



City Research Online

City St George's, University of London

Citation: Montesano, G., McKendrick, A. M., Turpin, A., Brusini, P., Oddone, F., Fogagnolo, P., Perdicchi, A., Johnson, C. A., Lanzetta, P., Rossetti, L. M., et al (2021). Do Additional Testing Locations Improve the Detection of Macular Perimetric Defects in Glaucoma?. *Ophthalmology*, 128(12), pp. 1722-1735. doi: 10.1016/j.ophtha.2021.06.012

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/26573/>

Link to published version: <https://doi.org/10.1016/j.ophtha.2021.06.012>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Do additional testing locations improve the detection of macular perimetric defects in glaucoma?

Giovanni Montesano^{1,2}; Allison M. McKendrick³; Andrew Turpin⁴; Paolo Brusini⁵; Francesco Oddone⁶; Paolo Fogagnolo⁷; Andrea Perdicchi⁸; Chris A. Johnson⁹; Paolo Lanzetta¹⁰; Luca M. Rossetti⁷; David F. Garway-Heath²; David P. Crabb¹

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

Corresponding author: David P. Crabb
Email: david.crabb.1@city.ac.uk
Phone number: +44 (0)20 7040 0191
Fax number: None
Address: City, University of London
Northampton Square, London
EC1V 0HB, United Kingdom

Meeting presentation: this work was presented during the ARVO Online Meeting 2021

Financial support(s): The contribution of IRCCS G. B. Bietti Foundation to this work was supported by the Italian Ministry of Health and by Fondazione Roma. The original data collection was sponsored by CenterVue, Padua, Italy. 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..

Key words: macula, glaucoma, perimetry, visual field

Conflict of interest: G.M, A.M.M., A.T., P.B., F.O., P.F., C.A.J., P.L., L.M.R., D.F.G.-H and D.P.C are or were Consultants for CenterVue, Padua, Italy.

Running head: combined perimetric grids to detect macular defects in glaucoma

Abstract

Purpose: to evaluate the ability of additional central testing locations to improve detection of macular visual field (VF) defects in glaucoma.

Design: prospective cross-sectional study.

Participants: 440 healthy people and 499 patients with Glaucomatous Optic Neuropathy (GON) were tested with a fundus tracked perimeter (CMP, CenterVue, Italy) using a 24-2 grid with 12 additional macular locations (24-2+).

Methods: GON was identified based on expert evaluation of optic nerve head photographs and optical coherence tomography scans, independently of the visual field (VF). We defined macular defects as locations with measurements outside the 5% and 2% normative limits on Total Deviation (TD) and Pattern Deviation (PD) maps within the VF central 10 degrees. Classification was based on the total number of affected macular locations (*overall detection*) or on the largest number of affected macular locations connected in a contiguous cluster (*cluster detection*). Criteria based on the number of locations and cluster size were used to obtain equivalent specificity between the 24-2 and the 24-2+, calculated using false detections in the healthy cohort. Partial Areas Under the detection Curve (pAUCs) were also compared at specificities $\geq 95\%$.

Main Outcome Measure: matched specificity comparison of the ability to detect glaucomatous macular defects between the 24-2 and 24-2+ grids.

Results: at matched specificity, *cluster detection* identified more macular defects with the 24-2+ compared to the 24-2. For example, the mean (95% confidence interval) increase in percentage of detection was 8 (5, 11)% and 10 (7, 13)% for TD-5% and PD-5% maps, respectively, and 5 (2, 7)% and 6 (4, 8)% for the TD-2% and PD-2% maps respectively. There was good agreement between the two grids. The improvement measured by pAUCs was also significant, but generally small. The percentage of eyes with macular defects ranged from 30 to 50%. Test time for the 24-2+ was longer (21% increase). Between 74% and 98% of defects missed by the 24-2 had at least one location with sensitivity < 20 dB,

Conclusions: VF examinations with additional macular locations can modestly improve the detection of macular defects in GON without loss of specificity when appropriate criteria are selected.

1 Glaucoma is an optic neuropathy with damage to Retinal Ganglion Cells (RGCs) and
2 progressive loss of the visual field (VF) ^{1, 2}. The use of Standard Automated Perimetry (SAP) to
3 measure the VF is central to management of glaucoma. SAP stimuli are organised in a fixed
4 grid of regularly spaced locations, typically covering the central 30 degrees with locations 6
5 degrees apart. However, this coarse uniform sampling could be insufficient for the macular
6 region, which contains more than 40% of all RGCs³. Testing patterns employed by some
7 perimeters, such as the Octopus (Haag-Streit, Köniz, Switzerland), might be less affected by
8 this problem. However, the widely used 24-2 test program of the Humphrey Field Analyser
9 (HFA, Zeiss Meditec, Dublin, CA) assesses 54 points but only 12 (less than one quarter) are
10 located within the central 10 degrees of the VF. Some reports suggest denser perimetric
11 sampling of the macular region to be more effective in identifying early glaucomatous defects,
12 using the HFA 10-2 test program for example⁴⁻⁶. These findings are of great importance both
13 for the correct diagnosis of patients and for the identification of *de novo* sight threatening
14 central defects in patients with confirmed glaucoma. Other very recently reported data
15 suggests there is little evidence that a 10-2 test will reveal central VF loss not already detected
16 by a 24-2⁷. This conflicting evidence is unhelpful for those managing glaucoma. Of course,
17 performing both tests is often not practical in a clinical setting. One alternative could be the
18 use of SAP grids that combine 24-2 with additional testing locations in the macular area, like
19 the 24-2C program for the HFA ⁸⁻¹¹.

