Detecting the Progression of Eye Disease: CUSUM Charts for Assessing the Visual Field and Retinal Nerve Fiber Layer Thickness

Johannes Ledolter¹ and Randy Kardon²

¹ Department of Management Sciences/Department of Statistics and Actuarial Science, University of Iowa, and Department of Veterans Affairs Medical Center, Iowa City, Iowa

² Department of Ophthalmology and Visual Sciences, University of Iowa, and Department of Veterans Affairs Medical Center, Iowa City, Iowa

Correspondence: Johannes Ledolter, University of Iowa, Department of Management Sciences, S352 PBB, Iowa City, IA. e-mail: johannes-ledolter@uiowa.edu.

Received: 26 February 2013

Accepted: 29 July 2013

Published: 12 September 2013

Keywords: glaucoma; optical coherence tomography; statistical process control; visual fields

Citation: Ledolter J, Kardon R. Detecting the progression of eye disease: CUSUM charts for assessing the visual field and retinal nerve fiber layer thickness. *Trans Vis Sci Tech.* 2013;2(6):2, <http://tvstjournal.org/doi/full/10.1167/tvst.2.6.2>, doi: 10.1167/tvst.2.6.2

Purpose: The cumulative sum (CUSUM) is proposed and tested in a group of glaucoma patients and healthy subjects as a method for monitoring disease progression and for identifying clinically significant step changes in visual structure or function.

Methods: The CUSUM procedure is the recommended method for the timely detection of small step changes in manufacturing process control. The CUSUM procedure is discussed and compared with traditional approaches for the detection of change in the status of the visual system over time. The CUSUM approach is used to monitor over time visual field (VF) mean deviations and optical coherence tomography (OCT) measurements of retinal nerve fiber layer (RNFL) thickness in 53 healthy subjects and 103 patients with glaucoma.

Results: The CUSUM method detects VF progression for 35 of the 103 glaucoma patients (34.0%), and OCT RNFL reductions for 20 of the 103 glaucoma patients (19.4%).

Conclusions: The CUSUM method is effective in detecting small level changes. This method can be used to monitor the progression of disease and it benefits the clinician who must decide, on the basis of a time series of variable data, whether a change has occurred.

Translational Relevance: A cumulative sum chart helps the clinician decide whether a step change has taken place, and it does so as quickly as possible. This approach is particularly effective for detecting small step changes, which very likely are unnoticed with currently used change detection approaches.

Introduction

Assessment of visual function using best corrected visual acuity, visual field (VF) quantification, and light-evoked responses of the eye and brain have been used extensively for monitoring the onset and the progression of eye diseases such as glaucoma, retinal degeneration, diabetic retinopathy, and other optic neuropathies. Similarly, assessment of retinal and optic nerve structure give valuable information about the onset and the progression of morphologic changes that lead to eventual loss of vision, and such measurements are easily determined with today's measurement technology using digital fundus imaging and spectral-

domain optical coherence tomography (SD-OCT). Detecting changes in either functional (i.e., VF) or structural (i.e., thickness of the retinal nerve fiber layer [RNFL]) measurements is important to the medical practitioner as such information affects a patient's treatment. This is certainly true for glaucoma, but also for other diseases that are known to change the visual function and the thickness of the RNFL.

Measurements on the VF (functional data) and the thickness of the retinal nerve fiber layer (structural data) are usually taken every 6 to 12 months to check for the onset of disease and to monitor disease progression. The timing of the onset and the change in progression of loss is unknown and may vary greatly

between patients, prompting the need for a statistical method to detect the onset and the change in progression as quickly as possible while keeping the rate of false positives low. In this paper, we provide evidence that cumulative sum (CUSUM) procedures are ideally suited for this purpose and apply these methods on summary statistics of the VF (mean deviation) and the thickness of the RNFL (average thickness). Instead of basing a decision on just the most recent deviation from a target, the CUSUM aggregates the most recent as well as previous deviations from the target and bases the decision about a change from the target on their CUSUM. Cumulative sum procedures are not only useful in ophthalmology, but in all areas of medicine where physicians must decide, usually on the basis of a short time sequence of measurements, whether or not a significant change has occurred.

