Thomas Lindenberg
Andrea Peters
Folkert K. Horn
Berthold Lausen
Matthias Korth

Diagnostic value of multifocal VEP using cross-validation and noise reduction in glaucoma research

T. Lindenberg (✉) · F. K. Horn · M. Korth
Augenklinik mit Poliklinik,
Friedrich-Alexander-Universität
Erlangen-Nürnberg,
Schwabachanlage 6, 91054 Erlangen,
Germany
e-mail: Tomlin@web.de
Tel.: +49-9131-483431
Fax: +49-9131-483346

A. Peters · B. Lausen
Institut für Medizininformatik, Biometrie
und Epidemiologie,
Friedrich-Alexander-Universität
Erlangen-Nürnberg,
Waldstrasse 6, 91054 Erlangen, Germany

Abstract *Background:* The diagnostic value of multifocal visual evoked potentials (mf VEP) in glaucoma research is still under debate. Several previous studies proclaim it to be a useful tool for clinical applications, but according to other studies, different problems (low specificity, poor records, and interindividual variation) still retard its clinical use. The aim of the present study was to examine whether the mf VEP data obtained with the RETIscan system are appropriate for formulating a classification rule for glaucoma. *Method:* We examined and evaluated 65 eyes in 38 advanced glaucoma patients and 27 normal subjects, using four occipital gold cup electrodes (cross layout) for bipolar recording and a CRT monitor (display diameter 60 °, chequerboard pattern reversal, 60 segments in dartboard layout) for stimulation. In each case, eight cumulative measurements (77 s each) were made. The data of the 60 segments were cross-correlated with a RETIscan-internal VEP norm ("VEP finder"), combined in 16 sectors, and evaluated via the classification technique "double-bagging" and the Wilcoxon U-test. *Results:* In three out of the 16 sectors, the VEP amplitudes of the patients were significantly reduced (Wilcoxon U-test). Applying double-bagging on the cross-correlated data (with VEP finder) resulted in a sensitivity of 75% and a specificity of 71%, and the estimated misclassification rate was 27%. For uncorrelated data (without VEP finder), the same analysis achieved a sensitivity of about 60% and a specificity of 40%. *Conclusions:* Estimated sensitivity and specificity suggest that by using the RETIscan system for recording, a classification of the VEP data—i.e. a separation between normal and glaucoma subjects—is possible.

Introduction

The diagnostic value of multifocal visual evoked potentials (mf VEP) in glaucoma research is still under debate. Previous studies give considerable breadth of data concerning its sensitivity and specificity.

Goldberg, Graham, and Klistorner [6, 7, 12] propose different classification rules, which achieve sensitivities between 74% and 100% and a specificity of about 97%. Although these results are not verified based on an independent sample of glaucoma and normal subjects, or via appropriate error estimation (e.g. 10-fold cross-validation), they indicate the diagnostic value of mf VEP for glaucoma. Goldberg et al. examined whether the mf VEP has the potential to detect glaucomatous damages earlier than static perimetry, the gold standard in glaucoma diagnostics, and concluded that the mf VEP is a useful tool for clinical applications.

Results of Bengtsson [1] give a sensitivity of 68% (reaching 81% if only eyes with visual field losses are considered) and a specificity not exceeding 58%. According to this study, the mf VEP still needs to be upgraded for reliable clinical use. Hood et al. also mention limitations of the mf VEP, e.g. "poor VEP producers",

low signal-to-noise ratio (SNR), and fixation and refractive errors[8].

It has to be emphasized that in the above mentioned studies, the mf VEP, recorded with AccuMap [1, 6] and VERIS [7, 8, 12], was primarily examined concerning its correlation with visual field defects, i.e. its usability in objective perimetry. Moreover, the authors developed so-called classification rules that assign to normal or to glaucoma based on a given data set and do not verify misclassification results on independent observations. These stated sensitivities and specificities might be overestimated [14].

In the present study we used mf VEP data obtained with the RETIscan system, which includes a cross-correlation method ("VEP finder") for noise reduction.

The aim of the study was to evaluate whether the present mf VEP data with the given glaucoma severity are suitable to distinguish between normal and glaucoma subjects and whether an automated classification rule can be formulated and verified with a reliable estimation of sensitivity and specificity. For these purposes we applied the rather new classification method of "double-bagging". Double-bagging combines two common classification techniques, bagging and linear discriminant analysis (LDA), and has performed well in many applications [10]. We consider double-bagging to be a tool that reflects the discriminant potential of mf VEP data. Misclassification results were evaluated via 10-fold cross-validation.

