

Anatomy of a Supergroup: Does a Criterion of Normal Perimetric Performance Generate a Supernormal Population?

Andrew John Anderson and Chris A. Johnson

PURPOSE. To interpret individual results from automated perimeters, a normative database must be developed. Typically, a set of criteria determines those subjects that may be included in the database. This study examined whether a criterion of normal performance on an established perimeter generates a subgroup with supernormal perimetric performance.

METHODS. The right-eye perimetric results of 100 subjects were analyzed. Subjects had visual acuities of 6/12 or better, no history of eye disease, and normal slit lamp biomicroscopic and ophthalmoscopic examinations. Subjects performed test-retest visual field examinations on a Humphrey Field Analyzer (HFA) 24-2 test (Zeiss Humphrey Systems, Dublin, CA), and on a custom frequency-doubling (FD) perimeter with targets spaced in the same 24-2 pattern.

RESULTS. Test-retest correlation (Spearman rank correlation coefficients, r_s) for mean defect (MD) and pattern SD (PSD) were 0.65 and 0.40 (HFA), and 0.82 and 0.39 (FD perimeter). Three subjects with HFA MDs in the lower 5% had similarly low MDs on retest, whereas no subject was common between the test and retest for the lower 5% of HFA PSD. Correlation between the HFA and FD test results were 0.41 (MD) and 0.05 (PSD). Based on these correlations, the bias introduced into perimetric probability limits were determined, by using Monte Carlo simulations.

CONCLUSIONS. Although a criterion of a normal MD may produce a subpopulation with supernormal perimetric performance, a criterion of a normal PSD is less likely to do so. Also, a criterion on one test type is less likely to create a supernormal group on a different test type. The bias introduced into perimetric probability limits is small. (*Invest Ophthalmol Vis Sci.* 2003;44:5043-5048) DOI:10.1167/iovs.03-0058

A fundamental requirement of automated perimetry is the generation of a normative database to which individual results can be compared. Typically, investigators use inclusion criteria to determine those subjects that may be included in the database.^{1,2} These criteria may include a negative history of ocular disease, normal findings in slit lamp biomicroscopic and ophthalmoscopic examinations, visual acuity better than a prescribed limit, and a restricted range of refractive errors. It is also possible to establish a perimetric criterion based on a subject's performance on an established clinical perimeter²⁻⁵—for example, mean defect (MD) and pattern SD (PSD)

within the 95% limits of normality, on the Humphrey Visual Field Analyzer (HFA; Carl Zeiss Meditec, Inc., Dublin CA), where the index MD provides a measure of uniform loss or loss involving a large fraction of the visual field, and PSD provides a measure of local irregularity. Such perimetric criteria may be useful in detecting visual pathway disease not manifest on ophthalmoscopy (e.g., vascular⁶ and compressive⁷ lesions), or early ocular disease in which the ocular fundus is not frankly abnormal (e.g., early glaucoma). Both the original HFA analysis package⁸ and the newer Swedish interactive test algorithm (SITA) (both achromatic⁹ and short-wavelength automated perimetry [SWAP]⁵) are based on analyses of subject groups from which those with abnormal visual field results were excluded. It should be noted, however, that the presence of an abnormal field result does not necessarily mean that eye disease is present. Indeed, 5% of the visual fields of healthy eyes should be judged abnormal when 95% probability limits are used. It is important, therefore, to appreciate that a distinction exists between visual fields from healthy eyes and visual fields that are statistically normal, with the latter being a subset of the former. Although in this study we examined perimetry specifically, the issue of normative databases pervades ophthalmology, particularly in functional testing (e.g., normative limits for contrast sensitivity charts) and in imaging (e.g., normative limits for optic nerve head parameters in retinal tomography). Because of this, it is important to appreciate the factors underlying normative databases, as well as any limitations that may result.

Investigators have stressed the importance of specifying study populations when evaluating clinical diagnostic tests,^{10,11} and subject inclusion criteria provide a means through which to do this. Inherent in the use of inclusion criteria for subjects in a study of normal observers is the assumption that a classification of "normal" is equivalent to that of "disease-free." Unfortunately, "normal" and "disease" may be part of a continuum, as in a disease process such as hypertension, thereby making the distinction between the two categories less clear. To avoid making this distinction, it is possible to create a perimetric database without any specified inclusion criteria for subjects. Ignoring the possible effects of unintentional recruitment bias,¹² the resultant probability limits give the likelihood of a particular index value arising from the population as a whole (i.e., disease and disease-free observers), rather than from a group of normal observers. Such limits, however, would have a reduced sensitivity for detecting ocular disease, particularly once the prevalence of disease rose above the probability limit defining an abnormal result. Because the prevalence of glaucoma alone reaches above 5% (a commonly accepted limit for "normality") in older populations,¹³ the use of a criterion-determined normal population to create perimetric databases is important for maximizing a test's sensitivity for detecting ocular disease.

