Glaucoma

Progression of Patterns (POP): A Machine Classifier Algorithm to Identify Glaucoma Progression in Visual Fields

Michael H. Goldbaum,[1] *Intae Lee,*[1] *Giljin Jang,*[*,1] *Madhusudhanan Balasubramanian,*[1] *Pamela A. Sample,*[1] *Robert N. Weinreb,*[1] *Jeffrey M. Liebmann,*[2] *Christopher A. Girkin,*[3] *Douglas R. Anderson,*[4] *Linda M. Zangwill,*[1] *Marie-Josee Fredette,*[4,5] *Tzyy-Ping Jung,*[1] *Felipe A. Medeiros,*[1] *and Christopher Bowd*[1]

PURPOSE. We evaluated Progression of Patterns (POP) for its ability to identify progression of glaucomatous visual field (VF) defects.

METHODS. POP uses variational Bayesian independent component mixture model (VIM), a machine learning classifier (MLC) developed previously. VIM separated Swedish Interactive Thresholding Algorithm (SITA) VFs from a set of 2,085 normal and glaucomatous eyes into nine axes (VF patterns): seven glaucomatous. Stable glaucoma was simulated in a second set of 55 patient eyes with five VFs each, collected within four weeks. A third set of 628 eyes with 4,186 VFs (mean ± SD of 6.7 ± 1.7 VFs over 4.0 ± 1.4 years) was tested for progression. Tested eyes were placed into suspect and glaucoma categories at baseline, based on VFs and disk stereoscopic photographs; a subset of eyes had stereophotographic evidence of progressive glaucomatous optic neuropathy (PGON). Each sequence of fields was projected along seven VIM glaucoma axes. Linear regression (LR) slopes generated from projections onto each axis yielded a degree of confidence (DOC) that there was progression. At 95% specificity, progression cutoffs were established for POP, visual field index (VFI), and mean deviation (MD). Guided progression analysis (GPA) was also compared.

RESULTS. POP identified a statistically similar number of eyes (P > 0.05) as progressing compared with VFI, MD, and GPA in suspects (3.8%, 2.7%, 5.6%, and 2.9%, respectively), and more eyes than GPA (P = 0.01) in glaucoma (16.0%, 15.3%, 12.0%, and 7.3%, respectively), and more eyes than GPA (P = 0.05) in PGON eyes (26.3%, 23.7%, 27.6%, and 14.5%, respectively).

CONCLUSIONS. POP, with its display of DOC of progression and its identification of progressing VF defect pattern, adds to the information available to the clinician for detecting VF progression. (*Invest Ophthalmol Vis Sci.* 2012;53:6557–6567) DOI:10.1167/iovs.11-8363

Glaucoma is a blinding but treatable disease that affects up to 91 million individuals worldwide, 6.7 million of whom have bilateral blindness secondary to glaucoma.[1,2] The time course of glaucomatous deterioration is generally years. The goal of management is to detect the disease in the early stage and to intervene to prevent progression at any stage. To manage glaucoma successfully, a glaucoma specialist needs to know whether an eye with glaucomatous damage is stable or progressively deteriorating and, if so, the rate of that deterioration.

The visual field (VF), tested by standard automated perimetry, is a ubiquitous test of the severity of glaucomatous damage in an eye. However, several causes, especially damage to retinal ganglion cells, increase variability in the sensitivity of parts of the retina to a light stimulus in the VF test.[3–7] This variability is noise that can mask a weak signal of progression in serial VF tests.

A number of change detection algorithms applied to perimetry, such as progression by visual field index (VFI),[8] mean deviation (MD), or guided progression analysis (GPA), are statistical methods that use linear classification methods to represent the rate and magnitude of change or use analysis of variance to identify change outside the limits of short term variability.[8,9] These statistical methods distinguish between two classes of eyes, stable glaucoma and progressing glaucoma.

The shape of the boundary that best separates these two classes of eyes is generally constrained by linear statistical methods, and that constraint can lead to many stable eyes being identified as progressing (false positives) and progressing eyes being identified as stable (false negatives). Theoretically, machine learning classifiers (MLC) can reduce these errors because they learn from the data how to generate better separating surfaces.

The primary goal of our current research is to improve the detection of eyes with progressing glaucomatous damage manifested by glaucomatous VF defects. For this purpose, we developed Progression of Patterns (POP) based on a novel MLC, variational Bayesian independent component analysis mixture model (VIM), previously developed at the University of California at San Diego (UCSD).[10,11] The premise we are testing is that this rigorous mathematical method will detect more eyes with progression of glaucomatous VF patterns over time than current rules derived primarily from clinical experience.

METHODS

Methods adhere to the tenets of the Declaration of Helsinki and to the Health Insurance Portability and Accountability Act and were approved by the institutional review boards of the University of California at San Diego (UCSD), The New York Eye and Ear Infirmary (NYEE), The University of Alabama at Birmingham (UAB), and the University of Miami Miller School of Medicine. All of these institutes provided data used in the current study. All participants gave written informed consent.

Inclusion and Exclusion of Participants

Participants came from the Diagnostic Innovations in Glaucoma Study (DIGS), and the African Descent and Glaucoma Evaluation Study (ADAGES). The eyes were included if, at baseline, they had open angles, a best corrected visual acuity of 20/40 or better, and a refractive error less than or equal to 5.0 diopters (D) sphere and 3.0 D cylinder. We required at least one good quality stereoscopic pair of disk photographs. Both eyes were included, except in cases where only one eye met the study criteria. All participants were over 18 years of age.

Participants were excluded if they had a history of intraocular surgery (except for uncomplicated cataract surgery or glaucoma surgery), secondary causes of glaucoma, other systemic or ocular diseases known to affect the VF, significant cognitive impairment, history of stroke, an inability to perform VF exams reliably ($< 33\%$ false positives, fixation losses, or false negatives not explained by severity of defect), or a life threatening disease that precluded retention in the study. The inclusion and exclusion criteria were re-evaluated annually.

Examination

Each participant underwent a complete ophthalmologic examination at baseline and at least annually thereafter, which included best-corrected visual acuity, slit lamp biomicroscopy, gonioscopy, Goldmann applanation tonometry, central corneal thickness measurement, dilated indirect ophthalmoscopy examination, stereoscopic ophthalmoscopy of the optic disc with a 78 D lens, VF testing, and simultaneous stereoscopic disc photography. We monitored all systemic and ocular procedures and medications, and any concurrent conditions that might affect vision.

Visual Fields

All standard automated perimetry (SAP) fields were obtained with Humphrey Visual Field Analyzers (Iii; Carl Zeiss Meditec, Inc., Dublin, CA) using the Swedish Interactive Thresholding Algorithm (SITA) standard program 24-2. All fields were processed through the UCSD-based Visual Field Assessment CenTer (VisFACT). VisFACT personnel reviewed only the VFs and were masked to study, patient identity, diagnosis, and other test information. Abnormal SAP was defined as pattern standard deviation (PSD) at the 5% probability value or worse, or a glaucoma hemifield test (GHT) outside normal limits on at least two consecutive exams.[12-14]

Simultaneous Stereoscopic Optic Disc Photographs

Color stereoscopic photograph pairs were simultaneously recorded with a camera (Nidek Stereo Camera 3-DX; Nidek Inc., Palo Alto, CA) through maximally dilated pupils. All stereoscopic photographic evaluations were performed with the Asahi Pentax Stereo Viewer II (Pentax of America, Inc., Montvale, NJ), illuminated with a color corrected fluorescent light bulb. Certified photograph graders from the UCSD Optic Disc Reading Center evaluated all photographs. Each stereoscopic photograph was graded by two independent graders according to a set protocol using a standard set of photographs as reference. Each grader was masked to the participant's identity, diagnostic status, study, race, and other results. In cases of disagreement, a third senior grader adjudicated.

