Detection of Functional Change Using Cluster Trend Analysis in Glaucoma

Stuart K. Gardiner, Steven L. Mansberger, and Shaban Demirel

Devers Eye Institute, Legacy Research Institute, Portland, Oregon, United States

Purpose. Global analyses using mean deviation (MD) assess visual field progression, but can miss localized changes. Pointwise analyses are more sensitive to localized progression, but more variable so require confirmation. This study assessed whether cluster trend analysis, averaging information across subsets of locations, could improve progression detection.

Methods. A total of 133 test–retest eyes were tested 7 to 10 times. Rates of change and P values were calculated for possible re-orderings of these series to generate global analysis ("MD worsening faster than x dB/y with $P < y$"), pointwise and cluster analyses ("n locations [or clusters] worsening faster than x dB/y with $P < y$") with specificity exactly 95%. These criteria were applied to 505 eyes tested over a mean of 10.5 years, to find how soon each detected "deterioration," and compared using survival models. This was repeated including two subsequent visual fields to determine whether "deterioration" was confirmed.

Results. The best global criterion detected deterioration in 25% of eyes in 5.0 years (95% confidence interval [CI], 4.7-5.3 years), compared with 4.8 years (95% CI, 4.2-5.1) for the best cluster analysis criterion, and 4.1 years (95% CI, 4.0-4.5) for the best pointwise criterion. However, for pointwise analysis, only 38% of these changes were confirmed, compared with 61% for clusters and 76% for MD. The time until 25% of eyes showed subsequently confirmed deterioration was 6.3 years (95% CI, 6.0-7.2) for global, 6.3 years (95% CI, 6.0-7.0) for pointwise, and 6.0 years (95% CI, 5.3-6.6) for cluster analyses.

Conclusions. Although the specificity is still suboptimal, cluster trend analysis detects subsequently confirmed deterioration sooner than either global or pointwise analyses.

Keywords: perimetry, analysis, progression

Clinicians and researchers use perimetry to measure peripheral visual function in glaucoma, since this relates directly to quality of life and activities of daily living. Functional testing also forms a key part of clinical trials. Indeed, the United States Food and Drug Administration has indicated that structural measures would only be accepted as primary outcomes of clinical trials in glaucoma if they "show a strong correlation ($R^2 \approx 0.9$) to current vision or future vision."[1] Most commonly, functional testing is performed by using white-on-white standard automated perimetry (SAP). SAP estimates the pointwise sensitivity at different locations in the visual field, with various test patterns being available on different instruments. These are then summarized by one or more global indices such as mean deviation (MD) and the visual field index (VFI). These measures are used to interpret the test results and assess the rate of glaucomatous progression in an eye, reducing problems related to the substantial variability present in pointwise measurements.[2-6] We have recently shown that MD detects change sooner than other global indices (e.g., VFI) in early glaucoma, for equal specificity.[7]

However, glaucoma can commonly result in localized visual field defects, which global indices (e.g., MD) are not designed to detect. For example, in the early stages of the disease, a defect spanning five locations could be considered clinically significant; yet this represents less than 10% of the locations tested in the 24-2 visual field. Global indices may miss the development and/or progression of such a defect, because changes are obscured by the variability of the remaining 90% of unaffected locations. Furthermore, some global indices can be affected by generalized functional deterioration that can be associated with cataract or systemic conditions, rather than being specific to glaucoma.[8]

An alternative approach is to use pointwise analyses, considering each visual field location separately. In particular, pointwise linear regression assesses the statistical significance of change at each location[9] and has been used to assess the likelihood of progression in clinical trials.[10] However, this approach lacks specificity owing to the high test–retest variability of pointwise sensitivities, especially at locations where visual field damage has occurred. Therefore, most protocols require that the rate of change be worse than a predefined value[11] and have confirmation of changes by subsequent visual fields before determining functional progression.[12]

To better detect localized changes, without the signal being dampened by pointwise variability, software has been developed to detect change in clusters of locations. The Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Inc., Dublin, CA, USA) presents the Glaucoma Hemifield Test (GHT),[13] which categorizes eyes on the basis of differences between the sensitivities within 10 predefined clusters of locations, and so can aid diagnosis of early damage. EyeSuite software developed to accompany the Octopus perimeter (Haag-Streit, Inc., Bern,

Switzerland) calculates whether the average sensitivity within each of 10 predefined clusters is deteriorating significantly more rapidly than would be expected from normal aging.[14] De Moraes et al.[15] have recently presented a review of the different current strategies to detect functional loss and deterioration.

In this study, we compared these three techniques (global analyses, pointwise analyses, and cluster analyses) for detection of visual field change in a large cohort with a long period of longitudinal follow-up. We tested the hypothesis that cluster trend analysis provides a good compromise between global analyses and pointwise analyses, allowing rapid detection of functional changes without excessive false-positive determinations. This would then allow researchers and clinicians to detect visual field progression with higher sensitivity and a decreased number of visual field tests.

Methods

Participants: Test–Retest Cohort

A test–retest cohort was acquired from the Assessing the Effectiveness of Imaging Technology to Rapidly Detect Disease Progression in Glaucoma (RAPID) study, performed at Moorfields Eye Hospital, London, United Kingdom. The RAPID study was undertaken in accordance with good clinical practice guidelines and adhered to the tenets of the Declaration of Helsinki. It was approved by the North of Scotland National Research Ethics Service Committee, and NHS Permissions for Research were granted by the Joint Research Office at University College Hospitals NHS Foundation Trust.

Inclusion and exclusion criteria for the test–retest cohort were chosen to match those for the United Kingdom Glaucoma Treatment Study,[16,17] including a diagnosis of primary open angle glaucoma. Participants underwent testing approximately weekly for an average of 10 weeks, including SAP with the same test procedures and exclusion criteria given above for the longitudinal cohort.

Participants: Longitudinal Cohort

This was a retrospective cohort study. Participants in the Portland Progression Project (P3) were recruited to a tertiary glaucoma clinic at Devers Eye Institute. Inclusion criteria were simply a diagnosis of primary open angle glaucoma and/or likelihood of developing glaucomatous damage, as determined subjectively by each participant's physician, in order to reflect current clinical practice. Exclusion criteria were an inability to perform reliable visual field testing, best-corrected visual acuity at baseline worse than 20/40, cataract or media opacities likely to significantly increase light scatter, or other conditions or medications that may affect the visual field. All protocols for this study were approved and monitored by the Legacy Health Institutional Review Board, and adhered to the Health Insurance Portability and Accountability Act of 1996 and the tenets of the Declaration of Helsinki. All participants provided written informed consent once all of the risks and benefits of participation were explained to them.

Participants in the longitudinal cohort are tested every 6 months with a variety of structural and functional tests. SAP was performed with an HFAII perimeter, using the 24-2 test pattern, a size III white-on-white stimulus, and the SITA Standard algorithm.[18] Only tests with <15% false positives and <33% fixation losses were used. For this study, only eyes with series of at least five reliable tests by these criteria were included in the analysis.