The use of inclusion criteria, however, means that the resultant database no longer reflects the performance of the general population, but rather that of a criterion-determined subpopulation.¹² Using a perimetric-based criterion raises an

From Discoveries in Sight, Devers Eye Institute, Portland, Oregon. Supported by National Eye Institute Grant EY03424 (CAJ). Submitted for publication January 21, 2003; revised May 5 and June 25, 2003; accepted June 30, 2003.

Disclosure: **A.J. Anderson**, None; **C.A. Johnson**, 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: Chris A. Johnson, Discoveries in Sight, Devers Eye Institute, 1225 NE Second Avenue, Portland OR 97232; cajohnso@discoveriesinsight.org.

interesting question, however: Is it appropriate to use an inclusion criterion (perimetric performance) that is based on the variable for which normal limits are being determined? In particular, what is the meaning of normative probability limits of 2%, 1%, and 0.5%, when they are based on a group from which the lowest 5% was removed? For example, using 5% probability levels as a criterion for normality results in a database that contains only the top 95% of normal performers. Furthermore, requiring subjects in the database to have two normal indices (e.g., MD and PSD) at the 5% level makes things worse, resulting in only the top 90% (0.95 squared, assuming complete independence of the indices) of performers. Such a database would produce a high false-positive rate for detecting abnormal visual fields, as the subject group who formed the database had perimetrically "supernormal" performance.

This example assumes that an otherwise normal observer with an abnormal MD or PSD always returns an abnormal index on subsequent testing. This is unlikely to be true, however. Variability in both indices would result in a variety of subjects periodically returning abnormal fields, rather than a fixed 5% of the otherwise normal population.

Furthermore, even if perimetric indices could be determined without variability, the supernormal phenomenon should only manifest if there is a good correlation between the perimeter used for the inclusion criteria and the perimeter whose normative database is being created, in normal observers. This correlation is distinct from that which exists between two tests in subjects with ocular disease¹⁴⁻¹⁶ and is likely to be lower, given the comparatively restricted range of test indices returned by normal observers.¹⁷ Although it may be expected that good correlation exists among perimeters that have similar test parameters, this may not be true among perimeters designed to measure different visual functions (e.g., frequency-doubling [FD] perimetry,¹⁸ or SWAP³). Previous work has failed to find a significant correlation between the MD index for conventional increment-threshold perimetry and FD perimetry in a group of normal observers, despite the presence of a strong correlation when a similarly sized group of glaucomatous observers was used.¹⁵

Therefore, two important factors determine whether using a perimetry-based inclusion criterion generates a supernormal group of perimetric observers: the variability of perimetric indices in normal observers and the correlation between the perimetric indices of two tests (the established perimetry test and the new perimetry test) in normal observers. In the current study, we investigated how well a normal subject's perimetric performance predicts his or her performance on subsequent perimetric examinations. In addition, we examined the correlation between the performance of normal observers for two different types of perimetry: increment-threshold perimetry (HFA) and FD perimetry.¹⁹ Based on our empirical findings, we performed a Monte Carlo investigation of the effects of inclusion criteria on perimetric normative database probability limits.

MATERIALS AND METHODS

We retrospectively gathered visual field results from data collected for a previous study. The study was approved by an institutional review board and complied with the tenets of the Declaration of Helsinki. All subjects gave informed consent before testing.

For experiments 1 and 2, we analyzed the right eye results of 100 subjects recruited from staff, associates, and patients of the Ophthalmology Department of the University of California, Davis. All subjects had normal biomicroscopic and ophthalmoscopic examination results, visual acuities equal to or better than 20/40 (6/12), spectacle refraction not greater than ± 5.00 DS and ± 2.00 DC, and no history of ocular disease or systemic disease known to affect vision, and none was taking

TABLE 1. Distribution of Subjects by Age

Age Range (y)	Frequency (N = 100)
25-34	7
35-44	18
45-54	25
55-64	13
65-74	14
75-84	22
85-94	1

any medications known to affect vision. Subjects had at least one prior visual field examination (HFA II 24-2 full threshold) at a session separate from the main study, at which time both MD and PSD indices were normal ($P \geq 5\%$). Because of this, our subjects were not naïve perimetric observers for the data presented in this study and so may not demonstrate the same improvement in performance with serial field testing (the "learning effect") expected in a naïve sample. The significance of this is discussed in the following sections.