20
21 Studies investigating improvement in the detection of macular defects with additional testing
22 locations (or with a 10-2 grid) all have limitations. For example, increasing the number of
23 tested locations for the central 10 degrees is bound to make finding abnormal locations more
24 likely but will increase false positive detection reducing the specificity of the assessment.
25 West et al. sensibly used matched-specificity criteria to compare the central values of the 24-2
26 grid and the 10-2 grid⁷, yet other investigators failed to do this and reported the detection of
27 macular defects using fixed rules (usually three contiguous affected locations) independently
28 of the grid^{4, 11}. This has led to conflicting results. Another source of disagreement lies in the
29 selection bias that might have affected previous reports, since the presence or absence of VF
30 defects in the 24-2 test was used to define the study groups (glaucoma or healthy controls).

31
32 We have previously presented the results of the main outcome of a clinical trial¹² comparing
33 the diagnostic precision of the Compass fundus perimeter (CMP, CenterVue, Padova, Italy)
34 with the HFA. The study was done in a large cohort of healthy people and patients with
35 Glaucomatous Optic Neuropathy (GON), defined independently of VF loss; an advantageous
36 design since it minimises selection bias towards the type of VF test. A secondary outcome of
37 the study was to evaluate whether additional macular testing locations improved the
38 detection of macular defects in patients with GON. To this aim, the testing grid used for the
39 CMP was designed to include all the usual locations of a 24-2 grid with 12 additional locations
40 within the central 10 degrees. A prospectively planned study in a large number of people such
41 as this represents an incredible opportunity to address the controversy surrounding the
42 benefit of more accurate testing of the macular region. We adopt matched-specificity criteria

43 to quantify the improvement in the detection of macular defects offered by the additional
44 testing locations over the conventional 24-2.

45 **Methods**

46 **Data collection**

47 This was a cross-sectional case-control study conducted in accordance with the Declaration of
48 Helsinki after written informed consent was acquired from each participant. This study
49 received approval from the ethics committee (International Ethics Committee of Milan, Zone
50 A, 22/07/2015, ref: Prot. n° 0019459) and was registered as a clinical trial
51 (ISRCTN13800424). Participants were recruited (14/09/2015 - 31/07/2017) at eight
52 different study sites¹². The primary aim of the study was to compare the relative diagnostic
53 performance of CMP and HFA perimeters in their ability to distinguish a large number of
54 people with healthy vision from those with GON. All participants in the study were tested with
55 both the HFA and the CMP, in randomised order. The main results have already been
56 presented elsewhere¹².

57 A secondary aim of the trial was to quantify the impact on detection of macular defects in the
58 glaucoma cohort brought by adding testing locations to the macular region of the VF. This will
59 be the focus of this work. To this aim, the CMP was equipped with a custom grid containing all
60 the 52 locations of a 24-2 grid (excluding the two blind spot locations) and 12 additional
61 macular locations. This grid, called 24-2+ ("New Grid" in our previous report¹², **Figure 1**), was
62 used for all participants. The additional 12 testing locations were excluded from the previous
63 analysis¹².

64 Selection criteria have been previously detailed¹² and are available from the published
65 protocol¹² (<http://www.isrctn.com/ISRCTN13800424>). Participants (consecutive adults)
66 eligible for inclusion had both eyes examined but only one eye per subject was used in the
67 final analysis, chosen randomly if both eyes were eligible. Each subject underwent complete
68 ophthalmological evaluation involving biometry to measure axial length (AL), Spectral
69 Domain Optical Coherence Tomography of the ONH and circumpapillary Retinal Nerve Fibre
70 Layer (cp-RNFL), perimetric demonstration (only for subjects naïve to perimetry); one
71 examination with HFA 24-2 grid to both eyes (data not used for this analysis, except to define
72 inclusion criteria as explained below) and one examination with CMP 24-2+ to both eyes;
73 colour fundus photo with CMP. The reference standard to diagnose GON was clinical
74 evaluation by an expert based on cp-RNFL SD-OCT and/or optic nerve head photography
75 acquired during the protocol examination. Experts were required to evaluate OCT scans for
76 RNFL thinning and ONH colour pictures for disc cupping, rim narrowing, focal notching or
77 peripapillary haemorrhages. The rationale for this reference standard was to avoid any
78 classification based on VF testing that could bias the results towards either perimeter. For the
79 current work, this also reduced the bias towards the 24-2 grid, since participants with GON
80 were included regardless of their VF.

81 **Details of the visual field examination**

82 The CMP uses continuous infrared imaging of the retina designed to track and compensate for

83 eye movements during the test¹²⁻¹⁴. Threshold acquisition used a Bayesian testing strategy, an
84 adaptation of the Zippy Estimation by Sequential Testing (ZEST)^{15,16}. No near correction was
85 needed because the CMP is equipped with auto-focusing. All locations in the 24-2+ were
86 tested independently in randomised order. Therefore, unlike the HFA SITA (Swedish
87 Interactive Thresholding Algorithm) test strategy, the CMP test does not use spatial
88 correlations between neighbouring locations or specific spatial patterns¹⁵; this was a
89 noteworthy convenience for our analyses because it allows the evaluation of the isolated
90 contribution of additional macular locations without this confounding effect, unavoidable
91 with SITA strategies. Differently from the 24-2C adopted by the HFA, the 24-2+ was designed
92 as a general-purpose pattern and not specifically to detect macular defects from glaucoma¹¹.
93 VF examinations were considered unreliable if the frequency of false positive errors was >
94 18% or the Blind Spot response frequency was > 25%. Same criteria were applied to the HFA
95 examination¹². If either the HFA or the CMP VF was deemed unreliable, the eye was excluded
96 from the analysis and was therefore not present in the final dataset.

97 **Statistical analysis**

98 Only locations within the central 10 degrees from fixation were considered (**Figure 1**). The
99 24-2+ contained all central 24-2 locations (N = 12) plus 12 additional macular locations (total
100 N = 24).