Analysis: Deriving Criteria for Change by MD

We first derived criteria for change that have exactly equal specificity in a test–retest data set, using the same method as in our recent publication that compared global indices with each other.[7] The first five visual field tests performed on participants in the test–retest cohort were assigned artificial "test dates" at 6-month intervals, to match the typical inter-test interval in the longitudinal cohort. Then, the rate of change of MD was calculated by linear regression. If the rate of change was in the direction of apparent worsening of the visual field over time, then the significance of the rate of change (the P value from a two-sided t-test) was recorded; otherwise, the series was assigned $P = 1.000$. All analyses in this study were performed by using the R statistical programming language.[19]

This process was repeated for all 120 possible re-orderings of the first five tests per eye. Hence there are $(120 * N)$ P values, where N represents the number of eyes in the test–retest cohort. The fifth percentile of these P values, that is, the 798th smallest value (based on 120 permutations for each of 133 eyes), was defined as the "critical value" for MD, labeled $Crit_{MD}$. Therefore, 5% of these artificial series for MD were worsening with $P < Crit_{MD}$ in this test–retest data set in which we know that no true change is likely to have occurred, such that this criterion has specificity equal to 95% in this data set.

Some previous studies have also imposed a minimum slope criterion when defining "progression." That is, instead of a criterion of the form "deteriorating with $P < P_{Crit}$," they instead used criteria of the form "deteriorating with rate worse than x dB/y that is also significant with $P < P_{Crit}$." Most notably, pointwise linear regression has commonly been applied by using criteria of the form "locations deteriorating with rate worse than -1 dB/y that are significant with $P < 1\%$."[9,11,12,20] Therefore, we defined a series of such criteria for different values of x. The value of P_{Crit} here was defined as the 798th smallest P value among series that had a rate worse than x, ensuring that each of these criteria still had 95% specificity. If fewer than 798 series had rate worse than x dB/y, then no criterion could be derived for that value of x. Note that setting $x = 0$ gives the criterion $Crit_{MD}$ described in the previous paragraph.

It may be predicted that the optimum criteria could depend on series length. Therefore, the above procedure was repeated to derive equivalent criteria with 95% specificity in series of length 7 tests, based on all 5040 possible re-orderings of the first seven tests per eye in the test–retest cohort. Equivalent criteria were also derived with 95% specificity in series of length 10 tests, but in this case based on 500 randomly chosen re-orderings of the first 10 tests per eye (in the 116 eyes out of 133 that had at least this many tests in their series); it was considered impractical and unnecessary to use all 3.6 million possible re-orderings of series of this length.

Analysis: Deriving Criteria for Change by Pointwise Analysis

For each series of visual fields, the pointwise total deviation values (i.e., the difference from age-matched normal subjects) were regressed against time, and the associated P value was recorded, for each of the 52 non-blindspot locations in the 24-2 visual field. The 52 resulting P values for each eye (one per location) were sorted starting with the smallest. This allowed criteria to be defined in the same way as for MD, but based on different numbers of locations. The fifth percentile out of the list of smallest P values was labeled $Crit_{1Loc}$; hence, a criterion of "at least one location deteriorating with $P < Crit_{1Loc}$" has

specificity equal to 95%. Similarly, the fifth percentile out of the list of second-smallest P values was labeled $Crit_{2Loc}$; a criterion of "at least two locations that are both deteriorating with $P < Crit_{2Loc}$" also then has specificity equal to 95%. Naturally, $Crit_{2Loc}$ will be larger than $Crit_{1Loc}$, since there are equal numbers of series meeting each of these criteria. Similar criteria were defined by increasing the required number of deteriorating locations. As for the global analyses, for each number of locations, criteria of the form "n locations each worsening at a rate worse than x dB/y with $P < Crit_{nLoc.x}$" were defined for different rates x, and based on series of length 5, 7, or 10 in the test–retest cohort.

Analysis: Deriving Criteria for Change by Cluster Trend Analysis

The cluster analysis used cluster definitions from the EyeSuite software (Haag-Streit) (shown in Fig. 1), which divides the total deviation values into 10 clusters. Researchers created these clusters to represent paths of retinal nerve fiber bundles, in order to detect patches of glaucomatous visual field abnormality. Similar clusters are used in the GHT on the HFA perimeter.[13] For each series of tests, the average deviation within a cluster was regressed against time, and the associated P value was recorded (hence, if this P value was below 0.05, the cluster would be significantly deteriorating according to conventional definitions and would be displayed as such on the commercial EyeSuite software). As with the pointwise analysis, the 10 resulting P values for each eye (one per cluster) were reordered starting with the smallest. The fifth percentile out of the list of smallest P values was labeled $Crit_{1Cl}$; hence, a criterion of "at least one cluster deteriorating with $P < Crit_{1Cl}$" has specificity equal to 95%. The fifth percentile out of the list of second smallest P values was labeled $Crit_{2Cl}$; giving a criterion of "at least two clusters deteriorating with $P < Crit_{2Cl}$" with specificity equal to 95%. Similar criteria were defined by larger numbers of clusters, and incorporating additional rate of change criteria.

Analysis: Detection of Change

For each eye in the longitudinal cohort, change in MD, and change using pointwise and cluster trend analyses were assessed by the criteria derived above, using the first five tests in their series, then the first six tests, and so forth. The eye was labeled as "significantly deteriorating" by a criterion on the earliest test date at which this criterion was met; if the criterion was never attained then the last test in the series was recorded as the censoring date. These dates were then framed in terms of "years since baseline," using the first test date in the series for an eye.

To explore the likelihood that change would be confirmed at the next test, the analysis was repeated adding two more tests to the series after the initial date at which the series showed significant deterioration, in order to see whether this extended series still met the criterion. Only series with at least two tests available after the first determination of deterioration were included in this analysis. This exercise was also performed by adding four extra tests to the series, in eyes for which they were available.

Analysis: Comparison of Criteria

Kaplan-Meier survival curves were plotted to show how soon each criterion detected "significant deterioration" for eyes in the longitudinal cohort. The lower quartile and median survival times were found (the first dates at which $\geq 25\%$ or $\geq 50\%$ of eyes, respectively, had shown significant deteriora-
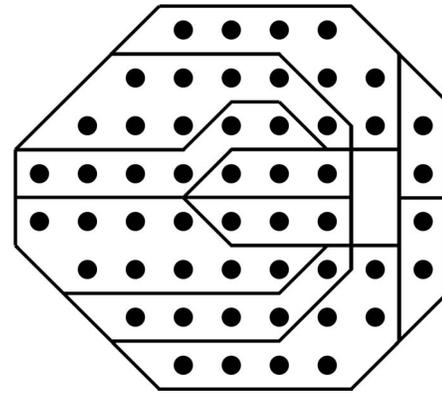
FIGURE 1. The 10 predefined clusters used for cluster trend analysis, as used in the EyeSuite software for the Octopus perimeter. Total deviations are averaged within each of the 10 clusters, and these averages are then regressed against time. The visual field is presented in right eye orientation.

tion), together with 95% confidence intervals (CIs) using standard errors based on Greenwood's formula.[21] Differences between the survival curves were assessed by using a stratified Cox proportional hazards model,[22] with strata identifying clusters for fellow eyes of the same individual. Including strata is equivalent to using generalized estimating equations in a linear model to adjust for intereye correlations.[23] To determine whether the relative performance of criteria was consistent across disease stages, given that our longitudinal cohort mostly have only early visual field loss, subanalyses were performed in which the best global, pointwise, and cluster criteria were compared within the subset of the longitudinal cohort that had MD > 0 dB on their first visit; and the subset that had MD ≤ 0 dB on their first visit. In order to assess which criterion has better sensitivity for detecting rapid change, the survival analysis was repeated by using only the first 5 years of testing for each eye, on the basis that more rapidly changing eyes should show "significant deterioration" earlier in their series.