If a perimetric test is used purely as a screening procedure on naïve subjects, then it would be appropriate for a learning effect to be accounted for so that false-positive results do not arise simply through subject inexperience. Such a test, however, would have its sensitivity to detect disease compromised, due to the increased variability in results (and, correspondingly, the increased width of the 95% limits) obtained with naïve observers.² A test that is used primarily to monitor patients should not account for a learning effect, however, as it is expected that subjects taking the test are either not perimetrically naïve or are naïve and so may require training to achieve consistent results. We believe that it generally is undesirable to have a database that is influenced by a learning effect and that it is preferable to be aware that some naïve subjects may require training. Our approach is consistent with that used in the development of the Humphrey Visual Field Analyzer, in which only subjects experienced with perimetry were included. As noted by Hiejl et al.,² "... [I]f a model of the normal visual field were to be based on subjects without any previous experience in visual field testing, the normal variability would be very large and nonrepresentative for many clinically examined patients."

Subjects performed test and retest sessions (right and left eyes) on both the 24-2 HFA II (full threshold) and a customized FD perimeter using the same 24-2 test pattern,¹⁹ with testing spaced over four visits and interspersed with other perimetric tests (not analyzed in this study). We performed customized calculations²⁰ of the indices MD and PSD for each eye and for each test type, for both test and retest sessions, using a linear model for the effect of aging^{2,19} and the formulas used by the commercial HFA device.²¹ Both test indices also are available on the newer SITA test algorithm for the HFA perimeter.⁹ Percentile limits were calculated empirically, using linear interpolation. Table 1 shows the distribution of subjects, by age.

We found that both test and retest MD distributions for the HFA were not significantly different from normal (Kolmogorov-Smirnov test, $P = 0.59$ and 0.24 , respectively), although test and retest distributions of HFA PSD were significantly different from normal ($P = 0.03$ and $P < 0.001$, respectively). The distributions of these indices are given in Figure 1. Because of these departures from normality, we used Spearman rank correlation coefficients (r_s) to assess the monotonicity of relationships between tests nonparametrically, with a value of 1 indicating a perfect monotone relationship among ranks.²² Correlation analyses typically are inappropriate for comparing test methods, as the level of correlation depends on the range of intersubject variability in the study.¹⁷ This presents significant problems when assessing correlation between perimetric indices in ocular disease, because the correlation obtained depends, in part, on the range of disease severity included in the study. In our study on normative subjects, however, the range of intersubject variability is essentially fixed and so this problem is avoided.

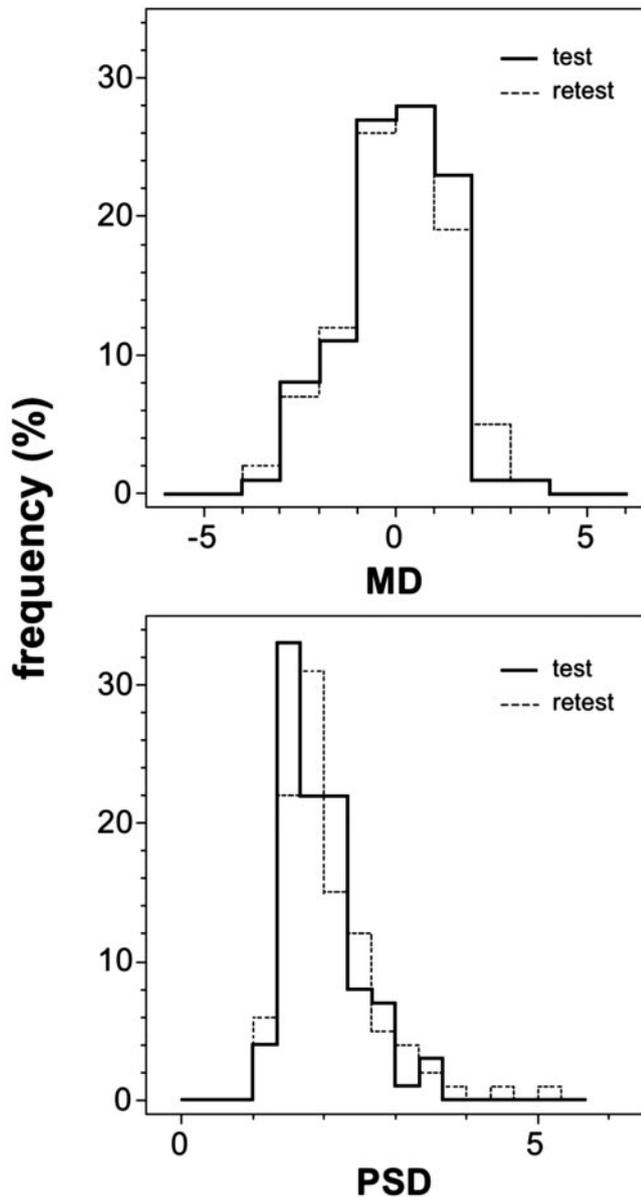


FIGURE 1. Distribution of HFA perimetric indices for both MD (top) and PSD (bottom), for both test (solid lines) and retest (dashed lines). The upper limit of each range was applied inclusively and the lower limit exclusively.

