Reducing Variability in Visual Field Assessment for Glaucoma Through Filtering That Combines Structural and Functional Information

Lisha Deng, Shaban Demirel, and Stuart K. Gardiner

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

Correspondence: Stuart K. Gardiner, Devers Eye Institute, Legacy Research Institute, 1225 NE 2nd Avenue, Portland, OR 97232, USA; SGardiner@deverseye.org.

Submitted: December 20, 2013

Accepted: June 6, 2014

Citation: Deng L, Demirel S, Gardiner SK. Reducing variability in visual field assessment for glaucoma through filtering that combines structural and functional information. *Invest Ophthalmol Vis Sci*. 2014;55:4593–4602. DOI:10.1167/iovs.13-13813

PURPOSE. To reduce variability and improve measurements of true change signal in visual field (VF) assessments through the use of filters that combine functional and structural test results.

METHODS. Humphrey VF data (Swedish Interactive Thresholding Algorithm [SITA] Standard, 24-2) and confocal scanning laser ophthalmoscopy (Heidelberg Retina Tomograph [HRT]) data from 1057 eyes of 637 participants were used to derive a filter. Another dataset, consisting of VF and HRT data from 112 eyes of 62 participants each with ≥ 5 visits, was used to test the filter. At each VF location per eye, the trend over time was modeled by a linear model (LM), and a nonlinear model (NLM), using filtered or unfiltered data, but with the last visit excluded. The SD of residuals from the trends, and prediction errors (PE) for the last visit were compared between filtered and unfiltered data. The filter was reconstructed and analyses were repeated after truncating VF data so that thresholds < 19 dB were replaced by 19 dB to reduce noise.

RESULTS. The SD of the residuals at all 52 VF locations for all analyses was reduced by filtering ($P < 0.001$). The PE was reduced by filtering at 43 and 47 VF locations ($P < 0.05$) for LM analyses on observed and truncated data, and all 52 VF locations ($P < 0.05$) for both NLM analyses. Truncating data before filtering reduced variability ($P < 0.01$) at 41 and 40 VF locations for LM and NLM analyses.

CONCLUSIONS. Filtering can reduce variability about trends in longitudinal sequences of VF data, and improves the accuracy of predicting the next test result.

Keywords: automated perimetry, variability, CSLO

Glaucoma is a progressive optic neuropathy that eventually can lead to irreversible vision loss and blindness. However, the Ocular Hypertension Treatment Study¹ demonstrated that early medical treatment can delay or prevent the onset of primary open-angle glaucoma (POAG) among high risk individuals with statistically raised IOP. In addition, it is known that eyes with damaged visual fields are more likely to display more rapid progression.^{2,3} Therefore, it is important that glaucoma is identified at an early stage and managed appropriately.

Evaluation of patients with glaucoma is based in part on results from functional testing and structural imaging of the optic nerve head (ONH) and/or retinal nerve fiber layer (RNFL). Examination of visual function usually is performed via standard automated perimetry (SAP) to examine sensitivity at a predefined set of visual field (VF) locations. Examination of the ONH/RNFL normally is conducted through direct observation of the ONH and via imaging techniques, such as confocal scanning laser ophthalmoscopy (Heidelberg Retina Tomograph [HRT]; Heidelberg GmbH, Heidelberg, Germany) or spectral-domain optical coherence tomography (SD-OCT). It has been reported frequently that structural abnormality can be detected before functional damage in eyes with POAG,^{4–11} but sometimes functional damage can occur without apparent structural damage assessed using conventional methods.^{7,12,13} However, this observed lack of concordance may be due to variability in

test results rather than a true time lag alone.¹⁴ Although the association between functional damage (as measured in SAP) and structural damage (with RNFL thickness and rim area [RA] measurements) is modest at best,^{7,15,16} it is important as clinicians put great emphasis on mutually confirming functional and structural test results.

The SAP measurements are known to be quite variable, especially when VF damage becomes evident.^{17–20} Variability obscures the identification of true glaucomatous progression in VF and makes detection of progression challenging. Therefore, reducing variability is particularly important¹⁹ as the change in a glaucomatous VF must be greater than the noise about the measurement before it becomes statistically distinguishable.^{21,22} Reducing variability may be possible through improved methods of data acquisition, or by postprocessing currently available data.²³ One attractive feature of postprocessing of existing data is that it requires no additional testing time. One approach to processing existing data is to exploit the associations known to exist between sensitivities at neighboring VF locations, and between regions of the ONH and certain VF locations through a process of filtering.