Methods

Shewhart Charts

The usual event analysis approach to the detection of the onset of VF or RNFL problems assesses the likelihood that the measurement at time n , Y_n , is consistent with a certain in-control value μ_0 . A warning flag gets raised if the difference between the measurement and the in-control value exceeds a critical threshold that is determined from data on healthy (in-control) patients. It is common to compare the measurement with a reference distribution that comes from healthy subjects, express the measurement as a percentile of that distribution, and raise a flag signaling a reduction if the measurement Y_n represents a percentile of order 100α or smaller (where α is a value such as 0.05, 0.025, or 0.01). For normal distributions with known in-control mean μ_0 and SD σ , this approach compares the observation Y_n with $\mu_0 + z_\alpha\sigma$; z_α is the percentile of the standard normal distribution such as $z_{0.05} = -1.65$, $z_{0.025} = -1.96$, and $z_{0.01} = -2.33$.

This approach is identical to the Shewhart control chart used in manufacturing quality control where measurements on items sampled sequentially in time are compared with “three-sigma limits” (see Shewhart,¹ Deming,² Ledolter and Burrill³). A flag gets raised whenever an observation exceeds the in-control value by more than a multiple of the observations’ SD. For a multiple of three, three sigma control limits correspond to percentiles of order 0.135% and 99.865%, assuming a normal distribution. A Shewhart chart that monitors

the process for a reduction is equivalent to raising a flag whenever the measurement Y_n represents a very small (0.135%) in-control percentile.

The choice of z_α , or the percentage point α , affects the properties of the procedure. It is common to characterize a Shewhart chart by its implied average run length (ARL). The run length is the number of observations it takes to conclude that a reduction has occurred, and the ARL is its expected value. One wants the ARL large if there has been no change, and one wants it small if the process has changed to a new lower level. The in-control ARL of the Shewhart chart is $ARL(\mu_0) = 1/\alpha$; see Ledolter and Burrill.³ For example, the Shewhart chart with $\alpha = 0.025$ and $z_{0.025} = -1.96$ implies $ARL(\mu_0) = 40$; the chart with $\alpha = 0.01$ and $z_{0.01} = -2.33$ implies $ARL(\mu_0) = 100$.

Alternatively, one can characterize a Shewhart chart by its probability of signaling a change within the next n observations,

$$\begin{aligned} P[\text{Signal}] &= 1 - P[\text{NoSignal}] \\ &= 1 - P[Y_1 \geq \mu_0 + z_\alpha\sigma \cap Y_2 \geq \mu_0 + z_\alpha\sigma \\ &\quad \cap \dots \cap Y_n \geq \mu_0 + z_\alpha\sigma] \\ &= 1 - P[Y_1 \geq \mu_0 + z_\alpha\sigma]P[Y_2 \geq \mu_0 + z_\alpha\sigma] \\ &\quad \dots P[Y_n \geq \mu_0 + z_\alpha\sigma] = 1 - (1 - \alpha)^n. \quad (1) \end{aligned}$$

For given n and 5% error of false positives, one can solve the equation $0.05 = 1 - (1 - \alpha)^n$ for $\alpha = 1 - 0.95^{1/n}$ and z_α . For the data that we use in the Results section of this paper (53 healthy and 103 glaucoma patients in a University of Iowa/Veterans Administration study) the average number of monitoring periods is $n = 7$, sampled about every 6 months. With $n = 7$, $\alpha = 1 - 0.95^{1/7} \approx 0.01$, and $z_{0.01} = -2.33$. With cutoff -2.33 , groups of seven consecutive observations are falsely rejected 5% of the time.

