording, were eliminated by the RETIscan software, and that part of the sequence was repeated to record a new response free of artifacts. Averages of responses recorded during 8 cycles were calculated for each subject for the 103 hexagons and were analyzed with the RETIscan software.

A standard m-sequence was recorded over 12 minutes for each eye and was divided into 8 short segments of 89 seconds for patient comfort. The recording protocol was chosen according to the International Society for Clinical Electrophysiology of Vision (ISCEV) guideline for mfERG.[40]

The mean simultaneous response component for the FOK and SOK was recorded. Implicit times (latencies) and response densities and the amplitude relative to their respective areas (nV/deg^2) of the first positive peak (P1) and the first negative peak (N1) were measured for the SOK. The P1 response amplitude was measured from the starting baseline to the peak of P1. The N1 response amplitude was measured from the peak of P1 to the trough of N1. The P1 and N1 implicit times were measured from the onset of the stimulus to the peak of P1 and the trough of N1, respectively. To retain a sense of sequence in the SOK, we substitute *a* for P1 and *b* for N1. The 4 SOK variables studied were a-amplitude, a-latency, b-amplitude, and b-latency.

MACHINE LEARNING CLASSIFIER ANALYSIS

Machine Learning Classifiers

Pattern recognition can use methods of MLCs. Supprt vector machine is a machine classification method that seeks the boundary that best separates the data into 2 classes.[11,13] Support vector machine kernels (different from mfERG kernels) can map the feature space (eg, a 103-dimensional pattern of measurements in 103 mfERG locations) into a high-dimensional transformed space. We selected a gaussian kernel because it surpassed the linear kernel with the mfERG data. The dimensionality of the transformed space with a gaussian kernel is infinite. In such high-dimensional space, a linear hyperplane can be an effective boundary. The samples that are closest to the separation boundary are the most difficult to classify. These samples, called support vectors, determine the location and orientation of the separation boundary. The margin is defined as the distance between the hyperplane and support vectors in the transformed space. The SVM learns by locating and orienting the hyperplane to maximize the margin between the support vectors in the 2 classes. The boundary based on the more-difficult-to-classify examples can separate the 2 classes more effectively than the boundary based on all examples. The hyperplane that is learned in the transformed space maps back to the feature space as a nonlinear separation boundary. Thus SVM learning adapts well to separate the data and often outperforms other classifiers. The support vector method was implemented by using Platt's sequential minimal optimization algorithm in commercial software (MatLab, version 7.0;

MathWorks, Natick, Massachusetts). For classification of the mfERG SOK amplitudes and latencies, gaussian (nonlinear) kernels of various widths were tested, and the chosen gaussian kernel width was the one that gave the highest area under the receiver operating characteristic (AUROC) curve using 10-fold cross-validation.

Training and Testing Machine Learning Classifiers

Cross-validation uses a *k*-partition rotating resampling technique to seek an unbiased evaluation of MLC performance by using test data separate from the training data. This study used 10-fold cross-validation, which randomly split each class into 10 equal subsets. The classifier was trained on a set that combined 9 of the 10 partitions, and the 10th partition served as the testing set. This procedure was performed 10 times, with each subset having a chance to serve as the test set.

Performance Measure of Trained Machine Learning Classifiers

Receiver operating characteristic (ROC) curves displayed the discrimination (sensitivity and specificity) of each classifier as the separation threshold was moved from one end of the data to the other. We used the AUROC as a single measure of classifier performance to compare different classifiers; the larger the AUROC, the better the classifier. A performance equal to chance will have an AUROC = 0.5, whereas the perfect separation of chasses will give an AUROC = 1.0. The method of DeLong and associates[41] tested the null hypothesis (*P* value) for comparing the AUROCs of classifiers.

We trained and tested SVM to distinguish mfERG measurements from normal eyes (N) and eyes in the high-CD4 group (H) and between normal eyes and eyes in the low-CD4 group (L). The 4 measurements were a-amplitude, a-latency, b-amplitude, and b-latency. For each of these measurements, we trained the SVM with a gaussian kernel with the full feature set (SVM full) and a performance-peaking subset of near optimal features derived by a method of backward elimination (SVM back). The AUROC for each of these measurements was compared to the AUROC for chance. To achieve a chance arrangement, we randomly split the normal (N) group and the high (H) or low (L) HIV group evenly and mixed the split N with the split H or the split N with the split L. With this arrangement, we obtained the AUROC for chance (which, as expected, came out to be 0.5), the standard deviation, and the curve correlation necessary for the method of DeLong and associates.[41]