Filtering is used widely in image processing to improve the quality of digital information. Crabb et al.^{24,25} and Fitzke et al.²⁶ first applied this technique in a novel fashion to VF data in the mid-1990s. Unlike the pixel values in digital images, VF sensitivities have associations that are shaped by the anatomy

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

Investigative Ophthalmology & Visual Science

FIGURE 2. Examples of the shape of the filter.

(Heidelberg Retinal Tomograph [HRT]; Heidelberg GmbH) and data were split into 30° ONH sectors. Both tests were performed within 1 month of each other. Only visits with reliable SAP results ($\leq 15\%$ false positives, $\leq 30\%$ false negatives and fixation losses) were used. Sensitivities at each of the 52 non-blindspot VF locations were used from SAP. Rim areas in twelve 30° sectors were taken from the HRT scans. A recent study³⁰ suggested using 30° ONH sectors with the first sector centered on the fovea-ONH axis for general clinical use, since it is robust to the effects of small inaccuracies in mapping individual VF locations to the ONH, and imperfections in imaging and VF measurements.

A “longitudinal dataset” was constructed using SAP and HRT data from the Portland Progression Project (P3). None of the P3 cohort was included in the cross-sectional dataset. The inclusion criteria for the longitudinal dataset were the same as for the cross-sectional dataset. To be included in the longitudinal trend analysis, a minimum of five eligible visits were required for each eye, so that a reliable measure of the rate of change could be obtained. The longitudinal sequences had a mean follow-up time of 5.3 years, ranging from 1.9 to 11.5 years. All participants in the P3 study gave informed consent after having the risks and benefits of the study

explained to them. The study protocol was approved by the Legacy Health Institutional Review Board, and adhered to the tenets of the Declaration of Helsinki.

At each of the 52 VF locations, the filter at location $j(\tilde{y}_j)$ was constructed in three steps. Firstly, a linear regression model using the cross-sectional dataset was fitted to predict sensitivity at location j , namely $(\hat{y}_j^{\text{xsec}})$; see Equation 1:

$$\hat{y}_j^{\text{xsec}} = \alpha + \sum_{i=1}^{37} \beta_i x_i^{\text{xsec}} + \varepsilon_j, \quad (1)$$

where $\varepsilon_j \sim N(0, \sigma_j^2)$ is a normally distributed error term with mean zero and constant variance and the x_i^{xsec} s are predictors, including the sensitivities at the other 25 VF locations within the same upper/lower hemifield and the logarithms of the rim area at each of the 30° sectors. Logarithms of the rim area were used because recent studies^{31–33} on structural and functional tests have suggested that the change in sensitivity when expressed on a logarithmic scale in decibels (dB) appears to be proportional to the logarithm of the percentage loss of retinal ganglion cells, for which rim area is a structural surrogate. All 12 sectors of the optic disc were used, instead of restricting the analysis to one-half of the optic disc (as was done for the VF

TABLE 2. The SD of Residuals From LM and NLM Analyses, Compared Between Filtered and Unfiltered Data, With Filters Constructed From Either Observed or Truncated Data at Each of 52 Nonblindspot 24-2 VF Locations

Hypothesis	Base Data	Model	Wilcoxon Signed-Rank Test
SD of residuals of filtered data <	Observed data	LM	52 of 52 $P < 0.001$
		NLM	52 of 52 $P < 0.001$
SD of residuals of unfiltered data	Truncated data	LM	52 of 52 $P < 0.001$
		NLM	52 of 52 $P < 0.001$

Linear model analysis for Unfiltered and Filtered data

FIGURE 3. Boxplots of the SD of residuals excluding outliers from the trend line using a linear model of change, for unfiltered (*black*) and filtered (*red*) data, at each of the 52 nonblindspot 24-2 VF locations. For example, at (+3, +15), which is superior-temporal, the interquartile range (IQR) for unfiltered data is 1.91 times the IQR for the filtered data.

locations), to allow for the possibility that VF locations do not necessarily correspond solely with one half of the optic disc, in particular when near the horizontal raphe³⁴ or when between the optic disc and the fovea.³⁵ The regression coefficients β_s were constrained to be non-negative based on the assumption that if sensitivity at one particular location is high then it should not predict that the sensitivity at another location will be low. The estimation of β_i , denoted by $\hat{\beta}_i$, was determined by least absolute shrinkage and selection operator (LASSO).³⁶ The LASSO often is used for automatic variable selection when predictors are highly dimensional and correlated. The method shrinks the coefficients of noninformative variables to zero by penalizing on the sum of the absolute size of the regression coefficients. To simplify the filter and avoid overfitting, a subset of the eight best predictors was selected through LASSO for each VF location.