Photographic pairs were graded for quality and evidence of glaucomatous optic neuropathy (GON) at baseline. GON was defined by evidence of excavation, neuroretinal rim thinning or notching, localized or diffuse retinal nerve fiber layer (RNFL) defect, or asymmetry of the vertical cup-to-disc ratio greater than 0.2 between eyes.

The baseline and the most recent stereoscopic photographs of an eye were assessed for progression of glaucomatous optic neuropathy (PGON) by two observers, based on a decrease in the neuroretinal rim thickness, appearance of a new RNFL defect, or enlargement of a pre-existing RNFL defect. Observers were masked to the patient identification, diagnosis of glaucoma, and temporal order of the photographs. Any disagreement in assessment between these two observers was adjudicated by a third observer. PGON was considered both indicative of glaucoma and evidence of progression.

Brief Review of VIM

To analyze progression, POP uses the axis environment created by VIM to represent the distribution of SAP-SITA fields. A summary of VIM applied to SAP is provided below. For additional details, see the Appendix in this manuscript and the Appendix in Goldbaum et al.[15]

In brief, the VIM was developed in earlier work by our group from a cross-sectional analysis of 1,146 eyes with normal SAP-SITA fields and 939 glaucoma eyes. From this cloud of 2,085 normal and glaucoma fields, VIM segregated the fields into three clusters (mostly normal fields, fields with mild glaucomatous defects, and fields with moderate to severe glaucomatous defects), and in each cluster, VIM oriented statistically independent axes through the cluster mean. This means that the axes represented patterns of glaucomatous VF defects that differed greatly from each other (Fig. 1). The field defect patterns increased in severity in the positive direction of each axis by expansion or deepening of the field defects making up the patterns. The scale of severity along each axis was made comparable by using VIM units (VU) of SD. VIM derived nine such axes in the three clusters. Seven axes represented glaucomatous defect patterns, and two axes represented normal field patterns. It is on the seven glaucoma axes that POP detects progression in new data. Since each axis represents a VF pattern, it is possible to see the pattern that is progressing (Fig. 1).

Progression of Patterns

This section provides a brief description of POP. For a detailed mathematical description, please see the Appendix.

The POP method works by projecting each field in a sequence of fields of an individual eye in 53-dimensional (53D) space (described
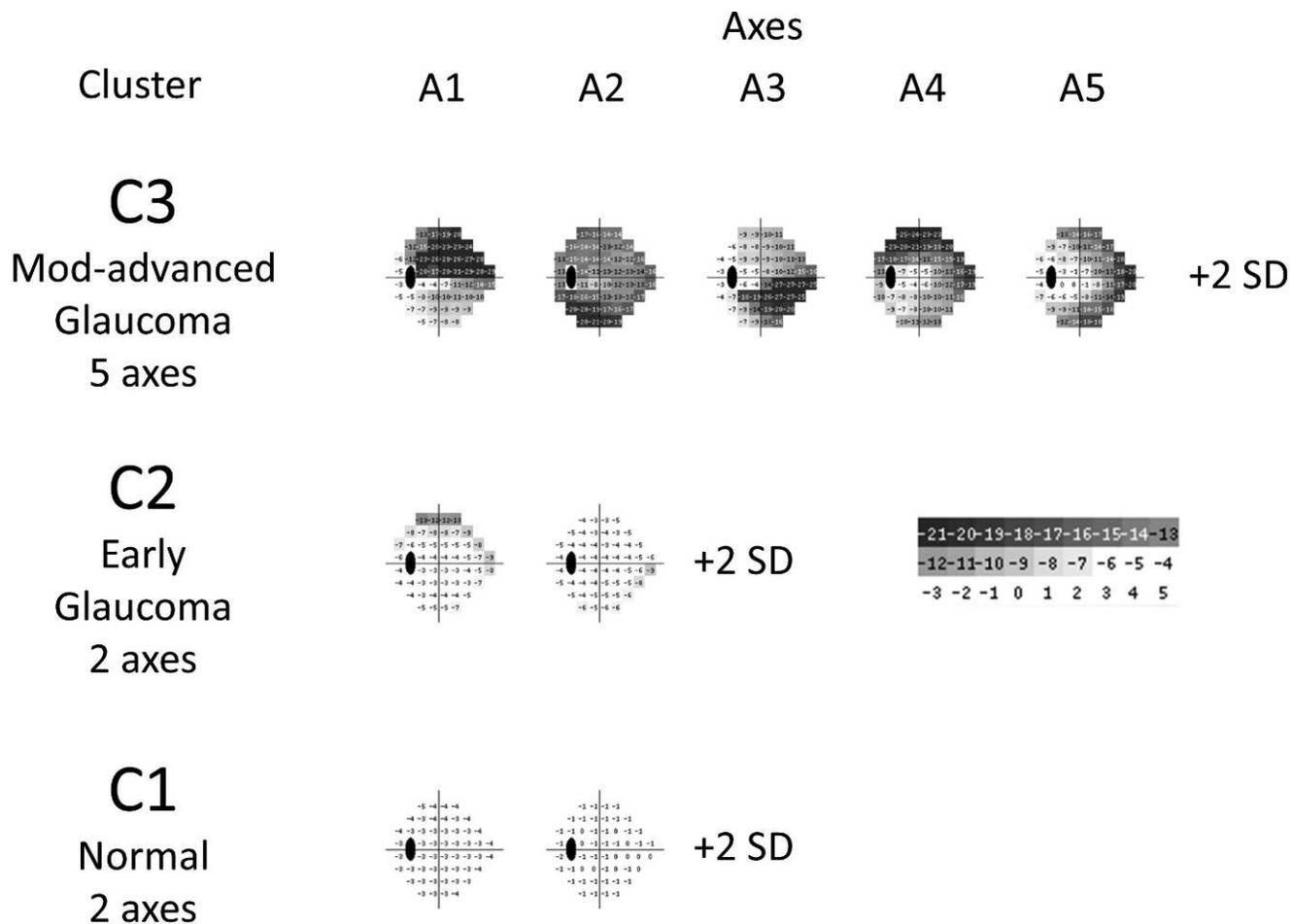
FIGURE 1. Discrete tone plots that demonstrate the VF patterns represented by each of the VIM clusters and axes. The presentation simulates a total deviation plot generated at +2 VIM units (SDs) on each axis. Field sensitivity less than normal is displayed in *graded steps of gray.* The numbers disclose deviation from normal in decibels.

below) onto each of the seven predefined VIM glaucoma axes (Fig. 2). Linear regression (LR) identifies the rate of change along each axis, and the axis with the maximal change is selected to display the rate of change (Fig. 3). To accomplish this method of detecting progression, we must (1) apply VIM to represent the normal and glaucomatous axes in a normative database of SITA fields, (2) establish the acceptable amount of variability in stable data from glaucomatous eyes, and (3) institute a definitive process that indicates when glaucomatous damage is progressing in an eye that is being tested. To do these three tasks requires three distinct databases.

POP was Developed Using Three Independent Visual Field Datasets. *Dataset 1, VIM Training Set of Single VFs.* The absolute sensitivity of each VF location (with blind spot locations omitted) constituted a dimension in classifier input space. An additional feature, age, increased the input space to 53D. Each VF was located in VIM's 53D input space.

Any new VF, whether single or part of a temporal sequence, was located in VIM's input space and was projected onto each VIM axis from the field's point in space to the axis (Fig. 2). POP projected a sequence of VF from an individual eye onto each axis (defect pattern), and POP assessed the movement over time along each axis; from this change over time, POP determined whether or not the eye was progressing in severity along each pattern of field loss.