Optimization by Feature Selection

A high ratio of the number of samples (*n*) to the number of features (*f*) reduces the uncertainty of the location of the separation boundary between the classes. The *n*/*f* ratio can be increased either by enlarging the number of samples or by reducing the number of features. We reduced the number of features with feature selection. The feature selection ranked the features and removed the least effective ones. Two prominent methods of feature selection are forward selection and backward elimination. We chose backward elimination to identify small subsets with the best mfERG locations, because our previous experience with automated perimetry data found backward elimination to work better than forward selection.[6] Backward elimination started with the full feature set. Each feature was given the opportunity to be the one removed by determining its effect on the SVM performance. The feature that, when removed, either maximally increased or minimally decreased the SVM performance was removed, and the process was repeated iteratively down to one feature. As the number of features decreased, close to the best feature set was determined by the number of features yielding the peak performance.

RESULTS

Table 1 presents group demographics for ease of comparison. Observation of the age, gender proportions, racial proportions, visual acuity, and refractive error indicates that the low-CD4, high-CD4, and normal groups were similar. Table 2 displays the class means for the amplitudes and latencies, the AUROCs, and the *P* values of selected comparisons of the 4 mfERG parameters, a-amplitude, a-latency, b-amplitude, and b-latency, for L vs N and for H vs N. In general, the amplitudes were reduced and the latencies were prolonged in the HIV classes. Figure 1 enables visual comparison of the ROC curves along their entire lengths. Figure 2 exhibits the location of the top 10 out of 103 hexagons for b-latency.

Of interest first was whether the ROC curves for each of the 4 parameters differed significantly from chance for L vs N and H vs N. Second was how amplitude compared to latency and how a-parameters compared to b-parameters. Third was whether removing ineffective mfERG locations improved the separation of L from N and H from N. Last was whether eyes from presumably healthier high-CD4 patients differed less from normal than eyes from presumably sicker low-CD 4 patients.

First, b-latency was the only parameter that consistently separated L from N ($P = .001$ b-latency against chance) and H from N ($P = .0001$ b-latency against chance). Second, in general, b-parameters were more predictive than a-parameters ($P = .008$ for b-amplitude over a-amplitude and $P = .007$ for b-latency over a-latency in L vs N), and latency was more predictive than amplitude ($P = .030$ for b-latency over b-amplitude and $P = .009$ for a-latency over a-amplitude in H vs N). For this reason, we concentrated on b-latency and b-amplitude. Third, optimization of SVMs by backward elimination for b-latency significantly improved the separation of H from N ($P = .0001$) and L from N ($P < .00005$), compared to training the SVMs with the full feature set. The number of features in the peak set was 21 out of 103 for both H vs N and L vs N. The top 10 out of 103 mfERG locations, displayed in Figure 2, did not form a discernible pattern for either H vs N or L vs N. Last, b-amplitude separated L from N ($P = .003$ against chance), but did not significantly separate H from N ($P = .053$ against chance), whereas b-latency nearly equally separated L from N (AUROC = 0.730 ± 0.044, yielding $P = .001$ against chance) and H from N (AUROC = 0.732 ± 0.038, yielding $P = .0001$ against chance).

TABLE 2. SELECTED CLASS COMPARISONS OF MFERG PARAMETERS

a-amplitude	b-amplitude	a-latency	b-latency	b-lat optimized	chance	
			Low CD4 HIV eyes			
1.84±0.38 nV/deg^2	5.89±0.91 nV/deg^2	18.86±0.64 msec	37.82±0.45 msec			Mean ± SE
			Low CD4 HIV eyes vs Normal eyes			
0.631±0.048	0.703±0.047	0.594±0.050	0.730±0.044	0.869±0.030	0.500±0.051	←AUROC ± SD
						↓*P* value
	.008	**.490**	*.096*	*<.00005*	**.072**	a-amp
		.037	**.65**	*.002*	***.003****	b-amp
			.007	*<.00005*	**.20**	a-latency
				<.00005	***.001****	b-latency
					<.00005*	b-lat opt
			High CD4 HIV eyes			
2.44±0.45 nV/deg^2	8.02±1.0 nV/deg^2	18.51±0.41 msec	37.84±0.34 msec			Mean ± SE
			High CD4 HIV eyes vs Normal eyes			
0.595±0.045	0.617±0.043	0.710±0.040	0.732±0.038	0.859±0.029	0.500±0.045	←AUROC ± SD
						↓*P* value
	.45	***.009***	*.012*	*<.00005*	**.12**	a-amp
		.039	***.030***	*<.00005*	**.053**	b-amp
			.60	*.0005*	***.0004****	a-latency
				.0001*	***.0001****	b-latency
					<.00005*	b-lat opt
			Normal eyes			
2.65±0.38 nV/deg^2	8.30±0.91 nV/deg^2	18.82±0.56 msec	36.53±0.34 msec			Mean ± SE