Simulation

In experiment 3, we examined the effect of different correlations between tests on the probability limits for perimetric indices, using a Monte Carlo simulation. We modeled the distribution of perimetric indices with a Gaussian distribution, although it should be noted that the outcomes of our model are derived from ranks data and so are independent of the underlying test distribution. Because of this, our model is appropriate for the index PSD, despite its non-Gaussian distribution (Fig. 1).

To generate two distributions with known correlation, we combined each element in two independent Gaussian distributions (X and Y ; mean = 0, SD = 1) using the following equation:

$$Z_i = \alpha \cdot X_i + (1 - \alpha) \cdot Y_i \tag{1}$$

where α gives the proportion of each distribution contributing to Z and can vary between 0 and 1. It should be noted that α is not identical

with Pearson's coefficient of determination (r^2) nor the Spearman rank correlation coefficient (r_s). The variance of the distribution Z is:

$$\sigma_z^2 = \alpha^2 \cdot \sigma_x^2 + (1 - \alpha)^2 \cdot \sigma_y^2 \tag{2}$$

where σ_x^2 and σ_y^2 are the variances of distributions X and Y , respectively. The SD of distribution Z could then be normalized by dividing each element by the root of this variance:

$$Z(\text{norm})_i = \frac{Z_i}{\sigma_z} \tag{3}$$

Based on these equations, we produced a normally distributed set of test indices, X , and another set of normally distributed indices, $Z(\text{norm})$, with a known correlation between the two sets. We simulated 2000 indices for each of the two distributions X and $Z(\text{norm})$, using two combined multiplicative congruential random number generators, as implemented by Press et al.,²³ giving a period of approximately 2.3×10^{18} . Serial correlations were removed by using a Bays and Durham shuffle.²³

RESULTS

Experiment 1: Correlation between Indices from the Same Test

Figure 2 plots the test (abscissa) and retest (ordinate) scores of the group of 100 normal observers for the HFA indices MD (top) and PSD (bottom). Test and retest MD values correlated moderately, with three subjects (95% CI for proportion: 0.6%–8.5%) found in the lower fifth percentile on both test and retest (gray box). Test–retest correlation for PSD was poorer, with no subjects (97.5% CI, i.e., one-way: 0%–3.6%) common between the lower fifth percentile on test and retest. The 95% confidence intervals for the MD and PSD correlation coefficients overlapped.

There was a moderate correlation between the MD and PSD on the test result for the HFA (Fig. 3). No subject (97.5% CI, 0–3.6%) had both a PSD and MD value in the lower fifth percentile.

Table 2 shows the correlation coefficients for a similar analysis on the FD perimeter results. Correlations were higher than with the HFA, although the same pattern of results was preserved, with MD showing a greater test–retest correlation than PSD. The correlation coefficient for MD was significantly higher than for PSD, as evidenced by the lack of overlap between the 95% confidence intervals.

Experiment 2: Correlation between Indices from Different Test Types

Figure 4 plots the test HFA (abscissa) versus the test FD (ordinate) summary indices. For the MD index (top), correlation was low, and only one subject (95% CI, 0.03–5.4%) was in the lower fifth percentile on both tests (gray box). For the PSD index, there was no significant correlation between the tests, and no subject (97.5% CI, 0–3.6%) was in the lower fifth percentile on both tests.

Experiment 3: Monte Carlo Simulation of Probability Limits

Figure 5 plots the shift in the nominal 5% (circle) and 1% (square) probability limits of a new perimetry test when an established perimetry test, of correlation as given on the abscissa, is used to screen subjects for the new test's database. Subjects were required to be greater than or equal to the fifth percentile performance on the established perimeter to be