101 **Calculation of probability maps**

102 Data from the healthy cohort collected in this study represent the only available normative
103 database for the CMP and were therefore used to define normative values. Total Deviation
104 (TD) and Pattern Deviation (PD) maps were calculated as previously described¹². In short, a
105 linear regression for each test location was used to model the normal sensitivity decay with
106 age using the healthy cohort. The TD is simply the difference between the observed sensitivity
107 and the expected age corrected value at each location. PD is then calculated by subtracting the
108 General Height (GH) of the field from the TD map. Following the definition of the Imaging and
109 Perimetry Society (IPS)¹⁷, the GH was the value corresponding to the 7th highest location TD
110 map. This was calculated separately for the 24-2+ and the 24-2 grid, considering all locations
111 (64 and 52 respectively). Normative limits were calculated using quantile regression of TD
112 and PD values to account for age-related changes in variability as previously described¹². For
113 this analysis we calculated which sensitivity values in each VF were within or outside the 5%
114 or 2% normative limits for each map. For healthy subjects, the normative limits were
115 calculated using a leave-one-out procedure, so that each healthy subject was excluded from
116 the normative database when their maps were calculated. The whole normative dataset was
117 used to calculate the maps for patients with GON. Mean Deviation (MD) reported in this study
118 is the mean of total deviation values. MD was only used as a descriptive measure and was
119 calculated for the central 24-2 locations, the central 24-2+ locations and the whole 24-2 grid.

120 **Calculation of the matched specificity criteria**

121 Macular defects were identified using the probability maps. Hence, we obtained a defect
122 identification for the TD and the PD maps, each one with two possible probability thresholds
123 (5% or 2% normative limits). The macular defect for each probability map was identified
124 when a certain number of locations fell outside their normative limit (*abnormal locations*). We

125 used two defect definition strategies:

- 126 1) **Overall detection:** the central locations from either the 24-2 grid or the 24-2+ were
127 treated as a whole. Identification of the defect depended only on the total number of
128 locations outside the normative limits in each grid. This analysis is similar to what has
129 been proposed by West et al.⁷.
- 130 2) **Cluster detection:** the central locations outside the normative limits were pooled
131 together only if belonging to a cluster of contiguous points. The number used for
132 detection is then the largest cluster size identified in the central VF. Defining
133 contiguous locations is trivial in a regular grid such as the 24-2 but poses a challenge
134 for grids with irregular spacing, such as the 24-2+. We defined a neighbourhood
135 system for the 24-2+ based on a nearest-neighbour triangulation (**Figure 1**). The
136 locations were not allowed to connect across the horizontal midline. Hence, the
137 maximum size of one cluster was six locations for the 24-2 and 12 locations for the 24-
138 2+ (i.e. half the total number of central locations). The details of the computation are
139 reported in the **Appendix**. This method resembles a typical definition of glaucomatous
140 VF defects, used for example by De Moraes et al.⁴ and Phu et al.¹¹ in a similar analysis,
141 but allows for a flexible selection of the cut-off for the cluster size to match specificity
142 between the two grids (see below).

143

144 The objective of this study was to compare the detection rate of macular defects between 24-
145 2+ and 24-2 at the same specificity. To achieve this, the number of locations used for detection
146 needed to be different for the two grids. Specificity is defined based on the False Positive Rate
147 (FPR) as $S = (1 - FPR)$. We considered all locations outside the normative limits detected in
148 the healthy cohort as false positives. From this assumption, we could calculate the specificity
149 of increasing threshold criteria on the number of abnormal locations for each probability map
150 (**Figure 2**). For each map, we selected the two threshold criteria (one for each grid) that
151 yielded the same specificity. Given the discrete nature of the criterion, however, an exact
152 match in specificity is unlikely. For our analysis, we selected the smallest pair of threshold
153 criteria that provided a specificity $\geq 95\%$ and a difference in specificity between the two grids
154 $\leq 0.5\%$, assuming that any smaller difference would be clinically irrelevant. This calculation
155 was performed for both grids with all probability maps and for both detection strategies
156 (*overall detection* and *cluster detection*, **Figure 2**). This was meant to provide a practical
157 evaluation of the performance with criteria that could actually be applied. We also performed
158 a comparison of the partial Area Under the detection Curve (pAUC) limited to specificities \geq
159 95%, to provide an analysis that was not affected by these practical limitations. Note that
160 these pAUCs are not meant to evaluate the diagnostic precision but only the detection of
161 macular defects.

162 **Analysis of the detection rate**

163 The matched-specificity criteria were then applied to the GON cohort to detect macular
164 defects. The two grids were compared for all maps in terms of detection rate. Confidence
165 Intervals (CIs) and p-values for the difference in detection rate were calculated via bootstrap
166 (N = 5000 samples). Note that the bootstrap procedure was performed by resampling only
167 eyes with GON, so that specificity was held constant at each draw. The agreement between the

168 two grids was also calculated and represented using Venn diagrams. The pAUCs were
169 calculated by interpolating between different threshold criteria and normalising their value
170 over 0.05, the maximum AUC achievable with the selected specificity range (95% - 100%).
171 The pAUCs were compared using the same bootstrap procedure explained above. We also
172 explored the spatial distribution of locations outside the normative limits by calculating their
173 frequency for each location of the 24-2+ within the GON cohort. We compared the four
174 quadrants (**Figure 1**) in terms of probability of finding an abnormal location, using a logistic
175 regression. For this latter comparison, the p-values were corrected for multiple testing using
176 the Bonferroni-Holm method and considered significant when < 0.05 . Finally, we reported the
177 distribution of the defect depth missed by either grid to assess their clinical relevance. For
178 this descriptive analysis, we quantified the distribution of TD values for all the locations
179 identified as abnormal by each probability map that met the criteria to detect a macular defect
180 in one grid but not the other (i.e. the cases in which the 24-2 and the 24-2+ were in
181 disagreement). We also defined as “deep” defects all locations with $TD \leq -20$ dB and quantified
182 them for each grid.
183 All statistical analyses were performed in R (R Foundation for Statistical Computing, Vienna,
184 Austria).