AUROC, area under the receiver operating characteristic curve; HIV, human immunodeficiency virus; mfERG, multifocal electroretinogram; SD, standard deviation; SE, standard error of the mean. Italic text indicates *P* values <.5. Bold text shows important *P* value comparisons. Nonbold text indicates *P* value comparisons of little interest. Asterisk indicates *P* value comparisons of particular interest with *P* values <.5.

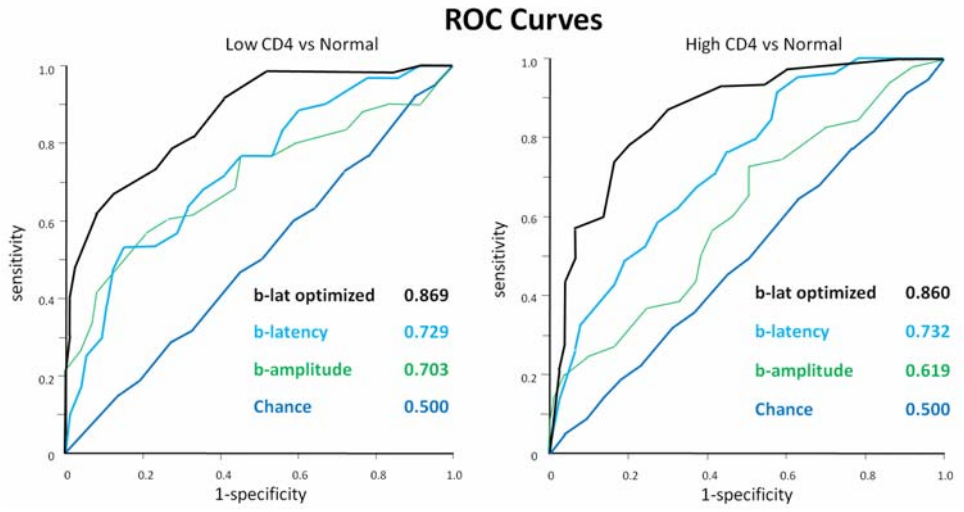

FIGURE 1

Receiver operating characteristic curves. b-lat optimized = b-latency with the support vector machine trained on the top 21 out of 103 hexagons.

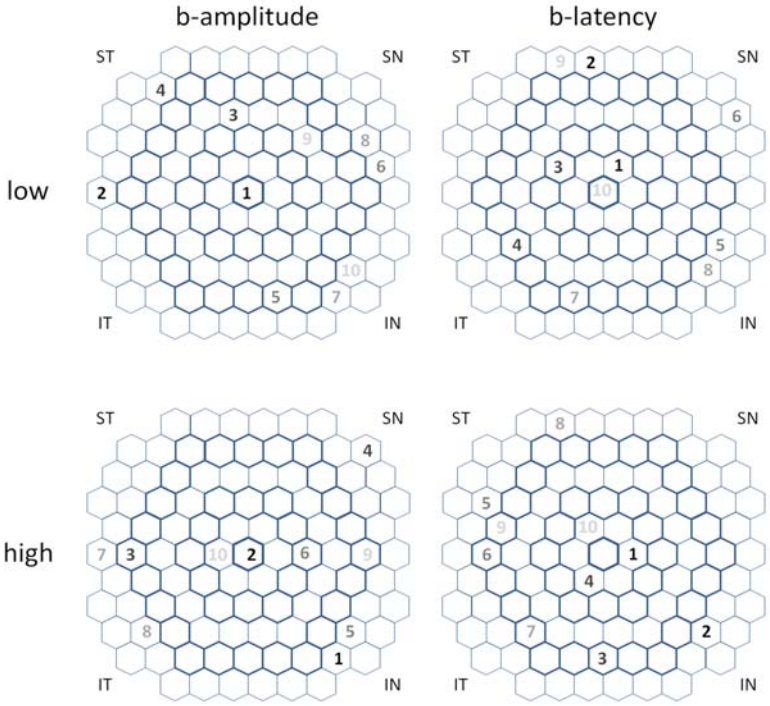

FIGURE 2

Map of the top 10 out of 103 hexagons of b-amplitude and b-latency for low CD4 vs normal and high CD4 vs normal.