RESULTS

The test–retest cohort contained series of at least seven reliable visual field tests from 133 eyes of 71 participants; and series of 10 reliable tests from 116 eyes of 63 participants. Fifty-nine percent were female; 69% were Caucasian, and 21% were Black. Fourteen had MD (on the last visit) greater than 0 dB; 74 had MD between 0 dB and −6 dB; and 45 had MD worse than −6 dB. The longitudinal cohort consisted of 505 eyes from 256 participants, once we had excluded eyes with fewer than five reliable tests. Fifty-eight percent were female; 95% were Caucasian. Most eyes had early visual field loss, with 207 eyes (41%) having MD on their most recent visit > 0 dB; 245 (49%) having MD between 0 and −6 dB; and 53 (10%) having MD worse than −6 dB. Table 1 summarizes other characteristics of the two cohorts.

For MD, the criterion for "significant deterioration" to achieve 95% specificity in series of five tests was "MD worsening with rate < 0.0 dB/y and $P < 0.101$." Twenty-five percent of eyes in the longitudinal cohort met this criterion within 5.0 years (95% CI, 4.7-5.3 years). The median time to meet this criterion was 8.6 years (95% CI, 7.4-10.8 years). When also requiring a rate of MD change worse than −0.1 dB/y, and adjusting the critical P value accordingly to maintain 95% specificity, the median time to detect significant deterioration increased to 9.1 years (95% CI, 7.6-12.7 years).

TABLE 1. Clinical Characteristics of the Participants in the Test-Retest and Longitudinal Cohorts

Demographics	Mean	Interquartile Range	Full Range
Test–retest cohort			
Series length, No. visits	10	10 to 10	7 to 10
Age, y	69	64 to 77	45 to 90
Most recent MD, dB	−4.8	−8.2 to −1.2	−19.8 to +1.6
Longitudinal cohort			
Series length, No. visits	14	10 to 18	5 to 24
Age, y	70	63 to 78	38 to 91
Most recent MD, dB	−1.6	−2.8 to +0.7	−27.0 to +2.7
Rate of MD change, dB/y	−0.20	−0.36 to +0.05	−3.20 to +1.89

Age and MD are at the date of their most recent visual field test. The rate of MD change in the longitudinal cohort is over their most recent six visual field tests, and so excludes the 18 eyes with only five visits. Both cohorts consist of participants who were diagnosed at baseline as having early or suspected glaucoma.

With even stricter rate of change criteria, fewer than 50% of the series ever showed significant deterioration before the end of their series, despite the specificity (in series of length 5 tests) still being 95%.

Table 2 shows the lower quartile of the times to detect significant deterioration, using pointwise analyses, together with 95% CIs, for a selection of different numbers of locations and different rate of change criteria, each with 95% specificity

TABLE 2. The Time for 25% of Eyes to Show "Significant Deterioration," by a Selection of Pointwise Analysis Criteria, Together With the Appropriate _P_ Value Criterion to Give 95% Specificity in Series of Five 6-Monthly Visual Fields

No. of Locations, _n_	Slope Criterion _x_		
	0	−0.5	−0.9
1	$P = 0.0055$	$P = 0.0055$	$P = 0.0055$
	4.9 y	4.9 y	6.0 y
	(4.5–5.0)	(4.6–5.0)	(5.2–7.3)
2	$P = 0.0154$	$P = 0.0154$	$P = 0.0154$
	4.5 y	4.9 y	8.1 y
	(4.2–5.0)	(4.5–5.1)	(6.5–12.3)
4	$P = 0.0467$	$P = 0.0467$	$P = 0.0467$
	4.2 y	4.6 y	11.9 y
	(4.1–4.9)	(4.2–5.0)	(9.0–NA)
6	$P = 0.0689$	$P = 0.0689$	$P = 0.0791$
	4.8 y	5.1 y	NA
	(4.3–5.0)	(4.8–5.6)	(11.2–NA)
9	$P = 0.1380$	$P = 0.1380$	$P = 0.1514$
	4.1 y	5.0 y	NA
	(4.0–4.5)	(4.4–5.5)	(NA–NA)
15	$P = 0.2152$	$P = 0.2382$	$P = 0.2783$
	4.9 y	6.1 y	NA
	(4.4–5.1)	(5.3–NA)	(NA–NA)

Criteria for "significant deterioration" are of the form "_n_ locations with total deviation deteriorating more rapidly than _x_ dB/y, with $P <$ $Crit_{nLoc.x}$," where the critical value $Crit_{nLoc.x}$ is calculated to give 95% specificity in series of length 5 visual fields in the test-retest cohort. Each cell shows the appropriate critical value $Crit_{nLoc.x}$ (top row), and the time for 25% of eyes in the longitudinal data set to meet this criterion in years, together with 95% confidence interval. For some criteria, fewer than 25% of series in the longitudinal cohort met the criterion before the end of their series, hence the time is shown as NA. Cohorts consist of participants who were diagnosed at baseline as having suspected or confirmed glaucoma.
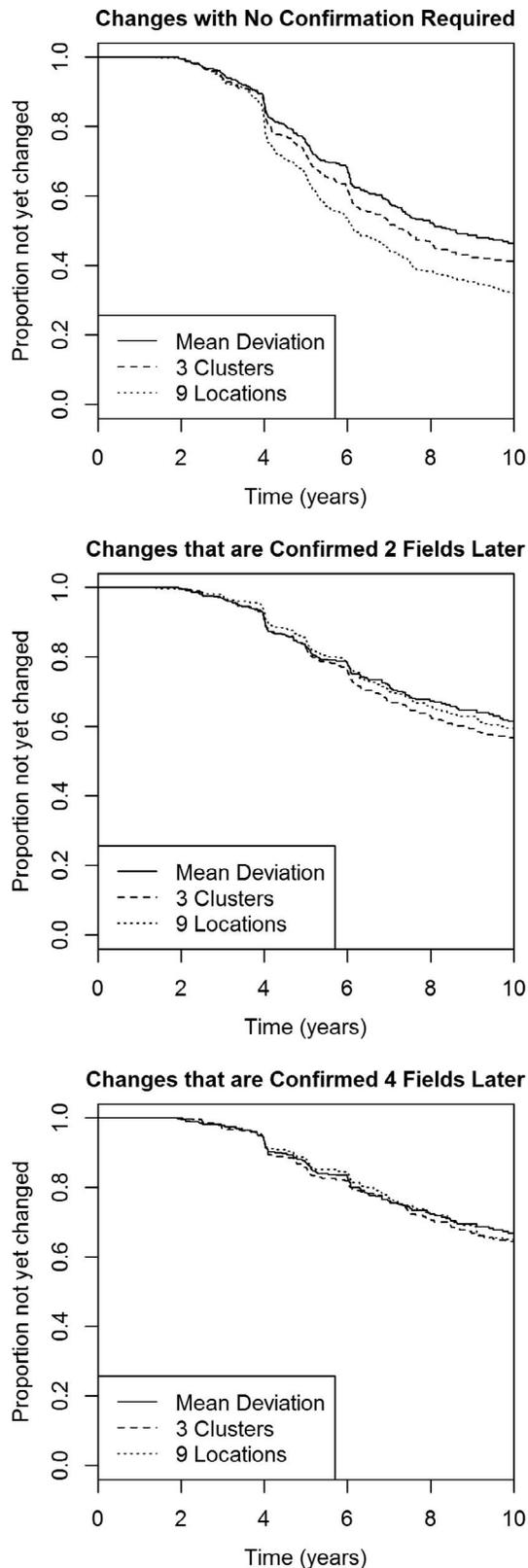
TABLE 3. The Time for 25% of Eyes to Show "Significant Deterioration," by a Selection of Cluster Trend Analysis Criteria, Together With the Appropriate _P_ Value Criterion to Give 95% Specificity in Series of Five 6-Monthly Visual Fields