185 **Comparing test times**

186 The duration of the test with the 24-2+ was recorded by the device. We obtained an accurate
187 estimate of the time taken to test the additional macular locations by analysing the recorded
188 history of each test extracted from the device, which reports the number of stimulus
189 presentations for each location. With this, we calculated the average time for each
190 presentation which was then used to estimate the time needed to test the 24-2 locations only.
191 (Note this is different from our first report¹², where the time estimate for the 24-2 was simply
192 derived as a proportion of the number of tested locations, not of the presentations
193 themselves.)

194 **Results**

195 **Description of the sample**

196 Of the 1249 people screened for eligibility, 177 did not satisfy the inclusion criteria and 59 did
197 not complete the examination protocol. Finally, 70 subjects were excluded because they had
198 at least one unreliable VF test and four healthy subjects were erroneously tested only with the
199 24-2 grid. Therefore, 440 healthy subjects and 499 patients with GON were included in the
200 final analysis. Descriptive statistics are reported in **Table 1**. As previously reported¹², despite
201 a significant gap in the average age between the healthy and GON cohort, the age range for the
202 healthy subjects was large (18 - 84 years) and allowed for reliable estimates of the normative
203 limits for all GON patients. The range of VF damage in the GON cohort was wide (range of MD
204 for the whole 24-2 grid: -27.85, +2.89 dB). MD for the central VF was very slightly higher
205 when measured with the 24-2+ compared to the 24-2 (**Table 1**). This difference was
206 statistically significant for people with GON ($p < 0.001$, paired Mann-Whitney test) but not for
207 the healthy cohort ($p = 0.99$).

208 Specificity analysis

209 The 440 subjects in the healthy cohort were used to calculate matched-specificity criteria for
210 the detection of macular defects. **Figure 2** shows how increasing the number of abnormal
211 locations required to identify a defect progressively increases the specificity. **Table 2** reports
212 the selected criteria and their respective specificity values for each map and detection
213 strategy, according to the overall and *cluster detection*.

214 Detection of macular defects

215 The 499 eyes with GON were used to test the performance of the two grids in detecting
216 macular defects with matched-specificity criteria. The improvement with the 24-2+ was
217 variable for the *overall detection*, with a significant improvement detected only for the 2%
218 probability maps (**Table 3**). The relative improvement was larger and significant for all maps
219 when contiguous *cluster detection* was employed. In this case, the largest improvement was
220 for the TD 5% map. The pAUC analysis partially replicated the results of the matched
221 specificity analysis, but the differences were in general much smaller (**Table 4**). All
222 differences were still significant when *cluster detection* was employed except for the TD 2%
223 map ($p = 0.068$). The *overall detection* was no longer statistically significant with the TD 2%
224 map but reached significance with the TD 5%. **Figure 3** shows the detection rate curves for all
225 methods and all maps, marking the operating points used for the matched specificity analysis.
226 The curves show that most of the difference between the two grids was obtained at very high
227 specificities.

228 Agreement between the two grids was good but not perfect, as shown by the Venn diagrams
229 (numerical values are reported in **Figure 4**). Interestingly, some macular defects were
230 identified by the 24-2 and not by the 24-2+. This was either a consequence of healthy
231 locations in the 24-2+ breaking up contiguous clusters or because the number of affected
232 locations was sufficient to meet the detection criteria with the 24-2 but not the 24-2+.
233 Two illustrative examples of the results obtained with the cluster detection with the selected
234 same-specificity criteria are reported in **Figure 5**. The distribution of the depth of the defects
235 missed by either grid for the cases of disagreement is reported in **Figure 6**, including the
236 number of locations with a deep defect. Many of the locations missed by the 24-2 had a deep
237 defect. It should be noted, however, that these locations came from a relatively small number
238 of eyes (largest $N = 54$, see **Figure 4**). The majority of the eyes that had a defect missed by the
239 24-2 but detected by the 24-2+ also had at least one location with a deep defect. This
240 percentage ranged between 74% and 98% when all 24-2+ locations were considered,
241 depending on the detection method and map, and between 47% and 90% using only locations
242 in common between the two grids. More details are available as **supplementary material**.

243 Spatial distribution of abnormal locations

244 The frequency of locations outside the normative limits in the GON cohort was higher in the
245 superior-temporal quadrant (**Figure 7**) for all probability maps. All pair-wise differences
246 between quadrants were significant except for the comparison between Quadrant 2 and 4
247 with the TD at 5% probability ($p = 0.329$). All other p-values were < 0.001 except for the
248 comparison between Quadrant 2 and 4 with the TD at 2% probability ($p = 0.002$). **Figure 7**
249 reports the predicted frequency and 95% CIs for each quadrant from the logistic regression.

250 Discussion

251 We evaluated the effect of additional perimetric testing locations on detecting macular defects
252 using prospectively collected data in a large cohort of patients with GON. Detection was based
253 on the number of abnormal locations in TD or PD probability maps, detected with matched-
254 specificity criteria derived from a large group of healthy eyes. We compared two different
255 detection strategies, where the abnormal locations were either pooled independently of their
256 spatial pattern (*overall detection*) or only if belonging to clusters of contiguous locations
257 (*cluster detection*). Improvement in the detection rate provided by the additional macular
258 locations was variable, ranging from an average of less than 1% to almost 10% depending on
259 the detection method used. All analyses were conducted at a specificity level that would be
260 clinically useful ($\geq 95\%$). When taken together, the size of these experimental effects suggest
261 perimetric grids with additional macular locations can modestly improve sensitivity of
262 detecting macular defects in glaucomatous optic neuropathy without loss of specificity,
263 provided appropriate criteria are applied to define a defect. Note, this is different from
264 comparing the diagnostic precision of the two grids. In fact, the 24-2 might be able to
265 correctly identify glaucoma cases based on peripheral damage while failing to detect central
266 defects. The pAUC analysis partially confirmed this view, showing a consistent improvement
267 with the *cluster detection*, less so with the *overall detection*. It should be noted that the
268 improvement in interpolated pAUC with the 24-2+ grid was modest for both methods,
269 indicating very similar overall performance (**Figure 3** and **Table 4**). However, it should be
270 stressed that this is not reflective of how the two grids would be used in practice, i.e. with
271 discrete cut-offs on the number of affected locations to identify a defect. This is why it was
272 important to define a clinically acceptable difference in matched specificity for our practical
273 evaluation. For example, many comparisons yielded a difference in detection rate between
274 8% and 10% (**Table 3**), which we believe is meaningful accepting a clinically insignificant
275 change in specificity $\leq 0.5\%$.