DISCUSSION

Previous reports have described defects in visual fields and thinning of the peripapillary retinal nerve fiber layer in the eyes of HIV-positive individuals in the absence of infectious retinitis in the HAART era. Those studies that considered the level of immune competence found that eyes from patients with reduced immune status differed more from normal eyes than eyes from patients who never were observed to have reduced immune status.[9-11] Prolonged reduction of immune status is a risk factor for HIV retinopathy.[1] Retinal trypsin digests of human autopsy eyes with HIV retinopathy without retinitis revealed vascular attenuation, increased ratio of endothelial cells to pericytes, and more microaneurysms compared to non-HIV controls.[42] We believe that cumulative damage in the retinal microvasculature of such individuals can lead to the structural changes observed in this subpopulation in contrast to seropositive individuals with good immune status.

We previously reported that the SOK of mfERG could discern the eyes of HIV-positive subjects from the eyes of individuals who were not HIV-positive. That study, which used linear classification techniques, also revealed that the eyes of patients with prolonged reduction in the CD4 counts differed more from normal eyes than the eyes of patients who never were observed to have reduced CD4 counts.[15] The current study used SVM, which is a strong classifier that is excellent at adapting to the properties of the data. The SVM analysis of the SOK parameters of the mfERG easily separated the eyes from both the low-CD4 and high-CD4 HIV-positive patients from normal eyes; with b-latency, the high-CD4 eyes and the low-CD4 eyes were equally different from normal eyes, a new finding. From this result, several conclusions are possible: (1) mfERG SOK is a sensitive detector of inner retinal abnormalities; (2) SVM is more effective at analyzing mfERG than linear classifiers; and (3) HIV-positive patients who, under HAART, have never been observed to have been compromised immunologically, as a group, have retinal processing abnormalities comparable to patients who have had prolonged immunosuppression.

The mfERG SOK has been effective at revealing retinal abnormalities in the early stage of conditions such as diabetes and glaucoma.[21,27,28,43] In a similar cohort of patients, SVM learned to distinguish between low-CD4 eyes measured with full-threshold standard automated perimetry and normal eyes (P = .0002). Standard automated perimetry analyzed with SVM was barely able to separate high-CD4 eyes from normal (P = .02).[10] Yet, with SOK b-latency, the groups of high-CD4 eyes and low-CD4 eyes were both significantly different from normal eyes (P = .0001 and P = .001, respectively). mfERG appears to be a good method of detecting early or minimal retinal response to disease. The comparison of various MLCs, including SVM, with linear classifiers applied to complex multidimensional visual field,[36] optic disc topography,[35] and retinal nerve fiber layer thickness data[37] has consistently demonstrated the superiority of the MLCs over linear classification methods. So it is not surprising that SVM applied to mfERG b-latency was effective at separating both low-CD4 eyes and high-CD4 eyes from normal eyes. In contrast, linear classification applied to b-latency was less able to separate high-CD4 eyes from normal.[15]

Another value of MLCs is that they learn from the data. No human expertise or input is needed. The SOK of mfERG is a good example of data that are difficult for human interpretation (Figure 3). This study shows the value of MLCs in extracting useful

information from complex data without the need for human skill or intervention.

The separation of groups was more obvious using analysis of b-latency rather than amplitude values. This finding correlates to studies that have shown a similar observation in early diabetic retinopathy, a condition, like HIV, evolving from a retinal microvasculopathy.[43] Fortune and associates[44] used mfERG to find increased implicit times and unchanged amplitudes in diabetic eyes even before retinopathy was visible. Similarly, Klemp and colleagues[45] found abnormalities in implicit times more than in amplitudes in SOK mfERGs in hyperglycemic patients without retinopathy. The finding that b-latency abnormalities were just as prominent in high-CD4 eyes as in low-CD4 eyes is difficult to interpret. A physical defect, such as the thinning of the retinal nerve fiber layer, is easy to conceptualize as damage. The significance of a breakdown in retinal adaptation to a stimulus is that this breakdown may be a sensitive indicator of current damage and damage to come. Alternatively, this measurable abnormality may not affect the quality of life of an individual and may not predict future physical damage. It is unlikely that the b-latency difference between the HIV eyes and the normal eyes is due to a confounding variable. Table 1 shows that the groups are similar in a number of important criteria. Since this study found retinal responses in the high-CD4 group that were just as severe as those in the low-CD4 group, it is possible that high-CD4 patients have retinal damage that will continue to accumulate. The existence of these retinal abnormalities implies that damage has been done elsewhere in the body and that this damage may progress. It is possible that a higher threshold of CD4 count between the low and high classes will find some protective effect of a better immune status. That can be a subject of a future study. The current information, however, indicates that people at risk of HIV infection cannot assume that HAART will protect them from cumulative damage.