Dataset 2, Time Series of Visual Fields in Stable Eyes with Glaucoma. This set consisted of eyes with five weekly serial VFs collected within four weeks provided by investigators at the Bascom Palmer Eye Institute at the University of Miami Miller School of

Medicine. The assumption was made that the glaucomatous defects in these eyes were not progressing over such a short time, and that any change noted would be due to the variability in the VFs measured in stable glaucoma. Each eye in the stable dataset was required to have reliable VFs at all 5 visits, and each eye had to have evidence of glaucoma based on ocular examination and the presence of repeated VF loss as defined under the Visual Fields header, above. These constraints resulted in the inclusion of 55 eyes from 55 participants (See Appendix). LR was applied to the sequence of five visits projected on each axis. Since the VFs were considered to be stable, the sequence of each of the 55 eyes was permuted (5! = 120) to yield 6,600 possible slopes (rates of change) for determining the confidence intervals (CI) for stable eyes on each axis. This dataset was used to estimate the variability of stable serial data used in POP analysis.

The interval between visits in the stable data set was one week. Patients with glaucoma are commonly followed at intervals between six months to one year. We assumed that each interval in the stable data set could be reset to one year to approximate the limits of stability of eyes measured at around one year intervals.

Dataset 3, Time Series of Visual Fields in Eyes to Be Analyzed. This test set included serial VFs to be classified for progression or stability from eyes of glaucoma suspects and patients enrolled in DIGS and ADAGES at three centers (UCSD, NYEE, UAB). For each sequence of fields, we required at least five reliable VFs but did not require the initial field to be abnormal. Participants who became ineligible and those with cataract surgery during the study period were excluded.
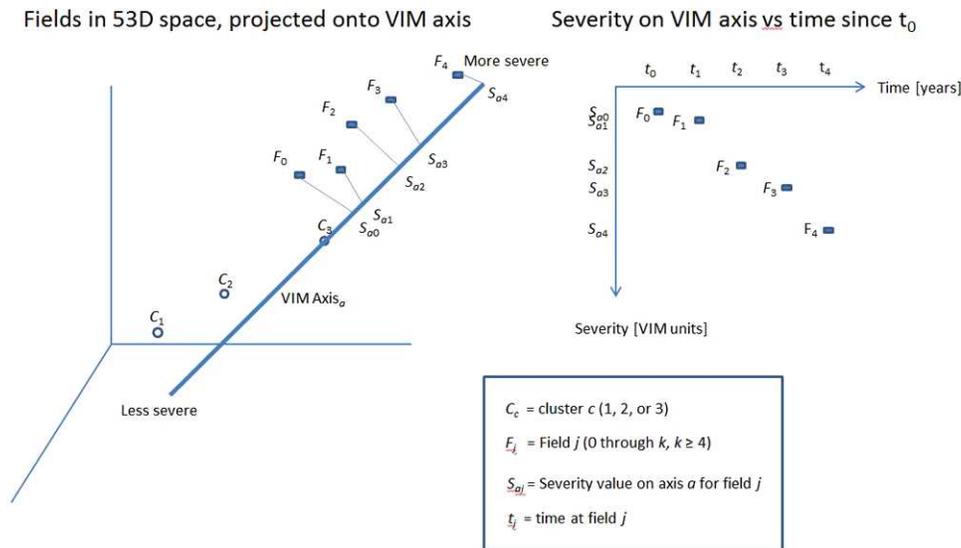
FIGURE 2. *Left:* Sequence of VFs from a sample individual eye projected onto an axis from cluster 3. This 3 dimensional graph illustrates what is occurring in the original 53D space of field location sensitivities plus age. F1 through F5 represent a sequence of VFs in 53D space, which are projected onto an axis. The projected point is the intersection of the appropriate line in 53D space from the field to the axis. *Right:* Severity along an axis versus time. POP reduces the sequence of fields in 53D to a sequence of severity points in one dimension along each of the seven glaucoma axes. Those severity points are represented by a *LR line*, as seen in the next figure.

These criteria resulted in 628 eyes of 418 participants tested with 4,186 VFs.

The dataset of serial fields was used for analyzing the performance of POP and for comparing POP with currently existing progression algorithms detailed later in the Methods section. Eyes in the serial dataset were divided into two mutually exclusive categories of glaucoma suspects and those diagnosed with glaucoma. For this study, suspect eyes had ocular hypertension (OHT, defined as repeatable measurements of untreated IOP $\geq$ 22 mm Hg with normal VFs and normal appearing optic discs by masked stereophotograph assessment), glaucomatous optic neuropathy (GON) with normal VFs regardless of IOP, or glaucomatous VF defects without GON regardless of IOP. The glaucoma group was composed of eyes with both glaucomatous VF defects and GON.

PGON (as described above) was considered as a gold standard for progression that differed from the test (VFs) being used to determine progression. Sensitivity of each progression detection method was also assessed in the PGON eyes.

POP Algorithm

POP projects the sequence of fields in an eye being evaluated for progression onto each of the seven glaucoma axes created by VIM. LR is applied to the sequence of severity values on each axis. The y value of the regressed line represents the overall offset of the severity value from the cluster mean at a particular time, and the slope represents the estimated rate of change of severity over time along the axis (VIM-defined VF pattern). We use the difference between the mx values on the regressed line at the initial and last visits, $m\Delta x$, as a surrogate for the slope of the regressed line (see Appendix).

Stable Data Analysis. To distinguish between progressing glaucomatous fields and stable glaucomatous fields, we determined the limits of change in severity that stable eyes can have. The mean of all the regression slopes on each axis approached zero, but due to the variability of the fields from visit to visit, the sequence of fields from some stable eyes had slopes upward, suggesting improvement, or downward, suggesting deterioration. From the stable data, we derived the 95% confidence limits (CL) of the $m\Delta x$ of these non-progressing eyes empirically by sorting the slopes and choosing the slope at the

95th percentile. Since POP sought only deterioration of VFs, POP used a single-tail CI in the direction of deterioration.

Identification of Change in Serial Data. For each series of VFs measured from an individual's eye, a decision of progression was made when a sufficient proportion of the eye's estimated change, $m\Delta x$, fell outside the lower bound of the CLs for the stable group (Fig. 3). The estimated rate of change was computed for each of the seven glaucoma axes, and the axis with the greatest rate of change (largest $m\Delta x$) was selected for display (Fig. 3).

Computation of the Degree of Confidence in Change (See Appendix). POP uses two probability density functions (PDFs) to compute the degree of confidence (DOC) of change on an axis, one derived from the inherent variability in VFs in stable glaucoma eyes and one modeled on the variability of an individual eye manifested as the distribution of the slope derived from LR in the eye being tested (Fig. 3). The LR line and its rate of change are actually the estimated mean of these values. The t-distribution of the rate of change in the test eye is calculated. Eyes with severity values at each visit that deviate little from the estimated mean slope have a narrow DOC distribution (Fig. 3), and the estimated mean rate of progression is more certain. Eyes with severity values scattered far from the slope have a broad DOC distribution, and the estimated rate of change is more uncertain. Both the CIs derived from the distribution of slopes of stable eyes and the t-distribution of slopes from an individual sequence of fields are used to determine the overall DOC of progression for an individual sequence of fields (Fig. 3). This combination of PDFs represents how certain we can be that the change really represents glaucomatous progression.

