

# Retesting Visual Fields: Utilizing Prior Information to Decrease Test–Retest Variability in Glaucoma

Andrew Turpin,<sup>1</sup> Darko Jankovic,<sup>2</sup> and Allison M. McKendrick<sup>2</sup>

**PURPOSE.** To determine whether sensitivity estimates from an individual's previous visual field tests can be incorporated into perimetric procedures to improve accuracy and reduce test–retest variability at subsequent visits.

**METHODS.** Computer simulation was used to determine the error, distribution of errors and presentation count for a series of perimetric algorithms. Baseline procedures were Full Threshold and Zippy Estimation by Sequential Testing (ZEST). Retest strategies were (1) allowing ZEST to continue from the previous test without reinitializing the probability density function [*pdf*]; (2) running ZEST with a Gaussian *pdf* centered about the previous result; (3) retest minimizing uncertainty (REMU), a new procedure combining suprathreshold and ZEST procedures incorporating prior test information. Empiric visual field data of 265 control and 163 patients with glaucoma were input into the simulation. Four error conditions were modeled: patients who make no errors, 15% false-positive (FP) with 3% false-negative (FN) errors, 15% FN with 3% FP errors, and 20% FP with 20% FN errors.

**RESULTS.** If sensitivity was stable from test to retest, all the retest algorithms were faster than the baseline algorithms by, on average, one presentation per location and are significantly more accurate ( $P < 0.05$ ). When visual fields changed from test to retest, REMU was faster and more accurate than the other retest approaches and the baseline procedures. Relative to the baseline procedures, REMU showed decreased test–retest variability in impaired regions of visual field.

**CONCLUSIONS.** The obvious approaches to retest, such as continuing the previous procedure or seeding with previous values, have limitations when sensitivity changes between tests. REMU, however, significantly improves both accuracy and precision of testing and displays minimal bias, even when fields change and patients make errors. (*Invest Ophthalmol Vis Sci* 2007;48:1627–1634) DOI:10.1167/iovs.06-1074

There are many clinical situations in which automated perimetry is conducted on the same patient on a regular basis—for example, in the management of people with glaucoma or glaucomatous risk factors. Consequently, for many

patients, there is existent information regarding their visual field sensitivity. It is not clear how such previous test information is used in most commercial perimeters. Either it is not used at all, or its usage is not clearly described in the scientific literature or perimetric manuals. One commercial implementation of a perimetric algorithm using prior patient information was the Full-Threshold from Prior Data strategy of the Humphrey Field Analyzer 1 (HFA; Carl Zeiss Meditec, Inc., Dublin, CA) which commenced a staircase strategy based on a prior test of the eye.<sup>1</sup> The Full Threshold from Prior Data strategy was found to have a test duration similar to that of the standard Full-Threshold algorithm and so became obsolete.<sup>1</sup> Modern Bayesian perimetric procedures may be more likely to benefit from the incorporation of prior data from a tested individual.

Many perimetric strategies incorporate population information regarding likely thresholds. For example, the Swedish Interactive Thresholding Algorithm (SITA) of the HFA maintains two probability functions based on population data: one representing the probability of each possible outcome, assuming the test location is abnormal, and the other assuming the location is normal.<sup>2</sup> When SITA terminates, the mode of the probability functions is returned as the sensitivity estimate. Similarly, the Zippy Estimation by Sequential Testing (ZEST) procedure commences with a probability density function (*pdf*) that typically represents the distribution of population sensitivity.<sup>3,4</sup> A notable exception is the implementation of ZEST within a perimeter (Matrix; Carl Zeiss Meditec, Inc.) which commences with a *pdf* that assigns equal probability to all possible outcomes, to avoid having the prior *pdf* bias the final sensitivity estimate.<sup>5</sup> Population information can also be used to influence the selection of the starting intensity for staircase procedures.

As well as incorporating population-based knowledge into perimetric test procedures, it may also be beneficial to include information from a given patient's previous tests. However, as perimetric results often display high test–retest variability, particularly in areas of visual field damage,<sup>6,7</sup> automatically using previous sensitivity estimates to seed subsequent tests may counterproductively increase test times or measurement error. Furthermore, if there is a change in visual field sensitivity, relying too heavily on previous estimates may hamper detection of such change. It is not readily apparent whether greater advantages arise from seeding perimetric tests with population information or individual previous test information.

We used computer simulation to explore test procedures developed for retesting patients. Although we investigated hundreds of procedure variations, we report only the best performers herein. Specifically, we used a Bayesian procedure (ZEST), as Bayesian-like procedures are common in perimetry (for example, the SITA family of algorithms used in the HFA, and the ZEST procedure itself, used in the Medmont perimeter [Medmont Pty., Ltd., Camberwell, Australia] and the Matrix [Carl Zeiss Meditec]). We compared the performance of retest strategies with the error and precision obtained by simply rerunning the initial test strategy. We also explored the utility of a retest modification of a combined suprathreshold–threshold procedure that we have described previously, Estimation Minimizing Uncertainty (EMU).<sup>8</sup> When modified for retesting

---

From the <sup>1</sup>School of Computer Science and Information Technology, RMIT University, Melbourne, Australia; and the <sup>2</sup>Department of Optometry and Vision Sciences, University of Melbourne, Carlton, Victoria, Australia.

Supported by an Australian Research Council QEII research fellowship (AT). The project was supported by Australian Research Council Discovery Project Grant DP0450820.

Submitted for publication September 10, 2006; revised November 26 and December 14, 2006; accepted February 15, 2007.

Disclosure: A. Turpin, None; D. Jankovic, None; A.M. McKendrick, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Allison M. McKendrick, Department of Optometry and Vision Sciences, University of Melbourne, Corner of Cardigan and Keppel Streets, Carlton, Victoria 3053, Australia; allisonm@unimelb.edu.au.

visual fields, it is called Retest Minimizing Uncertainty (REMU). We sought to find a retest algorithm that enables rapid determination of sensitivity estimates with significantly improved accuracy and reduced test-retest variability. Priority was given to reducing variability, in preference to simply saving test time. We sought for a similar average number of presentations per location as the SITA Standard algorithm, as SITA Standard is considered of generally acceptable test length in practice.