For all eyes in the longitudinal dataset, the predicted sensitivity at location j ($\hat{y}_j$) was defined as

$$\hat{y}_j = \hat{\alpha} + \sum_{i=1}^{37} \hat{\beta}_i x_i, \quad (2)$$

where $\hat{\alpha}$ and $\hat{\beta}_i$ s were obtained from the filter constructed on the cross-sectional data and predictors x_i s were defined in the same manner as in Equation 1. Finally, the filtered value at location j was defined as the average of the observed and predicted sensitivity, such that $\hat{y}_j = 0.5(y_j + \hat{y}_j)$.

To assess the performance of the filter, longitudinal trend analyses were performed on the filtered and unfiltered data, for

each location in the VF. Specifically, the trend over time for each eye was modeled by a linear model (LM), and a nonlinear model (NLM) with the following forms, respectively:

$$y_j^{eye} = a_j^{eye} + b_j^{eye} t^{eye} + e_j^{eye} \quad (3.1)$$

$$y_j^{eye} = a_j^{eye} + b_j^{eye} \exp(t^{eye}) + E_j^{eye}, \quad (3.2)$$

where e_j^{eye} and E_j^{eye} are normally distributed errors with mean zero and constant variance. In Equations 3.1 and 3.2, y_j^{eye} is the observed sensitivity for each eye at location j using the first to second last visits (visits 1: $[n - 1]$), and t is the time for each corresponding visit. Then, the LM and NLM analyses were repeated for filtered data for each eye. Use of a nonlinear model for trend analyses is supported by a recent study that found that mean deviation in dB appears to decline exponentially over time.³⁷ The SDs of the residuals from the trends over time served as a measure of variability that is unaffected by the possibility of progression over time for that eye. This was compared between observed and filtered data sequences. Prediction error (PE), which is defined as the deviation between the observed sensitivity at the last visit (visit n) and the sensitivity that would be predicted by extrapolating the trend from visits 1 to $[n - 1]$, also was compared. The PE has been used widely as a surrogate of the performance of new techniques to detect change in series of VF measures and was first suggested by McNaught et al.³⁸ As a formal comparison, a Wilcoxon matched-pairs one-sided test between the filtered

Linear model analysis for Unfiltered and Filtered data

FIGURE 4. Boxplots of the absolute prediction error ($|PE|$) excluding outliers using a linear model of change, for unfiltered (*black*) and filtered (*red*) data, at each of the 52 nonblindspot 24-2 VF locations, and a *blue asterisk* is marked at a VF test location if $|PE|$ are significantly different ($P < 0.05$) after filtering through a Wilcoxon matched-pairs 1-sided test at that VF location. For example, the IQR of the prediction error for unfiltered data is 1.3 times the IQR for the filtered data at VF location superior temporal ($-3, +3$).

and unfiltered data was performed on the SD of the residuals and the absolute values of the PE ($|PE|$).

Studies show that retinal ganglion cell (RGC) responses saturate at high contrasts, implying that the probability of detecting perimetric stimuli may not continue to increase as contrast is increased above a point where RGC response saturation has occurred. For example, sensitivities at VF locations that are estimated to be 2 dB during SAP testing may not really be different from locations that are estimated to be 10 dB by SAP testing, and perhaps should not be treated differently. A recent study³⁹ concluded that sensitivities estimated to be between 0 and approximately 15 to 19 dB during SAP perimetry are unreliable, and are only weakly correlated with true functional status within that range. Therefore as a secondary analysis, we used 19 dB as the lower limit of the reliable stimulus range of standard perimetry to improve the reliability of sensitivities. The filter was reconstructed using the cross-sectional dataset after setting thresh-

olds < 19 dB equal to 19 dB. Then all analyses were repeated after applying this change to the longitudinal dataset in an additional effort to reduce noise.