Pop Score. For each axis, the glaucomatous field defect in a test eye is classified as progressing if the cutoff proportion (see Matching Specificity below) of the DOC distribution of the LR line falls outside the 95% CL for stable glaucoma eyes. Of the axes that qualify for progression, the one with the greatest $m\Delta x$ is selected both for the overall DOC of progression and the rate of change for the eye. If no axis qualifies for progression, the selected axis is the one with the greatest $m\Delta x$. The selected axis shows us which VIM defined VF defect pattern has the most change. The POP score is the DOC of progression of the selected axis and is a surrogate for the probability of progression for the eye being tested.
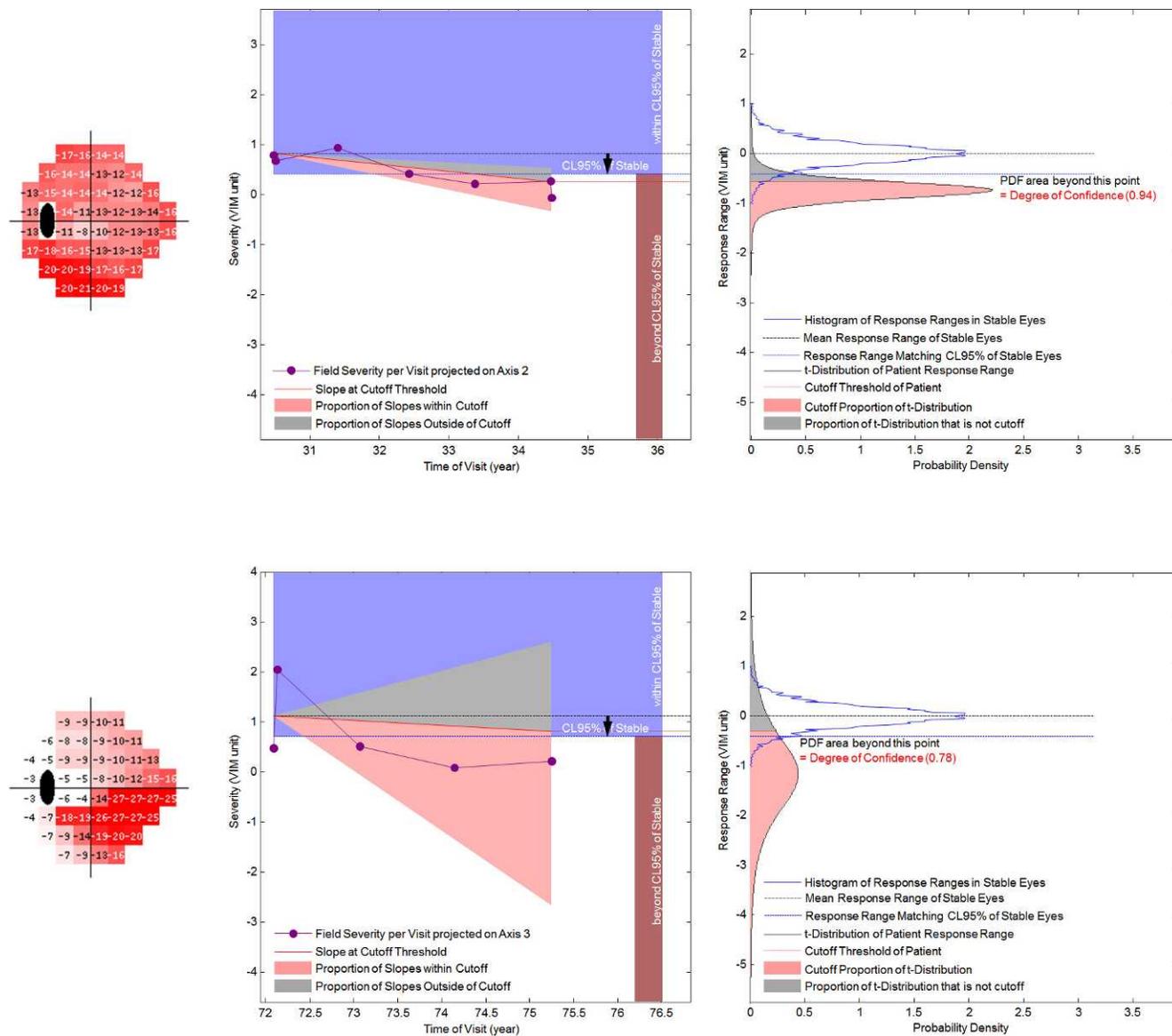
FIGURE 3. Display of the DOC and rate of progression in the axis with the maximal change in severity. The *upper panels* show an eye with little field variability; the *lower panels* show an eye with much field variability. The *left panels* are designed to display information about progression that can be easily assimilated. The *right panels* are the PDFs that are the basis of the *left panels*. *Far left* are gray scale displays of the selected axis. The *purple dots and lines* are the projected values of a sequence of visits in VIM units. The empirically derived PDF of response ranges in stable eyes is represented as a *blue bell shaped histogram* in the *right panels* and a *blue rectangle* in the *left panels*. The 95% CL is a *blue line* in the *right panels* and the inferior limit of the *blue rectangle* in the *left panels*. The *t*-distribution of the LR slopes of an individual eye's sequence of projected values on an axis is a *black bell shaped curve* in the *right panels*. The cutoff proportion is in *pink*, and the complement is in *gray*. The *triangle* in the *left panels* demonstrates the 95% CI of the slopes, with the cutoff proportion in *pink* and the complement in *gray*. The *red line* is the limit of the cutoff proportion. If the *red line* crosses the lower edge of the *blue rectangle*, the eye is classified as progressing (*upper panels*). If it remains within the *blue triangle*, the eye is classified as not progressing (*lower panels*).

Comparison of Pop to Available Change Measures: VFI, MD, and GPA

We compared the number of progressing eyes identified by POP with those identified by three familiar clinical measures of progression available in the Humphrey Visual Field Analyzer STATPAC software, VFI, MD, and GPA. VFI and MD each give continuous output that permits comparison of LR. POP assesses the rate and DOC of progression of the individual eyes' specific defect patterns. The VFI and MD scores are represented as a single global index of VF severity, which contain both signal (real change) and noise (non-progression field variability). GPA can signal likely progression no matter the location of the progressing test points, as long as at least three points

have progressed. In contrast, POP concentrates on the particular areas of the VF where there is the most change and eliminates noisy areas of the VF that have little or no real change, thereby improving the signal-to-noise ratio (SNR).

Matching Specificity Prior to Comparison of Change Algorithms

The POP, VFI, and MD progressions were determined for all VFs in the stable group. To reduce confounding variables, the same LR method and the same method for determining the progression DOCs were employed for POP, VFI, and MD. The cutoff value for the DOC of progression for each algorithm was determined based on a set

TABLE 1. Degree of Certainty Cutoff Values for 95% Specificity

C2-A1	C2-A2	C3-A1	C3-A2	C3-A3	C3-A4	C3-A5	VFI	MD
0.821	0.852	0.825	0.810	0. 935	0.822	0.855	0.501	0.500

The columns C3-A1 to C2-A2 are axis projections of VIM, with C3-A1, for example, meaning axis 1 of cluster 3. The DOC cutoff proportions for VIM axes are higher than the 50% proportion for VFI and MD, since VIM progression is designated if any axis out of seven is progressed. GPA is not included, because its method does not permit adjustment of specificity.

specificity of 95% (no progression) in eyes from the stable group. Because there are seven glaucoma axes (field defect patterns) in VIM, and progression is detected if any one axis is progressed, POP has seven chances to detect progression compared with only one chance for either VFI or MD. To compensate, the specificity of each axis was adjusted upwards to achieve an overall specificity of 95% for POP. This compensation resulted in larger DOC cutoff values for stability for the individual axes than those of the VFI and MD (Table 1). The cutoff areas for VFI and MD each remained at 50%. Equating for specificity prior to determining the percentage progressed minimized the effect of differences among the algorithm methods. Statistical methods are detailed in the Appendix.

In summary, the values of the 6,600 permuted regression slopes for the stable glaucoma eyes were distributed in a ranked list for each axis, and the boundary slope of the single-tail 95% CI was the CL of stability. The t-distribution of the rate of change around the estimated mean slope produced by regression accounted for the variability of the field in an individual eye. The percentage of area under the t-distribution curve of slopes for a test eye sequence that was outside of the limit of stability was the estimated DOC that the glaucomatous VF defect was deteriorating (Fig. 3). Any percentage could have been used to make the binary decision whether the glaucoma was progressing or not; for comparison purposes, progression was defined as the presence of a proportion of the t-distribution equal to or greater than the cutoff located outside the CL for stability.

