A Comparison between the Compass Fundus Perimeter and the Humphrey Field Analyzer

Giovanni Montesano, MD,1,2,3 Susan R. Bryan, PhD,2 David P. Crabb, PhD,2 Paolo Fogagnolo, MD,1 Francesco Oddone, MD, PhD,4 Allison M. McKendrick, PhD,2 Andrew Turpin, PhD,6 Paolo Lanzetta, MD,7 Andrea Perdicchi, MD,8 Chris A. Johnson, PhD,9 David F. Garway-Heath, MD, FRCOphth,3 Paolo Brusini, MD,10 Luca M. Rossetti, MD, FRCOphth1

**Purpose:** To evaluate relative diagnostic precision and test–retest variability of 2 devices, the Compass (CMP, CenterVue, Padova, Italy) fundus perimeter and the Humphrey Field Analyzer (HFA, Zeiss, Dublin, CA), in detecting glaucomatous optic neuropathy (GON).

**Design:** Multicenter, cross-sectional, case-control study.

**Participants:** We sequentially enrolled 499 patients with glaucoma and 444 normal subjects to analyze relative precision. A separate group of 44 patients with glaucoma and 54 normal subjects was analyzed to assess test–retest variability.

**Methods:** One eye of recruited subjects was tested with the index tests: HFA (Swedish interactive thresholding algorithm [SITA] standard strategy) and CMP (Zippy Estimation by Sequential Testing [ZEST] strategy), 24-2 grid. The reference test for GON was specialist evaluation of fundus photographs or OCT, independent of the visual field (VF). For both devices, linear regression was used to calculate the sensitivity decrease with age in the normal group to compute pointwise total deviation (TD) values and mean deviation (MD). We derived 5% and 1% pointwise normative limits. The MD and the total number of TD values below 5% (TD 5%) or 1% (TD 1%) limits per field were used as classifiers.

**Main Outcome Measures:** We used partial receiver operating characteristic (pROC) curves and partial area under the curve (pAUC) to compare the diagnostic precision of the devices. Pointwise mean absolute deviation and Bland–Altman plots for the mean sensitivity (MS) were computed to assess test–retest variability.

**Results:** Retinal sensitivity was generally lower with CMP, with an average mean difference of 1.85 ± 0.06 decibels (dB) (mean ± standard error, \( P < 0.001 \)) in healthy subjects and 1.46 ± 0.05 dB (mean ± standard error, \( P < 0.001 \)) in patients with glaucoma. Both devices showed similar discriminative power. The MD metric had marginally better discrimination with CMP (pAUC difference ± standard error, 0.019 ± 0.009, \( P = 0.035 \)). The 95% limits of agreement for the MS were reduced by 13% in CMP compared with HFA in participants with glaucoma and by 49% in normal participants. Mean absolute deviation was similar, with no significant differences.

**Conclusions:** Relative diagnostic precision of the 2 devices is equivalent. Test–retest variability of MS for CMP was better than for HFA. *Ophthalmology* 2019;126:242-251 © 2018 by the American Academy of Ophthalmology

Supplemental material available at www.aaojournal.org.

---

Standard automated perimetry is used to assess the visual field (VF) and is a key examination for detection, diagnosis, and follow-up in glaucoma. Standard automated perimetry typically uses stimuli of varying intensities to assess the differential light sensitivity at static locations across the VF. The examination demands strong cooperation1 from test subjects; they are required to maintain central fixation and respond timely and accurately to the presented stimuli. Fixation instability might be an unavoidable feature of a person’s vision, especially with advanced age and macular damage.2 One proposed solution has been to incorporate live fundus tracking in the macular perimetric exam to compensate for eye movements in unstable fixation.3

Recently, a novel instrument, the Compass fundus perimeter (CMP, CenterVue, Padua, Italy), has successfully used a live fundus tracking technology for wide-field (30 degrees) VF assessment yielding results comparable to the Humphrey Field Analyzer (HFA) in a preliminary study.4 The CMP captures images of the fundus during the perimetric examination using a scanning laser ophthalmoscope. This design feature is intended to afford compensation for eye movements when the stimuli are presented at predetermined test locations. Moreover, the instrument provides color images of the fundus and optic nerve that can be mapped to the final perimetric results, potentially providing clinically useful information about structure and function in 1 assessment.