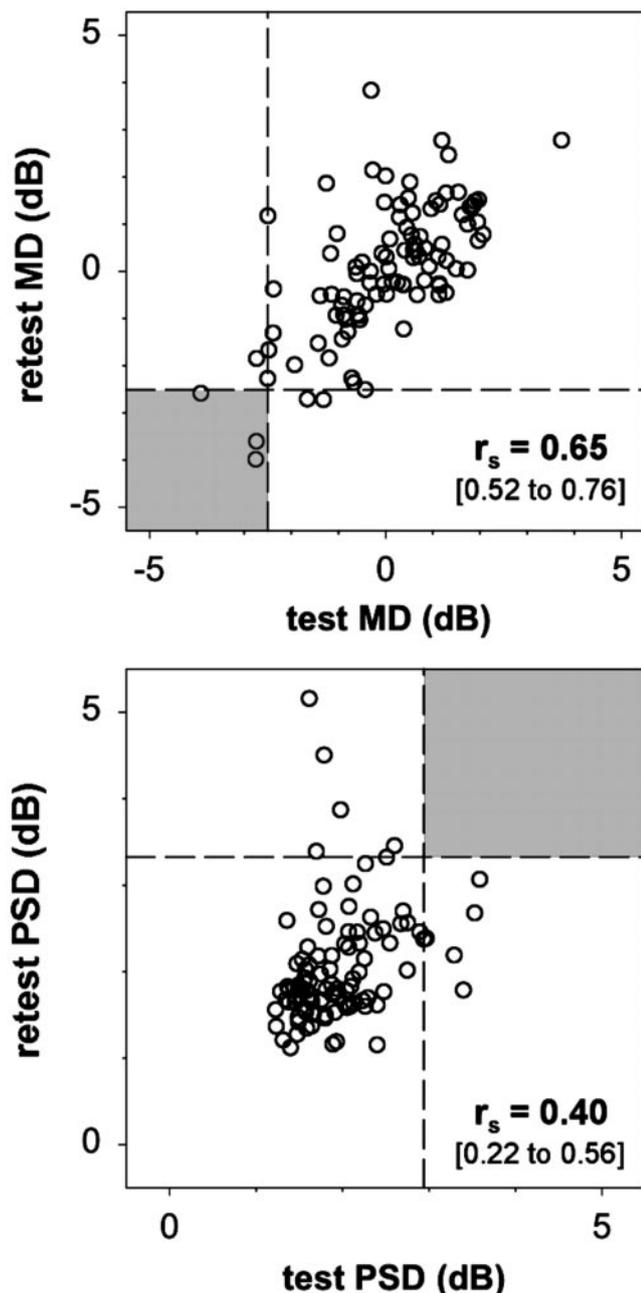


FIGURE 2. Test versus retest performance of 100 normal observers on the HFA for MD (top) and PSD (bottom). Dashed lines: fifth percentile limits. For both correlation coefficients, $P < 0.0001$.

included in the new perimeter's database. When there is no correlation between the two tests ($r_s = 0$), screening subjects on an established perimeter has no influence on the probability limits in the new database. When there is perfect correlation between the two tests ($r_s = 1$), the 5% and 1% probability limits exclude 9.75% ($5\% + 0.05 \times 95\%$) and 5.95% ($5\% + 0.01 \times 95\%$) of normal subjects when no criterion for prior performance on the established perimetry test is used. Between these two limits, both the 5% and 1% limits show an accelerating function with r_s .

DISCUSSION

We found only a moderate correlation between visual field indices on test-retest (Fig. 2), with a poorer correlation be-

tween indices from different test types (Fig. 4). We found that subjects falling in the lower 5% of test results did not necessarily remain in the lower 5% on retest (Fig. 2), which indicates that subjects with visual field indices of $P < 5\%$ do not represent a constant proportion of an otherwise normal population. Further support for this latter finding comes from the observation that six of our hundred subjects had an abnormal ($P \leq 5\%$) MD and/or PSD for their right and/or left eye on initial testing, as determined by the commercial HFA database, despite having normal ($P \leq 5\%$) indices at a prior screening visit (see the Methods section).

Based on our correlation findings, it is possible to predict by how much nominal probability limits would shift when a criterion of normal perimetric results is used in generating a perimetric database. We will take the example of generating a normative database for the FD perimeter, using only subjects with a normal ($P \geq 5\%$) MD index on the HFA. The correlation between the FD perimeter and the HFA was 0.41 (95% CI, 0.23–0.57; Fig. 4), suggesting that the 5% limits in our new database would actually exclude approximately 5.9% (95% CI, 5.5–6.7%) of the normal population not previously screened on the HFA (Fig. 5). Similarly, the 1% limits would exclude approximately 1.5% (95% CI, 1.2–1.8%) of the normal population. Such shifts are small compared with those expected if there were a perfect correlation between the two tests (9.75% and 5.95% for the 5% and 1% limits, respectively: see the Results section). As the correlation between any new test and the HFA should be less than the autocorrelation (i.e., test-retest) of the HFA, the test-retest correlation of the HFA should set an upper limit on what probability limit shifts are expected in practice. We found a test-retest correlation coefficient of 0.65 (95% CI, 0.52–0.76) for the HFA index MD (Fig. 2), which predicts an upper limit of 7.2% (6.4%–7.9%) and 2.0% (1.7%–2.6%) for the 5% and 1% probability limit shifts, respectively (Fig. 5). Given the uniformly lower correlations found for the index PSD, it is likely that a criterion based on PSD will cause a probability limit smaller than those expected with a criterion based on the index MD. It is possible that many or all the described probability limit shifts are smaller than those intro-