For GPA, progression was defined for the full field if change greater than the variability observed between two baseline tests was repeatable at three of the same points in three consecutive exams (i.e., a GPA result of likely progression),[9] regardless of specificity in the stable group.

The McNemar's test with continuity correction was used to compare the sensitivities (number of eyes identified as progressed in each experimental group) of POP, VFI, and MD, as well as GPA.

RESULTS

At baseline, participants in the stable group (Dataset 2) were older and had more severe VF defects, by MD and PSD, than the individuals in test Dataset 3 (Table 2). The severity of glaucoma, as indicated by MD and PSD in Table 2, as well as the age of the participants from UCSD, NYEE, and UAB were similar. A total of 628 eyes (4,186 fields) of the 418 participants in the test group were tested for progression, which resulted in a mean $\pm$ SD of 6.7 ± 1.7 VFs (range 5 to 13 fields) per eye followed for 4.0 ± 1.4 years (range 1.8 to 9.2) for an average interval of 0.7 years.

Figure 3 demonstrates how the two probability distributions (for stable eyes and for the eye being analyzed) are combined to give a DOC for progression. This display shows the CL for 95% specificity derived from the stable dataset and illustrates the effect of test variability on the DOC of progression. The proportion of the regression distribution for an individual sequence that is beyond the CL for 95% specificity represents the DOC of progression. Two example eyes progressing on axes 2 and 3 in cluster 3 show the effect of variability of the fields in a sequence. One eye with low variability (top) has 94% DOC of progression. This proportion is greater than the cutoff proportion (0.81), and the eye is classified as progressing. Another eye with high variability (bottom) has 78% DOC of progression. This proportion is less than the cutoff proportion (0.935), and the eye is classified as stable. The suspect and glaucoma columns in Table 3 were mutually exclusive. With the specificity fixed at 95%, the proportion of eyes detected as progressing was higher in the glaucoma group than the suspect group (Table 3).

The Venn diagrams (Fig. 4) compared the number of eyes detected by POP, VFI, and MD; GPA was not included in the diagram because it generally performed less well. Only 28% of the eyes detected by any method were detected by all methods (37% for PGON eyes). Of eyes detected by any method, 18% were detected by POP only (11% for PGON eyes), compared with 6% by VFI only, and 21% by MD only (7% and 17%, respectively, for PGON eyes). All these were statistically similar except that MD was better than VFI ($P = 0.05$) for all eyes tested.

McNemar's tests indicated that in the suspect group, LR of MD identified significantly more eyes as progressed than LR of VFI ($P = 0.01$) and GPA ($P = 0.03$). No other comparisons were significantly different (all $P \geq 0.10$). In the glaucoma group, LR of both POP and VFI identified significantly more eyes as progressed than GPA ($P = 0.01$ and $P = 0.02$, respectively); no other comparisons were significantly different (all $P \geq 0.06$). Finally, in the PGON group, LR of both POP and MD identified significantly more eyes as progressed than GPA ($P = 0.05$ and $P = 0.02$, respectively); no other comparisons were significantly different; all $P \geq 0.17$).

DISCUSSION

POP, a purely mathematical approach that learns from data without human intervention, performs similarly to VFI and MD,

TABLE 2. Age and Glaucoma Characteristics of Subjects

Dataset	Source	n	Age, y Mean $\pm$ SD	MD, dB Mean $\pm$ SD	PSD, dB Mean $\pm$ SD
2 (Stable)	Stable eyes	55	70.3 ± 10.0	-8.7 ± 6.6	7.4 ± 4.21
3 (Test)	Total test eyes	628	60.0 ± 12.2	-1.8 ± 2.9	2.7 ± 2.6
	UCSD	343	62.0 ± 12.5	-1.9 ± 2.6	2.5 ± 2.3
	NYEE	126	57.4 ± 11.0	-2.6 ± 3.5	3.4 ± 3.4
	UAB	159	57.7 ± 11.5	-1.0 ± 2.8	2.6 ± 2.4

The sources of the test eyes were the University of California at San Diego (UCSD), New York Eye and Ear Infirmary (NYEE), and the University of Alabama at Birmingham (UAB).

TABLE 3. Detection of Progression

		Suspects*	Glaucoma:† VF + GON	Total	PGON‡
n		478	150	628	76
			Mean ± SD		
Characteristics at baseline	Age	59.1 ± 12.0	62.9 ± 12.2		65.4 ± 9.91
	MD [dB]	−0.89 ± 1.86	−4.74 ± 3.73		−2.70 ± 3.22
	PSD [dB]	1.88 ± 1.14	5.42 ± 3.79		3.78 ± 3.39
			Number of Eyes (Percent of n) Rate of Decline in VIM Units		
Progression detected	POP	18 (3.8%), $r = -2.01$	24 (16.0%), $r = -2.93$	42	20 (26.3%), $r = -2.93$
	VFI	13 (2.7%), $r = -1.43$	23 (15.3%), $r = -1.60$	36	18 (23.7%), $r = -1.65$
	MD	27 (5.6%), $r = -1.72$	18 (12.0%), $r = -1.84$	45	21 (27.6%), $r = -1.89$
	GPA	14 (2.9%)	11 (7.3%)	25	11 (14.5%)

* Suspects: eyes with OHT, GON, or VF abnormality.
† Glaucoma: both glaucomatous VF defect and GON present.
‡ PGON: progression of glaucomatous optic neuropathy in sequential stereophotographs.

and better than GPA in early and moderate glaucoma and in eyes demonstrated photographically to be progressing (PGON). As advantages for POP, the progression is along glaucomatous field patterns that are recognizable to healthcare providers, and POP information can be presented in clinically useful displays.

The eyes in Dataset 3 were separated into suspect and glaucoma eyes to assess the action of POP and the other progression detection methods at these two levels of disease. Without an indication of progression, it was possible to compare the sensitivity of each progression detection method at fixed specificity, but it was unclear which method had the fewest false positives and false negatives. PGON was found in both suspect and glaucoma eyes. Since PGON represented evidence of glaucomatous progression in the optic nerve, a method different than VFs, higher sensitivity at fixed specificity in eyes with PGON would more likely be indicative of better detection of progression.

It has long been recognized that glaucomatous eyes have larger VF variability than eyes without VF defects.[3-7] Although increased VF variability masks progression, the widespread use of this test necessitates the development of effective methods that extract progression information from VFs. Unlike currently accepted techniques for detecting progression, POP seeks to optimize the change information in the VF and to account for test variability by focusing on the axis with the best SNR.

The SNR indicates the level of masking of the desired signal by noise. It is useful to consider the change in severity of glaucomatous field damage as the desired signal and the variability in VF measurements as noise masking this signal. As the VF deteriorates, not all regions of the VF are changing. The

FIGURE 4. Proportional Venn diagrams indicating that the three methods of detection frequently do not detect progression in the same eyes. The *left diagram* displays the distribution of eyes detected as progressing in all suspect and glaucoma eyes; the *right diagram* displays the distribution in PGON eyes.

damaged VF regions that are not changing have little of the desired signal; nevertheless, they are noisy and can degrade performance of global detectors that use the whole VF. POP uses machine learning pattern recognition procedures to concentrate on the patterns with highest change signal and to disregard the regions with little change signal, thereby improving the SNR.

Reduction in dimensionality improves the SNR. VIM segments the VF into patterns and POP concentrates on the pattern with the best SNR. The original 53D of VF locations plus age are initially reduced to seven dimensions (the seven glaucoma axes). The selection of the single pattern with the most change further reduces the dimensions to one, which ensures the highest SNR, and turns a complex analysis into a univariate analysis. The maximal statistical independence of the VF patterns in VIM justifies considering the analysis in POP to be univariate.

POP demonstrated that the worsening of glaucomatous field damage in the VIM environment manifests as widening and deepening of a VF defect pattern. POP seeks to detect glaucomatous progression in the shortest time, when the pattern with the most change is likely to remain the principal pattern of progressive damage. Because the affected VF defect pattern in VIM is visible and recognizable to glaucoma practitioners, they can develop an understanding of how the glaucoma is worsening.