Diagnostic accuracy studies are used to certify new examinations before they are brought into clinical practice. The CMP has not been scrutinized in this way, and this is
the main purpose of our investigation. Studies investigating relative diagnostic accuracy are at risk of bias because of shortcomings in design and conduct. For this reason, we designed our study to follow appropriate guidelines on this specific aim.6,7

Our cross-sectional and multicenter study was designed to evaluate and compare 2 index tests, namely, the CMP and the HFA. One objective was to evaluate and compare test–retest variability of the 2 index tests in healthy subjects and patients with glaucomatous optic neuropathy (GON). We hypothesized that the CMP could obtain a 20% reduction in test–retest variability on the measurement of the mean sensitivity (MS) of the VF. Another objective was to build a normative database for the CMP and analyze its relative discriminative ability, compared with HFA, in detecting subjects with GON. We specifically hypothesized that the 2 index tests will have equivalent relative diagnostic precision as assessed by partial area under the curve (pAUC) at >75% specificity, across a spectrum of disease severity. In both analyses, the reference assessment for GON was specialist evaluation based on the inspection of fundus photograph or spectral-domain OCT evaluation of the retinal nerve fiber layer (RNFL), independent of the VF. A further objective was to evaluate examination times for the CMP and HFA.

Methods

Data Collection for the Normative Database and Discrimination Analysis

People were recruited at 8 study sites. These were as follows: ASST - Santi Paolo e Carlo, Milan, Italy; Azienda Ospedaliero Universitaria Santa Maria della Misericordia of Udine, Udine, Italy; NIHR Clinical Research Facility at Moorfields Eye Hospital, London, United Kingdom; Department of Ophthalmology and Visual Sciences University of Iowa, 200 Hawkins Drive, Iowa City, Iowa; Department of Optometry & Vision Sciences, The University of Melbourne, Parkville, Australia; IRCCS Fondazione “G. B. Bietti,” Clinica Oculistica Università degli Studi di Roma “La Sapienza,” Rome, Italy; and Azienda Ospedaliera Sant’Andrea, Rome, Italy. Recruitment started on September 9, 2015, and concluded on July 31, 2017. Data collection was planned before the index test and reference standard were performed. The study was designed to achieve a target number of 1000 participants with glaucoma and 600 healthy participants for the normative database and discrimination analysis. However, these targets were not reached by the termination date of the study.

Participants eligible for inclusion were consecutive adults (aged 18–90 years) with the following:

- best-corrected visual acuity >0.8 (if ≤50 years old) or >0.6 (if >50 years old) in the study eye;
- refraction –10 diopters (D)/+6 D; astigmatism ±2 D;
- absence of systemic pathologies that could affect the VF; and
- no use of drugs interfering with the correct execution of the perimetric test.

Additional specific inclusion criteria for healthy subjects were as follows:

- normal optic nerve head (ONH) in both eyes (no evidence of excavation, rim narrowing or notching, disc hemorrhages, RNFL thinning);
- intraocular pressure less than 21 mmHg in both eyes; and
- no ocular pathologies, trauma, surgeries (apart from uncomplicated cataract surgery) in both eyes.

Additional specific inclusion criteria for those with glaucoma were as follows:

- GON defined as glaucomatous changes to the ONH or RNFL as determined by a specialist from fundus photograph or spectral-domain OCT, independently of the VF.
- Patients had to be receiving antiglaucoma therapy.
- Patients had to have no ocular pathologies, trauma, surgeries (apart from uncomplicated cataract surgery), other than glaucoma, in both eyes.

Eligible patients were identified on the basis of a clinical diagnosis of GON from the clinical registry of the glaucoma clinics in the recruiting centers. An expert clinician confirmed the diagnosis of GON using the imaging data (RNFL spectral-domain OCT or optic nerve photograph) acquired during the protocol examination (discussed later). Subjects were recruited consecutively. Because the VF metrics were not included in the identification of patients with GON, no stratification was planned according to disease severity.

Eligible healthy participants were identified among staff in the clinics, volunteer registries, patients’ spouses or partners, and patients attending the clinic for reasons other than glaucoma (e.g., for preoperative assessment for cataract in the fellow eye). If deemed eligible for the study, healthy subjects were recruited consecutively.