RESULTS

The cross-sectional dataset consisted of 1057 eyes of 637 participants, which were used to determine the informative predictors and their weights while building the filter. The longitudinal dataset consisted of 112 eyes of 62 participants with a minimum of five visits per eye for which eligible, reliable results were available for SAP and HRT. Table 1 shows the characteristics of both datasets. Among all longitudinal sequences at the 52 VF locations, 2.6% had a second to last VF sensitivity below 19 dB, and 2.5% had a last VF sensitivity below 19 dB.

Figure 1 shows HRT RA sectors that contributed to the filter for each of the 52 VF locations obtained by applying LASSO to

TABLE 3. Comparison of Absolute Values of PE ($|PE|$) From LM and NLM Analyses, Compared Between Unfiltered and Filtered Data, With Filters Constructed From Either Observed or Truncated Data

Hypothesis	Base Data	Model	Wilcoxon Signed-Rank Test
$ PE $ of filtered data <	Observed data	LM	43 of 52 $P < 0.05$
		NLM	52 of 52 $P < 0.05$
$ PE $ of unfiltered data	Truncated data	LM	47 of 52 $P < 0.05$
		NLM	52 of 52 $P < 0.05$

TABLE 4. Comparison Between the SF filter Described in this Paper, an F Filter Derived in the Same Manner, and the Previously Published Filters of Gardiner et al.²³ (SG), and Strouthidis et al.²⁸ (NS)

	SF vs. SG		SF vs. NS		SF vs. F	
	SG	SF	NS	SF	F	SF
SD	33/52	11/52	34/52	13/52	0/52	27/52
PE	1/52	47/52	3/52	17/52	0/52	6/52

The filters are compared based on the SD of residuals from the linear regression analysis and the PE in the observed data. In each case, the numbers represent the number of VF locations (of 52 nonblindspot 24-2 visual field locations) for which the chosen filter performed significantly better with $P < 0.05$.

the observed cross-sectional data. It can be seen that at peripheral VF locations, where correlations between sensitivities at neighboring VF locations tend to be stronger, fewer HRT sectors were included as predictors. At more central VF locations, for example, (+3, +3) and (+9, +3), generally a greater number of HRT sectors were included as predictors (i.e., within the best eight predictors), which is consistent with previous findings on the association between structural and functional test results.⁴⁰

The shape of the filter was examined at each VF location. Figure 2 illustrates the VF locations and HRT sectors (shaded in gray) that were within the best eight predictors and, therefore, used to construct the filters at three sample VF locations (−21, −3), (−9, +3), and (+3, +3) (marked by red circles in each case). The darker the shading for the location or sector the larger its effect was as a predictor. The remaining contribution to the filtered value comes from the observed value at the location itself. It can be seen that the filters at the three VF locations shown conform to the course of RGC axons within the RNFL. Note that when a stimulus is displayed in the superior nasal VF this corresponds to the temporal region of the retina, and, hence, the corresponding optic disc sectors are in the temporal ONH. It can be seen that for the three filters shown, the most influential location for predicting sensitivity at the locations shown in red, excluding the contribution of the location being assessed, always was an immediate neighbor.

Figure 3 shows boxplots of the SD of residuals about the trend from an LM analysis at each of the 52 non-blindspot VF locations for filtered (red) and unfiltered (black) data. It can be seen that there is a greater range of values for unfiltered data compared to filtered data, indicating that the filter is reducing variability about the trendline. The SD were significantly smaller for the filtered data than the unfiltered data at all 52 VF locations ($P < 0.001$). The mean and 95% confidence limits (in brackets) of SDs for the filtered and unfiltered data though LM analysis are 1.06 (0.25, 3.42) and 1.53 (0.37, 4.68). Overall the SDs were reduced by 0.46 by filtering on average, and the 95% confidence limits for the difference was (−0.27, 1.74). Similarly, boxplots (Supplementary Figs. S3.1–S3.3) which compared SD of the residuals between filtered and unfiltered data from a NLM analysis, and between filtered and unfiltered data when the filter was constructed using truncated data using either LM or NLM analyses, showed that the SD of residuals always were significantly smaller at all VF locations after filtering. A formal comparison using Wilcoxon signed-rank tests of matched pairs of SD between filtered and unfiltered data shown in Table 2

suggests that filtering reduces variability of residuals about the trend at all 52 VF locations ($P < 0.001$ for all analyses).