It is straightforward to assess how the Shewhart chart with given cutoff z_α responds to a reduction in the level to $\mu_1 = \mu_0 - r\sigma$ (here change is defined as a multiple, r , of the process SD). The probability of obtaining an out-of-control signal is $P[Y < \mu_0 + z_\alpha\sigma] = P[Z < z_\alpha + r]$, where the probability for the standard normal random variable Z can be looked up in statistical tables. The out-of-control ARL is $ARL(\mu_1) = 1/P[Z < z_\alpha + r]$ and the probability that the Shewhart chart signals a change within the next n observations is

$$\begin{aligned} P[\text{Signal}] &= 1 - P[Y_1 \geq \mu_0 + z_\alpha\sigma] \\ &\quad \times P[Y_2 \geq \mu_0 + z_\alpha\sigma] \dots P[Y_n \geq \mu_0 + z_\alpha\sigma] \\ &= 1 - \{P[Z \geq z_\alpha + r]\}^n \quad (2) \end{aligned}$$

Modified Shewhart Charts

Event analysis for VF data usually looks for multiple warning flags. Flags at three consecutive time periods are typically required before one concludes that a change has occurred. However, a cutoff $z_{0.01} = -2.33$ for each of three consecutive observations makes a signal unlikely and leads to a negligible probability of false positives among $n = 7$ consecutive observations and weak power of detecting an actual change. The cutoff needs to be selected smaller in absolute value in order to achieve the targeted 5% false positive rate. Simulations reported in the Results section show that for such modified charts $\alpha = 0.25$ and $z_{0.25} = -0.68$ lead to a 5% false positive rate among seven consecutive observations.

A change detection that raises a signal if three consecutive deviations from baseline are smaller than a certain threshold is very similar to Shewhart control charts with additional run rules (see Ledolter and Burrill³). Hence, our terminology referring to such charts as “modified” Shewhart charts.

CUSUM Charts

The standard Shewhart chart is very good at detecting large shifts, but it is known to be poor at detecting small changes. Shewhart charts supplemented with additional run rules are able to detect small step changes quicker, while still achieving an acceptable low false positive error rate. However, the “gold standard” for detecting the onset of a small step change is the CUSUM procedure. The CUSUM procedure has become a standard tool of manufacturing process control⁴ and is the recommended method for the timely detection of small step changes. Optimal theoretical results for the CUSUM procedure were shown by Moustakides⁵: Among all procedures with the same in-control ARL, the CUSUM minimizes the expected time until a change gets signaled once the process has shifted to the out-of-control state. One would like the ARL large if the process is in-control (that is, the subject stays free of disease or if there is no progression of disease), but wants it to be small if a shift to the out-of-control state (i.e., onset or progression of disease) has occurred. We describe the CUSUM chart in this section, and illustrate it with examples in the Results section.

For its implementation, one needs to specify:

- (1) The SD of the repeat measurement variability, σ .
- (2) The in-control value μ_0 and the magnitude of the

step change one wants to detect. The in-control value μ_0 comes from either the population average of healthy (in-control) patients or from subject-specific baseline information (such as the mean deviation and the mean RNFL thickness that is obtained from the first few observations on each subject). A shift of one SD of the repeat measurement variability is taken to represent the magnitude of a clinically relevant step change, but smaller shifts can be studied if they are thought to be more relevant. The adopted step change from the in-control value determines the out-of-control value μ_1 .

- (3) The ARL under the in-control situation. For example, $ARL = 100$ implies that a false positive signal (signaling a change when no change is present) occurs on average after 100 consecutive observations. For patients visiting the clinic every 6 to 12 months, a procedure constructed with $ARL = 100$ allows for few false positive signals during a patient's follow-up period of reasonable length. We use $ARL = 100$ in our examples. Recall that the Shewhart chart with $\alpha = 0.01$ and $z_{0.01} = -2.33$ attains $ARL = 100$.