Materials and methods

Subjects

All patients and control subjects were recruited from the Erlangen (Germany) Glaucoma Registry. The participants in the control group were recruited from the staff of the department and the university administration.

An ophthalmologist examined all individuals included in the study. They had open anterior chamber angles and clear optic media. Subjects with previous cataract surgery, any eye disorders other than glaucoma, general diseases (e.g. diabetes or vascular disease), or myopic refractive errors exceeding ±8 D were excluded. On the day of examination, intraocular pressure (IOP) was equal to or less than 21 mmHg in all subjects. IOP was measured using applanation tonometry in all subjects. The subjects were corrected for near vision, and pupils were not dilated.

The investigation followed the tenets of the Declaration of Helsinki, and informed consent was obtained from all participants once the nature of the study had been explained.

Normals

The right eyes of 27 normal subjects who met the above inclusion criteria were evaluated. They had normal IOP; normal findings in the ophthalmologic evaluation, which included slit-lamp examination, tonometry, perimetry, and ophthalmoscopy; and no family history of glaucoma or other eye diseases. Most of the normal subjects underwent only one visual field test (examination type G1)

Table 1 Summary statistics of normal subjects and glaucoma patients (*MD* mean deviation, *IOP* intraocular pressure)

	Normal	Glaucoma
Number	27	38
Age (in years)	49±14	59±9
Gender (male:female)	15:12	18:20
MD (dB)	1.4±1.4	8.3±4.9
MD (range/dB)	−0.49 to 4.47	2.12 to 20.34
IOP (mmHg)	13.0±3.0	14.1±3.4
Visual acuity	0.94±0.17	0.81±0.23

The normal subjects underwent only one visual field test (Octopus 500, G1). That might be the reason for the slightly higher MD values of some individuals within this group

on Octopus 500. The main focus of their classification as normals were the normal findings judged by ophthalmologists.

Patients

We evaluated 38 eyes of 38 patients (18 male, 20 female, age 59±9 years), including 17 primary and 5 secondary open angle glaucomas with IOP >21 mmHg in applanation tonometry, and 16 normal tension glaucomas with IOP <22 mmHg.

The glaucoma patients showed glaucomatous abnormalities of the optic disk and retinal nerve fibre layer, such as an abnormally small neuroretinal rim area in relation to the optic disk size, an abnormal neuroretinal rim shape, and localized or diffuse loss of the retinal nerve fibre layer. Reasons for optic disk damage other than glaucoma were excluded by neuroradiological examination. The diagnosis of normal tension glaucoma was based on normal IOP (<22 mmHg) in at least two 24-hour IOP curves without medical therapy.

All glaucoma patients had localized visual field defects, i.e. three or more adjacent points with a defect depth of 10 dB or more.

The MD was 8.3±4.9 dB, but a cut-off MD value was not defined. The patients had performed Octopus 500 fields (type G1) on two or more occasions in order to demonstrate reproducible visual field defects. In cases of differing glaucoma severity between eyes, the more affected eye was chosen; in cases of equality, we selected the eye with the better visual acuity.

Summary statistics of some characteristics of the study populations are given in Table 1.

Stimulation and recording

Stimulation and recording of the mf VEPs were performed with the RETIscan system (Roland Consult, Brandenburg, Germany). A dartboard pattern stimulus (diameter 60 °, Fig. 1a) was generated with a 19-inch CRT monitor that reversed m-sequence-based (length $2^{10}-1=1023$) at a rate of 17.5 reversals/s. It consisted of 60 segments with 16 checks, eight "white" (102 cd/m^2) and eight "black" (1.2 cd/m^2), thus the contrast was about 98%. The surround was 42 cd/m^2.

The recordings of the VEPs were performed in a four-channel bipolar-cross-layout as shown in Fig. 2 in order to cover their different allocations depending on location of the stimulus [12]. After the scalp was cleaned with abrasive skin prepping paste, four occipital gold cup electrodes were placed and fixed with electrode gel (impedance between electrodes <5 kΩ).