Both eyes were examined, but only 1 eye per subject was used in the final analysis, chosen randomly if both eyes were eligible. All patients gave their written informed consent to participate in the study. Ethics Committee approval was obtained (International Ethics Committee of Milan, Zone A, July 22, 2015, ref: Prot. no 0019459), and the study was registered as a clinical trial (ISRCTN13800424). This study adhered to the tenets of the Declaration of Helsinki.

Each subject had an ophthalmological evaluation after a standard operating procedure involving assessment of axial length measurement with the IOLMaster (Zeiss, Dublin, CA) biometer, spectral-domain OCT of the ONH and RNFL, perimetric demonstration (only for subjects naïve to perimetry); 1 examination with HFA 24-2 grid Swedish interactive thresholding algorithm (SITA) standard to both eyes, and 1 examination with CMP New Grid (discussed later), Zippy Estimation by Sequential Testing (ZEST) strategy to both eyes; and color fundus photo with CMP.

The reference standard to diagnose GON was clinical evaluation by an expert based on RNFL spectral-domain OCT or ONH photography. The rationale for this choice was to avoid any classification based on VF testing that could have affected the analysis of the relative discriminative power of the index tests. The 2 index tests were VF examinations with the HFA and the CMP. The order of CMP and HFA tests was randomized. The VF examination performed with the HFA used a 24-2 grid and the SITA standard algorithm. Near correction was used. Fixation was monitored with blind spot tests using the Heijl and Krakau method.3

The VF examination performed with the CMP used a testing grid termed “New Grid,” which differs from the HFA 24-2 grid (Fig S1, available at www.aaojournal.org). The new grid contains all the 52 locations tested with a 24-2, only 1 blind spot location (instead of 2 as in the 24-2) and 12 additional points in the macular region of the VF. The testing strategy was an adaptation of the ZEST.9,10 Because the CMP is equipped with autofocusing, no near correction was needed. Blind spot responses were monitored by projecting stimuli on the location of the ONH, identified manually by the operator on the baseline infrared fundus image captured at the beginning of the test. In all the analyses, only
the 52 locations in common between the 24-2 and the new grid were used.

For both devices, VF examinations were considered reliable if the false-positive frequency was \( \leq 18\% \) and the blind spot response frequency was \( \leq 25\% \). If either the HFA or the CMP VF was deemed unreliable, the eye was excluded from the analysis.

### Statistical Analysis

All analyses were based exclusively on the 52 locations in common between the 24-2 grid (HFA) and the new grid (CMP).

Differences between the 2 devices in terms of MS and its decrease with age in healthy subjects were analyzed. Because the same eyes were tested with both devices, a mixed model was used to account for repeated measurements.

Linear regression was used to estimate expected decrease in sensitivity with age in healthy subjects (decibels [dB]/years) at each VF location. Total deviation (TD) values for each VF in normal and glaucoma subjects were calculated as the deviation from the mean trend in the age model for each location. Mean deviation (MD) was calculated as the mean of all 24-2 grid TD values in each VF. Mixed models were used to compare MS and MD values between the 2 devices in both the glaucoma and normal groups. The MD values were only compared for the glaucoma group because subjects in the normal group were used to calculate the TD values and are bound to have a mean MD equal to zero with both devices.

Normative lower limits for each location were calculated for TD values using quantile regression\(^{11,12}\) to account for changes in normal variability with age. Because the variability of thresholds in healthy subjects is known to increase with age,\(^{12,13}\) we only allowed for negative slopes in quantile regression, meaning that normative limits could not shrink with age. Only the lower 5% and 1% limits for TD values were used in this analysis.

For a fair comparison, TD values and their normative limits were calculated in the same fashion for HFA and CMP, using the dataset of healthy subjects acquired with each respective device in this study. For each VF, we calculated the total number of TD values below the 5% and 1% limits, which we refer to as TD 5% and TD 1%, respectively.