## METHODS

### Computer Simulation

We used the Barramundi simulation model that we have described previously.<sup>3,8</sup> In brief, an input visual field is provided to the model, as well as error characteristics of the model subject and details of the test procedure. The model runs the test procedure responding as if the input sensitivities were the subject's true sensitivity, incorporating the appropriate error responses. The accuracy of the test procedure is assessed by comparing the sensitivity measures that are output by the simulation to those input, and the speed determined by counting the number of presentations required for the procedure to complete. The simulations assume a 24-2 spatial test pattern identical with that used in the HFA.

Two initial test procedures (Full-Threshold [FT] and ZEST) and three retest procedures (ZEST Continue [Z-Cont], ZEST Gaussian [Z-Gauss], and REMU) were incorporated in the simulation.

**Test Procedure 1: Full Threshold.** FT is a staircase algorithm that has been largely replaced by the SITA family of algorithms.<sup>1</sup> We included FT, as the full details of SITA are not available in the public domain. According to the developers of SITA Standard, it was designed to have similar test-retest characteristics as FT but to terminate more quickly, a development goal that appears to have been met based on clinical comparisons of these test procedures.<sup>6,7,9-11</sup> Consequently, we have included FT to provide a surrogate comparison to the test-retest performance of SITA Standard.

FT commences with 4-dB luminance changes until the first response reversal (seeing to nonseeing or vice versa). The step size is then reduced to 2 dB. After two reversals the procedure terminates and sensitivity is estimated as the "last seen" intensity. For each location, the starting estimate of FT is determined according to a "growth pattern." The growth pattern used herein was the same as we have described previously (illustrated in McKendrick and Turpin,<sup>8</sup> Fig. 1). If the measured estimate differs from the starting estimate of FT by more than 4 dB, a second staircase is commenced, using the first returned estimate as the starting intensity for the second staircase. In this situation, the HFA reports both estimates with no instructions on how they should be interpreted and does not use the second estimates when calculating the mean deviation (MD) or pattern standard deviation. Herein, we report the first estimate.

**Test Procedure 2: ZEST.** The ZEST procedure was identical with that described in detail previously.<sup>3,12</sup> The initial *pdfs* were a weighted combination of normal and abnormal sensitivities determined using empiric patient data from 541 normal and 315 glaucomatous visual fields (illustrated in Turpin et al.,<sup>3</sup> Fig. 2). The combined *pdf* represents sensitivities ranging from -10 to 50 dB, with values from -10 to -1 dB and 41 to 50 dB being assigned a small nonzero pedestal probability. The simulated perimeter is only able to present values of 0 to 40 dB, but the *pdf* is extended by 10 dB each side to enable ZEST to return values at the extremities of the 0- to 40-dB range. The ZEST procedure terminated when the standard deviation of the posterior *pdf* was less or equal to 1.6 dB, with the mean of the *pdf* being returned as the sensitivity estimate. The results of the ZEST procedure were used as seeding visual fields for the three retest procedures described in the next sections.

**Retest Procedure 1: Z-Cont.** Z-Cont recommences testing of a given location using the final posterior *pdf* of the initial ZEST as its prior *pdf*. Otherwise, the procedure runs the same as ZEST. Consequently, the initial *pdf* for Z-Cont must already have a standard deviation of less or equal to 1.6 dB. We explored a variety of termination criteria and report the case in which the standard deviation of the *pdf* was required to be  $\leq 0.8$  dB, as this provided a similar average number of presentations per location as SITA Standard.

**Retest Procedure 2: Z-Gauss.** Z-Gauss is a ZEST procedure in which the prior *pdf* for a given location is a Gaussian centered on the sensitivity estimate returned at the first test visit. Gaussian standard deviations from 2.0 to 3.5 dB in steps of 0.5 dB were tested. Z-Gauss was terminated when the standard deviation of the *pdf* was  $\leq 1.5$  dB. Details for a *pdf* standard deviation of 3.0 dB are reported as this provided a similar average number of presentations per test location as SITA Standard. A further modification of the Z-Gauss procedure involved a standard deviation that varied with the sensitivity estimate of the first test according to the "combined" formulas proposed in Table 1 of Henson et al.,<sup>13</sup> but capped to a maximum of 5 dB. This procedure is referred to as Z-Gauss-H.

**Retest Procedure 3: A Modification of the EMU Procedure for Retest (REMU).** The EMU procedure combines a screening strategy with ZEST.<sup>8</sup> EMU was designed to enable accurate estimates of threshold for situations where the neighboring locations are a poor predictor of true threshold, and to reduce test-retest variability in areas of reduced sensitivity. EMU utilizes a larger number of presentations in areas of visual field loss to improve accuracy and repeatability, while maintaining an acceptable total number of presentations across the visual field by adopting a multisampling screening strategy for normal locations. Table 1 shows the algorithm describing how this process is modified for retest to give REMU.

REMU uses a quick suprathreshold test to check whether the sensitivity has decreased from the previous test as shown in steps 4.2.1 through 4.2.4 of Table 1. Steps 1 through 3 in the procedure are important, because any change in general height since the previous

**TABLE 1.** The REMU Procedure to Retest a Patient's Visual Field in the Case of Existing Data

1. Sort all sensitivity values from the first test. Choose four points in different quadrants that are as close as possible to the 85th percentile of sorted points.
2. Fully threshold the four selected points using ZEST.
3. Compute the mean difference between the current sensitivities at the four points and their previous values, giving general height change: GH.
4. For all remaining 48 points in the visual field:
  - 4.1. If the sensitivity at this location in the previous test was less than the age matched normal value for this location, fully threshold this location using Z-Gauss, with the starting *pdf* corrected by GH
  - 4.2. Otherwise:
    - 4.2.1. Set a suprathreshold value ST equal to the previous measured threshold for this location, plus GH, less 2 dB.
    - 4.2.2. Present a stimulus of value ST dB two times.
    - 4.2.3. If ST is seen once and not seen once, present ST a third time.
    - 4.2.4. If ST is seen twice in total, report the sensitivity for this location as the previous measured value plus GH.
    - 4.2.5. If ST is not seen twice, fully threshold this location using Z-Gauss with the Gaussian centred on ST less 2 dB.

test would yield an inaccurate setting of the suprathreshold values used in Step 4.2.1. The check in step 4.1 filters out locations where either there was genuine damage or a mistake was made in the previous test resulting in low sensitivity. In either case variability is high, and a short suprathreshold test at that location is not advisable. In step 4.1, the Gaussian *pdf* is located around the previous sensitivity adjusted by GH, whereas in step 4.2.5, the *pdf* is located 2 dB below the suprathreshold stimulus value.