Disease severity positively correlates with the DOC of finding progression, as seen in Table 3. The eyes in the stable group had, on average, more glaucomatous field damage than most of the eyes being tested. Thus, the CL for stability was determined by eyes with more severe disease than most of the eyes tested. Setting the CL for stability in more damaged eyes could affect the accuracy of POP; however, the likelihood is that POP was rendered less sensitive for detecting progression in less severely damaged eyes, because the CL was set too conservatively. Nevertheless, POP performed similarly to VFI and MD and, in general, GPA performed less well than POP and MD.

A possible reason for the poorer performance of GPA could be its reliance on individual locations in sick areas that tend to be noisy. POP, VFI, and MD rely on groups of locations (POP) or all locations (VFI and MD). The grouping can reduce the effect of individually noisy locations. In addition, GPA progression is defined relative to the variability present in the first two baseline VFs. If the variability in individual locations between these two specific fields is large, it is unlikely that progression will be detected. This is not the case for LR of POP, VFI, or MD.

Dataset 3, tested for progression, had both single eyes and paired eyes from the same patient. POP was compared with
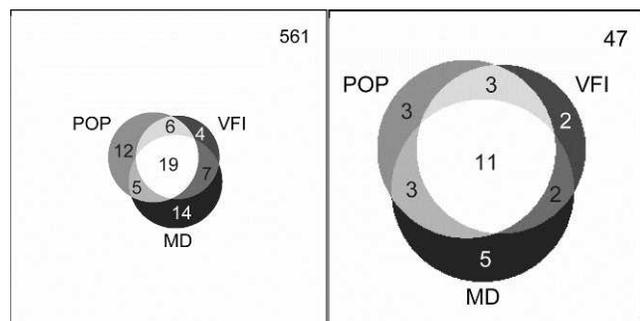
VFI, MD, and GPA. All four tests were presented the same set of eyes. It was assumed that the proportion of paired eyes would not affect the comparison.

Any method used to classify eyes as normal or glaucomatous can have false negatives and false positives. This principal is also true for determining GON or PGON. By requiring two out of three expert evaluators of stereoscopic photos to agree that the optic nerve head shows change in glaucomatous optic neuropathy attempts to reduce the errors. Whatever defects there may be in classification of PGON, those defects would be the same for all the classification methods assessed for progression of VFs.

Fixing the specificity at 95% for all tests made it possible to compare directly the relative effectiveness of POP, VFI, and MD at recognizing progression. GPA was not amenable to that approach. Whether the severity of field damage was in the suspect or glaucoma range, POP identified a similar number of eyes as progressing. In the PGON eyes, indicated to be deteriorating because of confirmed progressive disk or nerve fiber damage, similar results were observed. However, POP, VFI, MD, and GPA did not detect a particularly large percentage of PGON eyes. This probably is due in part to the use of optic disk progression to determine whether progression is occurring in VFs. Eyes appear to progress at different rates when assessed using functional and structural measurements.[16] Several studies have suggested less than ideal agreement between functional- and structural-based change detection.[17-19] This difference suggests that combining both field and structural tests to detect progression might find more progressing eyes. Additionally, statistically accounting for seven chances to detect progression in POP may have led to an overly conservative detection method. For example, assigning a cutoff proportion based on 95% specificity for each axis in POP classifies progression in many more eyes than VFI or MD, but the overall specificity of POP is reduced because there are seven chances to find progression in POP compared with one in VFI or MD.

The rate of progression was 1.7 and 1.6 times faster with POP than with VFI or MD, respectively, in eyes identified as progressing by PGON. The likely explanation is that POP concentrates on the defect patterns that progress the most and ignores the stable patterns. VFI and MD, on the other hand, are global tests that do not focus on progressing patterns.

Whereas the design of a method to separate classes (e.g., stable and progressing) based on human perception and reasoning can be successful, the dependence on human reasoning opens the method to the possibility of human bias and inadequate perception of the factors that can best separate these classes.[20] The lesser performance of GPA could be an example of these limitations. A good pattern recognition approach can learn from data how to approach the ideal separation of the classes given the available features, and it can do so without human input and, thus, without the risk of human bias. The indication that POP was as good a detector of progression validates a rigorous mathematical approach to separating progressing from stable eyes, giving POP the potential to be a useful tool for interpreting VF change.

In summary, POP was designed to maximize the SNR while identifying progressing eyes and make the best use of VF data while avoiding human biases during identification of progressing eyes. POP accounted for two major probability distributions and displayed the information in a manner that would bring understanding and intuition to the practitioner who has to decide whether the current management is working or if additional more aggressive treatment is necessary. POP shows promise in satisfying these clinical needs.

APPENDIX A

Variational Bayesian Independent Component Analysis Mixture Model (Vim)

Independent component analysis (ICA) finds a single set of axes or directions such that the distributions of the data projections onto the axes are as statistically independent as possible. ICA uses a measure of independence to align the axes to achieve that statistical independence. In contrast, the VIM, the mixture model of ICA, was constructed by us to permit more than one set of ICA axes.[10,11] VIM classifies multidimensional data into mutually exclusive clusters and, within each cluster, simultaneously uses ICA to extract the local features to create and align its own set of statistically independent axes. Hence, VIM separates the original distribution of data into clusters and axes that have their own distinct patterns. Our application of VIM represents SITA VF data as axes (VF defect patterns; Fig. 1 in main text) within clusters of normal and glaucomatous fields; yielding two axes for the normal fields and seven axes for seven distinct glaucomatous VFs.

Projection of Fields onto Vim Axes for Progression of Pop Analysis

The point in 53D space (absolute sensitivities in 52 perimetry locations plus the patient age) for each VF is projected onto each axis by a single appropriate line between the field and the axis (Fig. 2 in main text). To make the severity equivalent on all axes, the projections on each axis are then normalized to VU of SD by dividing the severity offset from the cluster mean by the severity value of one SD, computed from the cluster to which the corresponding axis belongs. Hence, a VU of 2 SD offset along an axis corresponds to pattern severity being 2 SDs away from the cluster mean. Figure 1 in the main text shows the
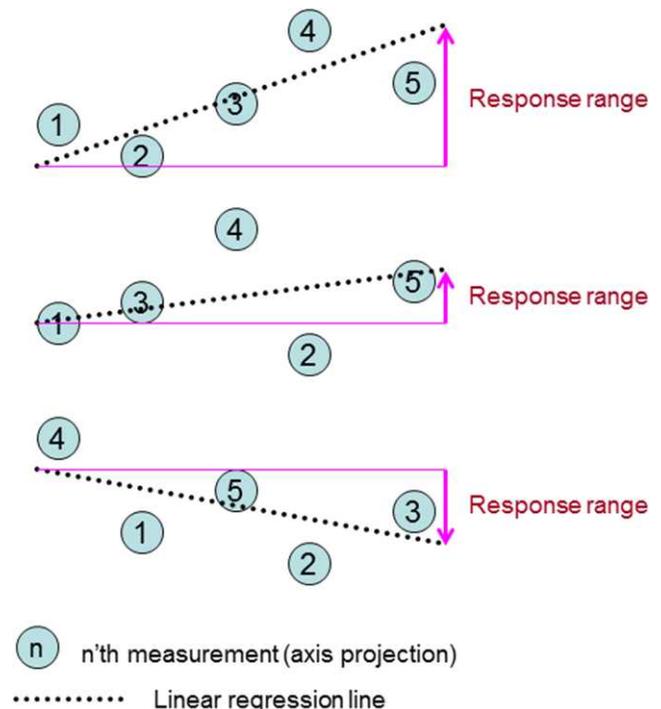

FIGURE A1. Three example permutations (out of 5! = 120) from the same test eye, yielding three different slopes and response ranges.

seven VIM-derived patterns of glaucomatous field defects located on the seven axes at 2 VU equals 2 SD from the cluster mean on the plus side of the axes, displayed in the style of a total deviation plot in Statpac. The absolute sensitivity plot of a pattern generated at some point on an axis is converted into a pattern simulating the total deviation plot by subtracting the absolute sensitivity plot of the pattern generated at the mean of the normal cluster (cluster containing mostly normal fields) from the generated absolute sensitivity plot on the axis.