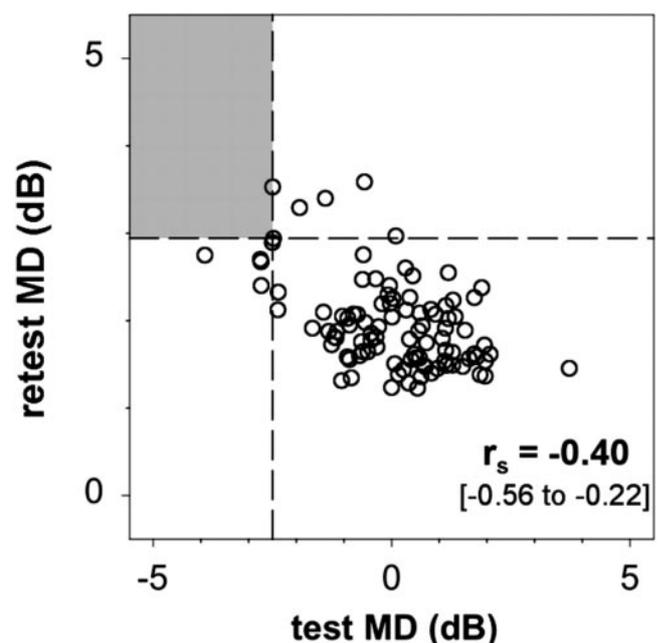


FIGURE 3. Test MD versus test PSD performance of 100 normal observers on the HFA. Dashed lines: fifth percentile limits. For the correlation coefficient, $P < 0.0001$.

TABLE 2. Spearman Rank Correlation Coefficients (r_s) for Various Combinations of Test-Retest Performance on the 24-2 Version of the Frequency-Doubling Perimeter

	r_s (95% CI)	P	Subjects Common to Lower 5th Percentile (95% CI for proportion; %)
Test vs. retest, MD	0.82 (0.74 to 0.87)	<0.001	3 (0.62%–8.5%)
Test vs. retest, PSD	0.39 (0.21 to 0.55)	<0.001	3 (0.62%–8.5%)
Test MD vs. test PSD	-0.30 (-0.48 to -0.11)	0.002	1 (0.03%–5.4%)

$N = 100$.