Discrimination ability of the 2 index tests was measured using MD, TD 5%, and TD 1% as classifiers. These classifiers were used to build receiver operating characteristic curves. Instead of comparing the whole receiver operating characteristic curve, we analyzed the partial receiver operating characteristic curve (pROC) down to a minimum specificity of 0.75 to avoid comparing the 2 devices at too low specificity values that would fall far outside a clinically useful range. The 95% confidence intervals (CIs) for pAUCs, and \( P \) values for differences were calculated via bootstrapping.\(^{14}\)

The normative data, used to calculate MD and TD metrics and their normative limits, were composed of the same set of healthy subjects used in the discrimination analysis to calculate pROC curves and their pAUCs. Therefore, they are only used here to compare the relative performance of the 2 devices and not to estimate or report their actual discriminative power.

To compare test times, CMP average time per location was calculated for each test and the result multiplied by the number of total points in a 24-2 grid (54 points). This made it comparable to the testing time read from the printout of the HFA.

### Data Collection for Test–Retest Variability

A separate group of glaucoma and healthy subjects was recruited to assess test–retest variability with the 2 devices. The target number was 56 subjects with GON and 56 healthy subjects. The sample size calculation for this part of the study was based on previously reported data for test–retest in healthy subjects and those with glaucoma.\(^{15,16}\) All subjects underwent the same examinations reported for the previous section, and the diagnosis of GON was again confirmed by expert evaluation of the RNFL on spectral-domain OCT images or photographs of the ONH. Subjects were sequentially recruited in the same way described for the previous part of the study. No stratification by disease (VF) severity was planned in the recruitment of glaucoma subjects. All subjects performed 4 VF tests: 2 with CMP with a 24-2 grid, ZEST strategy, and 2 with HFA with a 24-2 grid, SITA standard strategy, in randomized order. All examinations were done within a time span of 7 days.

### Statistical Analysis

Test–retest variability for the overall VF was assessed for MS using Bland–Altman plots and 95% limits of agreement. Any change in test–retest variability was evaluated by percentage reduction of the 95% interval of agreement of CMP over HFA. The 95% CIs for the percentage variation were estimated using a paired bootstrap procedure with 50,000 resamples. Mean absolute deviation was used to assess pointwise test–retest variability. Differences in mean absolute deviation, point-wise sensitivity, and MS were tested using \( t \) test statistics from linear mixed models with random effects to account for correlations between VF measurements from the same subject. All analyses were done using R version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

### Results

#### Normative Database

For this part of the study, 1249 people were screened for eligibility and invited to participate between September 14, 2015, and July 31, 2017. Of these, 177 subjects did not satisfy the inclusion criteria and 59 did not complete the examination protocol. Finally, 70 subjects were excluded because they had at least 1 unreliable VF test (48 with HFA, 20 with CMP, and 2 with both devices). Therefore, 444 healthy subjects and 499 glaucoma subjects (patients with GON) were included in the final analysis. Although no stratification by disease severity was planned, a wide spectrum of VF severity was obtained by the end of the recruitment. Glaucoma Staging System 2\(^{17}\) stage distribution for glaucoma participants is reported in Table 1 and depicted in Figure 1.

Subjects’ age distributions are reported in Table 1. Mean age (± standard deviation [SD]) was 48±16 years and 68±11 years for the normal and glaucoma groups, respectively.

Average MS was lower with CMP compared with HFA in healthy subjects (mean ± SD, 27.6±1.6 dB vs. 29.4±2.0 dB) and glaucoma subjects (20.5±6.7 dB vs. 21.9±6.9 dB), and these differences were both statistically significant (\( P < 0.001 \)). Comparison of the MD values in healthy subjects has not been performed because this group was used to calculate the normative average, and therefore they were bound to have zero means for both devices. The MD values from the 2 devices showed good agreement (Fig 2). Indeed, the average MD (±SD) for glaucoma subjects was –6.55±6.60 dB (median, –4.37 dB; interquartile range, 8.92 dB) with CMP and –6.50±6.63 dB (median, –4.73 dB; interquartile range, 9.19 dB) with HFA, and this difference was not statistically significant (\( P = 0.54 \)).