No. of Clusters, _n_	Slope Criterion _x_		
	0	−0.5	−0.9
1	$P = 0.0114$	$P = 0.0120$	$P = 0.0156$
	5.4 y	6.1 y	NA
	(5.0–6.0)	(5.6–6.8)	(11.5–NA)
2	$P = 0.0562$	$P = 0.0592$	$P = 0.0773$
	5.0 y	6.1 y	NA
	(4.5–5.2)	(5.2–8.0)	(NA–NA)
3	$P = 0.1168$	$P = 0.1244$	$P = 0.1636$
	4.8 y	10.8 y	NA
	(4.2–5.1)	(6.1–NA)	(NA–NA)
4	$P = 0.1797$	$P = 0.1929$	$P = 0.2823$
	4.8 y	NA	NA
	(4.2–5.1)	(NA–NA)	(NA–NA)
5	$P = 0.2514$	$P = 0.2779$	$P = 0.4889$
	4.8 y	NA	NA
	(4.2–5.1)	(NA–NA)	(NA–NA)
6	$P = 0.3374$	$P = 0.3896$	–
	4.8 y	NA	
	(4.4–5.1)	(NA–NA)	

Criteria for "significant deterioration" are of the form "_n_ clusters whose average total deviation deteriorates more rapidly than _x_ dB/y, with $P <$ $Crit_{nCl.x}$," where the critical value $Crit_{nCl.x}$ is calculated to give 95% specificity in series of length 5 visual fields in the test-retest cohort. Each cell shows the appropriate critical value $Crit_{nCl.x}$ (top row), and the time for 25% of eyes in the longitudinal data set to meet this criterion in years, together with 95% confidence interval. Several of the criteria were met by fewer than 5% of series in the test-retest cohort, hence no critical _P_ value can be defined for those cells. For other criteria, fewer than 25% of series in the longitudinal cohort met the criterion before the end of their series, hence the time is shown as NA. Both cohorts consist of participants who were diagnosed at baseline as having early or suspected glaucoma.

for series of length 5 tests. The table shows the time for 25% of series to meet each of the criteria. The same pattern was apparent when using the median time, but for many criteria fewer than 50% of series showed significant deterioration before the end of their series. Again, imposing a minimum rate of change criterion and adjusting the _P_ value criterion to maintain 95% specificity delayed detection of significant deterioration in all cases. One commonly used criterion for pointwise linear regression is sensitivity deteriorating with rate worse than −1 dB/y and $P < 5\%$,[11] which is the same as total deviation deteriorating with rate worse than −0.9 dB/y and $P < 5\%$. As seen in Table 2, requiring four such changing locations gave specificity 95%, but it took 11.9 years for 25% of longitudinal series to meet this criterion. Although several criteria were not significantly different from one another, the best pointwise criterion was "≥9 locations worsening with rate worse than 0 dB/y and $P < 0.138$." Using this criterion, 25% of series in the longitudinal cohort showed significant deterioration in 4.1 years (95% CI, 4.0–4.5); and 50% of series showed significant deterioration in 6.2 years (95% CI, 5.9–7.0).

Table 3 shows the lower quartile of the time to detect significant deterioration for a selection of cluster trend analysis criteria, again each with 95% specificity for series of length 5 tests. As for global and pointwise analyses, including a rate of change criterion delayed detection of significant deterioration for the same specificity. While a few different criteria were not significantly different from each other, the most rapid criterion to detect significant deterioration was

Changes with No Confirmation Required

Changes that are Confirmed 2 Fields Later

Changes that are Confirmed 4 Fields Later

FIGURE 2. The time to detect "significant deterioration" using the best global, pointwise, and cluster trend analyses, in a longitudinal cohort of participants with early or suspected glaucoma. The best global analysis criterion to give specificity 95% in series of five test–retest visual fields was "MD worsening with $P < 0.101$." The best pointwise criterion was "≥ 9 locations worsening with $P < 0.138$." The best cluster trend criterion was "≥ 3 clusters worsening with $P < 0.117$." *Top panel*: No

"≥ 3 clusters worsening with rate worse than 0 dB/y and $P < 0.117$." Using this criterion, 25% of series showed significant deterioration in 4.8 years (95% CI, 4.2–5.1); and 50% in 7.4 years (95% CI, 6.8–8.3).

The top panel in Figure 2 shows the survival curves for the best global, pointwise, and cluster trend analyses, that is, "MD worsening with $P < 0.101$," "≥ 9 locations worsening with $P < 0.138$," and "≥ 3 clusters worsening with $P < 0.117$," respectively. The cluster trend analysis detected significant deterioration significantly sooner than MD, with $P < 0.001$; but significantly slower than pointwise analysis with $P < 0.001$. Significant deterioration was detected within 5 years in 134 eyes, using MD; 148 eyes, using the best-performing cluster trend analysis; and 188 eyes, using the best-performing pointwise analysis. When only these first 5 years were considered, cluster trend analysis still detected change significantly sooner than MD (hazard ratio 0.859, $P = 0.012$) and later than pointwise analysis (hazard ratio 1.268, $P < 0.001$).

However, of the 293 eyes that had at least two subsequent visual fields after showing significant deterioration by the best-performing pointwise criterion, only 112 eyes (38%) still met the same criterion when including those two extra tests in the series. By contrast, while there were only 255 eyes that had at least two subsequent visual fields after showing significant deterioration by the best-performing cluster trend criterion, this change was confirmed when including those next two tests in 156 eyes (61%). Using MD, 232 eyes had at least two tests after significant deterioration was detected, and change was confirmed in 176 of those eyes (76%).

The second panel in Figure 2 shows survival curves for the same global, pointwise, and cluster trend analysis criteria as above for the detection of "confirmed significant deterioration," where eyes are only counted as deteriorating if they meet the same criterion after two more tests are added to the series (note that the date at which the eye first met the criterion is used for these survival curves, rather than the date that the deterioration was successfully confirmed). Twenty-five percent of eyes met this criterion, using MD after 6.3 years (95% CI, 6.0–7.2); using pointwise analyses after 6.3 years (95% CI, 6.0–7.0); and using cluster trend analyses after 6.0 years (95% CI, 5.3–6.6). The comparison between MD and cluster trend analysis had $P = 0.006$ for the entire series, and $P = 0.882$ when just the first 5 years were considered. The comparison between the pointwise and cluster trend analyses had $P = 0.186$ for the entire series, and $P = 0.078$ for the first 5 years.

The bottom panel in Figure 2 shows the equivalent results when deterioration had to be confirmed after the addition of four subsequent visual fields. Twenty-five percent of eyes met this criterion, using MD after 7.3 years (95% CI, 6.4–8.6); using pointwise analyses after 7.4 years (95% CI, 6.8–8.6); and using cluster trend analyses after 7.3 years (95% CI, 6.4–8.4). The comparison between MD and cluster trend analysis had $P = 0.10$ for the entire series, and $P = 0.27$ when just the first 5 years were considered. The comparison between the pointwise and cluster trend analyses had $P = 0.18$ for the entire series, and $P = 0.05$ for the first 5 years.