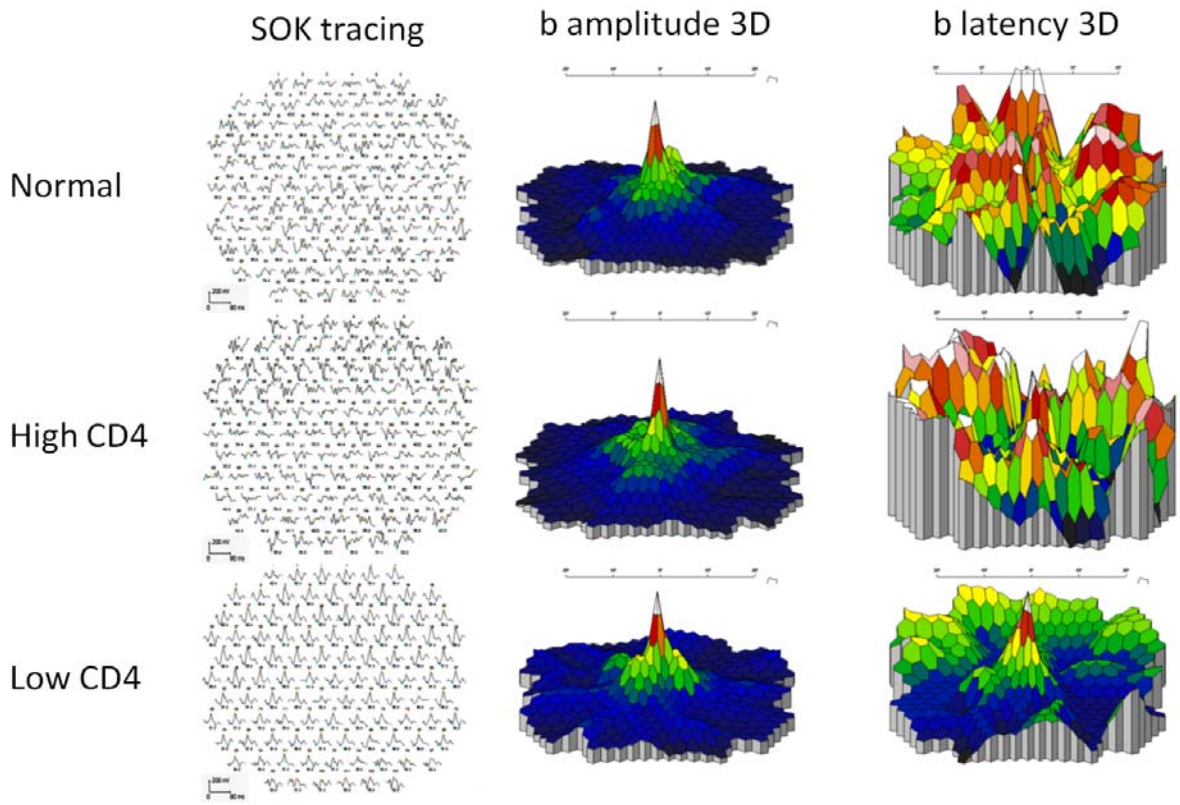

FIGURE 3

Examples of difficult-to-interpret second-order kernel tracings and 3D representations of b-amplitude and b-latency.

We did not detect a pattern in the 10 most predictive hexagon locations for separating either the high-CD4 group or the low-CD4 group from the normal group. One possible explanation is that the retinal abnormalities are small and randomly located in the central 28°. Also likely is that SOK parameters do not localize well and may obscure a real pattern of damage. Whereas the FOK displays the response of localized receptors and bipolar cells, the SOK is an adaptive change to the signal generated by the receptors and the bipolar cells, and the signal from this adaptation may be carried away from the source by horizontal and amacrine cells. Even if the SOK pattern is diffused or undetectable, the SOK analyzed with SVM remains a sensitive detector of retinal malfunction.

In summary, the mfERG is a sensitive detector of the effect of HIV on the retina, even in the high-CD4 group; MLCs can maximize the class information that can be obtained; HIV-positive patients with no history of immune suppression may be damaged

because of the presence of the HIV virus; and there is no observed pattern to the retinal effects from HIV detected by the SOK of mfERG.