The calculations for the CUSUM procedure are as follows. Assume that we are interested in detecting a decrease in the mean response, from an in-control value μ_0 to an out-of-control value μ_1 less than μ_0 . With consecutive observations $Y_1, Y_2, \dots, Y_n$, we compute signals $S_1, S_2, \dots, S_n$ according to the CUSUM recursion,

$$S_t = \min(0, S_{t-1} + [Y_t - k]), \quad \text{for } t = 1, 2, \dots, n \quad (3)$$

with starting value $S_0 = 0$. The constant $k = (\mu_1 - \mu_0)/2$ less than 0 is one-half of the difference between the out-of-control and in-control values, amounting to one-half of the decrease we want to detect. We conclude that a change has occurred when the signal S_t is smaller than a certain critical value h less than 0. Computer software is available to determine the critical value such that the CUSUM procedure achieves the desired in-control ARL. Average run lengths for specified alternatives can be calculated, assessing how long it takes on average to detect a change of a certain magnitude. Brook and Evans⁶ use a Markov chain approach to derive the ARLs for given critical value h ; a detailed discussion on how to do this is given in Hawkins and Olwell.⁷ This book and the webpage of Douglas Hawkins⁸ at the University of Minnesota provide useful and easy to use computer software.

Table 1. Percentage of Eyes Signaling Progression, Modeled Within Seven Observations: Shewhart (Single Event Analysis Compared With Baseline) and Modified Shewhart Charts (Three Consecutive Events Analyzed With Respect to Baseline)^a

	Shewhart (Single Signal)			Modified Shewhart: 3 Consecutive Signals
	$\alpha = 0.025;$ $z_{\alpha} = -1.96$	$\alpha = 0.01;$ $z_{\alpha} = -2.33$	$\alpha = 0.005;$ $z_{\alpha} = -2.58$	$\alpha = 0.25;$ $z_{\alpha} = -0.68$
No change	16.2%	6.9%	3.4%	6.2%
Reduction (0.5 σ)	40.8%	21.7%	12.4%	25.4%
Reduction (1 σ)	72.5%	49.7%	33.7%	58.7%

^aProperties of Shewhart charts as it relates to event change criteria for progression were determined theoretically. Properties of the modified Shewhart chart were determined from simulations. The first row is for the rate of false positive detection when no change exists. The second and third rows give the detection rate when the real amount of progression was set to be either small (0.5 SD) or larger (1.0 SD).

Consider observations with a repeat measurement SD σ and construct a CUSUM that monitors measurements for a reduction of 1 SD from in-control value μ_0 to out-of-control value $\mu_1 = \mu_0 - \sigma$. For illustration, we use $\sigma = 1$, $\mu_0 = 0$, and $\mu_1 = \mu_0 - \sigma = -1$. CUSUM signals are calculated from Equation 3, with $k = (-1 - 0)/2 = -0.5$. For in-control ARL of 100, the critical value is $h = -2.850$ and the ARL until detecting a shift to the out-of-control value ($\mu_1 = -1$) is 6.1. On average, the CUSUM detects a step change reduction of 1 SD six periods after the change has taken place (which amounts to 3 years in a typical case of a glaucoma subject being followed every 6 months). The critical value $h = -2.850$ and the out-of-control ARL 6.1 are obtained with statistical software; for example, with the program geth.exe from the webpage of Douglas Hawkins⁸ at the University of Minnesota, <http://users.stat.umn.edu/~dhawkins/> (the program is located under Software and Cumulative Sums). What if we wanted to detect a smaller change of half of an SD? Then $k = (-0.5 - 0)/2 = -0.25$ and $h = -4.418$, and the ARL at the out-of-control value ($\mu_1 = -0.5$) is 14.8. The smaller shift is more difficult to detect. On average, we detect a step change of half of an SD within 15 periods after the change has taken place.

Our illustration uses $\mu_0 = 0$ as deviations from the baseline should have mean zero if the process is in control. We consider $\sigma = 1$ and $\mu_1 = \mu_0 - \sigma = -1$, even though any other value of σ could have been used without affecting the ARL at the out-of-control value $\mu_1 = \mu_0 - \sigma$ and the time when the CUSUM exceeds the critical value. The only quantities that change with $\sigma \neq 1$ are the reference value k (it changes from $-1/2$ to $-\sigma/2$) and the critical value (it changes from $-h$ to $-h\sigma$).