276
277 Our findings are in partial agreement with those from previous reports. For example, De
278 Moraes et al.⁴ showed that more central defects were detected by the 10-2 when compared to
279 24-2 grid in patients with ocular hypertension (OHT), glaucoma or suspected glaucoma. The
280 largest number of missed central defects in their dataset was recorded within the OHT and
281 glaucoma suspect cohort. In these groups, they reported an improved detection of
282 approximately 12% with the 10-2, which is generally higher than we found in our dataset. The
283 results by De Moraes et al.⁴ could, however, be biased by their definition for OHT and suspects
284 requiring a normal 24-2 VF as an inclusion criterion. We had no such bias in our study design
285 because the inclusion criteria did not involve VF loss. Moreover, the analysis employed by De
286 Moraes et al.⁴ applied the same detection criteria to both the 24-2 and the 10-2 grid failing to
287 account for the loss of specificity introduced by the 10-2, which tests a higher number of
288 locations in a smaller area of the VF. The reported percentage of false positives with the 10-2
289 grid was 4.6% in their healthy dataset, but they could not assess specificity for the 24-2 grid
290 because all healthy subjects needed to have a normal 24-2 test result. Besides, in contrast to
291 our analysis, De Moraes et al.⁴ compared the ability of identifying any defect in the whole 24-2
292 grid versus the 10-2 grid; the percentage of missed central defects was then inferred from the

293 mismatch in the results from the two grids. This would not identify any occurrence when the
294 24-2 would identify glaucoma due to the presence of a peripheral defect but fail to highlight
295 macular damage (**Figure 5, A**). Instead, we focused our analysis only on the central region,
296 with the specific intent of evaluating the improvement in the detection of macular defects. We
297 think this makes for a fairer comparison. Of course, other reasons could explain the
298 differences between our results and those presented by De Moraes et al.⁴. For example, our
299 additional macular locations did not sample the central field at the same density as a 10-2
300 grid, possibly underestimating the number of central defects. Our analysis, however, directly
301 addresses the clinical question pertaining to the usefulness of combined VF grids, which are
302 more likely to be an acceptable compromise for everyday clinical practice.
303 More recently, West et al.⁷ performed a matched-specificity analysis with a method identical
304 to our *overall detection* to compare the 24-2 and the 10-2 grids. Their results are in close
305 agreement with ours, concluding for minimal improvement in the detection of macular
306 defects with the 10-2 grid when no spatial patterns are considered. Their analysis, however,
307 was performed on a much smaller sample (97 eyes with glaucoma and 65 controls) and
308 suffered from another bias, since glaucoma patients were required to have an early damage
309 on the 24-2 VF test for inclusion. In contrast, our recruitment scheme was originally designed
310 to compare the diagnostic ability of two perimeters and purposely avoided any inclusion
311 criterion based on VF tests, be it with a 24-2 or 10-2 grid¹².

312
313 Another novel aspect of our analysis was the use of an objective rule to detect contiguous
314 abnormal locations that was independent of the specific arrangement of the testing grid
315 (*cluster detection*). Differently from static criteria used previously^{4, 11}, our method allowed the
316 selection of matched-specificity thresholds in the same way as with the *overall detection*. This
317 is important because it is unusual for clinicians to consider abnormal locations in isolation.
318 Simply counting the number of abnormal values would likely deviate from clinical practice.
319 However, empirical combination rules are often applied to different grids without accounting
320 for changes in specificity^{4, 11}, which was instead central in our analysis. Interestingly, our
321 *cluster detection* provided the largest improvement with the 24-2+, which reached
322 significance in all probability maps; this suggests that additional testing locations are
323 beneficial in defining the spatial patterns of glaucoma damage, even when matched-specificity
324 criteria are applied. This may explain the differences observed between the results by De
325 Moraes et al.⁴ and West et al.⁷, since spatial continuity of abnormal locations were not
326 considered by the latter. Wu et al.¹⁸ also failed to show any significant improvement in the
327 detection of macular defects between 24-2 and 10-2 using only global metrics such as pattern
328 standard deviation.

329
330 Phu et al.¹¹ recently published a report comparing the 24-2 and the new 24-2C grid provided
331 by the HFA. The 24-2C has 10 additional testing locations derived from the 10-2 grid designed
332 to optimise the detection of macular glaucoma defects. Their main results showed no
333 significant differences between the two grids in detecting macular defects. Their study did not
334 select glaucoma patients based on VF criteria, hence removing the selection bias, but had
335 some technical limitations in the testing procedure. First, the HFA only allows the 24-2C test
336 with the SITA-Faster strategy. Moreover, all SITA strategies employ spatial correlations

337 between neighbouring locations¹⁵ to improve the speed of the test. This prevents an
338 independent evaluation of the isolated central locations of the 24-2 grid and of the
339 contribution of the additional locations in the 24-2C. Each location was instead tested
340 independently in the ZEST strategy implemented in the CMP. Although this might not be
341 optimal for practical perimetry, it provides a convenient and ideal condition for our
342 experiment. In fact, limitations imposed by the spatial correlations employed by SITA
343 strategies are common to all the aforementioned studies^{4, 7, 11} and would also greatly affect
344 any analysis aimed at isolating the central locations from a larger grid such as the 24-27,¹¹.
345 This therefore represents another strength of our study that allowed us to make additions to
346 the literature that would not be possible with conventional SITA methods.