A gold cup earclip at the right ear was used as ground. The cut-off-frequencies were 5 Hz and 50 Hz; no notch-filter was used. The sampling rate was 1 kHz. Each individual underwent a recording session of eight cumulative measurements (77 s each).
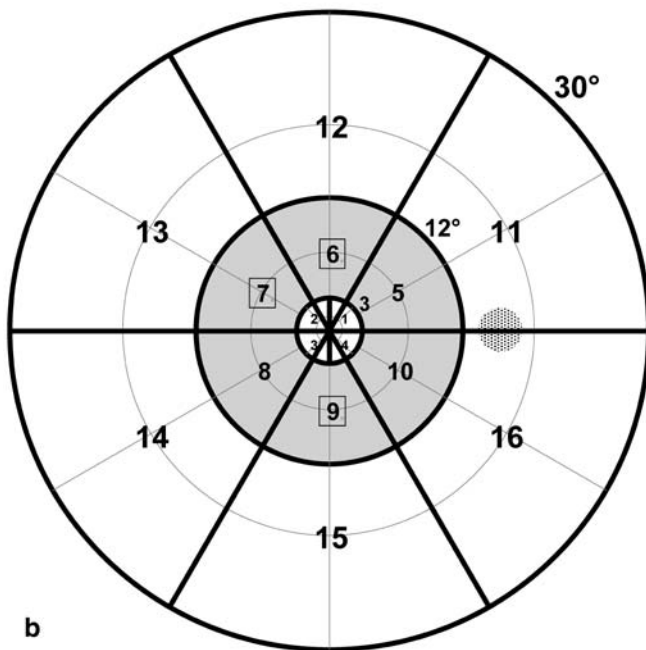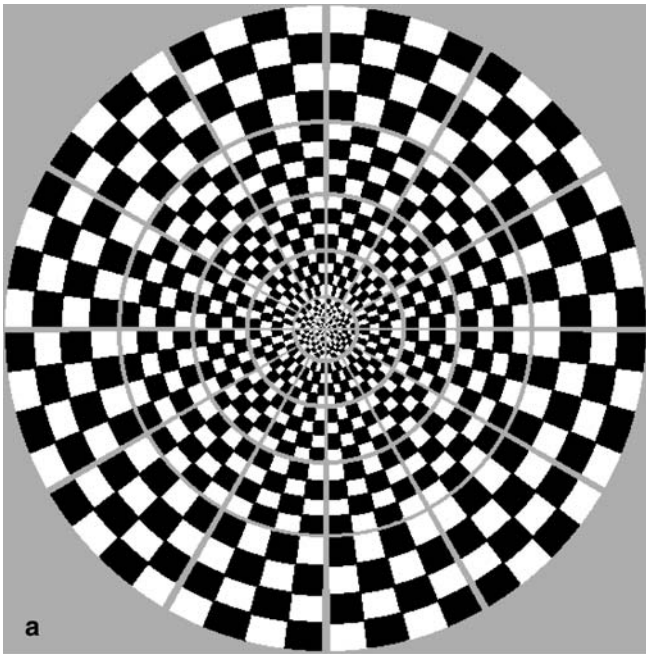
Fig. 2 The electrodes (gold cup) were placed in a four-channel bipolar layout around the inion, as introduced by Klistorner and Graham [12]

Fig. 1 a The dartboard stimulus used. Each of the 60 segments (surrounded by *grey lines*) underwent an m-sequence (1023 steps) controlled pattern reversal (rate 17.5 reversals/s). **b** Segments were combined in 16 sectors (12×4 segments+4×3 segments) as described by Hood et al. [8, 9]. The sector numbering is given for the right eye (note the blind spot); for the left eye, it is symmetrically inverse

Evaluation

The output variable for all evaluations in the present study was the peak-to-trough amplitude quantified by the vertical distance between lo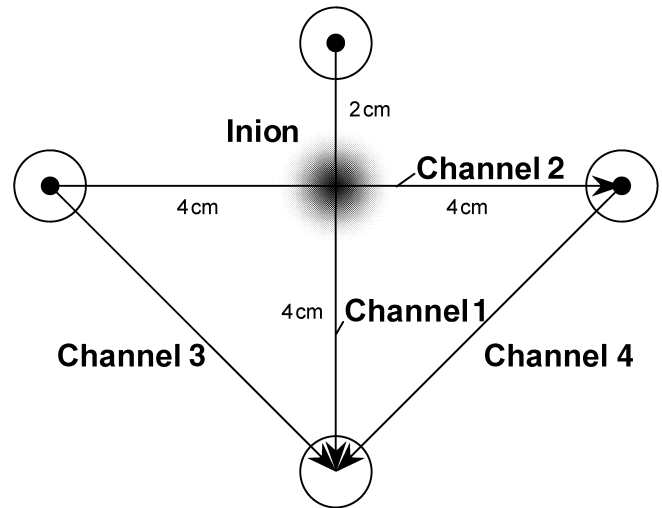cal maximum and local minimum (indicated by the cursors in Fig. 3b) in a time interval of 80–140 ms. The RETIscan system places the cursors automatically and measures their vertical distance regardless of how much the noise contributes to these local extrema. Failure to reduce the noise could lead to a poor signal-to-noise ratio.