Results

The first part of the Results section summarizes theoretical reasons why a CUSUM approach is ideally suited for detecting small shifts. The second part provides an illustration that uses CUSUM charts to monitor the VF data and OCT RNFL thickness of glaucoma and healthy patients.

Effectiveness of the CUSUM Chart

The effectiveness of competing procedures for detecting a change can only be compared if their proportions of false positives are the same. Table 1 shows how to control the proportion of false positives at close to 5% when monitoring seven consecutive deviations from baseline. For a Shewhart chart that monitors successive observations one at a time, the appropriate standardized threshold is $z_{0.001} = -2.33$, giving a 6.9% false positive rate of detecting a change when none exists ("No Change," row 1). For the modified Shewhart chart that concludes a reduction if three consecutive signals are below the threshold, the appropriate standardized threshold is $z_{0.25} = -0.68$, giving a 6.2% false positive rate of detecting a change when none exists. A comparison of the detection probabilities when the process mean has changed shows that the modified Shewhart chart detects a reduction from baseline more often than the Shewhart chart (58.7% vs. 49.7% detection of a 1-sigma reduction; row 3 of Table 1), while controlling the false positive rate at about the same level (6%).

Table 2 compares the CUSUM designed to detect a 1-sigma reduction and in-control ARL = 100 to the Shewhart chart with threshold $z_{0.01} = -2.33$ and ARL = 100, and the modified Shewhart chart that requires

Table 2. Proportions of Signaling a Reduction Within Seven Observations and Average Run Lengths: Shewhart Chart, Modified Shewhart Chart, and CUSUM Chart Designed to Detect Either a 0.5 Sigma or a One-Sigma Reduction^a

	Shewhart (Single Signal)		Modified Shewhart		CUSUM $k = -0.5; h = -2.850$
	$\alpha = 0.01; z_\alpha = -2.33$		(3 Consecutive Signals)		
	$\alpha = 0.25; z_\alpha = -0.68$				
	Proportion		Proportion	Proportion	
	Signaling Reduction	ARL	Signaling Reduction	Signaling Reduction	ARL
No change	6.9	100	6.2	4.9	100
Reduction (0.5σ)	21.7	29.1	25.4	28.7	16.1
Reduction (1σ)	49.7	10.7	58.7	73.4	6.1

^aProperties of the Shewhart chart and ARLs of the CUSUM chart are determined theoretically. Percentage of eyes (proportion) signaling disease progression (reduction) within seven observations for the modified Shewhart and the CUSUM charts are determined from simulations.

three consecutive signals with threshold $z_{0.25} = -0.68$. The three procedures have similar in-control properties (same in-control ARL = 100 for CUSUM and Shewhart charts, and similar proportions of false positives for a window of seven observations). Table 2 confirms that the CUSUM outperforms the other two procedures. The CUSUM has smaller out-of-control run lengths and higher detection proportions than either of the other two procedures.

Illustration of the CUSUM Chart

As an illustration, we consider VF data and OCT RNFL thickness of 103 glaucoma and 53 healthy patients that had been followed at the University of Iowa for several years (Variability in Perimetry (VIP) study funded by the Department of Veterans Affairs Rehabilitation Division; Michael Wall MD, Principal Investigator). The inclusion criteria for both patients and control subjects required corrected visual acuity of 20/30 or better, no or mild cataract, refractive error not exceeding 6 diopters (D) spherical and 3.5 D cylindrical, pupil diameter 3 mm, and mean deviation on standard automated perimetry (SAP) of better than 20 dB. The diagnosis of glaucoma was based on the evaluation of two glaucoma experts. The experts diagnosed glaucoma by using their clinical expertise in evaluating glaucomatous cupping of the optic nerve, corresponding VF with RNFL-associated pattern of loss, intraocular pressure, and patient history. Measurements for glaucoma patients were available at nine time periods, on average. Between 5 and 12 measurements were available, with a time period ranging from 1.5 to 5.2 years and spanning 4 years on average. On average, measurements for healthy patients were available at 9.4 time periods. Between

7 and 11 measurements were available for healthy patients, with a time period ranging from 2.5 to 5.0 years and spanning 4 years on average.