**Error Models Entered into the Simulation**

To assess the effect of erroneous responses on the test procedures, four error models were applied.

1. No-error observers: no errors are ever made. Any stimulus of lower luminance than the patient's threshold (higher dB) is *not seen*, whereas any stimulus presented at a higher luminance (lower dB) is *seen*. A stimulus equal to threshold has a 50% chance of being seen.
2. Typical FP observers: The threshold used to determine the response was randomly selected from a Gaussian of variable standard deviation depending on the input sensitivity according published formulas (Henson et al.,<sup>13</sup> Table 1) but capped to a maximum of 5 dB. In addition, there was a 15% chance of responding *seen* (FP) and 3% chance of responding *not seen* (FN), independent of the stimulus level presented.
3. Typical FN observers: Response variability was determined as for typical FP observers. In addition, there was a 15% chance of responding *not seen* (FN) and a 3% chance of responding *seen* (FP), independent of the stimulus level presented.
4. Unreliable observers: A 20% chance of FP and FN responses. Response variability was determined as for FP observers.

**Visual Fields Input to the Simulation and Change to Fields Incorporated at Retest**

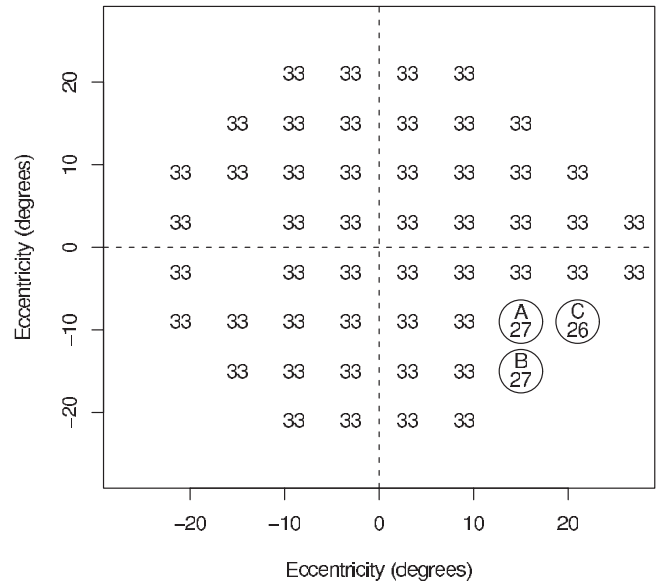
To examine the performance of the procedures on real visual fields, 265 normal visual fields and 163 glaucomatous visual fields were used as input (FT algorithm, 24-2 spatial pattern). The fields were collected for a previous study, at which time written informed consent, in agreement with the tenets of the Declaration of Helsinki, was obtained from the subjects to have their visual field data kept in a deidentified database for further research purposes. Normal patients were aged  $47 \pm 16$  years and glaucomatous patients were aged  $61 \pm 13$  years. Within the glaucomatous group, the visual field deficits ranged from mild to severe (median MD = -1.81 dB, 5th percentile = +2.14 dB, and 95th percentile = -22.55 dB). Visual fields were age corrected to 45 years, altered by 1 dB per decade. The locations adjacent to the blind spot ( $15, \pm 3^\circ$ ) were excluded from analysis.

Three change conditions were applied to the whole visual fields to assess the performance of the retest procedures:

1. No change: the visual field sensitivity was assumed to be stable so no change was incorporated between test and retest.
2. A uniform increase in sensitivity of 3 dB across the entire visual field at the retest visit.
3. A uniform decrease in sensitivity of 3 dB across the entire visual field at the retest visit.

We also assessed performance when the whole-field sensitivity varied by  $\pm 2, \pm 4,$  and  $\pm 6$  dB. The results for the intermediate  $\pm 3$ -dB case are reported herein. The ability of the algorithms to cope with diffuse change was assessed as diffuse variation in sensitivity is common and may reflect causes such as media opacification and nonvisual factors such as anxiety, learning effects, fatigue, or attention.

Performance was also assessed for locations within a simulated deepening scotoma. An artificial visual field was used (Fig. 1) with a scotoma that progressively deepened at each visit. We have demonstrated that FT and Staircase-Quest (an algorithm incorporating those aspects of SITA that appear in the public domain) have increased



**FIGURE 1.** Initial visual field used to simulate an isolated deepening scotoma. The sensitivity of the locations labeled A, B, and C was decreased by 3 dB per simulated test visit.

variability when the starting estimate provided to the procedure is inaccurate,<sup>3</sup> which is most likely to arise on the edge of a scotoma. The artificially progressing fields were not intended to model glaucomatous progression per se, but to explore performance for a known situation where FT (and presumably SITA) performs suboptimally. The field consisted of a true sensitivity of 33 dB for all locations except those labeled A, B, and C in Figure 1. The sensitivity at these three locations was decreased by 3 dB per visit to create a sequence of eight visual fields containing an isolated deepening scotoma. The sensitivity of the remainder of the visual field was stable.

Simulations were run 1000 times for each test procedure, error response model, and visual field change per visit.