347
348 Other important differences with the work by Phu et al.¹¹ pertain to the definition of VF
349 defects. They also used a cluster based approach, but specificity was not calculated in a
350 healthy cohort and was not matched between the two grids. Moreover, they did not define the
351 neighbourhood used in the 24-2C to identify contiguous locations. Importantly, their main
352 results relied on the definition of “additive” detection, where the contribution of additional
353 locations was only evaluated in terms of improvement over the 24-2. This would fail to
354 highlight all the instances where the additional normal locations would break up contiguous
355 clusters found in the 24-2 (**Figure 5, B**) and hence reducing the detection rate. They reported
356 only two such cases in their dataset. However, they also separately reported, in the same
357 paper, the detection of defects based only on the central locations of the two grids with
358 different cluster sizes, where this effect more clearly shows, i.e. some macular defects were
359 only detected with the 24-2 and not with the 24-2C. This analysis is similar to ours and the
360 results largely agree, showing that the vast majority of the defects were detected by both
361 grids, with some improvement in detection with the additional macular locations. In their
362 case, these differences were not significant, but their sample size (N = 64) was much smaller
363 than ours. It is also important to notice that in our specific case, the missed identification of
364 macular defects with additional central locations could also be due to the different criteria (i.e.
365 more abnormal locations required) used for the 24-2+ to maintain the same specificity as the
366 24-2. It is also worth noting that the many of the affected locations not identified as macular
367 defects by the 24-2 showed a deep defect (**Figure 6**). This might be of clinical relevance
368 despite the small number of eyes in which there was disagreement between the two grids.
369 However, in practice, it is unusual to consider probability thresholds in isolation and the
370 depth of the defect is often taken into account. In fact, many of these deep defects were also
371 detected by locations in common with the 24-2 grid, indicating that accounting for the
372 magnitude of loss might improve the detection with the 24-2 despite its lower spatial
373 resolution. Future work should focus on the development and validation of criteria that
374 combine different probability levels from TD and PD maps. These exist for the 24-2¹⁹ but have
375 not yet been formalised for other grids.

376
377 Our study has other important strengths. Due to our recruiting strategy, no stratification was
378 planned for VF damage. Despite this, the range of overall VF loss was large. However, most of
379 the eyes with GON had early to moderate central damage (**Table 1**), possibly as a
380 consequence of the lower bounds imposed on visual acuity for recruitment. This constitutes a

381 very convenient setting to evaluate the performance of the two grids in detecting macular
382 defects which in some cases could be subtle. Moreover, all locations were tested within the
383 same session and in no preferred order within the test. In all previous studies subjects were
384 tested with the 24-2 or the alternative grid (either the 10-2 or the 24-2C) in separate
385 sessions^{4, 7, 11}. Moreover, we employed a fundus tracked perimeter that compensates for eye
386 movements. This can be particularly valuable when closely spaced locations are tested near
387 fixation, removing the effect of eye movements on the spatial resolution of the grid^{20, 21}.
388 Another important aspect is that the GH was chosen as the 7th highest value in the TD and
389 could therefore be different for the two grids. We believe this is more reflective of how the
390 calculation would be performed in clinical practice, where all the available data-points would
391 be used. Moreover, this is also in line with previous studies involving the 10-2 test, for which
392 the GH would be calculated in a similar way. Also note that, differently from the 24-2, this
393 value for the GH is not exactly the 85th percentile for the 24-2+¹⁷.

394
395 Our study has some limitations. Our main analysis only focussed on the detection
396 improvement provided by the central locations. This could be a limitation for the *cluster*
397 *detection* method, since it would force the clusters to be fully contained within the central 10
398 degrees. For example, this approach may miss any macular defect identified by isolated
399 central locations that connect to larger clusters outside the central 10 degrees. This choice
400 was specifically meant to increase the specificity for macular defects which would likely lie on
401 nerve fibre bundles fully contained within the central 10 degrees²². In response to this
402 potential limitation the **supplementary material** reports a secondary analysis allowing
403 clusters to extend to the whole grid and detecting a macular defect if any of these clusters
404 invaded the central 10 degrees. As expected, more macular defects were identified with both
405 grids and the results leaned more towards an equivalence between the 24-2 and 24-2+.
406 However, we reiterate that such a liberal approach to cluster connectivity might not be
407 specific for anatomically plausible macular defects and should be interpreted with caution. It
408 should also be noted that the 24-2+ grid was not specifically designed to reflect the topology
409 of macular defects from glaucoma, unlike the 24-2C. However, while patterns tailored to
410 glaucomatous defects might be desirable for this specific application, they make the test less
411 generalisable, possibly compromising the detection of other sight threatening diseases such
412 as age-related macular degeneration. Finally, our healthy cohort represented the only
413 available normative database for CMP. We accounted for this by using a leave-one-out
414 approach in the calculation of the normative limits for the healthy cohort. One limitation of
415 our definition of the healthy cohort is that only one grader evaluated the structural data to
416 exclude the presence of GON. However, given our additional constraints on the intraocular
417 pressure and the relatively low prevalence of GON in the general population, misclassification
418 of eyes with GON as healthy is unlikely. The opposite misclassification (healthy as GON) would
419 only dilute the percentage of detected macular defects but would not compromise the same
420 specificity comparison between the two grids. Future evaluations might also benefit from
421 more detailed characterisation of structural damage with dense macular OCT maps, not
422 available for this cohort. The lack of structural confirmation also prevented us from assessing
423 the actual prevalence of macular damage in eyes with GON. Hence, we could not report the

424 sensitivity of the methods but only the relative difference in detection rate (proportional to
425 sensitivity).

