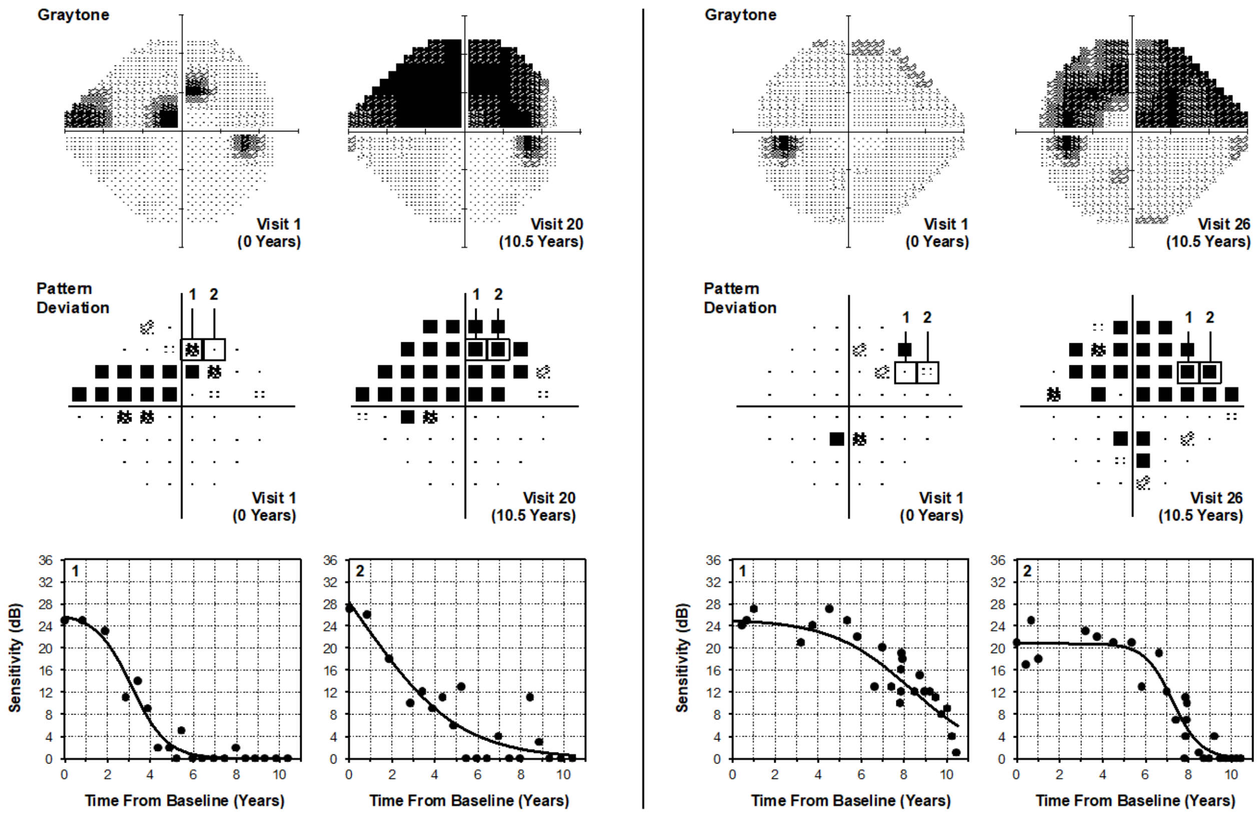
SUPPLEMENTARY MATERIAL 1

Computer Simulations of Visual Fields

As an overview, the reconstruction of "real-world" visual field tests was achieved by obtaining longitudinal estimates of changes in point-wise visual field sensitivity (which provides an estimate of the "true" sensitivity component) and estimates of visual field variability (representing the "noise" component) from the cohort of glaucoma eyes followed in this study. This method has been published in detail elsewhere (Wu Z, Medeiros FA. Development of a Visual Field Simulation Model of Longitudinal Point-Wise Sensitivity Changes from a Clinical Glaucoma Cohort. Trans Vis Sci Tech 2018;7(3):22).