Noise reduction

All data were preprocessed by a "VEP finder" (Roland Consult, Brandenburg, Germany), a tool to suppress cortical and electrical noise. It cross-correlates each recorded VEP response with the internally stored "VEP norm" based empirically on the knowledge of standard norm VEP responses (Fig. 3a).

The cross-correlation approaches zero for constant offsets, noise, and other non-VEP signals. The more the actual measured VEP response matches the internal norm, the higher the absolute result VEP_F (t) of the cross correlation, which is calculated using the following formula:

$$VEP_{\mathbf{Finder}}(t) = \frac{1}{n}\sum_{i=1}^{n} VEP_{\mathbf{R}espotnse}(i+t) \cdot VEP_{\mathbf{Norm}}(i)$$

where t is the time-shifting (80 ... 140 ms) between the norm curve VEP_N and the recorded response VEP_R at sample i (i=1, 2, ..., n; n=number of samples=256).

Figures 3b and 3c show the effect of the cross-correlation on the mf VEP responses.

The noise is largely reduced, while the polarity of the responses is not changed or influenced by the cross-correlation as long as the response peaks have "normal" latencies. In cases of pathological latencies, the amplitudes are more reduced and increasingly phase-shifted, which has the useful side effect that pathological latencies are automatically included in the amplitude evaluation.

Combining channels and sectors

Because the allocation of the cortical VEP responses among the four channels depends on the location of the corresponding stimulus, a single channel recording can never reflect the entire map of all 60 segments. Thus, the four channel records were combined into