The average number of presentations (± SD) per location in CMP was 3.02±0.55 for healthy subjects and 3.70±1.09 for glaucoma patients. Corrected test duration for CMP and test
Table 1. Age Distribution of Healthy Subjects and Patients with Glaucoma Enrolled for this Study

<table>
<thead>
<tr>
<th>Group</th>
<th>Age (Mean ± SD)</th>
<th>Age Cluster Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>47.7 ± 16.4</td>
<td>151 (34%)</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>67.5 ± 11.2</td>
<td>12 (2%)</td>
</tr>
<tr>
<td>GSS2 Staging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0</td>
<td>83 (17%)</td>
<td>S1 (55 (11%))</td>
</tr>
<tr>
<td>S1</td>
<td>39 (8%)</td>
<td>S2 (80 (16%))</td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td>S3 (77 (15%))</td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td>S4 (94 (19%))</td>
</tr>
<tr>
<td>S4</td>
<td></td>
<td>S5 (71 (14%))</td>
</tr>
</tbody>
</table>

GSS2 = Glaucoma Staging System 2; SD = standard deviation.

Except for the cluster of people aged >80 years, the different age clusters are well represented in the normative database. Distribution of the 499 glaucoma subjects in the different stages of the Glaucoma Staging System 2 is shown in bottom panel.

duration for HFA were similar in both the healthy and glaucoma subjects (Table 2).

Point-wise sensitivity was generally lower for CMP compared with HFA (Fig 3). The average mean difference was 1.85 ± 0.06 dB (mean ± standard error, P < 0.001) in healthy subjects and 1.46 ± 0.05 dB (mean ± standard error, P < 0.001) in patients with glaucoma. Similarly to the MD, such a difference was reduced when TDs were considered in glaucoma subjects (Fig 4), with 7 locations exceeding 1 dB difference.

The MS in the healthy group decreased with age in a similar fashion for both devices, with a small but statistically significant difference (−0.051 ± 0.005 dB/year for HFA and −0.027 ± 0.005 dB/year for CMP; mean ± standard error; P < 0.001 for slope difference). The rate of false-positives was 1.6% ± 4.0% for CMP and 1.6% ± 2.3% for HFA (mean ± SD).

**Discussion**

This study was designed to compare 2 index tests, the CMP and HFA, in terms of test–retest variability and relative discriminative power. We recruited a large cohort of 943 subjects (499 patients with glaucoma and 444 healthy subjects) for the discrimination analysis and 98 subjects (44 glaucomatous and 54 healthy) to compare test–retest variability. The reference standard used for the diagnosis of GON was independent of VF assessment, based on specialist assessment of ONH color photography or peripapillary RNFL thickness measured with spectral-domain OCT.

Figure 1. Glaucoma Staging System 2 plot showing the distribution of the 499 subjects with glaucomatous optic neuropathy (GON) in the different stages of the classification. The light grey lines indicate the boundaries for the different stages. Subjects are classified on the basis of their mean deviation (MD) and pattern standard deviation (PSD) values directly taken from the Humphrey Field Analyzer (HFA, Zeiss, Dublin, CA) printout. The distribution is approximately uniform across the different stages.
The primary objective was to show a reduction of test–retest variability in the MS of at least 20%. Such a reduction was achieved in healthy subjects (49%), but not in glaucoma subjects, in whom the reduction was of 13%. Several factors might have contributed to this result, such as a more pronounced perimetric learning effect with CMP. The mean difference in MS in CMP between the first and the second test was small but statistically significant, and this may be indicative of a learning effect in the glaucoma test–retest cohort. This effect was not seen in the HFA data. Indeed, despite all glaucoma subjects in our sample having had previous experience with standard automated perimetry, the new setup of a fundus perimeter might have created an unfamiliar testing condition for test takers. In fact, most of them were recruited from glaucoma clinics and were experienced with HFA. The different threshold acquisition strategies used by the 2 devices may also explain this difference. The SITA strategies incorporate spatial information between neighboring test locations. Such an approach allows for faster threshold estimation, but it has been shown to bias the estimates introducing correlations between neighboring points. However, the implementation of the ZEST strategy used in CMP tests each point independently. Moreover, test–retest variability is known to increase dramatically at lower sensitivities, and this effect may simply consume any improvements from adjusting for fixation stability afforded by the tracking in fundus perimetry. We speculate this is the reason we see a bigger improvement in test–retest reliability in the healthy subjects compared with the patients in this study. This is supported by the results shown in the Bland–Altman plots for pointwise sensitivities, where it can be observed that the CMP offers no advantage in test–retest variability compared with HFA at values below 15 dB. Indeed, the 95% limits of agreement between 11 and 14 dB were larger for CMP than for HFA. The difference here might be explained by the spatial smoothing and the use of growth pattern to seed the priors in the SITA strategy, which might play a large role in reducing the test–retest variability in this sensitivity range. However, the clinical utility of thresholds below 15 dB has been questioned. Indeed, recent evidence suggests that increasing perimetric contrast all the way to 0 dB may not be clinically useful, and sensitivities obtained at severely damaged VF locations (<15–19 dB) are unreliable and highly variable. It could be argued that improvements in test–retest variability in the upper range of