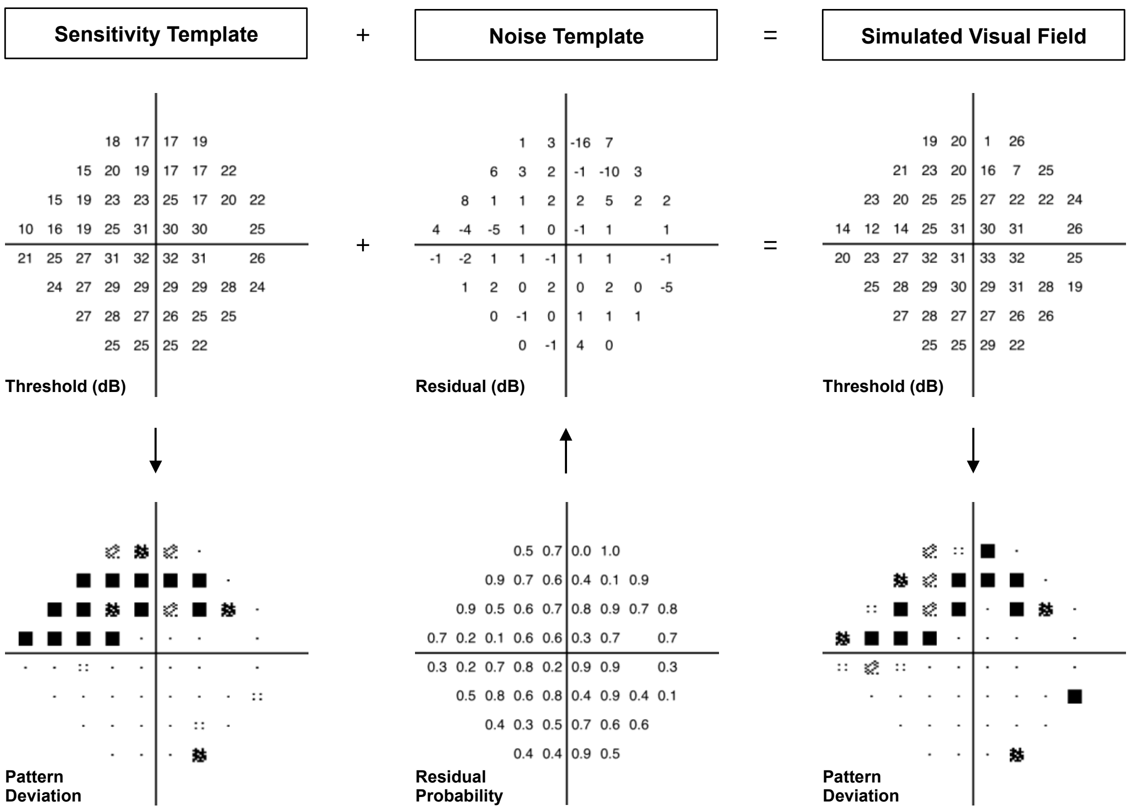
First, longitudinal estimates of point-wise visual field sensitivity changes for each eye were obtained by fitting a sigmoid regression model to the measured point-wise sensitivities. The sigmoid model can capture non-linear changes and has been shown previously to outperform linear based models. The sigmoid model allows for up to two asymptotes to be fitted, which takes into account an initial period of relative stability in visual field sensitivity (such as that seen when sensitivity is still near-normal) and at the perimetric floor. The expression of this model is as follows: $s = \gamma / (1 + e^{\alpha + \beta x})$, where s is the measured sensitivity in decibels (dB), γ is an estimate of the initial sensitivity, α is a coefficient that indicates of how soon a steep decline begins, β is a coefficient that indicates the steepness of this decline, and x represents the time from baseline. Fitting of this model was performed using an iterative feasible generalized nonlinear least squares method. Examples of different patterns of point-wise sensitivity change fitted with this sigmoid model are shown in Supplementary Figure 1. After fitting the sigmoid regression model to the existing visual field data from all glaucoma patients, the parameters of the sigmoid regression model could then be used to estimate the "true" visual field sensitivity for all locations of the visual field (or the "true" pattern of damage) at any given time point. We refer to these estimates as a "sensitivity template". Note that in this study, we only included eyes with an estimated MD at a time point 2 years from the last visit that was ≥ -15 dB to exclude patients with more advanced glaucoma.