ACKNOWLEDGMENTS

Funding/Support: Supported by grants EY13938 (M.H.G.) and EY07366 (W.R.F.) from the National Institutes of Health and by the David and Marilyn Dunn Foundation (M.H.G.).

Financial Disclosures: None.

Author Contributions: *Design and conduct of the study* (M.H.G., I.K., W.R.F); *Collection, management, analysis, and interpretation of the data* (M.H.G., I.F., J.H., D.-U.B.); *Preparation, review, or approval of the manuscript* (M.H.G., I.K., T.S., W.R.F.).

Conformity With Author Information: This study has IRB approval, the participants give consent for the participation in the study, and the data are managed to fit HIPAA guidelines.

REFERENCES

1. Freeman WR, Chen A, Henderly DE, et al. Prevalence and significance of acquired immunodeficiency syndrome-related retinal microvasculopathy. *Am J Ophthalmol* 1989;107:229-235.
2. Tenhula WN, Xu SZ, Madigan MC, et al. Morphometric comparisons of optic nerve axon loss in acquired immunodeficiency syndrome. *Am J Ophthalmol* 1992;113:14-20.
3. Plummer DJ, Sample PA, Arevalo JF, et al. Visual field loss in HIV-positive patients without infectious retinopathy. *Am J Ophthalmol* 1996;122:542-549.
4. Geier SA, Nohmeier C, Lachenmayr BJ, Klauss V, Goebel FD. Deficits in perimetric performance in patients with symptomatic human immunodeficiency virus infection or acquired immunodeficiency syndrome. *Am J Ophthalmol* 1995;119:335-344.
5. Geier SA, Kronawitter U, Bogner JR, et al. Impairment of color contrast sensitivity and neuroretinal dysfunction in patients with symptomatic HIV infection or AIDS. *Br J Ophthalmol* 1993;77:716-720.
6. Geier SA, Hammel G, Bogner JR, Kronawitter U, Berninger T, Goebel FD. HIV-related ocular microangiopathic syndrome and color contrast sensitivity. *Invest Ophthalmol Vis Sci* 1994;35:3011-3021.
7. Iragui VJ, Kalmijn J, Plummer DJ, Sample PA, Trick GL, Freeman WR. Pattern electroretinograms and visual evoked potentials in HIV infection: evidence of asymptomatic retinal and postretinal impairment in the absence of infectious retinopathy. *Neurology* 1996;47:1452-1456.
8. Sample PA, Plummer DJ, Mueller AJ, et al. Pattern of early visual field loss in HIV-infected patients. *Arch Ophthalmol* 1999;117:755-760.
9. Kozak I, Bartsch DU, Cheng L, McCatchan A, Weinreb RN, Freeman WR. Scanning laser polarimetry demonstrates retinal nerve fiber layer damage in human immunodeficiency virus positive patients without infectious retinitis. *Retina* 2007;27:1267-1273.
10. Kozak I, Sample PA, Hao J, et al. Machine learning classifiers are able to detect subtle field defects in eyes of HIV subjects. *Trans Am Ophthalmol Soc* 2007;105:111-120.
11. Kozak I, Bartsch DU, Cheng L, et al. Objective analysis of retinal damage in HIV-positive patients in the HAART era using OCT. *Am J Ophthalmol* 2005;139:295-301.
12. Kozak I, Bartsch DU, Cheng L, Freeman WR. In vivo histology of cotton-wool spots using high-resolution optical coherence tomography. *Am J Ophthalmol* 2006;141:748-750.
13. Kozak I, Bartsch DU, Cheng L, Freeman WR. Hyper-reflective sign in resolved cotton wool spots using high resolution optical coherence tomography and OCT ophthalmoscope. *Ophthalmology* 2007;114:537-543.
14. Shah KH, Holland GN, Yu F, Van Natta M, Nusinowitz S; Studies of Ocular Complications of AIDS (SOCA) Research Group. Contrast sensitivity and color vision in HIV-infected individuals without infectious retinopathy. *Am J Ophthalmol* 2006;142:284-292.
15. Falkenstein IA, Bartsch DU, Azen SP, Dustin L, Sadun AA, Freeman WR. Multifocal electroretinography in HIV-positive patients without infectious retinitis. *Am J Ophthalmol* 2008;146:579-588.
16. Hood DC. Assessing retinal function with multifocal technique. *Prog Retin Eye Res* 2000;19:607-646.
17. Sutter EE, Tran D. The field topography of ERG components in man—I. The photopic luminance response. *Vision Res* 1992;32:433-446.
18. Lai TYY, Chan W-M, Lai RY, Ngai JW, Li H, Lam DS. The clinical applications of multifocal electroretinography: a systematic review. *Surv Ophthalmol* 2007;52:61-96.
19. Kurtenbach A, Leo-Kottler B, Zrenner E. Inner retina contributions to the multifocal electroretinogram: patients with Leber's hereditary optic neuropathy (LHON). Multifocal ERG in patients with LHON. *Doc Ophthalmol* 2004;108:231-240.
20. Greenstein VC, Holopigian K, Hood DC, Seiple W, Carr RE. The nature and extent of retinal dysfunction associated with diabetic macular edema. *Invest Ophthalmol Vis Sci* 2000;41:3643-3654.
21. Bearse MA, Adams AJ, Han Y, et al. A multifocal electroretinogram model predicting the development of diabetic retinopathy. *Prog Retin Eye Res* 2006;25:425-448.