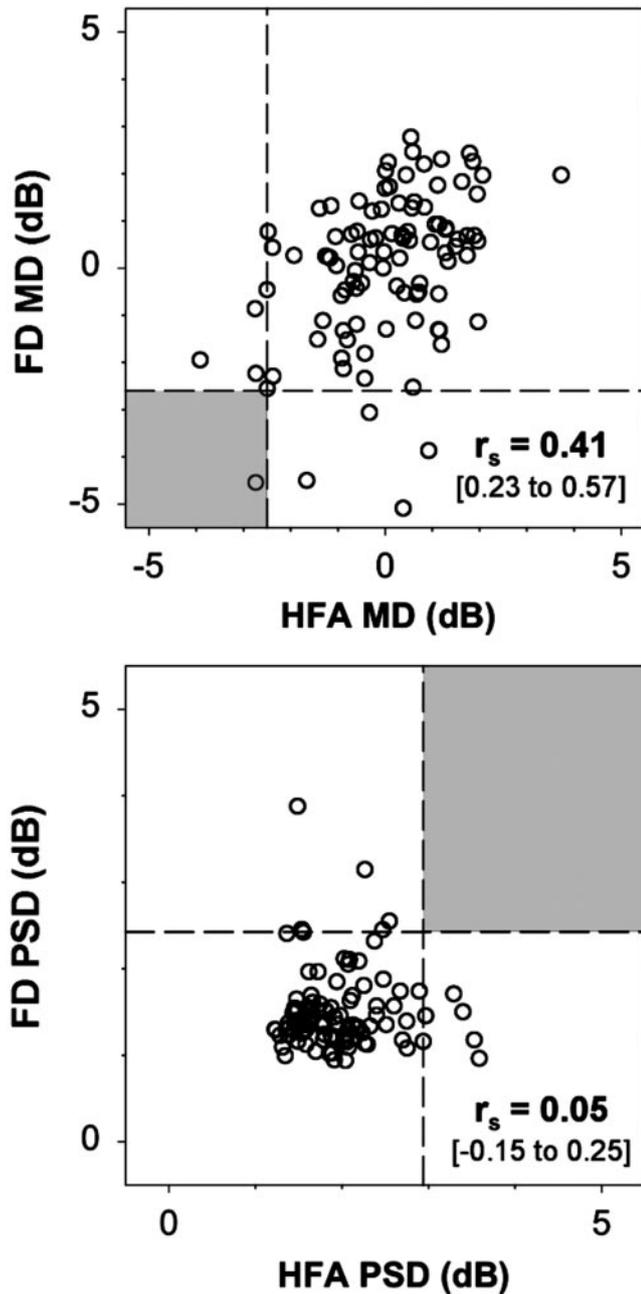


FIGURE 4. HFA versus FD perimetry performance of 100 normal observers for MD (top) and PSD (bottom). Dashed lines: fifth percentile limits. For the correlation coefficients, $P < 0.0001$ and $P = 0.59$ (top and bottom, respectively).

duced by inexact modeling of the change in sensitivity with age¹⁹ or by assuming that the variance of sensitivity distributions is constant with age.²

It could be argued that to exclude subjects with otherwise undetected disease reliably, it would be best to use a visual field index that has high test-retest repeatability. However, high repeatability will produce greater biases in the resultant database (Fig. 5). It must be remembered that the repeatability between normal observers, as assessed in this study, may not be the same as when repeatability is assessed in a group of normal and diseased observers.¹⁵ In particular, variability must be viewed in light of the range of index values encountered clinically. Because of this, a test index may show little or no repeatability among normal observers, but still be a useful diagnostic index provided diseased observers return test index values outside the normal range. For example, we found PSD to have poorer test-retest reliability than MD for normal observers, which could be interpreted that PSD would be the poorer

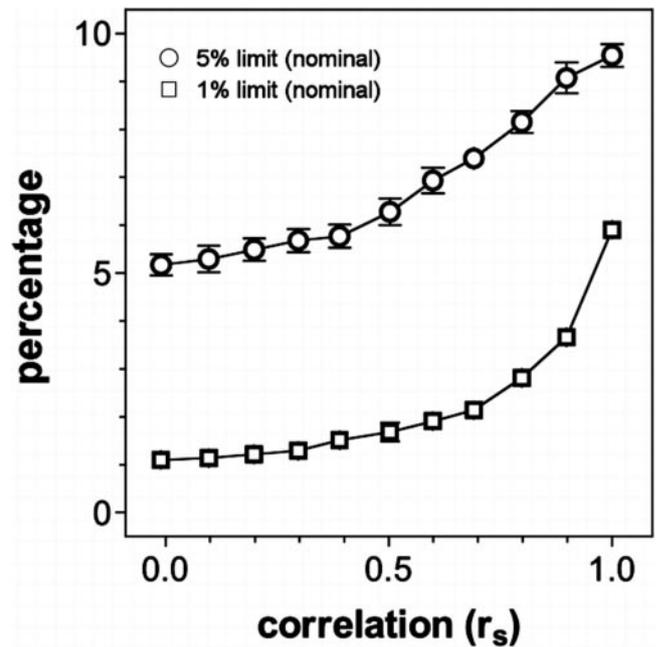


FIGURE 5. Shift in the nominal 5% and 1% probability limits of a new perimetry test when an established perimetry test, of correlation as given on the abscissa, is used to screen subjects for the new test's database. Subjects were required to return greater than, or equal to, the fifth percentile performance on the established perimeter. Probability limits were determined by examining the proportion of all normal subjects (i.e., no criterion on established perimetry) excluded by the new test. Data points are mean of five simulations, each using 2000 data points per test. Error bars, \pm SEM. The SEs along the abscissa (correlation) were always < 0.007 , and are not plotted. Values for α (equation 1) were 0, 0.10, 0.18, 0.25, 0.31, 0.38, 0.44, 0.50, 0.58, 0.68, and 1.00.

choice for detecting early disease. In contrast, though, previous work has found that PSD is superior to MD in detecting glaucomatous visual field damage²⁴ and that PSD (in an analogous form, corrected loss variance) is superior to MD for detecting the onset and progression of glaucomatous visual field damage.²⁵

If we use 5% limits of normality on conventional perimetry as a normative criterion, on average we expect to exclude 1 person in 20 from the database. If we use multiple test indices as a criterion for normality (for example, normal MD, PSD, and glaucoma hemifield test [GHT]), then the proportion of subjects rejected increases, and the specificity for disease detection decreases. By repeat testing of subjects with abnormal tests indices, test specificity may be improved,^{20,26} and some of the subjects initially falling outside the normative criteria may become eligible for inclusion in the database. If retesting is allowed, however, it is important that this be noted in the eligibility criteria.