Supplementary Figure 1: *Two examples illustrating how the sigmoid regression model fitted threshold sensitivity (in decibels [dB]) changes over time, at two different locations for each of the examples, indicated by numbers and lines on the pattern deviation map corresponding to the numbers on the top left corner of each graphs.*

Second, estimates of the visual field test-retest variability (or the “noise” component) were obtained using the residuals from the sigmoid regression model, which are the differences between the measured and fitted sensitivities. The residuals from the entire cohort were then binned based on the fitted sensitivities (to the nearest 1-dB), and the distributions of these residuals were used to generate “empirical probability distribution functions (PDFs)” for each fitted sensitivity bin. The residuals at each location for each eye during each visit were then converted into probabilities based on these empirical PDFs. This provides a normalized estimate of the level of deviation of the individual’s response from the estimated “true” sensitivity. The probabilities at all locations for a test thus collectively provide a template of patient performance that accounts for the correlation between measured point-wise sensitivities during each test (as a result of global visit effects such as patient vigilance; in a similar way to joint probabilities), and we refer to this set of probabilities as a “noise template”. These “noise templates” from all visits of all participants thus created a database of patient-level variability that was used for the simulations.

For a given sequence, “real-world” visual field results were then reconstructed by first deriving the “true” sensitivity at each location for each test at each time point using a “sensitivity template”. Then, measurement variability was added to the “true” sensitivity at each location for each test by sampling the residual from the empirical PDF based on the “true” sensitivity using the probabilities at each location according to a randomly chosen “noise template”. This is illustrated using an example in Supplementary Figure 2, showing how measurement variability is added to an eye estimated to have a superior hemifield defect based on the longitudinal modeling. For the randomly selected proportion of participants assumed to be “responders” to the new treatment (i.e. not exhibiting any further visual field progression), the baseline “sensitivity template” was used for all subsequent follow-up visits to provide a scenario where visual field sensitivities were stable over time.



***Supplementary Figure 2**: Illustration of the simulation of a single visual field test result, achieved by combining a “sensitivity template” (representing the “true” pattern of estimated sensitivity) with a* *"noise template", where the probabilities at each location were converted into estimates of measurement variability (or "noise") based on the empirical probability distribution function of the corresponding sensitivity bin at that location.*

Point-Wise Event-Based Analysis

To obtain test-retest limits to be used for the point-wise event-based analysis, visual fields at one time point were reconstructed to provide a test-retest scenario. Note that this method has also been published in detail elsewhere (Wu Z, Medeiros FA. Comparison of Visual Field Point-Wise Event-Based and Global Trend-Based Analysis for Detecting Glaucomatous Progression. Trans Vis Sci Tech 2018;7(4):20). For each eye included in this study, 100 sequences with three tests at baseline were simulated, and the difference between the third pattern deviation (PD) value and the average of the first two PD values at each location was determined. These differences were binned according to the averaged value after rounding to the nearest 1-dB. The 5th percentile of these differences for each bin was determined, and they were smoothed across sensitivity bins to provide a more precise estimate of the test-retest limits that would otherwise have been integers. Progression was then considered to have occurred if three or more locations showed a change exceeding these test-retest limits from the two baseline tests at three or more consecutive visits.