22. Feigl B, Brown B, Lovie-Kitchin J, Swann P. The rod-mediated multifocal electroretinogram in aging and in early age-related maculopathy. *Curr Eye Res* 2006;31:634-644.
23. Holopigian K, Seiple W, Greenstrein VC, Hood DC, Carr RE. Local cone and rod system function in patients with retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 2001;42:779-788.
24. Luu CD, Lau AM, Lee SY. Multifocal electroretinogram in adults and children with myopia. *Arch Ophthalmol* 2006;124:328-334.
25. Lai TY, Chan WM, Li H, Lai RY, Lam DS. Multifocal electroretinographic changes in patients receiving hydroxychloroquine therapy. *Am J Ophthalmol* 2005;140:794-807.
26. Hood DC, Greenstein V, Frishman L, et al. Identifying inner retinal contributions to the human multifocal ERG. *Vision Res* 1999;39:2285-2291.
27. Chan HL, Brown B. Multifocal ERG changes in glaucoma. *Ophthalmic Physiol Opt* 1999;19:306-316.
28. Palmowski AM, Sutter EE, Bearse MA, Fung W. Mapping of retinal function in diabetic retinopathy using the multifocal electroretinogram. *Invest Ophthalmol Vis Sci* 1997;38:2586-2596.
29. Falkenstein I, Kozak I, Kayikcioglu O, et al. Assessment of retinal function in patients with HIV without infectious retinitis by multifocal electroretinogram and automated perimetry. *Retina* 2006;26:928-934.
30. Vapnik V. *Statistical Learning Theory*. New York: Wiley; 1998.
31. Vapnik VN. *The Nature of Statistical Learning Theory*. 2nd ed. New York: Springer; 2000.
32. Goldbaum MH, Sample PA, White H, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci* 1994;35:3362-3373.
33. Sample PA, Boden C, Zhang Z, et al. Unsupervised machine learning with independent component analysis to identify areas of progression in glaucomatous visual fields. *Invest Ophthalmol Vis Sci* 2005;46:3684-3692.
34. Brigatti L, Hoffman BA, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol* 1996;121:511-521.
35. Bowd C, Chan K, Zangwill LM, et al. Comparing neural network and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci* 2002;43:3444-3454.
36. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci* 2002;43:162-169.
37. Zangwill LM, Chan K, Bowd C, et al. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci* 2004;45:3144-3151.
38. Han Y, Bearse MA, Schneck ME, Barez S, Jacobsen C, Adams AJ. Towards optimal filtering of "standard" multifocal electroretinogram (mfERG) recordings: findings in normal and diabetic subjects. *Br J Ophthalmol* 2004;88:543-550.
39. Fortune B, Schneck ME, Adams AJ. Multifocal electroretinogram delays reveal local retinal dysfunction in early diabetic retinopathy. *Invest Ophthalmol Vis Sci* 1999;40:2638-2651.
40. Marmor MF, Hood DEC, Keating D, Kondo M, Seeliger MW, Miyake Y; International Society for Clinical Electrophysiology of Vision. Guidelines for basic multifocal electroretinography (mfERG). *Doc Ophthalmol* 2003;106:105-115.
41. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-845.
42. Glasgow BJ, Weisberger AK. A quantitative and cartographic study of retinal microvasculopathy in acquired immunodeficiency syndrome. *Am J Ophthalmol* 1994;118:46-56.
43. Mita-Harris M. Changes in the second-order kernel component obtained by the techniques of the multifocal electroretinogram in early stages of diabetes mellitus. *Nippon Ganka Gakkai Zasshi* 2001;105:470-477.
44. Fortune B, Schneck ME, Adams AJ. Multifocal electroretinogram delays reveal local retinal dysfunction in early diabetic retinopathy. *Invest Ophthalmol Vis Sci* 1999;40:2638-2651.
45. Klemp K, Larsen M, Sander B, Vaag A, Brockhoff PB, Lund-Andersen H. Effect of short-term hyperglycemia on multifocal electroretinogram in diabetic patients without retinopathy. *Invest Ophthalmol Vis Sci* 2004;45:3812-3819.