Fig. 3 a The VEP finder norm curve (*strong line*) is based empirically on the knowledge of a standard norm VEP response (*dashed line*, simplified by sinusoidal extrapolation). **b** The effect of the VEP finder shown for a normal mf VEP (60 ° field, channel 1): The uncorrelated responses (*left*) show electrical and cortical noise, which is extensively removed by the VEP finder (*on the right*: the same responses cross-correlated with the VEP norm curve; the 60 records left and right correspond with the 60 segments of Fig. 1a). **c** Single responses from the 60 ° response field (areas marked with a *square* in **b**). On the *left* the responses are uncorrelated, and on the *right* they are cross-correlated
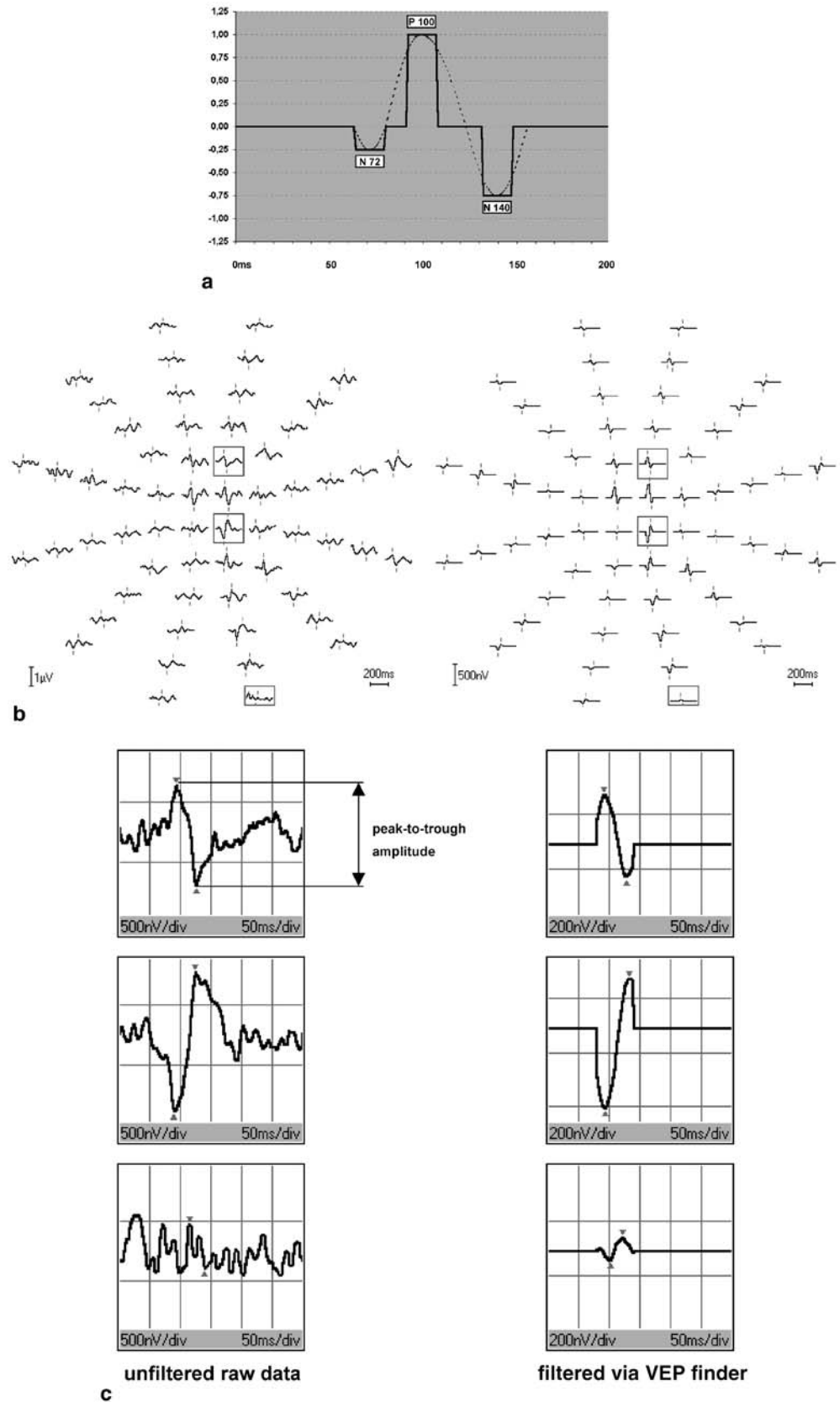
Table 2 Peak-to-trough VEP amplitudes (nV; mean and standard deviation) of normals and patients averaged in 16 sectors for unfiltered data (*Without VEP finder*) and filtered data (*With VEP finder*). The *p* values are the result of a Wilcoxon U-test. For the filtered data the difference between normals and patients was significant in three sectors (nos. 6, 7 and 9; *p* probability of error, n.s. not significant)

Location	Without VEP finder			With VEP finder		
	Normals	Patients	*p*	Normals	Patients	*p*
Sector 1	520±312	443±204	n.s.	217±127	184±98	n.s.
Sector 2	488±239	429±205	n.s.	301±95	167±74	n.s.
Sector 3	600±346	523±268	n.s.	287±181	229±122	n.s.
Sector 4	681±894	480±203	n.s.	286±337	231±119	n.s.
Sector 5	421±199	408±198	n.s.	156±63	145±74	n.s.
Sector 6	420±277	341±148	n.s.	153±82	111±42	0.001
Sector 7	485±272	381±151	n.s.	176±78	131±67	<0.001
Sector 8	499±264	417±190	n.s.	187±96	142±83	n.s.
Sector 9	549±338	447±231	n.s.	243±198	171±113	0.013
Sector 10	485±210	437±202	n.s.	177±66	170±89	n.s.
Sector 11	380±185	379±167	n.s.	113±54	118±40	n.s.
Sector 12	369±211	355±166	n.s.	108±68	105±42	n.s.
Sector 13	377±145	344±137	n.s.	116±47	107±39	n.s.
Sector 14	394±187	366±152	n.s.	137±62	120±57	n.s.
Sector 15	403±177	387±180	n.s.	151±68	133±58	n.s.
Sector 16	458±223	432±166	n.s.	171±82	159±68	n.s.

one topographic map by selecting the maximum response out of the four channels for every single segment [6, 8, 9, 12]. These maximum responses were averaged in the 16 sectors of Fig. 1b, and all further evaluations were based on these summarized data [8, 9].

