IMPORTANCE Standard automated perimetry is the current criterion standard for assessment of visual field (VF) loss in glaucoma. The 3 commonly used reliability indices to judge the quality of standard automated perimetry results are fixation losses (FLs) and false-positive (FP) and false-negative (FN) response rates. However, the influence of reliability indices, when within the manufacturer-recommended limits, on VF classification has been sparsely studied.

OBJECTIVE To evaluate the role of VF reliability indices in ruling out glaucoma.

DESIGN, SETTING, AND PARTICIPANTS A cross-sectional study of 291 eyes of 291 participants referred to a tertiary eye care facility by general ophthalmologists. The participants were suspected to have glaucoma based on optic disc appearance, but the eyes were judged to be normal with physiological cupping by glaucoma experts on masked evaluation of optic disc photographs. All participants underwent VF testing with the Swedish interactive threshold algorithm standard 24-2 program.

MAIN OUTCOMES AND MEASURES Logistic regression models were used to evaluate the associations between reliability indices and FP classifications on VF testing (glaucoma hemifield test as outside normal limits and pattern standard deviation with P < .05).

RESULTS Median FL, FP, and FN response rates were 7%, 1%, and 2%, respectively. Among the 241 participants with reliable VF results (FL <20% and FP response rate <15%), the VF classification was normal in 188 (78.0%) and glaucoma (FP) in 53 (22.0%). Probability of FP VF classification was associated with FN response rates (odds ratio [OR], 1.36; 95% CI, 1.25-1.48, P < .001) but did not appear to be associated with FLs (OR, 0.96; 95% CI, 0.90-1.03, P = .30) or FP response rates (OR, 0.96; 95% CI, 0.83-1.12, P = .64). Predicted probability of FP VF classification was 9% (95% CI, 6%-14%), 40% (32%-49%), and 82% (68%-91%) at FN response rates of 0%, 8%, and 16%, respectively.

CONCLUSIONS AND RELEVANCE This study suggests that FN response rates have an effect on the ability of automated VF assessments to rule out glaucoma. Since FN response rates are ignored by the manufacturer while flagging a test as unreliable, clinicians and researchers may benefit by realizing that FN response rates can lead to FP VF classification, even when their frequencies are small.
Standard automated perimetry (SAP) is the current criterion standard for assessment of visual field (VF) loss in glaucoma. However, because SAP is a subjective test, understanding and cooperation of the patient is essential for reliable results. The 3 commonly used reliability indices to judge the quality of SAP results are fixation losses (FLs) and false-positive (FP) and false-negative (FN) response rates. Manufacturers of the Humphrey field analyzer (Zeiss Humphrey Systems), one of the most commonly used SAP devices, had recommended a cutoff of 20% for FLs and 33% for FP and FN response rates for reliable results with the STATPAC algorithm.1 For the currently available Swedish interactive threshold algorithm (SITA), the manufacturers suggest a cutoff of 20% for FLs and 15% for FP response rates for reliable VF results.2 False-negative response rates are not considered while flagging a test result as unreliable. This was established following the results of multiple studies that found FN response rates were related more to the severity of glaucomatous damage than to the patient’s attentiveness.3–6 The estimation methods of the FP rates are also different with SITA compared with the STATPAC algorithm.7,8

One of the important groups of subjects seen in glaucoma practice is the group referred to as glaucoma suspects based on the optic nerve head appearance. Although a good clinical examination by a glaucoma expert can rule in or rule out glaucoma in most of these subjects, additional diagnostic tests are usually used to support or complement the clinical assessment. Standard automated perimetry is generally the first test to be performed in these situations, followed by optic disc imaging if available. We therefore hypothesized that the reliability indices of VFs—even when they are within the recommended limits—can affect the VF interpretation. The purpose of this study was to evaluate the role of VF reliability indices with SITA in ruling out glaucoma in subjects referred to a tertiary eye care facility by general ophthalmologists as glaucoma suspects based on the optic nerve head appearance. These subjects’ eyes were, however, judged as normal with physiological optic disc cupping by glaucoma experts on masked evaluation of optic disc photographs.

Methods

Study Population

This was an observational, cross-sectional study of subjects referred to a tertiary eye care facility by general ophthalmologists as glaucoma suspects based on the optic disc appearance. The study included new patients presenting for the first time between September 8, 2010, and November 23, 2012, and participants who came for routine follow-up examinations during this period. Written informed consent was obtained from all participants to participate in the study, and the institutional review board of L V Prasad Eye Institute approved the methods. All methods adhered to the tenets of the Declaration of Helsinki for research involving humans. For participants who had multiple VF assessments performed during the follow-up, the most recent assessment during the study period was selected for the analysis.