**RESULTS**

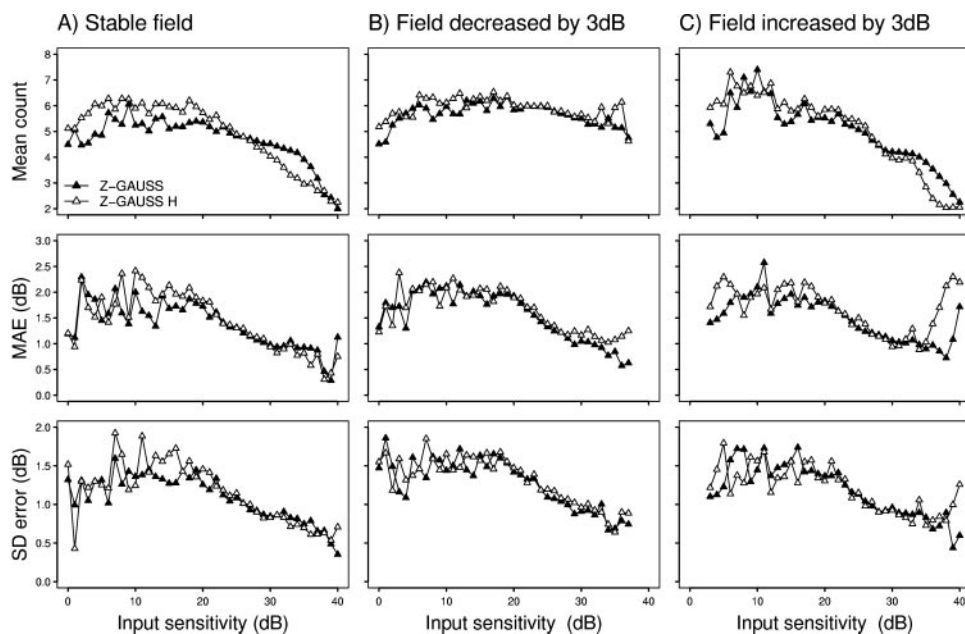
**Comparison of the Two Z-Gauss Procedures**

Figure 2 compares the performance of the two Z-Gauss procedures: Z-Gauss, which used a prior Gaussian *pdf* of standard deviation 3 dB, irrespective of the sensitivity estimate at the first visit; and Z-Gauss-H, which altered the width of the prior Gaussian *pdf* with sensitivity. The figure shows the mean presentation count, mean absolute error (MAE), and the standard deviation of the error as a function of input sensitivity for the no-error observer model. Figure 3A shows that when there was no change in visual field sensitivity from test to retest, Z-Gauss-H was faster than Z-Gauss at high sensitivity and slower to terminate at low sensitivity. For a whole-field decrease in sensitivity (Fig. 3B), there were no consistent differences between the procedures. In contrast, when the whole field increased in sensitivity, the narrow standard deviation of the Z-Gauss-H procedure resulted in rapid termination that was sometimes erroneous. Based on the results presented in Figure 2, the fixed standard deviation procedure (Z-Gauss) was chosen for subsequent experiments and to be used within the REMU procedure.

**Comparison of All Procedures When There Is No Change in Visual Field Sensitivity**

Figure 3 shows the performance of the algorithms when there was no change in visual field sensitivity between the first and





**FIGURE 2.** A comparison of the performance of the Z-Gauss and Z-Gauss-H procedures. Data are presented as a function of the sensitivity input to the simulation and shows mean presentation count (*top*), MAE (*middle*), and the standard deviation of the absolute error (*bottom*) for a stable field (A) and fields decreased (B) or increased (C) by 3 dB.

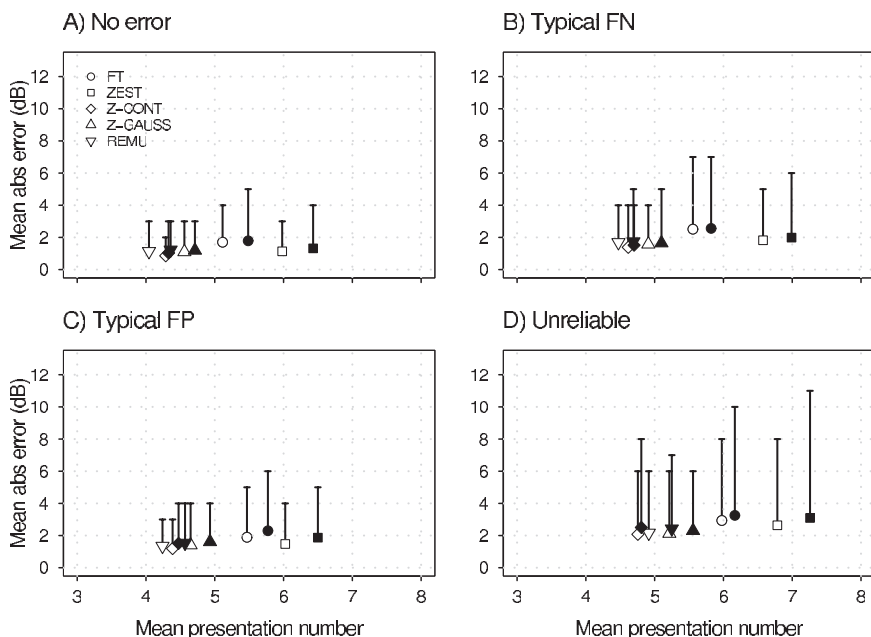
second visits. The absolute error averaged across all locations was plotted against the average number of presentations required for the procedure to terminate. The error bars show the 95th quantile of the absolute error. We were aiming to achieve results closest to the lower left corner of Figure 3, with the smallest error bars, as this represents the lowest average error, lowest spread of error (hence test-retest variability), and fastest test.