![Figure 2](image_url)

**Figure 2.** The 2 panels show the agreement of mean deviation (MD) (left) and mean sensitivity (MS) (right) values between the Compass (CMP, CenterVue, Padova, Italy) (vertical axis) and the Humphrey Field Analyzer (HFA) (horizontal axis). The black solid line indicates the ideal perfect agreement. The red dots represent the healthy subjects, and the green dots indicate glaucoma subjects. Differently from MS, MD values did not show important differences between the 2 devices.

<table>
<thead>
<tr>
<th>Table 2. Test Duration for Compass and Humphrey Field Analyzer and Number of Presentations for Compass in Healthy Subjects and Patients with Glaucoma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healthy</strong></td>
</tr>
<tr>
<td>Test duration (sec)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Presentations (52 locations)</td>
</tr>
</tbody>
</table>

CMP = Compass; HFA = Humphrey Visual Field Analyzer.

Normalized test duration, which compensated for the unequal number of test locations examined by the 2 devices, was similar with a statistically significant difference only in glaucoma subjects. However, the difference was small (11.1±126.3 sec, mean ± SD). Data are reported as mean ± SD. For the number of presentations, the 25th and 75th percentiles are reported in square brackets.
Sensitivity values could be more clinically relevant for progression detection. However, this is speculation because only analysis of long-term follow-up of glaucoma subjects with the CMP will allow the assessment of the real effect of such reduction in variability on earlier diagnosis of progression.

Figure 3. Average sensitivity (decibels [dB]) for each of the 52 locations considered in this analysis for Compass (CMP) (A) and Humphrey Field Analyzer (HFA) (B). The bottom panels report the average pairwise difference per location in the healthy subjects (C) and glaucoma patients (D).

Figure 4. Average total deviation (TD) value (decibel [dB]) for each of the 52 locations considered in this analysis for Compass (CMP) (A) and Humphrey Field Analyzer (HFA) (B). C, The average pairwise difference (CMP − HFA) in TD per location in the glaucoma subjects (in bold all differences exceeding 1 dB).
Additionally, Wyatt et al.\(^3\) identified gaze instability as a possible source of variability at the edges of scotomata, and tracking might help reduce this effect. However, their analysis was performed with a 10-2 grid, which has a finer spacing between locations (2 degrees). Thus, further investigation is needed to assess the effect of gaze instability in the estimation of edges on a typical testing grid, such as 24-2 or 30-2.

One limitation of our analysis is that the sample size of the glaucoma test–retest group was probably too small to reliably assess any differences, as shown by the large CIs calculated via bootstrapping (Fig S2, available at www.aaojournal.org). Post hoc power calculations based on bootstrap resampling estimated that 97 glaucoma subjects would have been needed to detect a 20% improvement at a significance level of 0.05 with 80% power. This is considerably above the initial estimates obtained from literature data\(^{15,16}\) used for designing of the study. Therefore, an additional investigation with longer test series on a larger sample might be needed to fully assess the effect of fundus tracking on test–retest variability.

Relative discriminative power for the 2 index tests (devices) was similar. When compared, pROC curves calculated using the number of abnormal points per field in the TD maps largely overlapped, with no evidence for any superiority of either index test (Fig 5). Statistically significant differences in pROC curves were observed when MD was used as a classifier, but such differences are too small to be likely relevant in clinical situations. These results are compatible with the fact that although the actual sensitivity estimates were lower for CMP compared with HFA, relative indices, such as the MD and TD values, showed only small differences in glaucoma subjects between the 2 devices, yielding similar diagnostic ability.