Linear Regression of VF Sequence Projected on an Axis

The directions of the axes are assumed to be the direction of changing severity. The projection of a sequence of fields from an eye onto the VIM axes represents the change in severity of the specific VF defect represented by the axis. LR is applied to each sequence of fields, and the slope and the intercept at the time of the first field in the sequence, denoted by m and b respectively, are obtained from

$$y_{tae} = \hat{m}_{ae}x_{te} + \hat{b}_{ae} + \hat{\varepsilon}_{tae} \qquad (A1)$$

where y_{tae} is the field severity (in VU) on axis a for the field obtained at the time of visit t projected onto axis a for sample e; x_{te} is the time at visit t for sample e; $\hat{m}_{te}$ is the estimated slope (of severity change) of the regression line for sample e on axis a; $\hat{b}_{ae}$ is the estimated field severity at baseline ($x_{1e} = 0$) for sample e on axis a; and $\hat{\varepsilon}_{tae}$ is the estimated offset from the regression line to y_{tae} for visit t. The offsets are errors that are minimized as a result of LR.

We define the change along the regression line as the response range by

$$\hat{R} = \hat{m}(x_s - x_1) \qquad (A2)$$

which equals $\hat{m}$ for stable data (Fig. A1). $\hat{R}$ is used as a surrogate for the expected mean of the slope of the regression line and fluctuation range for the rate of change or slope m of a stable eye. The response range is better than max range, ($y_{max} - y_{min}$) or the difference between the first and last visits ($y_5 - y_1$) in that the regression is a smoothing operation that is less dominated by outliers.

Stable Data Analysis

The stable dataset compiled by Anderson (see Dataset 2 in the Methods section in main text) is composed of five VF measurements collected weekly over a month to simulate stable severity. The assumption is that the variability of VF testing in this set should be due to factors other than glaucomatous progression. In the stable database, 55 eyes had at least two out of five tests identified to be glaucomatous by the Humphrey VF Analyzer methods of PSD triggered at 5% or worse or of a GHT result of outside normal limits. To provide a larger sample size and reduce quantization error, we permuted the original order of each sequence of fields and generated all the possible sequences from five fields. The number of all the permutations was 120 (5!) for each eye, so the total number of sequences for 55 eyes was 6600.

The original interval of the visits in the stable group was one week. As a surrogate for stable eyes in a clinical setting, the interval is set to be longer, for example, a year. Since the interval between visits is undefined in the application of the stable group, the time variable, x_{taep}, is set to be equally spaced between 0 and 1 (0.00, 0.25, 0.50, 0.75, and 1.00 for visits 1, 2, 3, 4, and 5, respectively), for five data points for permutation P, and the field severity, y_{taep}, is the permuted axis projection. For example, y_{tae} can be ordered by (y_{1ae}, y_{2ae}, y_{3ae}, y_{4ae}, y_{5ae}),

(y_{3ae}, y_{2ae}, y_{1ae}, y_{5ae}, y_{4ae}), and so on according to the permutation order.

Response ranges from all 6600 permutations of all the stable eyes are considered. The probability density function for stable data is obtained by ranking the response range from least to most severe (Fig. A2). The resultant histogram is different for each axis (Fig. A3). The single-tail 95% cutoff for stability is the response range at the 95th percentile (Fig. 3 in the main text). Since the 95% cutoff for stability represents 95% true negative rate, it also represents 95% specificity.

Test Eye Analysis: Probabilistic Progression Determination

Data Preparation. In addition to the probability distribution of rate of change for stable eyes (PDF of stability), there is also the distribution of the slope of the regression line for a given sequence of fields (test eye PDF), given the variability of the patient's response. A test eye's sequence of fields over the period of observation is projected onto an axis (Fig. 2), and LR is carried out on the sequence of projected values. The most probable regression line is the estimated mean regression line from the sequence of projections on an axis from each field in the sequence over the time of the observations, assuming that the severity change is linear. A t-distribution PDF of the response range (representing the slope) is generated from the LR. The approximated distribution (Fig. 3) is wider when the serial data points are more scattered (larger εs), and narrower when the data are more closely aligned to the estimated mean regression line (smaller εs). The distribution surrounded by 95% CLs of the response range is called a 95% prediction interval (triangle, Fig. 3), which is used to display the variability of the slope for an individual test eye in Figure 3. The proportion of this generated PDF outside of the stability cutoff
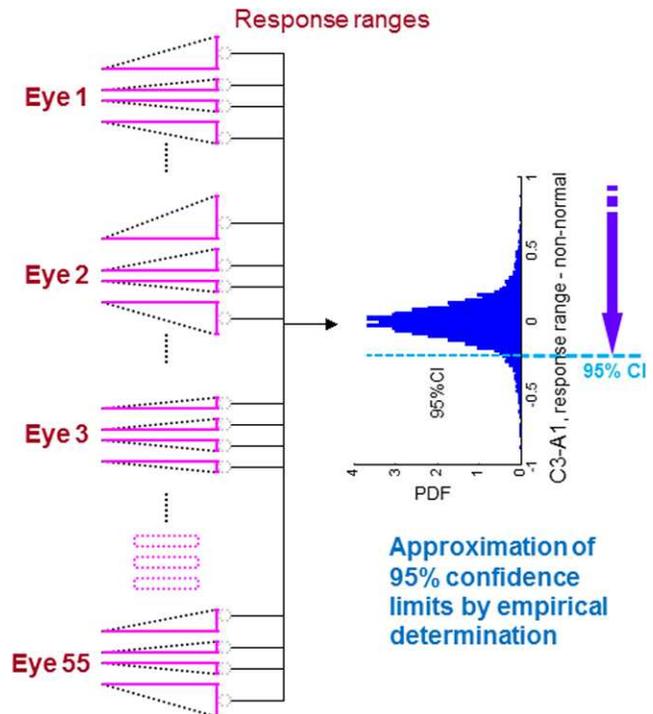

FIGURE A2. Obtaining CLs of response ranges. Response ranges from all 6600 permutations from the 55 stable eyes are distributed in a histogram, and the 95% CL is obtained.
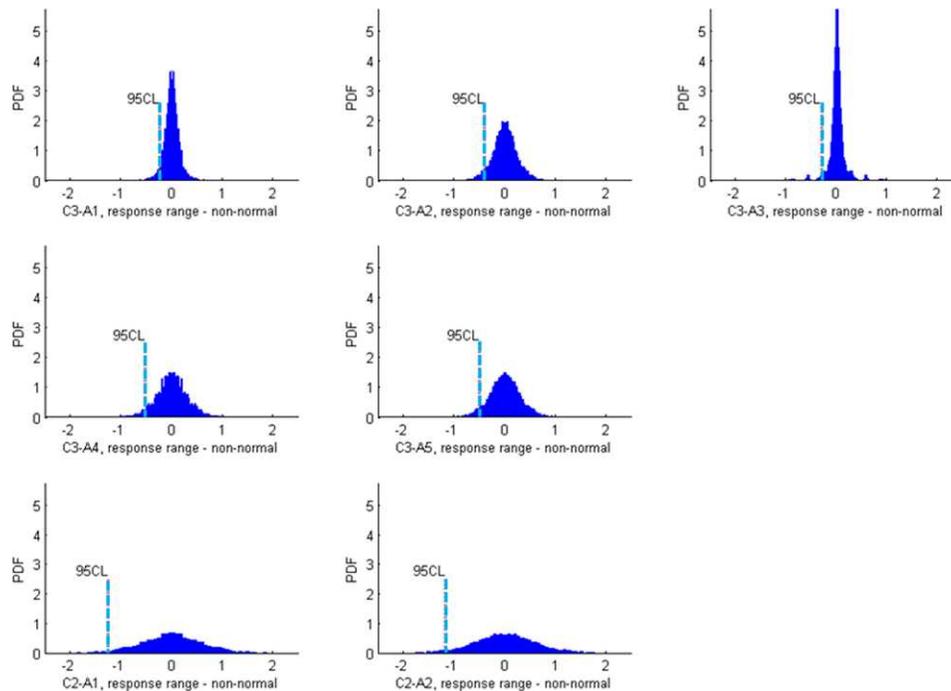
FIGURE A3. Histogram plots of response ranges, for all seven axes from C_3 and C_2. Their 95% CLs are shown by *blue lines*. The mean of response ranges is zero on each axis, because any permutation of a stable sequence has its reverse sequence.

represents the certainty level of progression (pink part of triangle in Fig. 3).