Figure 3 demonstrates that when sensitivity was stable from one test to the next, all the retest procedures were faster than the test procedures, with some reduction in the MAE and spread of error. Statistical comparison (ANOVA) resulted in significant differences in the MAE (defined as  $P < 0.05$  on post hoc Holm-Sidak testing) between almost all the test procedures (no-error condition: all procedures significantly different with the exception of Z-Gauss and REMU for patients with glaucoma, and ZEST and REMU for normal subjects; typical FN

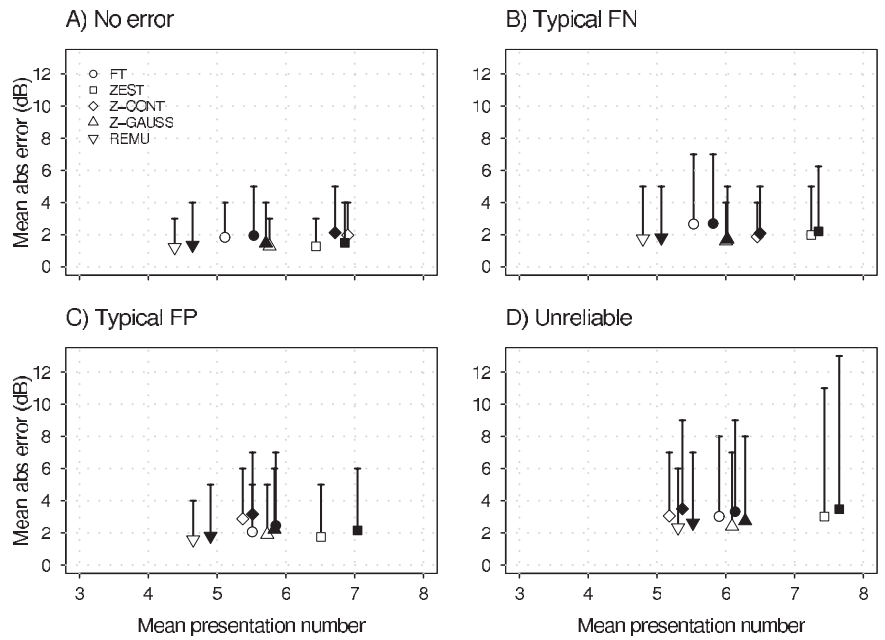
errors: all procedures different within both groups; typical FP errors: all different except ZEST and REMU for glaucoma group; unreliable: all significantly different except Z-Cont and REMU for glaucoma and Z-Gauss and REMU for normal subjects). Inspection of Figure 3 demonstrates that, in most cases, although statistically significant, the magnitude of the differences was small.

**Comparison of All Procedures for a Uniform Change in Sensitivity across the Whole Visual Field**

Figures 4 and 5 show the performance of the test and retest strategies when the entire visual field sensitivity was either reduced (Fig. 4) or elevated (Fig. 5) by 3 dB. The figures are plotted in the same format as Figure 3.



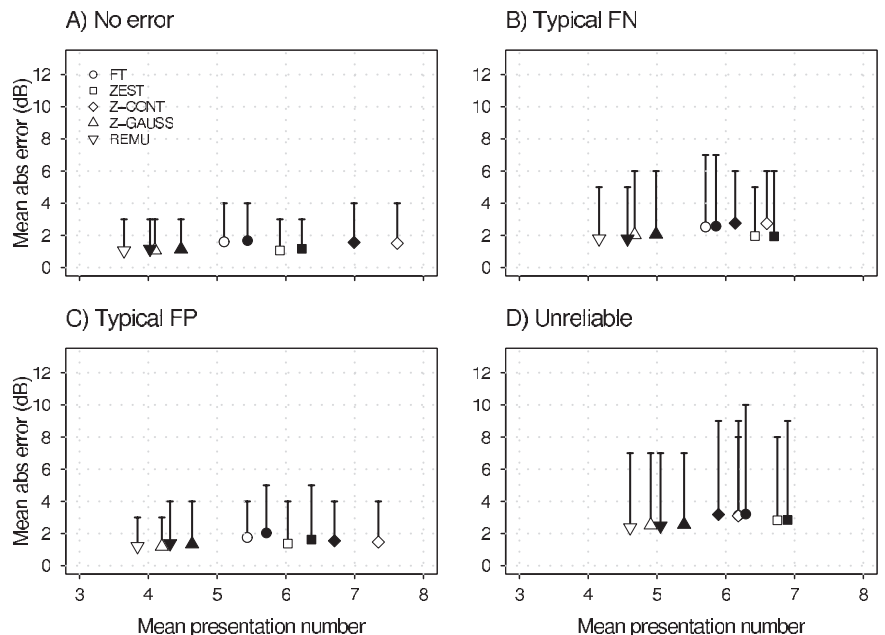
**FIGURE 3.** MAE plotted against the mean presentation number for each of the perimetric procedures—(A) no error, (B) typical FN, (C) typical FP, and (D) unreliable—when the visual field remains unchanged. *Filled symbols*: glaucomatous visual fields; *open symbols*: normal visual fields. Error bars, 95th quantile of absolute error.



**FIGURE 4.** MAE plotted against the mean presentation number for each of the perimetric procedures—(A) no error, (B) typical FN, (C) typical FP, and (D) unreliable—when the sensitivity of the entire visual field is decreased by 3 dB at the retest visit. Symbols are the same as in Figure 3.

The retest strategy most biased by the original test result was Z-Cont; thus, this procedure should be most affected by a mismatch between the initial result and the sensitivity at retest. Z-Cont is represented by the diamonds in Figures 3 to 5. When there was no change in visual field sensitivity from test to retest (Fig. 3) Z-Cont performed well. However, this was not the case when there was a shift in the overall sensitivity of the visual field. When the field was uniformly decreased by 3 dB (Fig. 4), Z-Cont was slow to terminate for the no-error and FN conditions. Z-Cont terminated more quickly when FP or unreliable responses were made; however, the magnitude of the error was greater than the other retest procedures. When the whole field increased in sensitivity (by 3 dB, see Fig. 5), Z-Cont was on average slower than the test procedures (FT and ZEST) and had a similar error distribution profile. Hence, Figures 4 and 5 demonstrate that Z-Cont was not a successful retest strategy when there was a change across the whole visual field.