We tested the hypothesis that the optimal criteria to detect change with 95% specificity in a test–retest cohort would depend on the series length. We repeated the analysis by using series of 7, and 10, visual fields in the test–retest cohort. As the

confirmation of detected deterioration is required. *Middle panel*: Deterioration had to be confirmed after the inclusion of two subsequent visual fields. *Bottom panel*: Deterioration had to be confirmed after the inclusion of four subsequent visual fields.

TABLE 4. The Criteria That Minimized the Time for 25% of Eyes to Show "Significant Deterioration," Out of All Tested Global, Pointwise, and Cluster Trend Analysis Criteria, for 95% Specificity in Series of Five, Seven, or Ten 6-Monthly Visual Fields

	95% Specificity in Test–Retest Series of Length		
Criterion	5 Fields	7 Fields	10 Fields
Best global analysis criterion			
P value criterion	$P = 0.1010$	$P = 0.1012$	$P = 0.1018$
Time for unconfirmed change	5.0 (4.7-5.3)	5.0 (4.7-5.3)	5.0 (4.7-5.3)
Time for confirmed change	6.3 (6.0-7.2)	6.3 (6.0-7.2)	6.3 (6.0-7.2)
Best pointwise analysis criterion			
P value criterion	$P = 0.1380$	$P = 0.1212$	$P = 0.1199$
Time for unconfirmed change	4.1 (4.0-4.5)	4.7 (4.2-5.0)	4.7 (4.2-5.0)
Time for confirmed change	6.3 (6.0-7.0)	6.8 (6.3-7.3)	6.8 (6.3-7.3)
Best cluster analysis criterion			
P value criterion	$P = 0.1168$	$P = 0.1106$	$P = 0.1083$
Time for unconfirmed change	4.8 (4.2-5.1)	4.9 (4.4-5.1)	4.9 (4.5-5.1)
Time for confirmed change	6.0 (5.3-6.6)	6.1 (5.7-6.7)	6.1 (5.8-6.8)

Criteria for "significant deterioration" were of the form "mean deviation deteriorating more rapidly than x dB/y, with $P < \text{Crit}_{MD}$," "n locations whose total deviation deteriorates more rapidly than x dB/y, with $P < \text{Crit}_{nLoc.x}$," and "n clusters whose average total deviation deteriorates more rapidly than x dB/y, with $P < \text{Crit}_{nCl.x}$," where the critical values Crit are calculated to give 95% specificity in series of length 5, 7, or 10 visual fields in the test-retest cohort. Only the best such criteria are shown; in each of these optimal cases the minimum slope was $x = 0$. Each cell shows the appropriate critical value Crit (top row); the time in years for 25% of eyes in the longitudinal data set to meet this criterion in years, together with 95% confidence interval (middle row); and the time in years for 25% of eyes to meet this criterion with that change being subsequently confirmed after the addition of two more fields to the series (bottom row).

series lengthens, the proportion of very rapid rates of change diminishes as the CI around the slope estimate narrows. This means that, for example, with a criterion of the form "three clusters each worsening at a rate worse than -1 dB/y with $P < \text{Crit}_{nCl.x}$," the value of $\text{Crit}_{nCl.x}$ will increase with series length in order to achieve 5% false positives (95% specificity). However, for criteria of the form "three clusters each worsening at a rate worse than 0 dB/y with $P < \text{Crit}_{nCl.x}$," this reduction in the magnitude of the slopes has no effect, and so $\text{Crit}_{nCl.x}$ will be comparatively independent of series length. Therefore since the best criteria for global, pointwise, and cluster analyses are all of the form "...rate worse than 0 dB/y...," these criteria vary little with series length, as shown in Table 4. The times for 25% of eyes to meet these criteria (with or without requiring subsequent confirmation) are very similar, especially for the cluster and global analyses, as seen in Table 4. The times to detect change are shown in the survival curves in Figure 3 (using test-retest series of length 7) and Figure 4 (using test-retest series of length 10). We therefore conclude that it is reasonable to use a constant criterion regardless of series length, so long as the chosen criterion is based on statistical significance without imposing a non-zero minimum rate of change.

We then tested whether the relative performance of these optimal criteria differed by disease stage. Figure 5 shows the time to meet the same criteria as before, based on series of length 5 in the test-retest cohort, but with the eyes split according to whether the MD at the start of the series was >0 dB (left column, $n = 326$) or ≤0 dB (right column, $n = 179$). The time until detectable change may be slightly longer when the initial MD was ≤0 dB, but that is likely because those eyes are being treated more aggressively and hence are less likely to progress rapidly. The main conclusions did not vary with disease stage; however, the benefits of the cluster trend technique were more apparent at the earliest stages of functional loss. Without requiring confirmation, in both cases cluster trend analysis detected change sooner than MD ($P = 0.044$ for MD > 0 dB; $P = 0.001$ for MD ≤ 0 dB) but slower than pointwise analysis ($P < 0.001$ for both subsets). However when requiring that deterioration must be con-
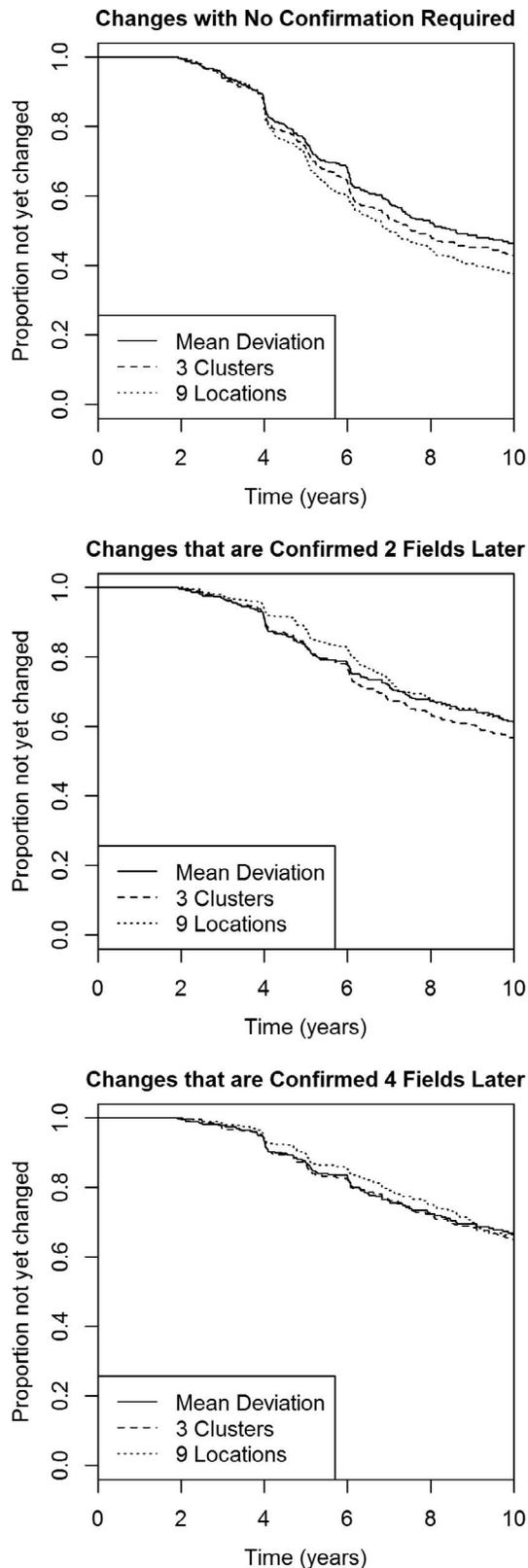
firmed after two subsequent visual fields, there were no significant differences for initial MD ≤ 0 dB ($P = 0.71$ for clusters versus MD; $P = 0.58$ for clusters versus pointwise); while for initial MD > 0 dB, cluster trend analysis detected change significantly sooner than either MD ($P = 0.001$) or pointwise analysis ($P = 0.005$).