Statistical methods

The aim of the analysis was to assess whether the present mf VEP data support the diagnosis of glaucoma. Hence, a classification technique that combines several statistical methods and that results in small misclassification rates is desirable.

We chose a classification technique called double-bagging [10] as a statistical tool to classify a subject as glaucoma or normal. Error rate estimation was performed via 10-fold cross-validation. The misclassification or error rate is the proportion of misclassified subjects, including normal subjects misclassified to glaucoma as well as glaucoma subjects misclassified to normal; that is, it summarizes sensitivity and specificity.

Double-bagging. Double-bagging combines bootstrap aggregated classification trees (bagging) and LDA. A classification tree [5] is a set of yes-no questions leading to a partition of the multivariate sample space; i.e. it assigns a subject to normal or to glaucoma. The tree is based on a set of explanatory variables; the present data set uses the peak-to-trough amplitudes of the 16 sectors summarized as described above. A recursive evaluation of all possible binary splits of every explanatory variable leads to the final classification tree [11]. Classification trees are sensitive to small changes in the learning samples; for example, removing or adding some observations may lead to large changes in the resulting tree. Breiman [2, 3, 4] proposes a procedure called bagging (bootstrap aggregation) to stabilize classification trees.

Bagging works as follows: First, a bootstrap sample is drawn from the original data—a random sample with replacement of size *n* out of *n* observations. Second, a classification tree is constructed based on the bootstrap sample. This procedure is repeated 50 times, resulting in 50 single trees. A new subject was assigned to glaucoma if most of the trees predicted glaucoma, and to normal otherwise.

The combination of LDA and bagging works in the following manner: Approximately two-thirds of the full set of observations are included in the described bootstrap sample, and the remaining observations are called the "out-of-bag" sample [3]. Based on a given learning sample, the classification rule is constructed in two steps: First, an LDA is performed on the out-of-bag sample corresponding to each bootstrap sample. The LDA computes a set of additional variables called "discriminant variables". Second, the bagging procedure is calculated based on the original variables and the discriminant variables. A new subject is assigned to either normal or glaucoma, following the same procedure as described above. We applied double-bagging on the set of combined VEP amplitudes of the 16 sectors. Hence, the classifier is trained on 16 variables and 65 observations (65 being the total number of subjects).

Because to our knowledge it is not yet clear how to calculate ROC curves for machine learning classifiers such as neural networks, double-bagging, or others, we set the calculation of ROC curves aside. Furthermore, the 16 variables per patient would result in 32 ROC curves, and we had only 65 observations. Consequently, such an ROC analysis would be doubtful because it does not account for this problem of dimensionality and could lead to highly varying results.

Error rate estimation. We estimated error rates of the different classification techniques via 10-fold cross-validation: The original data were divided into 10 subgroups of equal size. Each subgroup was used as a control, and the classifiers were trained on the nine remaining subgroups. Finally, the resulting 10 estimations of misclassification rate, sensitivity, and specificity were averaged.

Wilcoxon U-test. In addition to double-bagging, we tested the difference between normals and patients for each of the 16 sectors using the Wilcoxon U-test.

Results

The main results of the present study are presented in Tables 2 and 3. Table 2 shows the VEP amplitudes±standard deviation of the 16 sectors tested obtained from normals and patients. The raw data are shown on the left side, and the filtered data on the right side. It can be seen that the amplitudes of the filtered records are reduced by more than 50% compared with the original records. Likewise, their standard deviations are reduced by a comparable amount after cross-correlation. Thus, the relative standard deviations were not influenced by the VEP finder.

As expected, the patients showed reduced VEP responses for both the filtered and the unfiltered data in 15

Table 3 Estimated error rates, sensitivities, and specificities of double-bagging on filtered and unfiltered VEP data

	Error rate	Sensitivity	Specificity
Filtered	26.88%	74.72%	71.07%
Unfiltered	50.35%	59.57%	39.29%

out of 16 sectors. The Wilcoxon U-test gave significant results in the sectors 6, 7, and 9 of the circumfoveal area (area from 3 ° to 12 °) using the filtered data. In the outer sectors (11–16), the unfiltered amplitudes were almost equal, and in one case (sector 11) the amplitudes of the patients were slightly higher than those of the normals. For the unfiltered data the differences were not significant in all 16 sectors.