Boxplots of the absolute values of PE from an LM analysis at each of the 52 VF locations for filtered (red) and unfiltered (black) data are shown in Figure 4. It can be seen that for the majority of VF locations, such as central locations (+3, +3), (−3, +3), (+9, +3), (+9, −3), (−3, −3), and so forth, the PE are smaller for filtered data than unfiltered data. The one-sided Wilcoxon matched pair tests showed that at 43 of 52 VF locations, PE were smaller for the filtered data than the unfiltered data ($P < 0.05$). Through LM analysis, the mean and 95% confidence limits (in brackets) of PE for the filtered and unfiltered data are 1.67 (0.05, 7.12) and 1.86 (0.05, 7.55). The overall PE were reduced by 0.20 dB on average by filtering, and the 95% confidence limits for the difference was (−1.28, 2.02 dB). Similarly, analyses using truncated data suggested that PE were reduced at 47 of 52 VF locations after filtering (Supplementary Fig. S4.1; Table 3). Using NLM analyses, with either observed or truncated data, PE were significantly reduced ($P < 0.05$) at all 52 VF locations after filtering (Supplementary Figs. S4.2, S4.3; Table 3).

Additionally, our new structure-function (SF) filter was compared to two previously published filters from Gardiner et al.²³ (SG) and Strouthidis et al.²⁸ (NS), and a function-only (F) filter, which was derived similarly but using functional data only rather than combining structural and functional data. Results of comparison based on the reduction of SDs of residuals from the LM analysis and the PE were shown in Table 4. It is seen that including structural measures in the filter (i.e., SF versus F) improves fit at 27 locations and PE at 6 locations. Model fits and PE were not made worse by applying filtering at even a single VF location. Perhaps more interestingly, it is seen that while the filters of SG and NS may provide smaller residuals (indicating that they fit the data better) at many locations, the new filter generally has smaller PE, which implies that our new filter is not merely fitting the noise, but is revealing more of the true change signal.

To evaluate the effectiveness of truncating the data at 19 dB, SD of the residuals from the trends were compared between observed and truncated data (using filtered data in both cases). Figure 5 shows that the SD of the residuals using LM on filtered data were smaller for truncated data than for observed data at some VF locations; for example, central locations (+3, +3), (+3, +9), and outer regions in the superior field. Similar results can be seen for the NLM analysis (Supplementary Fig. S5.1). Results of Wilcoxon signed rank tests shown in Table 5 also suggest that truncating the data significantly reduced the SD of

TABLE 5. Comparison of SD of Residuals From LM and NLM Analyses Using Filtered Data, Comparing Between Observed and Truncated Data

Hypothesis	Base Data	Model	Wilcoxon Signed-Rank Test
SD of residuals of truncated data	Filtered	LM	41 of 52 $P < 0.05$
< SD of residuals of observed data		NLM	40 of 52 $P < 0.05$

FIGURE 5. Boxplots of the SD of residuals excluding outliers from the linear trend line for filtered data, using all the observed data (*black*) and after truncating the data at 19 dB (*red*), at each of the 52 nonblindspot 24-2 VF locations, and a *blue asterisk* is marked at a VF test location if the SD are significantly different ($P < 0.05$) for filtered data after truncating through a Wilcoxon matched-pairs one-sided test at that VF location.

residuals ($P < 0.01$) at 41 and 40 VF locations for the LM and NLM analyses, respectively.

DISCUSSION

Recent studies^{41–43} that integrated structural and functional test results to improve the detection of glaucoma tended to adopt a Bayesian approach. Unlike the conventional frequentist approach of using ordinary least squares (OLS) to perform VF trend analyses, the Bayesian approach has the flexibility of adding more variables to an inference model and strengthens posterior beliefs (the reader's belief in the parameters conditional on the data) when prior beliefs (the reader's belief in an uncertain parameter before evaluating the results of a study) are informative. Medeiros et al.⁴¹ used a joint multivariate mixed-effects model within a Bayesian hierarchical modeling framework to integrate information from the SAP VF index (VFI) and average RNFL thickness from scanning laser polarimetry (SLP) for a cohort of participants. Their study used noninformative priors to model the random effects and updated the posterior estimate in each patient through information from the whole sample. It found that the method improved detection of glaucoma progression compared to standard OLS linear regression of VFI measurements alone. Russell et al.⁴² applied a Bayesian normal error linear regression on a sample that consisted of patients with ocular hypertension (OHT). The model used individual imaging measurements of global neuroretinal rim area from HRT, as