PEER DISCUSSION

DR. LEONARD A. LEVIN: Goldbaum and colleagues used a powerful supervised learning technique to distinguish multifocal ERG (mfERG) findings of human immunodeficiency- (HIV-) positive patients from normal subjects. They also differentiated HIV-positive patients that had low CD4 counts from those with high CD4 counts. They concentrated on the second-order kernel, which likely has more contributions from the inner retina than the first order kernel, which is mostly generated in the outer retina, and which this group has previously shown is maintained in HIV-positive individuals.

The supervised learning technique they used is called a support vector machine (SVM), and is based on the concept that groups can be classified by separating them in a multidimensional space with a multidimensional plane. The multiple dimensions correspond to the multiple parameters associated with each member of a group, in this case the amplitudes and latencies of mfERG components. Determining the plane that separates the groups is done by computer, using data from patients. They then evaluated the quality of the classification by testing it on 10% of the patients that had not been used to generate the plane. This was done 10 times, with a different group of training and testing patients for each trial.

What did they find? The mfERG processed by a SVM was able to separate HIV-positive from presumably HIV-negative subjects, at a level much better than chance. They do not provide specificity and sensitivity values for "interesting" points along the ROC curve,

so it is difficult to compare the SVM directly with traditional methods (e.g. using a criterion of a mean b-wave latency one standard deviation above the mean of normal subjects). Nonetheless, it is clear that this technique was able to do a fairly good (but not stellar job) of differentiating the HIV subjects from normal controls.

What does this mean? By itself, these data do not implicate the inner retina as being different in patients with HIV. Although that is a natural assumption, based on the fact that the second order kernel was studied, it is not proven. Moreover, it is possible that other confounding variables could be responsible for the changes in the mfERG that led to the SVM differentiating the groups. Examples include drug treatment, other infections more commonly seen in HIV-positive patients, or even factors related to susceptibility to HIV infection. It also is unclear how well the low CD4 count patients can be differentiated from the high CD4 count patients. Future work could see if the hyperplane that (presumably) separates those groups is parallel to the plane that separates normal from HIV patients. Nonetheless, it is clear that this is an excellent example of how advances in machine learning techniques can be applied to difficult problems in classification of ophthalmic disease states.

ACKNOWLEDGMENTS

Funding/Support: None

Financial Disclosures: None.

DR. MICHAEL H. GOLDBAUM: Thank you for that very nice analysis. I would like now to touch on some of those points. This separation is not perfect, because we are at the extremes of trying to find a difference. We are at the early extreme and are trying to find the smallest defects possible and therefore, the system begins to breaks down. Nevertheless, this method seems to be better than other methods that have been applied, such as multivariate analysis, in discerning a difference between these groups. A great deal of work is contained in the paper, but I did not have an opportunity to present it on the screen. We assessed several parameters to determine if there was a difference between the HIV group and the normal control group. We matched the groups for race and age with several different methods to determine if there was any difference between them, and there were many similarities. There may be something hidden that we do not yet know, but we did make a great effort to assure that our groups were similar in every way possible, except for the HIV and the CD4 counts. We did have limitations in the refraction. Whether there may be other factors that cause changes in the inner retina we do not know. We are not even sure what the b latency findings mean. We do see a difference, but do not know if this means these patients will acquire something noticeable with time, or whether this is something that only we can detect and that the patients will never notice. Trying to find a difference, even extreme differences, gives us some idea of what is going on in the disease process or at least helps us to better understand it, and that was the purpose of the study. We did not use sensitivity and specificity as they are used in the detection of glaucoma, for example, because we tend to focus on the ROC curves at the 90% specificity level where we find the most clinical relevance in that condition. We are not using this test in these HIV-positive patients to make a diagnosis. We are using this test to determine if there is a difference between the eyes, and do not know at what cut off of sensitivity and specificity this would be best achieved. We did not address that issue and evaluated the entire curve. Thank you.