Figures 4 and 5 demonstrate that Z-Gauss (up triangles) performed better than Z-Cont and, in general, displayed better performance (similar or reduced error and faster test time) than simply repeating the test strategy ZEST. Z-Gauss performed similarly to REMU when there was a whole-field improvement in visual field sensitivity (Fig. 5), but was substantially slower to terminate when there was a reduction in sensitivity (Fig. 4). The asymmetry in the performance of Z-Gauss was due to the truncation of the *pdf* at the top end of the dynamic range (40 dB). Statistical comparison of the presentations required to terminate demonstrated that REMU terminated faster on average than Z-Gauss, and that Z-Gauss terminated faster than ZEST in all error conditions, for both whole-field increases and decreases in sensitivity (ANOVA, Holm-Sidak post hoc testing;  $P < 0.05$ ). Although statistically significant, it is important to consider whether the magnitude of the difference is likely to be clinically significant. When averaged across



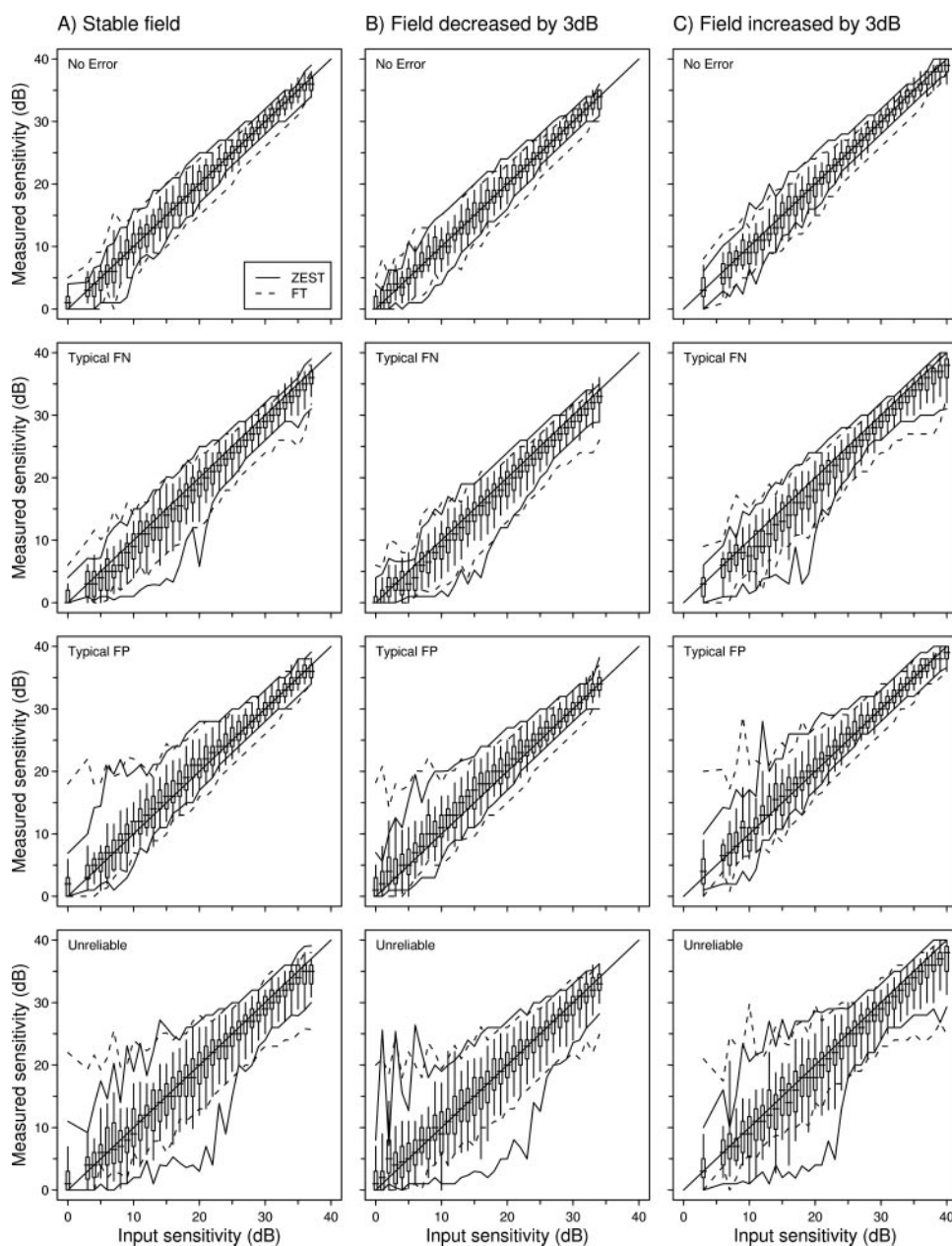
**FIGURE 5.** MAE plotted against the mean presentation number for each of the perimetric procedures—no error, (B) typical FN, (C) typical FP, and (D) unreliable—when the sensitivity of the entire visual field is increased by 3 dB at the retest visit. Symbols are the same as in Figure 3.

all error conditions, when the whole field decreased by 3 dB, ZEST required approximately seven presentations to terminate, Z-Gauss approximately six, and REMU approximately five. When sensitivity increased by 3 dB, ZEST required approximately 6.5 presentations on average, Z-Gauss approximately 4.5, and REMU approximately four. For comparison, FT terminated using between five and six presentations for both the whole-field increase and decrease in sensitivity. For most error conditions, the MAE of REMU and Z-Gauss was not different ( $P > 0.05$ ). REMU was more accurate on average than Z-Gauss in the presence of FN errors or unreliable performance if there was a whole-field decrease in sensitivity ( $P < 0.05$ , Holm-Sidak post hoc testing). The magnitude of the difference was unlikely to be of clinical significance.