an informative prior, to predict changes in perimetric mean sensitivity over time. Compared to the OLS linear regression approach for mean sensitivity, the Bayesian method suggested VF progression rates can be estimated more accurately. However, both methods used global indices/measurements, such as VFI and mean sensitivity for trend analyses, and did not attempt to exploit the spatial correlations between VF locations. VF evaluations based on global indices are more likely to overlook focal damage than evaluations based on pointwise data. Zhu et al.⁴³ used RNFLT from SLP to predict pointwise VF sensitivities from SAP using a BRBF (radial Basis function customized under a Bayesian framework) approach, which improved predictions of pointwise VF sensitivities compared to OLS linear regression. Nonetheless, those Bayesian approaches rely on structural measurements that relate closely to function. The filtering method proposed in this study, which leverages the physiological relations between VF locations and ONH regions, appears to be advantageous in reducing variability and improving predictability, even when the structure–function association appears poor.

It should be emphasized that the data used to build the filter were collected at a tertiary specialty glaucoma clinic. Therefore, the filter will accentuate glaucomatous VF defects, and would be inappropriate to use if nonglaucomatous patterns of VF damage, such as those brought about by nonophthalmic neurological pathology, are suspected. However, de-emphasizing nonglaucomatous defects may be appealing if a physician is trying to determine whether a patient has glaucoma, or whether their glaucoma is progressing. Note that

FIGURE 6. Plot of the number of predictors used in the construction of the structure-function filter against the absolute prediction error ($|PE|$) obtained using the filtered values through the linear regression analysis, averaged across the 52 nonblindspot 24-2 VF locations.

the filter was derived using data collected with an HRT II instrument, but was tested using longitudinal data collected with an HRT I instrument, which may cause an additional source of variability. Furthermore, the study used rim areas as a structural measure. The filter could be refined further by using measurements from structural tests that correlate better with functional testing, such as RNFL thickness (RNFLT) from SD-OCT. It is important to note that though this post hoc filter can reduce noise and improve the accuracies of future predictions, improving data acquisition techniques which will allow more accurate change detection still are preferred.

In the construction of the SF filter for each VF test location, LASSO was used to select the eight most informative predictors of the possible 37. However, LASSO should be used with caution, because in the special case when two predictors are highly correlated, it will randomly pick one variable rather than both and set the other variable to zero, which might omit the important predictors. However, while this defect of LASSO might affect the shape of the filter, the filtered values should not be affected. Another point to emphasize is that the choice of the number of predictors in the construction of the filter, with eight being used in the current work, is somewhat arbitrary. Including extra predictors in the filter will reduce the PE, but eventually the benefit derived from including additional predictors will be diminished. In Figure 6, as the number of predictors increased, the PE was initially reduced significantly. However, the reduction in PE with the inclusion of additional predictors then levels off when the number of predictors reaches approximately six. This association between PE and number of included predictors will vary between VF locations, and so we chose to use eight predictors per location in this study to safeguard against omitting useful information from the filter while also avoiding over-fitting. When the number of

predictors included in the filter is increased, naturally it will increase the number of HRT sectors included. Though altering the LASSO criteria will alter the shape of the filter at some test locations, the filter still improves the SD of residuals and PE compared to the unfiltered values.

Filtering works on some patients better than others. The LM analysis suggested that there were 15 of 112 eyes that demonstrated a reduction in SD of residuals less than 0.35 through filtering, while seven eyes had a reduction in SD greater than 0.75. One possible explanation is the heterogeneity among eyes, which has been blurred out by our “one size fits all” filter. This may suggest a place for “personalized” filters, which are customized for each individual eye. Some studies^{30,44,45} have used a computational model that combines an individual eye’s structural and functional test data with simple biometric data to derive individually customized maps relating each VF location to the ONH. It is possible to refine the filter developed in this paper to a “personalized” filter based on individual customized mapping of VF locations to the ONH. However, it recently has been reported that the benefit of such personalization on the structure-function relation is marginal (Ganeshrao S, et al. *IOVS* 2014;155:ARVO E-Abstract 962). Another possible reason for the variability in how different eyes benefit from filtering could be that some individuals in our longitudinal dataset have more follow-up tests than others. To evaluate the impact of longitudinal sequence length on filtering, two groups of data were created, one with sequence length between five and seven tests, the other with eyes having sequence length greater than seven (to balance the number of eyes within each group, so 61 eyes have shorter sequence and 51 eyes have longer sequence). The overall mean of the SDs of residuals for the shorter sequences was 1.61 using unfiltered data and 1.13 using filtered data, while for longer sequences

these values were 1.44 and 0.99, respectively. Notably, filtering improved the results regardless of sequence lengths. Another possible explanation for the differences between individuals is that perhaps some individuals are worse test takers than others, providing less reliable measurements.