We illustrate the CUSUM approach with VF mean deviations on two patients from the glaucoma group. The first two measurements, taken within several days, are used to establish a baseline value for the patient. The subsequent observations, obtained about every 6 months, are used to assess whether a 1 SD decrease from the baseline has taken place. The raw data and the deviations from the baseline are given in Table 3. Graphs of the two time sequences, with time periods reflecting 6-months intervals, are shown in Figure 1.

All event analysis approaches to change detection (Shewhart, modified Shewhart, and CUSUM charts) require that one knows the in-control SD σ . A 5-week repeatability study (5 determinations, once a week) with 34 glaucoma and 22 healthy patients at the University of Iowa found that the SDs among average thickness measurements of the RNFL (Stratus OCT3; Zeiss Meditec, Dublin, CA) are similar for glaucoma ($\sigma = 3.01 \mu\text{m}$) and healthy subjects ($\sigma = 2.99 \mu\text{m}$). The (decibel) SD of the mean deviation of the VF measurements (Humphrey SITA 24-2) for glaucoma patients, $\sigma = 0.98 \text{ dB}$, was found to be larger than the SD for healthy patients, $\sigma = 0.55 \text{ dB}$. These are population averages. While we know that there is substantial variability among σ from one subject to another, we do not have enough data to estimate a subject-specific σ at the time a person enters the clinic. The initial two measurements on a subject taken in brief succession give us a rough estimate for his/her baseline, but any estimate of a SD from just two

Table 3. Data for the Two Glaucoma Patients in Figure 1: Consecutive Visual Field Mean Deviations, Deviations From the Baseline (Obtained by Subtracting the Average of the Two Baseline Values From the Measurements), and CUSUM Signals Designed to Detect a Reduction of One Standard Deviation $\sigma = 1^a$

Patient 17 (Glaucoma Group)					Patient 18 (Glaucoma Group)			
Time	Time, y	Y(t) Mean deviation (db)	Y(t) Deviations from Baseline	S(t)	Time, y	Y(t) Mean deviation (db)	Y(t) Deviations from Baseline	S(t)
B1	0.00	−3.21	0	0	0.00	−2.29	0	0.00
B2	0.01	−3.76	0	0	0.04	−2.28	0	0.00
1	0.60	−3.77	−0.285	0	0.56	−4.66	−2.375	−1.875
2	1.12	−5.00	−1.515	−1.015	1.11	−2.16	0.125	−1.250
3	1.54	−4.21	−0.725	−1.240	1.62	−0.67	1.615	0.00
4	2.06	−6.55	−3.065	−3.805	2.39	−2.08	0.205	0.00
5	2.55	−8.04	−4.555		2.89	−4.19	−1.905	−1.405
6	3.02	−8.61	−5.125		3.64	−2.14	0.145	−0.760
7	3.69	−12.46	−8.975		4.52	−4.70	−2.415	−2.675
8	4.19	−10.25	−6.765					

^aCUSUM signals designed to detect a reduction of 1 SD, $\sigma = 1$. CUSUM statistics, using Equation 3 with $k = -\sigma/2 = -0.5$, are compared with threshold $h = -2.85\sigma = -2.85$. The signal in boldface is smaller than the threshold, and indicates a reduction for patient 17.

Plots of Consecutive Visual Field Mean Deviations (dB)

Figure 1. Consecutive visual field mean deviations for two glaucoma patients. The baseline average is determined from the first two measurements taken within a few days. The other measurements are taken approximately every 6 months.