The current EyeSuite software that accompanies the Octopus perimeter determines whether the mean total deviation within each cluster is deteriorating with $P < 5\%$. Requiring different numbers of clusters to meet this criterion, the closest to a specificity of 95% that could be achieved was "two clusters deteriorating with $P < 5\%$," which had a specificity of 94.4% in series of length 5 tests, 95.3% in series of 7 tests, and 95.4% in series of 10 tests. This criterion detected deterioration in 25% of eyes in 5.1 years (95% CI, 4.7-5.3), and in 50% of eyes in 8.1 years (95% CI, 7.3-9.5).

DISCUSSION

Glaucoma often creates localized visual field defects such as paracentral and arcuate scotomas. Our results support the hypothesis that cluster trend analysis is more sensitive than MD for detecting deterioration, since it can detect the development and/or worsening occurring within scotomas. In every analysis presented here, MD took longer to detect significant deterioration than the best-performing cluster trend criterion, for the same specificity, whether or not confirmation was required.

In a simple survival analysis, as seen in the top panel of Figure 2, pointwise analyses appeared to detect significant deterioration sooner than cluster trend analysis. However, this result is potentially misleading. In more than half of the cases, the "deterioration" detected by pointwise analysis was not confirmed when two or four more tests were added to the series. Relying on pointwise analyses may result in overcalling of progression unless confirmation of changes is performed,[12] because of the greater test-retest variability that is present when single locations are considered. When only deterioration that was subsequently confirmed was counted, pointwise

Changes with No Confirmation Required

Changes that are Confirmed 2 Fields Later

Changes that are Confirmed 4 Fields Later

FIGURE 3. The time to detect "significant deterioration" using the best global, pointwise, and cluster trend analyses, in a longitudinal cohort of participants with early or suspected glaucoma, when criteria were selected to give specificity 95% in series of seven test–retest visual fields. The best global analysis criterion was "MD worsening with $P < 0.101$." The best pointwise criterion was "≥ 9 locations worsening with $P < 0.121$." The best cluster trend criterion was "≥ 3 clusters

analysis no longer detected deterioration any sooner than cluster trend analysis; indeed, the time for 25% of eyes to show deterioration was actually slightly, albeit not significantly, slower when using pointwise analysis than cluster trend analysis. A criterion of "≥ 3 clusters deteriorating with $P < 0.117$" detected deterioration quickly without generating excessive false-positive determinations of progression. While not completely optimal, a criterion of "≥ 2 clusters deteriorating with $P < 5\%$" had specificity of very near 95% and detected deterioration almost as quickly, and can be evaluated by using the currently available EyeSuite software.

In a clinical trial, and in clinical care, it would be desirable to detect progression in less than the 4.8 years taken for 25% of eyes to show deterioration by the best criterion used in this study. Indeed, various techniques may decrease this time, such as altering the frequency or temporal distribution of testing.[24,25] This does not mean though that the cluster trend analysis is insensitive to deterioration. More likely is that most individuals in the longitudinal cohort truly were relatively stable during the first few years of their series. The cohort examined in this study consisted mainly of relatively early disease. Indeed over a quarter showed a slightly positive rate of change of MD over their series, with this positive rate likely being due to noise rather than true improvement, but still supporting the idea that these eyes were essentially stable.

In reality, all eyes are technically progressing, since everyone's visual system deteriorates with age. Perimetry helps clinicians and researchers determine whether a given eye is deteriorating more rapidly than normal aging by an amount that requires a change in management. Management decisions are based on multiple factors such as the rate of functional change, the health of the other eye, and the visual needs and anticipated residual life expectancy of the patient. This means that any predefined binary classification into "progressing" versus "stable" that does not take these factors into account, as used in this study, is inherently limited for clinical use. However, such binary classifications are invaluable as endpoints in clinical trials. Importantly, this study demonstrates that cluster trend analysis provides a more reliable and useful measure of the rate of change than global or pointwise analyses.

Researchers have proposed different methods to account for the spatial correlation between locations when detecting functional changes.[26–29] One big advantage of using the cluster trend technique over such approaches is that it is easy to visualize and comprehend the clusters, both for clinicians and patients, compared with the "black box" approach of more advanced statistical techniques such as neural network models. A second advantage is that cluster analysis could produce an intuitive associated measure of the rate of change, even though the best metric of rate is not yet clear (e.g., whether to use the third-fastest changing cluster, or the average of the three fastest changing clusters). A disadvantage is the reliance on predefined clusters, which may not accurately map out a defect in an individual patient. Such defects may be better detected by the development of techniques using overlapping clusters of locations.

Most individuals in both cohorts tested in this study had MD < 0 dB by the end of their series, despite undergoing regular testing that would be expected to increase sensitivities as

worsening with $P < 0.111$." *Top panel*: No confirmation of detected deterioration is required. *Middle panel*: Deterioration had to be confirmed after the inclusion of two subsequent visual fields. *Bottom panel*: Deterioration had to be confirmed after the inclusion of four subsequent visual fields.
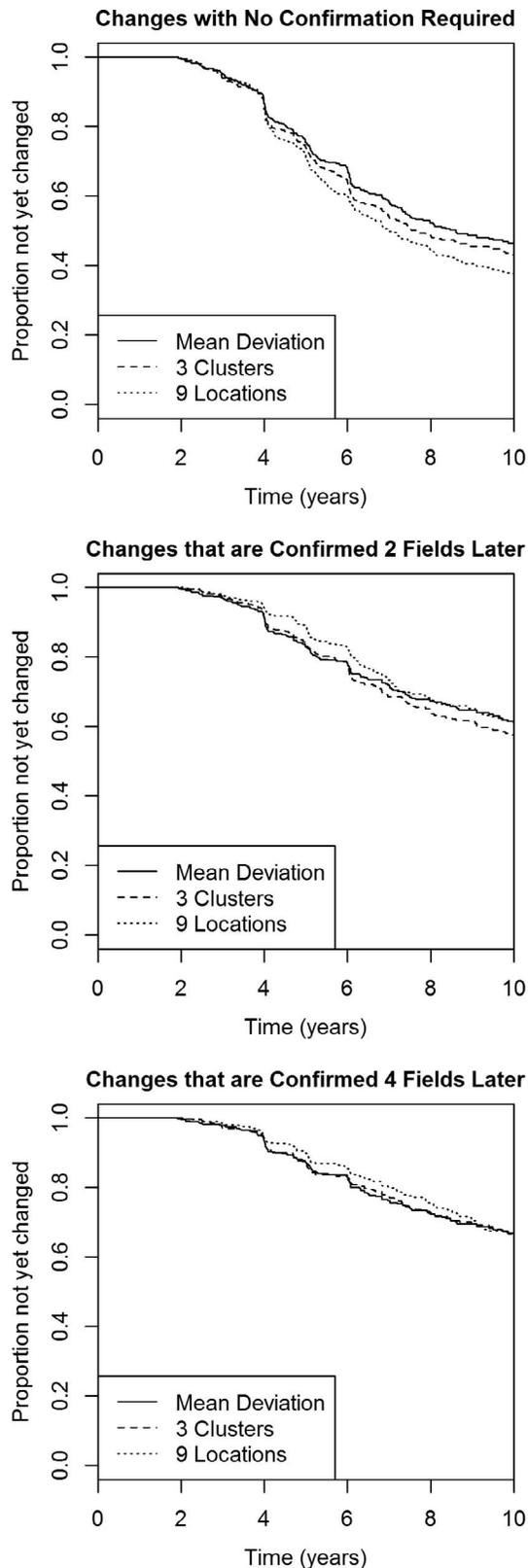
FIGURE 4. The time to detect "significant deterioration" using the best global, pointwise, and cluster trend analyses, in a longitudinal cohort of participants with early or suspected glaucoma, when the criteria for deterioration had specificity 95% in series of 10 test-retest visual fields. The best global analysis criterion was "MD worsening with $P <$ 0.101." The best pointwise criterion was "≥9 locations worsening with $P <$ 0.120." The best cluster trend criterion was "≥3 clusters