Inclusion criteria were best-corrected visual acuity of 20/40 or better and refractive error within ±5 diopter (D) sphere and ±3 D cylinder. Exclusion criteria were presence of any media opacities that prevented good-quality optic disc photographs and any retinal or neurologic disease that could confound the VF results. All participants underwent a comprehensive ocular examination that included a detailed medical history, best-corrected visual acuity measurement, slitlamp biomicroscopy, Goldmann application tonometry, gonioscopy, SAP, dilated fundus examination, and digital optic disc photography.

VF Assessment

Standard automated perimetry was performed using the Humphrey field analyzer, model 750i, with the SITA standard 24-2 algorithm. Experienced technicians explained the procedure to the participants in their local language before the test. Pupils were dilated if the pupillary diameter was less than 3 mm. Reliability parameters noted from the VF assessment printouts were FL, FP, and FN response rates. These parameters have been explained in detail elsewhere.2,8 Fixation losses, which are provided on the printout as fractions, were converted to percentages. Visual field results with FLs of more than 20% or FP response rates of more than 15% were classified as low reliability according to the manufacturer’s recommendation.2 Reliable VF results were classified as glaucomatous if the pattern standard deviation had a P value of less than .05 and the glaucoma hemifield test result was outside normal limits.2 Visual field results were classified as normal otherwise.

Digital optic disc photographs (FF 450i with VISUPAC 4.2.2, Carl Zeiss Meditec Systems GmbH) were obtained by trained technicians. Photographs consisted of a 50° image centered on the optic disc, a similar image centered on the macula, a 30° image centered on the optic disc, and a 20° image centered on the disc. All images also consisted of 1 color and 1 red-free image each. Each photograph was evaluated by 2 of 4 experts (H.L.R., U.K.A., N.S.C., and S.S.) independently, who were masked to the clinical examination results of the participants and the VF classification and other eye examination results. Experts graded the presence or absence of the following features on disc photographs: superior and inferior neuroretinal rim thinning, superior and inferior rim notch, superior and inferior disc hemorrhage, and superior and inferior wedge-shaped retinal nerve fiber layer (RNFL) defect. Any of these findings about which the experts were unsure were graded as suspicious. Discrepancies between the 2 experts were resolved by consensus. Eyes for which a consensus could not be reached were excluded from analysis. Eyes for which a feature was graded as suspicious by either of the experts were also excluded from the analysis. Experts also made an overall classification of glaucoma and nonglaucoma based on the above features. Eyes for which a classification to either the glaucoma or nonglaucoma group was not possible by either or both the experts (true disc suspects) were also excluded from the analysis.

For the current study, all eyes whose optic discs were graded as nonglaucomatous by the experts were included. For the analysis, 1 eye of participants for whom both eyes were eligible for inclusion was randomly chosen.

Statistical Analysis

Descriptive statistics included mean (SD) for normally distributed variables and median (interquartile range [IQR]) for non-
normally distributed variables. Characteristics of the eyes for which the VF classification was glaucoma (FP) were evaluated. Linear regression models were used to evaluate the factors associated with the reliability indices, and logistic regression models were used to evaluate the associations between reliability indices and FP classifications on VF assessment. Two other independent factors used in the above multivariable models were the age of the subject and the number of VF assessments performed by the subject during follow-up. Statistical analyses were performed using commercial software (Stata, version 11.2; StataCorp).

Results

During the study period, 941 eyes of 532 participants referred by general ophthalmologists for glaucoma evaluation were reviewed. Of these, 5 eyes with poor-quality disc photographs and 501 eyes classified either as glaucoma or glaucoma suspect on masked evaluation of disc photographs were excluded, leaving 435 eyes of 291 participants eligible for inclusion in the study. The overall agreement between experts for optic disc classification as normal was 94%. Remaining optic discs were classified as normal by consensus. One eye of participants for whom both eyes were eligible was randomly chosen, leaving 291 eyes of 291 participants for the final analysis. Median age of the participants was 52.5 years (IQR, 41.8-61.2 years). Median mean deviation, pattern standard deviation, and VF index values were –2.23 dB (IQR, –3.94 to –0.97 dB), 1.86 dB (IQR, 1.51-2.38 dB), and 98% (IQR, 95%-99%), respectively.