### Performance of REMU as a Function of Input Sensitivity

The data in Figures 3 to 5 are pooled across the entire visual field; however, it is well known that test-retest variability of

visual field algorithms is greatest when visual field sensitivity is reduced.<sup>6,7</sup> Increased variability is expected, as it has been established that the slope of the psychometric function for white-on-white perimetric stimuli decreases with reducing perimetric sensitivity.<sup>13</sup> Accurate and precise thresholds can be determined in the presence of such variability; however, a larger number of presentations are required.<sup>8</sup> This fact is critical to the design of the REMU test procedure: a minimum number of presentations are expended in normal test locations, to enable more presentations to be used in abnormal locations. To explore this more thoroughly, Figure 6 shows box plots of the mean errors of our (on average) best performing retest algorithm (REMU) as a function of the true input to the simulation for the situation where the whole field was stable (Fig. 6A) and where there was either a uniform increase or decrease in sensitivity. The *x*-axis of Figure 6 represents the known true sensitivity of the patient. This figure is a little different from similar figures reported in clinical studies, in which it is typical to plot the sensitivity measured at visit 1



**FIGURE 6.** Box plot of the distribution of sensitivities returned at retest as a function of the known true sensitivity input to the simulation. Data are pooled for both glaucomatous and normal visual fields. Boxes are not present for sensitivity values that were represented  $<20$  times within the visual field database. The performance of REMU (*box and whiskers*) is compared with that of ZEST (*solid lines*) and FT (*dashed lines*). The *boxes* represent the 25th quantile, median, and 75th quantile, whereas the *whiskers* represent the 5th and 95th quantiles of the distribution of absolute errors for a stable field (A) and fields decreased (B) or increased (C) by 3 dB.

against the sensitivity measured at visit 2, hence incorporating errors in both dimensions. For the ZEST and FT procedures, the errors derived from clinical test-retest should, on average, be double those derived from measured versus actual, as the error distribution will be the same at test and retest. However, for REMU, the error for the initial visit will be that of ZEST, with a different error distribution at the retest visit when the REMU procedure is used.

Figure 6 shows that when patients made errors in response (FP or FN), the distribution of the error of REMU was much more consistent in magnitude across the sensitivity range than that of either ZEST or FT. In our model, patient variability was increased with decreasing sensitivity, but was kept fixed for all low sensitivities, as a floor effect created difficulties in empirically establishing the relationship at low sensitivity. Hence, if the procedures were performing well, within our model there should have been a consistent level of test-retest variability for sensitivities below approximately 20 dB. This was true for REMU but not for ZEST or FT. Of importance, REMU returned sensitivity estimates that were not biased in either direction.

**Performance of the Retest Procedures When There Is a Localized Decrease in Sensitivity**

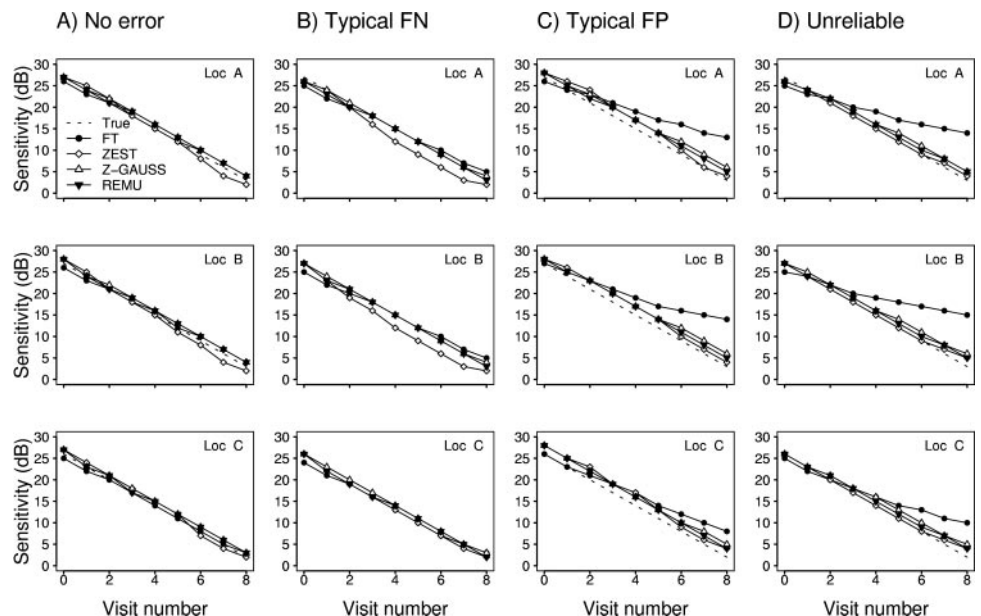
The previous figures compare performance of the procedures when used twice (test then retest). A simulated localized scotoma was modeled that decreases in sensitivity by 3 dB on each of eight visits. At each visit, the retest procedures were seeded by the previous visit, enabling observation of whether compounding errors result when the visual field is changing. Figure 7 shows the performance of the test (FT and ZEST) and retest (Z-Gauss and REMU) procedures. It is important to note that the x-axis in Figure 7 should not be interpreted as necessarily linear or evenly spaced in scale. Z-Cont is not shown here, as this procedure was shown to perform poorly for whole-field sensitivity change. As shown previously,<sup>8</sup> Figure 7 demonstrates that when patients make FP errors, the starting estimate for the FT procedure becomes an increasingly poorer predictor of true sensitivity each visit, hence the error increases. ZEST, Z-Gauss, and REMU are all able to track the change in visual field sensitivity, even when the patient is responding unreliably.

**DISCUSSION**

Current perimetric procedures such as ZEST and SITA use population information in the prior *pdfs* to enable rapid and often accurate determination of thresholds. When patients are retested, their previous results may be used to bias the prior *pdfs*. We have demonstrated that a naive application of this principal does not give good results for perimetry (Z-Cont); but more sophisticated approaches can result in both faster and more accurate tests, even when the visual field sensitivity changes from one visit to the next.

REMU demonstrated the best overall retest performance. In our simulations, a ZEST procedure was initiated when the suprathreshold test in REMU was not passed. There is no reason why other thresholding algorithms cannot be substituted in this step. REMU is designed to minimize presentations in areas of previously measured normal sensitivity, yet expends more presentations in areas of sensitivity loss. These additional presentations enable accurate and repeatable estimates in areas of visual field loss. In particular, REMU demonstrated a substantially narrower range of outcomes for each true sensitivity than did either ZEST or FT when the true sensitivity was below approximately 15 dB (see Fig. 6). This observation has several important implications. First, it predicts that REMU will have reduced clinical test-retest variability when compared with current commercially available algorithms, thereby enhancing detection of visual field change. Second, it demonstrates that a considerable proportion of the wide distribution of test-retest variability measured with current strategies is likely to result from the algorithms used to measure threshold. In our model, response variability was increased with decreasing sensitivity but was fixed in magnitude for all low sensitivities. Consequently, if performing well, the procedures should display a fixed level of test-retest variability across all low sensitivities as was demonstrated by REMU (see Fig. 6). This was not the case for either FT or ZEST.