Table 3 shows the results of the double-bagging analysis. They are expressed by estimations of sensitivity, specificity, and the total misclassification rate via 10-fold cross-validation. A misclassification rate of 50% accords to a random assignment to normal or to glaucoma.

For the uncorrelated data we found an estimated misclassification rate of 50.3%, which results in a sensitivity of 59.6%, and a specificity of 39.3%, meaning that without using VEP finder, the VEP data cannot be classified better than random using double-bagging. For the cross-correlated data the error rate was 26.9%; correspondingly, the sensitivity was 74.7% and the specificity was 71.1% (Table 3).

Discussion

The two major issues of the present investigation were (1) to test the usefulness of "VEP finder", a cross-correlation-based noise reduction procedure built in the RETIscan system, and (2) to apply a statistical evaluation called "double-bagging", which has performed well in different data structures in previous research [10, 11]. The main results of the present study are reproduced in Tables 2 and 3. Table 2 indicates that in most stimulus sectors, no significant difference exists between normals and glaucoma patients. Only after noise reduction are the amplitudes obtained from the patients significantly lower in sectors 6, 7, and 9, i.e. in sectors belonging to the Bjerrum area, which is the area most susceptible to glaucomatous perimetric damage. This result suggests that some method of noise reduction is very important for data analysis.

Different procedures of filtering out noise from mf VEPs have been described in the past. Thus, a rather arbitrary procedure of scanning the raw data in real time was used in order to reject segments contaminated by a high level of noise or eye movements [12]. Another procedure used a normalization of the VEP signals to underlying electroencephalographic activities determined by Fourier transform [1, 6, 13]; this supposedly reduces intersubject variability by removing α- and electrocardio-

gram contaminations. However, the usefulness of these procedures in discriminating between normal and glaucoma has not yet been demonstrated.

The present data show that using uncorrelated data, neither a classifier that classifies better than random nor significant differences in sectors could be found, whereas cross-correlation-based noise reduction improves significance levels (see Table 2) and classification rates (see Table 3). Compared to the other noise reduction method, the VEP finder might be a simple method because it does not require any further analysis. But consequently, it does not account for individual peculiarities, so it cannot reduce the intersubject variation expressed by the relative standard deviations in Table 2. A more detailed comparison of noise reduction methods is not possible as long as the methods are not tested on the same VEP records. Due to individual peculiarities (e.g. malfixation and constitution of the visual cortex), the mf VEP quality of six normals and ten patients was so low that they were not included in the study.

It might appear problematic that the MD values (mean, SD, and range) of the normals are slightly higher than commonly used in control groups (Table 1). We assume the reason can be found in the fact that, unlike the glaucoma patients, most of the normals underwent only one visual field test. This assumption is confirmed by a study by Zulauf et al., in which normal volunteers had an MD range up to 4.1 dB (both phases), although "all subjects were familiar with automated perimetry" [15]. As mentioned above, all the other findings, especially the ophthalmological ones, were without any abnormalities.

Age effects on the mf VEP have been studied previously, and no significant effects have been found [6, 13]. Our results calculated from the average of the 16 sectors confirm this finding (Spearmans rho=−0.104, $p>0.05$) and thus obviate the need for an age stratification of the patients.

An important criterion for the usefulness of a new experimental method is its ability to correctly classify patients as patients (sensitivity) and normals as normals (specificity). Table 3 indicates values of 74.72% and 71.07%, respectively, for the filtered data. These figures are somewhat lower than some reported in the literature. But in previous studies, classification rules were often constructed that optimised results in one study population and were evaluated on the same study group again; therefore, reported results may be highly overestimated. We estimated error rates, sensitivities, and specificities via 10-fold cross-validation. This prevents underestimating the error rates, i.e. overestimating sensitivities and specificities.

In the present study, a topographic resolution of local amplitude reductions and a correlation with corresponding perimetric defects was not intended and cannot be performed with double-bagging because it is applied to the whole data set. Instead, this method of analysis performs a global evaluation of responses obtained from all

stimulus segments. Double-bagging combines two common classification techniques, bagging and LDA, and thus avoids a model selection.