Figure 1 shows the distribution of the number of VF examinations performed during follow-up and the reliability indices of these eyes: 202 participants (69.4%) performed 1 VF examination, while 55 (18.9%) performed 3 or more examinations during follow-up. Median FL, FP, and FN response rates were 7% (IQR, 0-14.3%), 1% (IQR, 0%-4%), and 2% (IQR, 0%-6%), respectively.

Table 1 shows the results of multivariable regression models evaluating the factors associated with FP and FN responses in the study participants. False-positive responses increased significantly with increases in FLs and FN responses, while FN responses increased significantly with increases in FP responses. Although both FP and FN responses were statistically significantly related to each other, the association was weak ($R^2 = 0.1$).

Visual field classifications were flagged as low reliability in 50 participants (17.2%). Reasons for the classifications being unreliable were GL greater than 20% in 46 participants, FP response rate greater than 15% in 1 subject, and both FL greater than 20% and FP response rate greater than 15% in 3 participants. Results of the logistic regression model evaluating the factors associated with the unreliable VF classifications are shown in Table 2. Probability of unreliable classifications was higher in older participants. Predicted probabilities of unreliable VF classifications were 6% (95% CI, 3%-14%), 12% (95% CI, 7%-18%), 21% (95% CI, 16%-27%), and 35% (95% CI, 21%-51%) at 20, 40, 60, and 80 years of age, respectively, according to the logistic regression model. Probability of unreliable VF classifications did not appear to be associated with the number of VF assessments a subject had performed during follow-up.

Among the 241 participants with reliable VF results, the classification was normal in 188 (78.0%) and glaucoma in the remaining 53 (22.0%). Results of the multivariable logistic re-
Reliability indices of the VF are popular measures used both in clinical practice and research studies to assess how well a subject has performed the test. In this study, we analyzed the pattern of reliability indices and their role in ruling out glaucoma in a group of participants who were referred by general ophthalmologists as glaucoma suspects based on the optic disc appearance. The participants’ eyes were, however, classified as normal after masked evaluation of their optic disc photographs by 2 glaucoma experts independently.

Analyzing the relationship between the reliability indices, we found that the FP and FN response rates were statistically significantly associated with each other. False-positive response rates increased as the FN response rates increased. False-positive response rates were also significantly associated with FLs. Similar associations using the STATPAC algorithm were reported by Reynolds et al in patients with glaucoma. Both the FP and FN response rates were not associated with the total number of VF examinations the subject had performed during the follow-up period. Similar results were reported in a longitudinal study by Katz et al. A prospective 3-year study by Johnson and Nelson-Quigg also found very little change in FP and FN response rates in healthy participants followed up annually with VF examinations. Fixation losses showed a small decline on examinations during the follow-up period in their study.

Discussion

Reliability indices of the VF are popular measures used both in clinical practice and research studies to assess how well a subject has performed the test. In this study, we analyzed the pattern of reliability indices and their role in ruling out glaucoma in a group of participants who were referred by general ophthalmologists as glaucoma suspects based on the optic disc appearance. The participants’ eyes were, however, classified as normal after masked evaluation of their optic disc photographs by 2 glaucoma experts independently.

Analyzing the relationship between the reliability indices, we found that the FP and FN response rates were statistically significantly associated with each other. False-positive response rates increased as the FN response rates increased. False-positive response rates were also significantly associated with FLs. Similar associations using the STATPAC algorithm were reported by Reynolds et al in patients with glaucoma. Both the FP and FN response rates were not associated with the total number of VF examinations the subject had performed during the follow-up period. Similar results were reported in a longitudinal study by Katz et al. A prospective 3-year study by Johnson and Nelson-Quigg also found very little change in FP and FN response rates in healthy participants followed up annually with VF examinations. Fixation losses showed a small decline on examinations during the follow-up period in their study.

Table 2. Factors Associated With Unreliable Visual Field Classifications in the Study Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.04 (1.01-1.06)</td>
<td>.01</td>
</tr>
<tr>
<td>VF examinations performed, No.</td>
<td>1.00 (0.87-1.15)</td>
<td>.98</td>
</tr>
</tbody>
</table>

Abbreviation: VF, visual field.
* R² of the model evaluating factors associated with unreliable VF classification was 3%.

Table 3. Factors Associated With the False-Positive Classification of Visual Field Results as Glaucoma in the Study Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation losses</td>
<td>0.96 (0.90-1.03)</td>
<td>.30</td>
</tr>
<tr>
<td>Age</td>
<td>1.04 (1.01-1.06)</td>
<td>.01</td>
</tr>
<tr>
<td>VF examinations performed, No.</td>
<td>1.00 (0.87-1.15)</td>
<td>.98</td>
</tr>
</tbody>
</table>

Abbreviation: VF, visual field.
* R² of the model evaluating factors associated with false-positive VF classification was 26%.