Table 4. Number and Proportion of Reductions When Using a CUSUM With In-Control ARL = 100^a

	Number of Patients	Number of Reductions	Proportion of Reductions
Healthy			
OCT change ($\mu\text{m}/\text{y}$)	53	7/53	13.2
VF change (db/y)	53	6/53	11.3
Glaucoma			
OCT change ($\mu\text{m}/\text{y}$)	103	20/103	19.4
VF change (db/y)	103	35/103	34.0

^aThe SDs were adjusted for the estimation of the baseline. Among the 53 healthy subjects, we observed three subjects with simultaneous OCT and VF signals, 43 without any signal, four with OCT but no VF signal, and three with VF but no OCT signal. Among the 103 glaucoma patients, we observed seven patients with simultaneous OCT and VF signals, 55 patients without any signal, 13 with OCT but no VF signal, and 28 with VF but no OCT signal.

observations is subject to sizeable sampling variability.

In the absence of subject-specific SDs we rely on population averages and use $\sigma = 1$ dB for monitoring the progression of the two glaucoma patients. We wish to learn whether the subsequent measurements indicate a reduction from the subject-specific baselines. The calculated CUSUM statistics, using Equation 3 with $k = -\sigma/2 = -0.5$, are given in Table 3. For in-control average run length 100, the critical cutoff is $h = -2.85\sigma = -2.85$. For patient 17, a reduction is signaled at time-period 4. For patient 18, the evidence for a reduction is insufficient, even though the CUSUM signal at time-period 7, $S_7 = -2.675$, is very close to the threshold $h = -2.85$. Continued reductions that extend the pattern established over the last periods would force the CUSUM below the threshold providing evidence for a reduction.

Finally, we apply the CUSUM approach that is illustrated in Table 3 for the VF mean deviations of the two glaucoma patients to the OCT RNFL and VF mean deviations of the 103 patients in the glaucoma group and the 53 subjects in the healthy group. The average of a subject's first two measurements is taken as the subject-specific baseline, and the subsequent deviations from the baseline are monitored for a possible reduction from $\mu_0 = 0$. In the absence of subject-specific SDs, we rely on the population averages that we obtained from the Iowa data. That is, $\sigma = 3$ μm for OCT RNFL for both healthy and glaucoma patients, $\sigma = 1$ dB for VF mean deviations of glaucoma patients, and $\sigma = 0.55$ dB for VIF mean deviations of healthy subjects.

While we expect a change for glaucoma patients who are subject to continual deterioration (even under best clinical treatment to help mitigate adverse changes), no changes are expected for healthy

patients. The CUSUM rejects about 10% of healthy subjects (13.2% for OCT RNFL and 11.3% for VF). It flags 34% (VF) and 19.4% (OCT) of the glaucoma patients. One could expect even better percentages if reliable subject-specific variability estimates had been available.

The results are shown in Table 4.

Discussion

A strength of the CUSUM test is that it does not require a fixed time frame to determine whether or not a change has taken place; a decision is made as soon as enough evidence for a change has accumulated. A cumulative sum chart helps the clinician decide whether a step change has taken place, and it does so as quickly as possible. Our analysis, both theoretically and using patient data, shows that the CUSUM is particularly effective for detecting small step changes, which are very likely unnoticed when using Shewhart charts and the traditional event analysis approach to change detection.

The specificity and sensitivity of the CUSUM test is dependent upon the subject-specific measurement variability of the test(s) being monitored, the ability of a particular test to detect a change in status of the disorder, and whether the patient being monitored is likely to change over time (how much true progression is anticipated). Newer approaches to measuring VF changes over time, which reduce measurement variability (i.e., increasing the size of the stimulus target; Wall, Doyle, Zamba, Artes, and Johnson⁹; Wall, Woodward, Doyle, and Artes¹⁰) and recent advances in reducing measurement variability of retinal structures over time (i.e., use of the baseline scan as a reference scan upon which subsequent scans

are aligned to scan the exact same retinal location, averaging of scan lines during image acquisition, and better segmentation algorithms of the retinal layers; Pemp, Kardon, Kircher, Pernicka, Schmidt-Erfurth, and Reitner¹¹) all will improve the sensitivity and specificity of the CUSUM approach. In fact, the CUSUM test can be used to model progression in order to estimate how much improvements in specific measurement variability and sensitivity of a given test will improve the ability to detect disease progression. It is also anticipated that in a more realistic clinical setting (and not a prospective well-controlled patient study group as was used here), the number of patients who would be detected as showing progression of disease such as glaucoma would increase even further.