It is almost certain that any kind of spatial filtering will reduce detectability of very small defects and VF change occurring with only one or two VF locations progressing. For example, a Gaussian filter applied to VF data often made it harder to detect small, gradual progressive field changes.²⁷ An example of a small nasal step defect is shown in Supplementary Figure S7. It can be seen in this particular example that our filter was able to capture VF deterioration occurring within a small, localized defect; however, there may be cases in which change occurring within a very small defect is “blurred out,” especially if the defect is surrounded by healthy, stable VF locations. Therefore, it is recommended that filtered and unfiltered data be examined, especially when the two approaches produce inconsistent results with respect to glaucomatous VF progression.

To conclude, the filter proposed in this study incorporated data from functional and structural tests, like others that have been designed previously,^{23–26,28} and provides better prediction of change in series of pointwise VF sensitivities when compared to using unfiltered or raw VF data alone. However, the true impact of filtering applied to series of VF data still remains untested. Filtering only improves predictions “on average.” Most VF series are very noisy, so any dampening of this noise also will improve the prediction on average. Since most VF series gathered during clinical testing are relatively stable with few rapidly deteriorating cases,⁴⁶ filtering is likely to yield clinical benefits by reducing the likelihood that a glaucomatous visual field will falsely be deemed progressing.

Acknowledgments

Supported by National Institutes of Health (NIH) Grants R01-EY020922 (SKG) and R01-EY019674 (SD). The authors alone are responsible for the content and writing of the paper.

Disclosure: **L. Deng**, None; **S. Demirel**, None; **S.K. Gardiner**, Allergan (C)

References

- Kass MA, Gordon MO, Gao F, et al. Delaying treatment of ocular hypertension: the ocular hypertension treatment study. *Arch Ophthalmol*. 2010;128:276–287.
- Chen P, Cady R, Mudumbai R, Ngan R. Continued visual field progression in eyes with prior visual field progression in patients with open-angle glaucoma. *J Glaucoma*. 2010;19:598–603.
- Gardiner SK, Demirel S, Johnson CA. Perimetric indices as predictors of future glaucomatous functional change. *Optom Vis Sci*. 2011;88:56–62.
- Medeiros FA, Alencar LM, Zangwill LM, Sample PA, Weinreb RN. The relationship between intraocular pressure and progressive retinal nerve fiber layer loss in glaucoma. *Ophthalmology*. 2009;116:1125–1133.
- Medeiros FA, Alencar LM, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Prediction of functional loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol*. 2009;127:1250–1256.
- Harwerth RS, Carter-Dawson L, Smith EL, Barnes G, Holt WF, Crawford ML. Neural losses correlated with visual losses in clinical perimetry. *Invest Ophthalmol Vis Sci*. 2004;45:3152–3160.
- Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120:701–713.
- Miglior S, Zeyen T, Pfeiffer N, Cunha-Vaz J, Torri V, Adamsons I. Results of the European Glaucoma Prevention Study. *Ophthalmology*. 2005;112:366–375.
- Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol*. 2005;123:464–470.
- Strouthidis NG, Scott A, Peter NM, Garway-Heath DE. Optic disc and visual field progression in ocular hypertensive subjects: detection rates, specificity, and agreement. *Invest Ophthalmol Vis Sci*. 2006;47:2904–2910.
- Leung CK, Cheung CYL, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a study on optical coherence tomography guided progression analysis. *Invest Ophthalmol Vis Sci*. 2010;51:217–222.
- Heijl A, Bengtsson B, Hyman L, Leske MC. Natural history of open-angle glaucoma. *Ophthalmology*. 2009;116:2271–2276.
- Medeiros FA, Zangwill LM, Alencar LM, Sample PA, Weinreb RN. Rates of progressive retinal nerve fiber layer loss in glaucoma measured by scanning laser polarimetry. *Am J Ophthalmol*. 2010;149:908–915.
- Gardiner SK, Johnson CA, Demirel S. The effect of test variability on the structure-function relationship in early glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 2012;250:1851–1861.
- Bowd C, Zangwill LM, Medeiros FA, et al. Structure-function relationships using confocal scanning laser ophthalmoscopy, optical coherence tomography, and scanning laser polarimetry. *Invest Ophthalmol Vis Sci*. 2006;47:2889–2895.
- Rao HL, Zangwill LM, Weinreb RN, Leite MT, Sample PA, Medeiros FA. Structure-function relationship in glaucoma using spectral-domain optical coherence tomography. *Arch Ophthalmol*. 2011;129:864–871.
- Chauhan B, House P. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology*. 1991;98:79–83.
- 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.
- Spry PG, Johnson CA. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol*. 2002;47:158–173.
- Wild J, Searle A, Dengler-Harles M, O'Neill E. Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmol*. 1991;69:210–216.
- Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci*. 1999;40:648–656.
- Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
- Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vision Res*. 2004;44:839–848.
- Crabb DP, Edgar DF, Fitzke FW, McNaught AI, Wynn HP. New approach to estimating variability in visual field data using an image processing technique. *Br J Ophthalmol*. 1995;79:213–217.
- Crabb DP, Fitzke FW, McNaught AI, Edgar DF, Hitchings RA. Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology*. 1997;104:517–524.