426

427 A definitive judgment on the usefulness of increased spatial sampling of the central VF in
428 glaucoma patients is hard to come by. With our results, we showed a consistently significant
429 increase in the detection rate of macular defects when spatially connected clusters are
430 considered. It would be therefore inappropriate to dismiss any improvement from denser
431 macular grids as a simple consequence of a loss in specificity⁷. Indeed, a previous report by
432 Grillo et al.⁵ showed that macular defects identified by the 10-2 grid and confirmed by
433 independent structural measurements, such as dense macular OCT scans, were missed by the
434 24-2 grid in 52% of the cases. Previous analysis on structural tests showed that many of the
435 missed bundle defects were within the central 4 degrees, not tested by the 24-2²². In our
436 dataset, the four testing points within 4 degrees from fixation in the 24-2+ identified at least
437 one damaged location in 56%, 39%, 35% and 28% of the subjects with GON for the TD-5%,
438 TD-2%, PD-5% and PD-2% maps respectively. This speaks to another important point of
439 discussion: similar detection power of central damage does not always correspond to
440 equivalent characterization of the defect itself. This could have important effects on the
441 definition of the spatial spread of glaucoma damage and on the accuracy of structure-function
442 relationship.

443

444 Improvement in detection seen in our results came at the cost of an estimated 21% average
445 increase in test time (**Table 1**). Whether this is clinically acceptable is outside the scope of the
446 questions we have asked in this study. Integration of structural metrics to guide the
447 acquisition of additional data from the macular region where damage is expected could be a
448 solution to this issue and could be the subject of future work. Such a strategy has proven
449 successful in reducing the test time and improving the overall efficiency of the test,^{13, 23-25} but
450 might also bias the estimates of VF sensitivity. This would not only reduce testing time but
451 also improve specificity. Another approach is to incorporate prior population based
452 knowledge about what locations are more likely to detect a macular defect. This is the
453 strategy chosen for the 24-2C grid implemented in the newest HFA perimeters⁸⁻¹¹, which
454 includes only few locations from the 10-2 grid. Our data indicate a spatial distribution of
455 macular defects in the GON cohort that largely reflects the expectation from previous
456 knowledge²². However, many abnormal locations were also distributed across the whole
457 central VF (**Figure 7**) and their detection would certainly be compromised by such an
458 approach. Additional locations might also be added dynamically based on the spatial features
459 of VF defects, providing customisable testing grids^{26, 27}.

460

461 In conclusion, VF tests with additional macular locations can modestly improve detection of
462 macular defects in glaucoma patients with minimal to no loss in specificity. However, the
463 overall difference between the two grids is small. A more tailored approach, possibly based on
464 structural evaluations, could help select people who are more likely to benefit from more
465 precise macular testing.

466 **References**

- 467 1. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in
468 early glaucoma. *Br J Ophthalmol* 1998;82(4):352-61.
- 469 2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a
470 review. *JAMA* 2014;311(18):1901-11.
- 471 3. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol*
472 1990;300(1):5-25.
- 473 4. De Moraes CG, Hood DC, Thenappan A, et al. 24-2 Visual Fields Miss Central Defects
474 Shown on 10-2 Tests in Glaucoma Suspects, Ocular Hypertensives, and Early Glaucoma.
475 *Ophthalmology* 2017;124(10):1449-56.
- 476 5. Grillo LM, Wang DL, Ramachandran R, et al. The 24-2 Visual Field Test Misses Central
477 Macular Damage Confirmed by the 10-2 Visual Field Test and Optical Coherence Tomography.
478 *Transl Vis Sci Technol* 2016;5(2):15.
- 479 6. Traynis I, De Moraes CG, Raza AS, et al. Prevalence and nature of early glaucomatous
480 defects in the central 10 degrees of the visual field. *JAMA Ophthalmol* 2014;132(3):291-7.
- 481 7. West ME, Sharpe GP, Hutchison DM, et al. Utility of 10-2 Visual Field Testing in
482 Glaucoma Patients with Early 24-2 Visual Field Loss. *Ophthalmology* 2020.
- 483 8. Lee GC, Monhart M, Callan T, et al. Performance of a modified 24-2 test pattern using
484 SITA Faster. *Investigative Ophthalmology & Visual Science* 2018;59(9):6032-.
- 485 9. Lee GC, Yu S, Callan T, et al. Diagnostic Efficacy of 24-2 and 24-2C SITA Faster Global
486 Summary Indices. *Investigative Ophthalmology & Visual Science* 2019;60(9):2455-.
- 487 10. Yu S, Lee GC, Callan T, et al. Comparison of SITA Faster 24-2C test times to legacy SITA
488 tests. *Investigative Ophthalmology & Visual Science* 2019;60(9):2454-.
- 489 11. Phu J, Kalloniatis M. Ability of 24-2C and 24-2 grids in identifying central visual field
490 defects and structure-function concordance in glaucoma and suspects. *Am J Ophthalmol* 2020.
- 491 12. Montesano G, Bryan SR, Crabb DP, et al. A Comparison between the Compass Fundus
492 Perimeter and the Humphrey Field Analyzer. *Ophthalmology* 2019;126(2):242-51.
- 493 13. Montesano G, Rossetti LM, Allegrini D, et al. Improving Visual Field Examination of the
494 Macula Using Structural Information. *Transl Vis Sci Technol* 2018;7(6):36.
- 495 14. Rossetti L, Digiuni M, Rosso A, et al. Compass: clinical evaluation of a new instrument
496 for the diagnosis of glaucoma. *PLoS One* 2015;10(3):e0122157.
- 497 15. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold
498 estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer
499 simulation. *Invest Ophthalmol Vis Sci* 2003;44(11):4787-95.
- 500 16. King-Smith PE, Grigsby SS, Vingrys AJ, et al. Efficient and unbiased modifications of the
501 QUEST threshold method: theory, simulations, experimental evaluation and practical
502 implementation. *Vision Res* 1994;34(7):885-912.
- 503 17. Sample PA, Dannheim F, Artes PH, et al. Imaging and Perimetry Society standards and
504 guidelines. *Optom Vis Sci* 2011;88(1):4-7.
- 505 18. Wu Z, Medeiros FA, Weinreb RN, Zangwill LM. Performance of the 10-2 and 24-2 Visual
506 Field Tests for Detecting Central Visual Field Abnormalities in Glaucoma. *Am J Ophthalmol*
507 2018;196:10-7.
- 508 19. Hodapp E, Parrish RK, Anderson DR. *Clinical decisions in glaucoma*: Mosby
509 Incorporated, 1993.
- 510 20. Matsuura M, Murata H, Fujino Y, et al. Evaluating the Usefulness of MP-3
511 Microperimetry in Glaucoma Patients. *Am J Ophthalmol* 2018;187:1-9.
- 512 21. Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated
513 with the gradient of visual sensitivity. *Vision Res* 2007;47(7):925-36.
- 514 22. Hood DC, Raza AS, de Moraes CG, et al. Glaucomatous damage of the macula. *Prog Retin*
515 *Eye Res* 2013;32:1-21.