In conclusion, we find that using a criterion of normal perimetric performance is unlikely to result in large biases in perimetric normative databases, particularly if the criterion test differs in type from the test whose database is being produced. In addition, we find that criteria based on the index PSD introduce smaller biases than those based on the index MD. Based on these findings, we recommend that future developers of perimetric databases might further minimize biases by adopting liberal criteria for those reference test indices showing strong correlation with the new test, and relatively stricter criteria for those indices showing poor correlation. For example, in our study, a normative criterion of MD \geq 1%, and a PSD \geq 5% on the HFA may be more appropriate than setting \geq 5% limits for both indices. Considering that the correlation between indices on the reference and new test typically is unknown when commencing data collection for a new database, it may be worthwhile to apply perimetric criteria post hoc once data collection is completed, correlation analyses performed, and the likely level of bias calculated.

References

- Adams CW, Bullimore MA, Wall M, Fingeret M, Johnson CA. Normal aging effects for frequency doubling technology perimetry. *Optom Vis Sci.* 1999;76:582-587.
- Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol.* 1987;105:1544-1549.
- Johnson CA, Adams AJ, Casson EJ, Brandt JD. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Arch Ophthalmol.* 1993;111:645-650.
- Casson EJ, Johnson CA, Nelson-Quigg JM. Temporal modulation perimetry: the effects of aging and eccentricity on sensitivity in normal. *Invest Ophthalmol Vis Sci.* 1993;34:3096-3102.
- Bengtsson B. A new rapid threshold algorithm for short-wavelength automated perimetry. *Invest Ophthalmol Vis Sci.* 2003;44:1388-1394.
- Wong AMF, Sharpe JA. A comparison of tangent screen, Goldmann, and Humphrey perimetry in the detection and localization of occipital lesions. *Ophthalmology.* 2000;107:527-544.
- Fujimoto N, Saeki N, Miyauchi O, Adachi-Usami E. Criteria for early detection of temporal hemianopia in asymptomatic pituitary tumor. *Eye.* 2002;16:731-738.
- Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. *Doc Ophthalmol Proc Ser.* 1987;54:153-168.
- Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand.* 1997;75:368-375.
- Harper R, Reeves B. Compliance with methodological standards when evaluating ophthalmic diagnostic tests. *Invest Ophthalmol Vis Sci.* 1999;40:1650-1657.
- Harper R, Henson D, Reeves BC. Appraising evaluations of screening/diagnostic tests: the importance of the study populations. *Br J Ophthalmol.* 2000;84:1198-1202.
- Anderson AJ, Vingrys AJ. Small samples: does size matter? *Invest Ophthalmol Vis Sci.* 2001;42:1411-1413.
- Weih LM, Nanjan M, McCarty CA, Taylor HR. Prevalence and predictors of open-angle glaucoma: results from the visual impairment project. *Ophthalmology.* 2001;108:1966-1972.
- Sponsel WE, Santiago A, Yolanda T, Mensah J. Clinical classification of glaucomatous visual field loss by frequency doubling perimetry. *Am J Ophthalmol.* 1998;125:830-836.
- 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.
- Horikoshi N, Osako M, Tamura Y, Okano T, Usui M. Comparison of detectability of visual field abnormality by frequency doubling technology in primary open-angle glaucoma and normal-tension glaucoma. *Jpn J Ophthalmol.* 2001;45:503-509.
- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician.* 1983;32:307-317.
- Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci.* 1997;38:413-425.
- Johnson CA, Cioffi GA, Van Buskirk EM. Frequency doubling technology perimetry using a 24-2 stimulus presentation pattern. *Optom Vis Sci.* 1999;76:571-581.
- Johnson CA, Sample PA, Cioffi GA, Liebmann JR, Weinreb RN. Structure and function evaluation (SAFE): I. Criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP). *Am J Ophthalmol.* 2002;134:177-185.
- Anderson DR, Patella VM. *Automated Static Perimetry.* St. Louis: CV Mosby; 1999:111-113.
- Hays WL. *Statistics.* New York: Holt, Rinehart and Winston, Inc. 1963:641-652.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C. The Art of Scientific Computing.* Cambridge: Cambridge University Press; 1992:278-282.
- Soliman MAE, de Jong LAMS, Ismaeil AA, van den Berg TJTP, de Smet MD. Standard achromatic perimetry, short wavelength automated perimetry, and frequency doubling technology for detection of glaucoma damage. *Ophthalmology.* 2002;109:444-454.
- Chauhan BC, Drance SM, Douglas GR. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci.* 1990;31:512-520.
- Khong JJ, Dimitrov PN, Rait J, McCarty CA. Can the specificity of the FDT for glaucoma be improved by confirming abnormal results? *J Glaucoma.* 2001;10:199-202.