26. Fitzke FW, Crabb DP, McNaught AI, Edgar DF, Hitchings RA. Image processing of computerised visual field data. *Br J Ophthalmol*. 1995;79:207-212.
27. Spry PG, Johnson CA, Bates AB, Turpin A, Chauhan BC. Spatial and temporal processing of threshold data for detection of progressive glaucomatous visual field loss. *Arch Ophthalmol*. 2002;120:173-180.
28. Strouthidis NG, Vinciotti V, Tucker AJ, Gardiner SK, Crabb DP, Garway-Heath DE. Structure and function in glaucoma: the relationship between a functional visual field map and an anatomic retinal map. *Invest Ophthalmol Vis Sci*. 2006;47:5356-5362.
29. Garway-Heath DE, Poinosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809-1815.
30. Denniss J, Turpin A, McKendrick AM. Individualized structure-function mapping for glaucoma: practical constraints on map resolution for clinical and research applications. *Invest Ophthalmol Vis Sci*. 2014;55:1985-1993.
31. Garway-Heath DE, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci*. 2000;41:1774-1782.
32. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. 2007;26:688-710.
33. Harwerth R, Wheat JL, Fredette MJ, DR A. Linking structure and function in glaucoma. *Prog Retin Eye Res*. 2010;29:249-271.
34. Vrabec F. The temporal raphe of the human retina. *Am J Ophthalmol*. 1966;62:926-938.
35. He L, Ren R, Yang H, et al. Anatomic vs. acquired image frame discordance in spectral domain optical coherence tomography minimum rim measurements. *PLoS One*. 2014;9:e92225.
36. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc B*. 1996;58:267-288.
37. Pathak M, Demirel S, Gardiner SK. Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma. *Invest Ophthalmol Vis Sci*. 2013;54:5505-5513.
38. McNaught A, Hitchings RA, Crabb D, Fitzke F. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 1995;233:750-755.
39. 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.
40. Gardiner SK, Johnson CA, Cioffi GA. Evaluation of the structure-function relationship in glaucoma. *Invest Ophthalmol Vis Sci*. 2005;46:3712-3717.
41. Medeiros FA, Zangwill LM, Girkin CA, Liebmann JM, Weinreb RN. Combining structural and functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol*. 2012;153:1197-1205.
42. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DE. Improved estimates of visual field progression using Bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci*. 2012;53:2760-2769.
43. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51:5657-5666.
44. Turpin A, Sampson GP, McKendrick AM. Combining ganglion cell topology and data of patients with glaucoma to determine a structure-function map. *Invest Ophthalmol Vis Sci*. 2009;50:3249-3256.
45. Denniss J, McKendrick AM, Turpin A. An anatomically customizable computational model relating the visual field to the optic nerve head in individual eyes. *Invest Ophthalmol Vis Sci*. 2012;53:6981-6990.
46. Saunders LJ, Russell RA, Kirwan JF, McNaught AI, Crabb DP. Examining visual field loss in patients in glaucoma clinics during their predicted remaining lifetime. *Invest Ophthalmol Vis Sci*. 2014;55:102-109.