Our results are based on a large sample of individuals from different centers. The different age clusters, except for people older than 80 years of age, were well represented (Table 1). This was sufficient to reliably conduct an analysis on relative discriminative power. It is important to note that, for both devices, all indices used in the discrimination analysis (MD, TD 5%, and TD 1%) and the normative limits for TD were recalculated in the same fashion from the raw sensitivities and are therefore comparable. However, because the normative limits have been derived from the same group of healthy subjects used in the discrimination analysis, the pAUCs are biased and can only be used to compare the relative discriminative ability of the 2 devices; they cannot be generalized to estimate

### Table 3. Sensitivity Values at Selected Specificity Values for Mean Deviation, Total Deviation 5%, and Total Deviation 1% for Both Compass and Humphrey Visual Field Analyzer

<table>
<thead>
<tr>
<th>Specificity (%)</th>
<th>CMP</th>
<th>HFA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>71</td>
<td>57.7 [48.9–67.8]</td>
</tr>
<tr>
<td>95</td>
<td>22.8 [16.4–36.9]</td>
<td>21.4 [13.3–32.2]</td>
</tr>
<tr>
<td>97.5</td>
<td>15.8 [8.6–21.6]</td>
<td>11.5 [6.3–19.8]</td>
</tr>
<tr>
<td>99</td>
<td>7.9 [2.7–14.6]</td>
<td>4.7 [0.0–10.6]</td>
</tr>
<tr>
<td><strong>TD 5%</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>79</td>
<td>76.2 [71.1–80.5]</td>
</tr>
<tr>
<td>95</td>
<td>64.7 [56.2–72.8]</td>
<td>58.6 [51.7–67.7]</td>
</tr>
<tr>
<td>97.5</td>
<td>49.2 [38.2–63.4]</td>
<td>45.5 [32.3–58.3]</td>
</tr>
<tr>
<td>99</td>
<td>36.8 [20.9–48.1]</td>
<td>31.7 [17.2–43.2]</td>
</tr>
<tr>
<td><strong>TD 1%</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>73.1 [69.1–77.1]</td>
<td>78.8 [74.8–82.4]</td>
</tr>
<tr>
<td>95</td>
<td>61.7 [55.9–67.7]</td>
<td>68 [59.7–73.8]</td>
</tr>
<tr>
<td>97.5</td>
<td>55.9 [48.3–61.8]</td>
<td>58.2 [48.9–66.4]</td>
</tr>
<tr>
<td>99</td>
<td>47.8 [35.6–55.9]</td>
<td>42.7 [21.6–58.1]</td>
</tr>
</tbody>
</table>

The confidence intervals (CIs) were obtained via Bootstrap.
CMP = Compass; HFA = Humphrey Visual Field Analyzer; MD = mean deviation; TD = total deviation.
the effective discriminative power of the CMP or the HFA in clinical practice.

Examination times for the 2 devices were similar. Both devices took, on average, 5 to 6 minutes to complete. Testing times had to be corrected before comparison because of the greater number of tested locations with the new grid used with CMP (65 locations) compared with the HFA 24-2 grid (54 locations). After corrections, no statistically significant differences could be detected between the 2 devices in healthy subjects. A statistical difference was observed in glaucoma subjects, but it is clinically irrelevant (approximately an 11-second difference on average).

Despite similarities in overall examination times, fewer presentations were needed to estimate thresholds in CMP when compared with HFA at the 52 matching locations. The number of presentations in healthy subjects was 157±28, which is lower than that reported for SITA standard in the literature (276 for 52 locations). Unfortunately, interpretation of the examination times of the 2 devices is difficult for a variety of reasons. For example, CMP uses catch trials, whereas HFA SITA algorithms use response times to estimate false-positive error rates. Moreover, the CMP does not project stimuli when the quality in the tracking signal is low, and this may increase overall examination time.

One limitation of our study is that the glaucoma subjects were not stratified according to disease severity, because VF data were not used in the diagnosis of GON. This could have resulted in an uneven representation of glaucoma stages. However, the range of VF damage was sufficient large to allow for a reliable evaluation across the whole spectrum of glaucoma damage (Table 1, Fig 1).