The severity of disease can also affect the ability to detect progression with approaches such as CUSUM. This is because some measures of visual progression such as VF testing are associated with an increase in measurement variability in perimetric locations as the severity of the disease increases. However, structural measures are not likely to show increases in measurement variability with increase in disease severity. Besides the effect of disease severity on test-retest variability, the magnitude of a change may also be affected (floor effect, as seen in severe perimetric VF loss of sensitivity or structural thinning of retinal layers to the point of loss of dynamic range of thickness; Hood, Anderson, Wall, Raza, and Kardon¹²).

The CUSUM test in this particular application is designed to identify patients who are already diagnosed with a progressive cause of visual loss (in the case of this study, glaucoma) in which detection of progression over time is of paramount interest to the clinician who is monitoring the patient to determine if any change in treatment is indicated. Once progression is detected using approaches such as CUSUM, and a change in treatment is instituted, then a new baseline test measurement can be reset to the new level and the process starts again for detecting subsequent change.

The same method could be applied to patients with other disorders in which the detection of progression of disease on a subsequent test is clinically important. Other examples include, but are not limited to multiple sclerosis (detection of retinal ganglion cell loss or axon loss over time), diabetic maculopathy, macular degeneration, retinitis pigmentosa, compressive optic neuropathy, and toxic disorders caused by medications such as hydroxychloroquine or ethambutal. Similarly, with newer treatments that can

improve vision in some causes of visual loss such as genetic treatments, drugs, and radiation, the CUSUM method may also be applied to detect improvement in visual function over time.

Acknowledgments

The authors thank Michael Wall, MD, Principal Investigator of the Variability in Perimetry (VIP) study funded by the Department of Veterans Affairs Rehabilitation Division, for allowing us to use visual field and OCT retinal nerve fiber layer thickness data on 103 glaucoma and 53 healthy patients that had been followed at the University of Iowa for several years.

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

References

1. Shewhart W. *Economic Control of Quality of Manufactured Product*. New York, NY: Van Nostrand; 1931.
2. Deming WE. *Out of the Crisis*. Cambridge, MA: MIT Press; 1982.
3. Ledolter J, Burrill CW. *Statistical Quality Control: Strategies and Tools for Continual Improvement*. New York, NY: John Wiley; 1999.
4. Montgomery DC. *Introduction to Statistical Quality Control*. New York, NY: John Wiley; 2008.
5. Moustakides GV. Optimal stopping times for detecting changes in distributions. *Ann Stat*. 1986; 14:1379–1387.
6. Brook D, Evans DA. An approach to the probability distribution of CUSUM run length. *Biometrika*. 1972;59:539–549.
7. Hawkins DM, Olwell DH. *Cumulative Sum Charts and Charting for Quality Improvement*. New York, NY: Springer; 1997.
8. Hawkins DM. Available at: <http://users.stat.umn.edu/~dhawkins/>. Accessed 27 August, 2013.
9. 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.
10. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with

- stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci.* 2009;50:974–979.
11. Pemp B, Kardon RH, Kircher K, Pernicka E, Schmidt-Erfurth U, Reitner A. Effectiveness of averaging strategies to reduce variance in retinal nerve fibre layer thickness measurements using spectral-domain optical coherence tomography. *Graefes Arch Clin Exp Ophthalmol.* 2013;251:1841–1848.
 12. Hood DC, Anderson SC, Wall M, Raza AS, Kardon RH. A test of a linear model of glaucomatous structure-function loss reveals sources of variability in retinal nerve fiber and visual field measurements. *Invest Ophthalmol Vis Sci.* 2009;50:4254–4266.