Going by the manufacturer-recommended limits, 17.2% of VF classifications in our study were flagged as low reliability. Fixation losses were the most common cause of low reliability of VF classifications. Forty-nine of these 50 unreliable classifications were due to FLs greater than 20%. Earlier studies have also reported FLs to be the most common cause of unreliable VF classifications. Probability of unreliable VF results was significantly higher in older participants. Past experience of a subject in VF examination was not associated with the probability of a low-reliability result.

When evaluating the influence of reliability indices in ruling out glaucoma on VF assessments, we found that FL and FP response rates had no effect on the false classification rates, while FN response rates had a significant effect. In a previous study using the STATPAC algorithm of the Humphrey field analyzer, Katz and Sommer evaluated the classification of automated perimetry against the criterion standard of manual perimetry and found that FN response rates significantly affected the specificity of automated perimetry classifications. The FP automated VF classification rates reported by Katz and Sommer were 7% at an FN response rate of 0%, 10% to 30% at an FN response rate of 1% to 19%, 30% to 40% at an FN response rate of 20% to 32%, and 40% to 60% at an FN response rate of more than 33%. These FP automated VF classification rates were significantly lower than that found in our study, possibly because of the difference in the criterion standards used in the 2 studies. We used structural abnormality on disc photographs as the criterion standard, while Katz and Sommer used functional abnormality on manual perimetry (which is supposed to agree better with automated perimetry) as the criterion standard.

Learning effect is an important confounder in all the analyses of our study. To account for the learning effect, we included the total number of VF examinations a subject had performed during follow-up as an independent variable in
all the analyses. Surprisingly, the influence of learning effect on the reliability indices, proportion of low-reliability VF classifications, or proportion of FP classification of VF results was nonsignificant. This outcome may be related to the detailed explanation of the test provided by the technicians to participants performing the test at our center; it may also be related to the low number of participants in our study who had an experience of performing more than 1 VF examination.

The digital optic disc photographs used in this study were 2-dimensional. Although simultaneous stereoscopic optic disc photographs are considered better than 2-dimensional photographs in evaluating subtle features, such as excavation of the neuroretinal rim, earlier studies have shown similar agreement between experts under both 2-dimensional and stereoscopic conditions in parameter estimation, such as cup to disc ratio, and in classifying optic discs as glaucomatous.13,14 The participants in our study were referred by general ophthalmologists as glaucoma suspects based on their optic disc appearance. Therefore, a possible limitation of our study is the inclusion of a few early glaucoma cases (misdiagnosed as normal) in the analysis. This situation might have increased the FP classification rates of VF results. This eventualty is, however, less likely as 2 glaucoma experts independently identified the optic discs as nonglaucomatous and the RNFL as normal. There was no ambiguity in the glaucoma experts’ classification. Therefore, optic discs included in the control group, although referred as suspects for glaucoma, were not true suspects but were discs with large physiological cups that caused a diagnostic uncertainty among general ophthalmologists. We excluded such true suspects (optic discs that were unable to be classified into the glaucoma or nonglaucoma group by 1 or both of the experts) from the analysis. Such true suspects would require a longitudinal study to look for progressive structural changes and to definitively classify them into the glaucoma or nonglaucoma group.15 We believe that including a control group that is likely to cause some amount of diagnostic uncertainty is more meaningful and mimics the real-life clinical situation than a control group with no suspicious findings of the disease. We have earlier used such a control group for evaluating the diagnostic ability of imaging technologies in glaucoma.16–20

Conclusions

Clinicians and researchers should realize that FN response rates are important even when their frequencies are small. This finding may contribute to a VF result erroneously being classified as glaucomatous in a patient who otherwise has normal physiological cupping.

ARTICLE INFORMATION
Submitted for Publication: May 29, 2014; final revision received July 29, 2014; accepted July 31, 2014.

Author Contributions: Dr Rao had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Rao. Acquisition, analysis, or interpretation of data: All authors. Drafting of the manuscript: Rao. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: Rao. Administrative, technical, or material support: All authors. Study supervision: Rao.

Conflicts of Interest Disclosures: Drs Rao and Garudadri are consultants to Allergan, and Dr Garudadri is also a consultant to Alcon and Merck. Dr Garudadri reports receiving a research grant from Optovue. No other disclosures were reported.

REFERENCES