Our recruitment of healthy subjects was not population based, and this is another potential limitation of our study. The
main design bias potentially recruiting “super-normals” in studies of diagnostic precision is to recruit the healthy control group using restriction criteria related to the outcome of interest, for example, requiring the healthy controls to have normal VFs. We explicitly avoided this bias. Nevertheless, volunteers to clinical studies may be healthier than an unselected population. This is hard to avoid, because participants need to volunteer. However, when we analyzed the MD values from the HFA printouts of the 444 healthy subjects, whose calculation is based on the independent internal normative database built in the device, we found that our sample did not show important deviations from the normative values. Indeed, the average MD was $-1.12\pm1.64$ dB (median, $-0.91$ dB; interquartile range, 1.97 dB).

Finally, the design of this study only allowed for a relative comparison of discriminative power. Evaluation of the actual diagnostic accuracy would need a further validation on an independent dataset to assess how much these findings can be extracted on the general population. Furthermore, such an evaluation should be conducted on a set of subjects before the reference test (the clinical diagnosis of GON) is performed, because case-control scenarios are known to produce biased estimates in discrimination analyses. One option might be to test glaucoma suspects with the CMP before they are diagnosed as healthy or as having glaucoma.

References

Footnotes and Financial Disclosures

Originally received: April 17, 2018.
Final revision: August 2, 2018.
Accepted: August 3, 2018.

Available online: August 13, 2018.

1 University of Milan — ASST Santi Paolo e Carlo, Milan, Italy.
2 City, University of London, Optometry and Visual Sciences, London, United Kingdom.
3 NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom.
4 G. B. Bietti Eye Foundation-IRCCS, Rome, Italy.
5 University of Melbourne, Department of Optometry and Vision Sciences, Melbourne, Australia.
6 University of Melbourne, School of Computing and Information System, Melbourne, Australia.
7 Department of Medical and Biological Sciences, Ophthalmology Unit, University of Udine, Udine, Italy.
8 Ophthalmology Unit, St. Andrea Hospital, NESMOS Department, University of Rome “Sapienza,” Rome, Italy.
9 Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, Iowa.
10 Department of Ophthalmology, “Città di Udine” Health Center, Udine, Italy.

Presented in part at the Association for Research in Vision and Ophthalmology Meeting, April 29, 2018, to May 3, Honolulu, Hawaii.

Financial Disclosure(s):

CenterVue, Padua, Italy, sponsored the study. The sponsor or funding organization had no role in the design or conduct of this research, data analysis, interpretation of the data, or preparation, review, or approval of the manuscript. The contribution of IRCCS G. B. Bietti Foundation to this work was supported by the Italian Ministry of Health and by Fondazione Roma.

HUMAN SUBJECTS: Human subjects were included in this study. The human ethics committees at the International Ethics Committee of Milan, Zone A approved the study. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent.

No animal subjects were used in this study.

Author Contributions:
Conception and design: Montesano, Fogagnolo, Oddone, McKendrick, Turpin, Lanzetta, Perdicchi, Johnson, Garway-Heath, Brusini, Rossetti
Data collection: Montesano, Fogagnolo, Oddone, McKendrick, Turpin, Lanzetta, Perdicchi, Johnson, Garway-Heath, Rossetti
Analysis and interpretation: Montesano, Bryan, Crabb, McKendrick, Turpin, Garway-Heath, Rossetti
Obtained funding: N/A
Overall responsibility: Montesano, Bryan, Crabb, Fogagnolo, Oddone, McKendrick, Turpin, Lanzetta, Perdicchi, Johnson, Garway-Heath, Brusini, Rossetti

Abbreviations and Acronyms:
CI = confidence interval; CMP = Compass; D = diopeters; dB = decibels; GON = glaucomatous optic neuropathy; HFA = Humphrey Field Analyzer; MD = mean deviation; MS = mean sensitivity; ONH = optic nerve head; pAUC = partial area under the curve; pROC = partial receiver operating characteristic curve; PSD = pattern standard deviation; RNFL = retinal nerve fiber layer; SD = standard deviation; SITA = Swedish interactive thresholding algorithm; TAU = total deviation; VF = visual field; ZEST = Zippy Estimation by Sequential Testing.

Correspondence:
David P. Crabb, PhD, City, University of London, Optometry and Visual Sciences, Northampton Square, London, England EC1V 0HB UK. E-mail: David.Crabb1@city.ac.uk.