Pop Algorithm. The POP algorithm computes the DOC that the response range of the measurement sequence of a patient's eye differs from the stable group. POP score serves as a surrogate for the probability of the progression of POP. LR is defined as progression detection by LR of values on an axis from projection of a sequence of fields. Recall the regression equation A1, where $\hat{m}_{ae}$ is the estimated slope of the current sequence of axis projections; μ_m, S_m is the estimated mean and standard error of slope m_{ae}; and $T = x_{last} - x_{first}$ is the time difference of the first and the last measurements in a particular sequence. In PDF, we have the overall PDF of stable eyes empirically derived from stable data and the individual PDF estimated for each test eye modeled as a t–distribution. The steps are:

1. Determine stable PDF: acceptable rate of change in severity for 95% of stable eyes built from the empirical distribution of rates of change in 6600 permutations in the stable group.
2. Account for test eye PDF: generate a t-distribution of regression slope built from a given eye's variability (in 5 to 20 visits), modeled by Student's t-distribution
3. Compute the POP score and DOC using stable and test eye PDFs:

$$DOC_{a,CL@95\%}(eye) = Pr(\hat{m}_a T < R_{a,CL@95\%})$$
$$= F_{t,v=n-2}\left(\frac{R_{a,CL@95\%} - \mu_{m,a}T}{s_{m,a}T}\right) \quad (A3)$$

where $DOC_{a},CL@95\%$ is the DOC using 95% CL for axis a for stable eyes; $R_{a},CL@95\%$ is the designated CL of response range from stable data for axis a; $F_{t,v=n-2}$ is the cumulative t-distribution function; and v is degrees of freedom, which is $n - 2$ for the t-distribution in LR.

4. Progression determination: test eye is designated progressed if, on any axis, $DOC_{a,CL@95\%}(eye)$ greater than

$DOC_{a,cutoff}$. The proportion of the PDF of the test eye that is outside the 95% CL for stable eyes must be equal to or greater than the cutoff in the DOC (Table 1 in main text) for the particular axis (Figs. 3, A3).

Normalization of the Doc Cutoff Value Computation from Stable Group for Comparison of Pop, Vfi, and Md

We compared POP with two conventional glaucoma severity measures: VFI,[8] and MD score [12] Please see the Methods section of the main text for details about VFI and MD. For each VF test, POP generated the severity value on each of the seven POP axes, while MD scoring and VFI calculation generated a single global severity. Since POP could look for progression on seven axes compared with one measure for VFI and MD, POP would have more opportunity to find progressing eyes than MD and VFI. Setting each of the seven axes at 95% specificity lowered the overall specificity for POP, which increased the number of eyes classified as progressing.

To make a fair comparison, it was desirable to set the overall specificity of POP to be the same as VFI and MD. We used stable data to set the overall specificity of POP. We can assume that the stable group eyes are all stable, so the specificity of each of the axes was adjusted such that POP identified 95% of the stable eyes as stable and the proportion of stable eyes identified as stable remained the same for each axis.

Out of total 6600 permutations (55 eyes $\times$ 120 permutations/eye), 3300 were used to compute the 95% CL for each axis (teaching stable set), and the other 3300 were chosen for specificity matching (test stable set). Starting at 95% specificity for each axis, the specificity of each axis was raised equally until the overall specificity for POP was 95%. Since the number of sequences used for specificity matching was 3300, the number of permutations selected as stable had to be 3135 to achieve overall specificity of 95% for POP. The cutoff proportion of the t-distribution that had to be outside the

95% CI for each axis (Table 1 in main text) was derived from the teaching set of stable data to identify 95% of the test set of stable data as stable (95% specificity) for VFI, MD, and the overall POP score. Whereas the *t*-distribution cutoffs for VFI and MD were 0.50, the cutoffs for the POP axes range from 0.81 to 0.935.

References

1. Goldberg I. How common is glaucoma worldwide? In: Weinreb RN, Kitazawa Y, Krieglstein G, ed. *Glaucoma in the 21st Century.* London, England: Mosby International; 2000: 3–8.

2. Quigley HA. Number of people with glaucoma worldwide. *Br J Ophthalmol.* 1996;80:389–393.

3. Flammer J, Drance SM, Fankhauser F, Augustiny L. Differential light threshold in automated static perimetry. Factors influencing short-term fluctuation. *Arch Ophthalmol.* 1984;102: 876–879.

4. Flammer J, Drance SM, Zulauf M. Differential light threshold. Short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol.* 1984;102:704–706.

5. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1989;108:130–135.

6. Lewis RA, Johnson CA, Keltner JL, Labermeier PK. Variability of quantitative automated perimetry in normal observers. *Ophthalmology.* 1986;93:878–881.

7. Wilensky JT, Joondeph BC. Variation in visual field measurements with an automated perimeter. *Am J Ophthalmol.* 1984; 97:328–331.

8. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol.* 2008;145: 343–353.

9. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology.* 1999;106:2144–2153.

10. Goldbaum MH. Unsupervised learning with independent component analysis can identify patterns of glaucomatous visual field defects. *Trans Am Ophthalmol Soc.* 2005;103:270–280.

11. Goldbaum MH, Jang GJ, Bowd C, et al. Patterns of glaucomatous visual field loss in sita fields automatically identified using independent component analysis. *Trans Am Ophthalmol Soc.* 2009;107:136–144.

12. Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol.* 1999;117:573–583.

13. Sample PA, Girkin CA, Zangwill LM, et al. The African Descent and Glaucoma Evaluation Study (ADAGES): design and baseline data. *Arch Ophthalmol.* 2009;127:1136–1145.

14. Racette L, Liebmann JM, Girkin CA, et al. African Descent and Glaucoma Evaluation Study (ADAGES): III. Ancestry differences in visual function in healthy eyes. *Arch Ophthalmol.* 2010; 128:551–559.

15. Goldbaum MH, Sample PA, Zhang Z, et al. Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci.* 2005;46:3676–3683.

16. Chauhan BC, McCormick TA, Nicolela MT, LeBlanc RP. Optic disc and visual field changes in a prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography with conventional perimetry and optic disc photography. *Arch Ophthalmol.* 2001;119:1492–1499.

17. Chauhan BC, Nicolela MT, Artes PH. Incidence and rates of visual field progression after longitudinally measured optic disc change in glaucoma. *Ophthalmology.* 2009;116:2110–2118.

18. Strouthidis NG, Scott A, Peter NM, Garway-Heath DF. Optic disc and visual field progression in ocular hypertensive subjects: detection rates, specificity, and agreement. *Invest Ophthalmol Vis Sci.* 2006;47:2904–2910.

19. Xin D, Greenstein VC, Ritch R, Liebmann JM, De Moraes CG, Hood DC. A comparison of functional and structural measures for identifying progression of glaucoma. *Invest Ophthalmol Vis Sci.* 2011;52:519–526.

20. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science.* 1974;185:1124–1131.