compared with naïve individuals.[30,31] The proportion that would be defined as "glaucoma" rather than "glaucoma suspect" is highly subjective and dependent on the criterion chosen,[32] and there are several different staging criteria that could be used, but it is reasonable to contend that most eyes had early glaucomatous damage, as evidenced by the mean MD of −1.63 dB. As such, we can only speculate about the performance of the different analysis techniques in more severe glaucoma. At later stages of the disease, visual field damage becomes more widespread, potentially improving the performance of global indices such as MD. A further complication is that perimetric sensitivities become increasingly more variable and unreliable in later disease,[33] which would have a greater influence on analyses that use fewer locations. Therefore, it could be hypothesized that global analyses could be at least as good as cluster trend analyses in eyes with more severe functional damage. Indeed, the benefits of cluster trend analysis appeared to be strongest at the earlier stages of glaucoma, as seen in Figure 5. It has been suggested that it may be preferable to switch to larger stimuli in such eyes, in order to reduce variability[34] and extend the dynamic range of perimetry[35,36]; while there is no reason to expect our conclusions to differ using larger stimuli, this has not been tested.

The development or worsening of cataract can also cause a reduction in perimetric sensitivities.[37,38] This loss would be expected to be generalized, rather than localized, and so would likely affect global indices such as MD to a greater extent than it would affect cluster or pointwise analyses. The fact that cluster and pointwise analyses still detected deterioration sooner than global analysis in the longitudinal cohort makes it unlikely that our main conclusions are driven by cataract rather than glaucomatous progression.

The primary results are based on test-retest data consisting of series of five visual fields. However, the same conclusion was reached using longer series of test-retest data. Crucially, the actual criterion that gives 95% specificity for cluster trend analysis changes remarkably little, based on the series length. This is because it is based on the statistical significance of whether the cluster is deteriorating faster than 0 dB/y; and statistical significance implicitly takes the series length into account. Criteria that included a stricter minimum rate of deterioration would alter more with series length in order to maintain a constant specificity, but these were found to detect deterioration slower in the longitudinal cohort. Since the best criterion did not alter substantially, it seems reasonable to use the same criterion for all series' lengths. Indeed, this is typical in existing clinical applications that flag locations or fields if they are deteriorating with $P <$ 5%, with this cutoff not changing with series length. A separate issue is that specificity will inevitably deteriorate with repeated assessment in a longitudinal series, for example, if a series is assessed using the first five fields, then assessed again using the first six fields, and so forth; however, this deterioration in specificity will affect all criteria equally and so would not affect our conclusions.

In conclusion, we found that cluster trend analysis detected progression of the visual field sooner than global indices, without the concomitant excessive reduction in specificity observed when using pointwise analyses, in patients with early

worsening with $P <$ 0.108." *Top panel*: No confirmation of detected deterioration is required. *Middle panel*: Deterioration had to be confirmed after the inclusion of two subsequent visual fields. *Bottom panel*: Deterioration had to be confirmed after the inclusion of four subsequent visual fields.
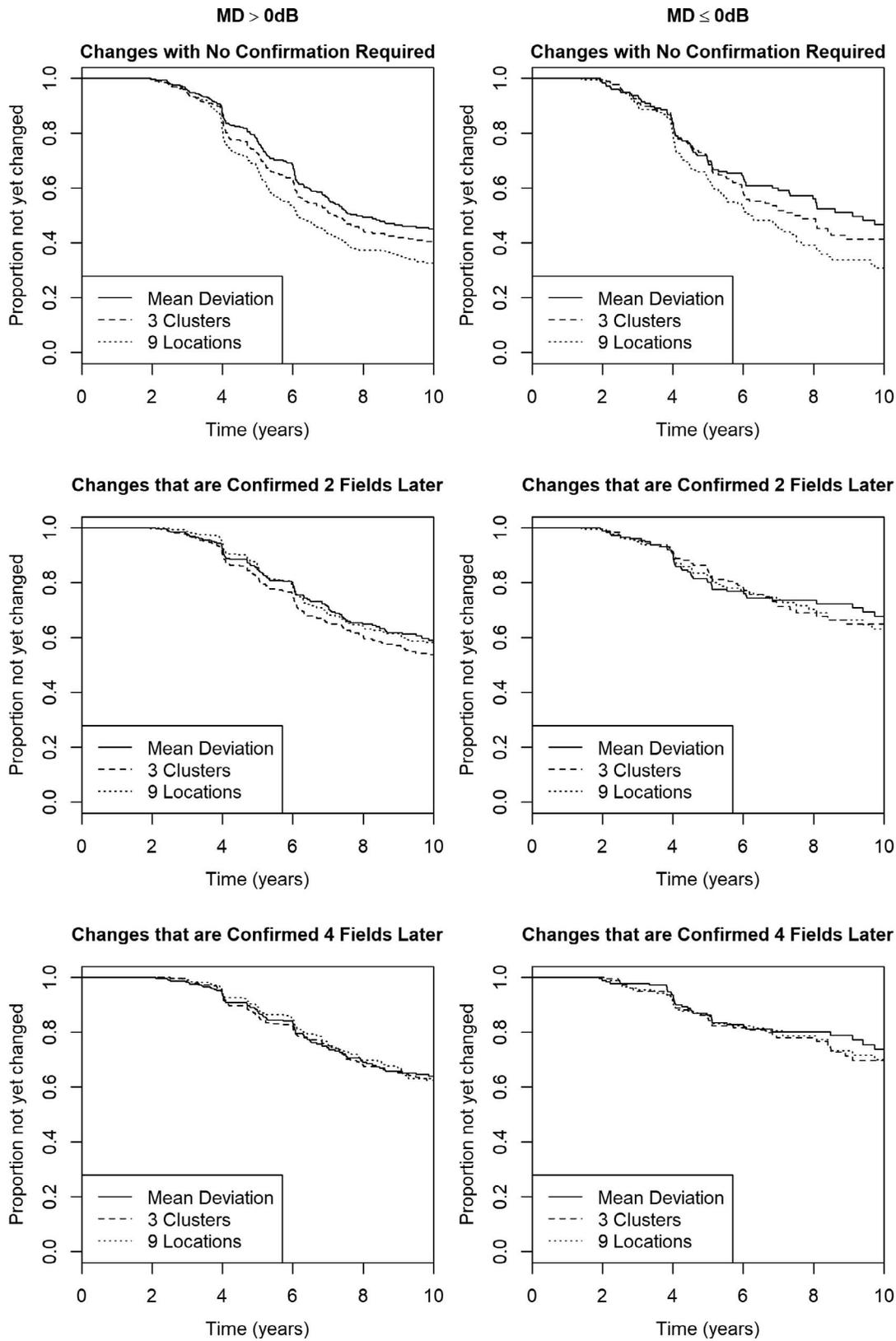
FIGURE 5. The time to detect "significant deterioration" using the best global, pointwise, and cluster trend analyses, in a longitudinal cohort of participants with early or suspected glaucoma split according to the degree of functional loss, when the criteria for deterioration had specificity 95% in series of five test–retest visual fields. The best global analysis criterion was "MD worsening with $P < 0.101$." The best pointwise criterion was "≥ 9 locations worsening with $P < 0.120$." The best cluster trend criterion was "≥ 3 clusters worsening with $P < 0.108$." *Top two panels*: No confirmation of detected deterioration is required. *Middle two panels*: Deterioration had to be confirmed after the inclusion of two subsequent visual fields. *Bottom two panels*: Deterioration had to be confirmed after the inclusion of four subsequent visual fields. *Left column*: MD at the start of the series > 0 dB. *Right column*: MD at the start of the series ≤ 0 dB.