- 516 23. Ballae Ganeshrao S, Turpin A, Denniss J, McKendrick AM. Enhancing Structure-
517 Function Correlations in Glaucoma with Customized Spatial Mapping. *Ophthalmology*
518 2015;122(8):1695-705.
- 519 24. Ballae Ganeshrao S, Turpin A, McKendrick AM. Sampling the Visual Field Based on
520 Individual Retinal Nerve Fiber Layer Thickness Profile. *Invest Ophthalmol Vis Sci*
521 2018;59(2):1066-74.
- 522 25. Ehrlich AC, Raza AS, Ritch R, Hood DC. Modifying the Conventional Visual Field Test
523 Pattern to Improve the Detection of Early Glaucomatous Defects in the Central 10 degrees.
524 *Transl Vis Sci Technol* 2014;3(6):6.
- 525 26. Turpin A, Morgan WH, McKendrick AM. Improving Spatial Resolution and Test Times
526 of Visual Field Testing Using ARREST. *Translational Vision Science & Technology*
527 2018;7(5):35-.
- 528 27. Muthusamy V, Turpin A, Walland MJ, et al. Increasing the Spatial Resolution of Visual
529 Field Tests Without Increasing Test Duration: An Evaluation of ARREST. *Translational Vision*
530 *Science & Technology* 2020;9(13):24-.
- 531

532 Figure legends

533 **Figure 1.** Schematic depiction of the 24-2 (red circles) and the 24-2+ (black dots), excluding the blind spot
534 locations. The additional macular locations were symmetric for all four quadrants. The coordinates for Quadrant
535 2 (X; Y) were [(1.5; 1.5), (1.5; 6), (6; 6)] degrees. The area within 10 degrees from fixation was used in this
536 analysis and is highlighted. On the right, neighbourhood relationships for the central region of the two grids (see
537 Methods).

538
539 **Figure 2.** Changes in specificity according to different threshold criteria. The criterion represents the minimum
540 number of locations outside the normative limits (at either 5% or 2% probability) used to define a macular
541 defect for the TD and PD maps. In the overall selection (left) the central field is considered as a whole (maximum
542 N is 12 for the 24-2 and 24 for the 24-2+). In the cluster selection (right) locations are pooled only if spatially
543 contiguous and the size of the largest cluster is then used (maximum N is 6 for the 24-2 and 12 for the 24-2+).
544 Black circles highlight the matched-specificity criteria selected for this analysis. These calculations were
545 performed with data from the healthy cohort (440 eyes), using a leave-one-out approach. FPR = False Positive
546 Rate.

547
548 **Figure 3.** Curves representing the detection rate at different specificity levels for the different maps and criteria
549 applied to the 24-2 and 24-2+ grids. The operating points for the matched specificity comparisons are marked by
550 a target.

551
552 **Figure 4.** Venn diagrams representing the agreement between 24-2 and 24-2+ in detecting macular defects with
553 different detection strategies. The area of the squares is proportional to the percentage of eyes for which a
554 macular defect was detected. The overlapping area represents the percentage of defects detected by both grids.
555 Percentages (number of eyes) are reported separately for each grid in colour-coded labels and in black for
556 simultaneous detections.

557
558 **Figure 5.** Panel (A) shows an example of a macular defect identified by the 24-2+ grid (outlined in red) but
559 missed by the 24-2 according to the *cluster detection*. Notice how the whole 24-2 grid would still detect a
560 peripheral defect (superior-temporal) and correctly identify this as a glaucoma case without detecting the
561 central damage. Panel (B) shows an opposite example, where the 24-2+ grid misses a defect identified by the 24-
562 2 because the cluster is interrupted by a healthy location. Both examples use the criteria for the TD-5% map. The
563 criterion refers to the minimum size of the largest cluster considered to identify a defect. The central locations
564 considered for the analysis are enclosed by the blue dashed outline. These maps were built using only the $P < 5\%$
565 and $P < 2\%$ symbols, reflecting the two maps used for this analysis.

566
567 **Figure 6.** Histograms of the distribution of the depth of the defect for all the abnormal locations missed by one
568 grid but detected by the other, i.e. the cases in which the 24-2 and the 24-2+ disagreed. The distribution of the
569 deep defects ($TD \leq -20$ dB) is outlined in black. The text-boxes report the total number of missed deep
570 defects/the total number of missed affected locations. TD = Total Deviation; PD = Pattern Deviation.

571
572 **Figure 7.** Spatial distribution of locations outside the 5% and 2% normative limits in the cohort of patients with
573 GON. The size of the circles is proportional to the frequency of values outside the normative limits for each tested
574 location. Estimated probability of finding abnormal values [95% CIs] are also reported for each quadrant, as
575 estimated by logistic regression. GON = Glaucomatous Optic Neuropathy; Q = Quadrant.