Similar to the present analysis, the VEP amplitude averaged over all segments has been used previously as a global measure of function [6], resulting in a sensitivity of 86% and a specificity of 95%. This discrepancy from the values of the present investigation could be caused by the overestimation of classification results in the previous study, since the classification rule was constructed and evaluated on the same population. Another reason for this may be that the 65 mf VEP data sets of our study population were in contrast to the comparative studies [1, 6, 12] not assigned to normal and glaucoma subjects before evaluation via double-bagging. The evaluation of preselected data can again result in overestimation of sensitivity and specificity [14].

Furthermore, major methodological differences between the two studies include different methods of noise reduction (discussed previously) and different applications of binary sequences for stimulation. While the present study uses one m-sequence for all stimulus segments, in the previous investigation [6] each stimulus site was modulated in time according to a different sequence. Other conditions, including stimulus display and method of recording, were rather similar in both studies and probably did not contribute to the differences in results.

On the other hand, if advantage is taken of the topographic resolution of the multifocal technique by comparing *local* amplitude reductions with corresponding perimetric defects in probability of abnormality plots, the sensitivity can be further improved up to 100% and the specificity up to 97% [12].

Although this result seems to be admittedly influenced by the fact that far-advanced glaucoma patients with rather high perimetric MD values have been studied, it suggests a superiority of the objective VEP perimetry over a global evaluation of VEPs obtained from all stimulus locations. Another study [1] using the multifocal objective perimetry method of Goldberg et al. [6] reported a sensitivity varying from 68% (all glaucoma states) to 81% (advanced glaucomas with visual field losses) and a low specificity of 58%, as well as no sufficient agreement between VEP results and visual field analysis. These figures are somewhat closer to the ones found in the present investigation.

Conclusions

We emphasize that the sensitivity and specificity values we found are valid only for the present data set and the given severity of glaucoma. Taking this into account, the mf VEP data show significant differences between normal and glaucoma even when a global analysis of the data is performed using the double-bagging method, provided the data are processed by a cross-correlation-based noise reduction algorithm (VEP finder). The sensitivities and specificities thus obtained differ somewhat from previous research, which might be explained by the use of different stimulation and noise reduction methods as well as by different patient selection and different statistical evaluation procedures causing underestimation of error rates.

References

1. Bengtsson B (2002) Evaluation of VEP perimetry in normal subjects and glaucoma patients. Acta Ophthalmol Scand 80:620–6
2. Breiman L (1996) Bagging predictors. Machine Learning 24:123–140
3. Breiman L (1996) Out-of-bag estimation. Technical report, Statistics Department, University of California at Berkeley, California
4. Breiman L (1998) Arcing classifiers. Ann Statistics 26:801–849
5. Breiman L, Friedman JH, Olshen RA and Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont, California
6. Goldberg I, Graham SL, Klistorner AI (2001) Multifocal objective perimetry in the detection of glaucomatous field loss. Am J Ophthalmol 133:29–39
7. Graham SL, Klistorner AI, Grigg JR, Billson FA (2000) Objective VEP perimetry in glaucoma: asymmetry analysis to identify early deficits. J Glaucoma 9:10–19
8. Hood DC, Greenstein VC (2003) Multifocal VEP and ganglion cell damage: applications and limitations for the study of glaucoma. Prog Retin Eye Res 22:201–251
9. Hood DC, Zhang X, Hong JE, Chen CS (2002) Quantifying the benefits of additional channels of multifocal VEP recording. Doc Ophthalmol 104:303–320
10. Hothorn T, Lausen B (2003) Double-bagging: combining classifiers by bootstrap aggregation. Pattern Recognition 36:1303–1309
11. Hothorn T, Lausen B (2003) Bagging tree classifiers for laser scanning images: data and simulation based strategy. Artif Intell Med 27:65–79
12. Klistorner A, Graham SL (2000) Objective perimetry in glaucoma. Ophthalmology 107:2283–2299
13. Klistorner AI, Graham SL (2001) EEG-based scaling of multifocal visual evoked potentials: effect on inter-subject amplitude variability. Invest Ophthalmol Vis Sci 42:2145–2152
14. Mardin CY, Hothorn T, Peters A, Jünemann AG, Nguyen NX, Lausen B (2003) New glaucoma classification method based on standard HRT parameters by bagging classification trees. J Glaucoma 12:340–346
15. Zulauf M, LeBlanc RP, Flammer J (1994) Normal visual fields measured with Octopus-program G1. II. Global visual field indices. Graefes Arch Clin Exp Ophthalmol 232:516–522