glaucoma. This study suggests that cluster-based endpoints may be preferable for determining glaucomatous progression.

References

1. Weinreb RN, Kaufman PL. Glaucoma research community and FDA look to the future, II: NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium: measures of structural change and visual function. *Invest Ophthalmol Vis Sci*. 2011; 52:7842–7851.

2. Demirel S, De Moraes CG, Gardiner SK, et al. The rate of visual field change in the ocular hypertension treatment study. *Invest Ophthalmol Vis Sci*. 2012;53:224–227.

3. Gardiner SK, Demirel S, De Moraes CG, et al. Series length used during trend analysis affects sensitivity to changes in progression rate in the ocular hypertension treatment study. *Invest Ophthalmol Vis Sci*. 2013;54:1252–1259.

4. Bleisch D, Furrer S, Funk J. Rates of glaucomatous visual field change before and after transscleral cyclophotocoagulation: a retrospective case series. *BMC Ophthalmol*. 2015;15:179.

5. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol*. 2013;91:406–412.

6. Song MK, Sung KR, Han S, et al. Progression of primary open angle glaucoma in asymmetrically myopic eyes. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:1331–1337.

7. Gardiner SK, Demirel S. Detecting change using standard global perimetric indices in glaucoma. *Am J Ophthalmol*. 2017;176:148–156.

8. Koucheki B, Nouri-Mahdavi K, Patel G, Gaasterland D, Caprioli J. Visual field changes after cataract extraction: the AGIS experience. *Am J Ophthalmol*. 2004;138:1022–1028.

9. Wild JM, Hutchings N, Hussey MK, Flanagan JG, Trope GE. Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthalmology*. 1997; 104:808–815.

10. Krupin T, Liebmann JM, Greenfield DS, Ritch R, Gardiner S; Low-Pressure Glaucoma Study Group. A randomized trial of brimonidine versus timolol in preserving visual function: results from the Low-Pressure Glaucoma Treatment Study. *Am J Ophthalmol*. 2011;151:671–181.

11. Fitzke FW, Hitchings RA, Poinoosawmy D, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80:40–48.

12. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci*. 2002;43:1400–1407.

13. Asman P, Heijl A. Glaucoma Hemifield Test: automated visual field evaluation. *Arch Ophthalmol*. 1992;110:812–819.

14. Naghizadeh F, Hollo G. Detection of early glaucomatous progression with octopus cluster trend analysis. *J Glaucoma*. 2014;23:269–275.

15. De Moraes CG, Liebmann JM, Levin LA. Detection and measurement of clinically meaningful visual field progression in clinical trials for glaucoma. *Prog Retin Eye Res*. 2017;56: 107–147.

16. Garway-Heath DF, Lascaratos G, Bunce C, et al. The United Kingdom Glaucoma Treatment Study: a multicenter, randomized, placebo-controlled clinical trial: design and methodology. *Ophthalmology*. 2013;120:68–76.

17. Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): a randomised, multicentre, placebo-controlled trial. *Lancet*. 2015;385:1295–1304.

18. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997;75:368–375.

19. R Development Core Team. *R: A Language and Environment for Statistical Computing*. 1.9.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2013.

20. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Visual field progression: comparison of Humphrey Statpac2 and pointwise linear regression analysis. *Graefes Arch Clin Exp Ophthalmol*. 1996;234:411–418.

21. Greenwood M. The errors of sampling of the survivorship tables: reports on public health and statistical subjects. 1926; 33:26.

22. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol*. 1972;34:187–220.

23. O'Quigley J, Stare J. Proportional hazards models with frailties and random effects. *Stat Med*. 2002;21:3219–3233.

24. Gardiner SK, Crabb DP. Frequency of testing for detecting visual field progression. *Br J Ophthalmol*. 2002;86:560–564.

25. Crabb DP, Garway-Heath DF. Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. *Invest Ophthalmol Vis Sci*. 2012;53:2770–2776.

26. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal function: Analysis With Non-Stationary Weibull Error Regression and Spatial Enhancement (ANSWERS). *PLoS One*. 2014;9:e85654.

27. Murata H, Araie M, Asaoka R. A new approach to measure visual field progression in glaucoma patients using variational Bayes linear regression. *Invest Ophthalmol Vis Sci*. 2014;55: 8386–8392.

28. Fujino Y, Murata H, Mayama C, Asaoka R. Applying "Lasso" regression to predict future visual field progression in glaucoma patients. *Invest Ophthalmol Vis Sci*. 2015;56: 2334–2339.

29. Yousefi S, Balasubramanian M, Goldbaum MH, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Trans Vis Sci Tech*. 2016;5(3):2.

30. Wild JM, Searle AE, Dengler-Harles M, O'Neill EC. Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmol (Copenh)*. 1991;69:210–216.

31. Gardiner SK, Demirel S, Johnson CA. Is there evidence for continued learning over multiple years in perimetry? *Optom Vis Sci*. 2008;85:1043–1048.

32. Ng M, Sample PA, Pascual JP, et al. Comparison of visual field severity classification systems for glaucoma. *J Glaucoma*. 2012;21:551–561.

33. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369.

34. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2013;54: 1345–1351.

35. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010; 128:570–576.

36. Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. The effect of stimulus size on the reliable stimulus range of perimetry. *Transl Vis Sci Tech*. 2015;4(2):10.

37. Heuer DK, Anderson DR, Knighton RW, Feuer WJ, Gressel MG. The influence of simulated light scattering on automated perimetric threshold measurements. *Arch Ophthalmol*. 1988; 106:1247–1251.

38. Chen PP, Budenz DL. The effect of cataract on the visual field of eyes with chronic open-angle glaucoma. *Am J Ophthalmol*. 1998;125:325–333.