The improved average performance of REMU comes at a small cost. REMU employs a suprathreshold check to see whether sensitivities that were previously in the age-matched normal range have decreased, and only fully determines these locations if they fail the suprathreshold test. If a location has sensitivity at or above age-matched normal and the sensitivity increases, then the suprathreshold test should be passed and



**FIGURE 7.** Mean sensitivity returned by the baseline procedures FT and ZEST and the retest procedures Z-Gauss and REMU for locations A, B, and C shown in Figure 1. The sensitivities of these locations were decreased by 3 dB per test visit. Data are shown for (A) no errors, (B) typical false-negative errors; (C) typical false-positive errors; and (D) unreliable results. Dotted lines: true thresholds.



the result reported as the previous value. That is, REMU is unlikely to detect localized improvements in sensitivity from an age-matched normal baseline. Whole-field improvements in sensitivity should be detected due to the general height check at stage 3 of the REMU algorithm. We see this as an acceptable tradeoff for the decrease in variability in areas of field loss. Of course, if a change in the visual field is expected, for example an improvement due to cataract extraction, then the standard algorithms can be run instead of REMU to create a new seeding field, and REMU can be run thereafter.

In this work we have explored ways of using previous sensitivity estimates to seed the current test. There is a wealth of prior test information that we have not used, such as individual location response sequences. How to make the best use of this information and whether it yields significant benefit is a topic of ongoing study in our laboratory. Other prior information includes the gradient of an individual's visual field. Many current test algorithms choose the starting estimate of threshold based on information from neighboring locations plus an eccentricity adjustment; for example, the growth pattern of both FT and SITA.<sup>1</sup> The eccentricity adjustment can be customized to an individual using the gradient of their previous visual field, thus resulting in a more accurate starting estimate of threshold. We have performed experiments with this approach (data not reported) but found that for quite a bit of effort, the gain in performance was minimal.

The utility of computer simulation ultimately depends on how closely the model represents human performance. To this end, we have used empiric visual field data, and known rates of response variability collected in clinical populations. Although simulation studies have limitations, in the absence of computer simulation, it is impossible to assess the accuracy of perimetric procedures, as a patient's true sensitivity can never be known with certainty. Simulation also enables us to explore in detail the performance of test procedures for situations that are less common, for which it is difficult to obtain large quantities of clinical data, yet which may have important implications if the algorithms perform poorly. Simulation is an essential precursor to clinical assessment of perimetric algorithm performance and has been used successfully for this purpose.<sup>2-4,8,14,15</sup> We are currently collecting real patient data on the best-performing procedures described herein.

The experiments predict that sensitivity estimates from previous test visits can be used to obtain more accurate and repeatable visual field assessment outcomes at subsequent tests. The obvious approaches to retesting visual fields, such as continuing the prior procedure or directly seeding with previous values, have performance limitations when sensitivity changes from one test to the next. REMU, however, significantly improves both accuracy and precision of retesting perimetric sensitivity, and furthermore displays minimal bias, even when fields change and patients make errors.

## Acknowledgments

The authors thank Chris Johnson (Devers Eye Institute, Portland, OR) for supplying the empiric visual field data for input into the simulation.

## References

1. Anderson DR, Patella VM. *Automated Static Perimetry*. 2nd ed. St. Louis: Mosby; 1999.
2. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry. *Acta Ophthalmol Scand*. 1997;75:368-375.
3. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from Full Threshold, ZEST, and SITA-like strategies, as determined by computer simulation. *Invest Ophthalmol Vis Sci*. 2003;44:4787-4795.
4. Vingrys A, Pianta M. A new look at threshold estimation algorithms for automated static perimetry. *Optom Vis Sci*. 1999;76:588-595.
5. Anderson AJ, Johnson CA, Fingeret M, et al. Characteristics of the normative database for the Humphrey Matrix perimeter. *Invest Ophthalmol Vis Sci*. 2005;46:1540-1548.
6. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654-2659.
7. Wild JM, Pacey IE, Hancock SA, Cunliffe IA. Between-algorithm, between-individual, differences in normal perimetric sensitivity: Full Threshold, FASTPAC, and SITA. *Invest Ophthalmol Vis Sci*. 1999;40:1152-1161.
8. McKendrick AM, Turpin A. Combining perimetric supra-threshold and threshold procedures to reduce measurement variability in areas of visual field loss. *Optom Vis Sci*. 2005;82:43-51.
9. Bengtsson B, Heijl A. Evaluation of a new perimetric strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand*. 1998;76:368-375.
10. Bengtsson B, Heijl A, Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. *Acta Ophthalmol Scand*. 1998;76:165-169.
11. Shirato S, Inoue R, Fukushima K, Suzuki Y. Clinical evaluation of SITA: a new family of perimetric testing strategies. *Graefes Arch Clin Exp Ophthalmol*. 1999;237:29-34.
12. McKendrick AM, Turpin A. Advantages of terminating Zippy Estimation by Sequential Testing (ZEST) with dynamic criteria for white-on-white perimetry. *Optom Vis Sci*. 2005;82:981-987.
13. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41:417-421.
14. Artes PH, Henson DB, Harper R, McLeod D. Multisampling suprathreshold perimetry: a comparison with conventional suprathreshold and full-threshold strategies by computer simulation. *Invest Ophthalmol Vis Sci*. 2003;44:2582-2587.
15. Bengtsson B. A new rapid threshold algorithm for short-wave-length automated perimetry. *Invest Ophthalmol Vis Sci*. 2003;44:1388-1394.