Optic disc classification by the Heidelberg Retina Tomograph and by physicians with varying experience of glaucoma.

ANDERSSON GEIMER, SABINA; Heijl, Anders; Bengtsson, Boel

Published in:
Eye (London, England)

DOI:
10.1038/eye.2011.172

2011

Citation for published version (APA):
Optic disc classification by the Heidelberg Retina Tomograph and by physicians with varying experience of glaucoma

Abstract

**Purpose** To compare the diagnostic accuracy of the Heidelberg Retina Tomograph’s (HRT) Moorfields regression analysis (MRA) and glaucoma probability score (GPS) with that of subjective grading of optic disc photographs performed by ophthalmologists with varying experience of glaucoma and by ophthalmology residents.

**Methods** Digitized disc photographs and HRT images from 97 glaucoma patients with visual field defects and 138 healthy individuals were classified as either within normal limits (WNL), borderline (BL), or outside normal limits (ONL). Sensitivity and specificity were compared for MRA, GPS, and the physicians. Analyses were also made according to disc size and for advanced visual field loss.

**Results** Forty-five physicians participated. When BL results were regarded as normal, sensitivity was significantly higher ($P < 5\%$) for both MRA and GPS compared with the average physician, 87%, 79%, and 62%, respectively. Specificity ranged from 86% for MRA to 97% for general ophthalmologists, but the differences were not significant. In eyes with small discs, sensitivity was 75% for MRA, 60% for the average doctor, and 25% for GPS; in eyes with large discs, sensitivity was 100% for both GPS and MRA, but only 68% for physicians.

**Conclusion** Our results suggest that sensitivity of MRA is superior to that of the average physician, but not that of glaucoma experts. MRA correctly classified all eyes with advanced glaucoma and showed the best sensitivity in eyes with small optic discs.

**Keywords:** glaucoma; optic disc; diagnostic accuracy; Moorfields regression analysis; glaucoma probability score; subjective assessment

Introduction

Assessment of the optic disc is included in the standard examination of patients with ocular hypertension or suspected or manifest glaucoma. Such evaluation is performed not only by glaucoma experts, but also by general ophthalmologists, ophthalmology residents, and ophthalmologists with special skills in areas other than glaucoma. Most studies of the diagnostic accuracy of subjective assessment of disc photographs have compared the abilities of glaucoma experts in that context and the results obtained have shown that even very experienced observers can find it difficult to discriminate between healthy and glaucomatous discs.

Various computerized quantitative imaging techniques have been developed to help doctors identify structural glaucomatous damage. Confocal scanning laser tomography using the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, GmbH, Heidelberg, Germany) was introduced in the beginning of the 1990s and has been further developed since then. Several investigators have examined the diagnostic performance of the HRT in comparison with subjective assessment, but their results have differed somewhat. In short,
some studies have shown similar performance of the HRT and subjective assessors,\textsuperscript{2,4} whereas others have indicated that either HRT classification\textsuperscript{5–8} or subjective assessment\textsuperscript{10–13} is superior. Thus, it is not entirely clear to what extent HRT measurements can replace subjective assessment in glaucoma practice.

The latest version of the Heidelberg instrument, HRT3, includes classification by both Moorfields regression analysis (MRA)\textsuperscript{10} and the glaucoma probability score (GPS).\textsuperscript{11} The results of research comparing MRA and GPS\textsuperscript{7,12–24} are to some extent conflicting, particularly regarding sensitivity. Of all published studies, about 30\% showed significantly or only slightly better sensitivity with MRA, around 50\% indicated better sensitivity with GPS, and the remaining 20\% demonstrated similar sensitivity for both methods. Considering specificity, a majority of the investigators showed that MRA was superior to GPS.

Thus, a number of studies have evaluated the diagnostic performance of the HRT, and some have compared HRT examination with subjective assessment of disc photographs conducted by glaucoma experts or clinicians with special interest in glaucoma, whereas few have compared the HRT with non-experts. In an investigation by Reus et al,\textsuperscript{4} the diagnostic results obtained by a limited number of graders with varying experience of glaucoma were compared with the corresponding results acquired using several other techniques, among them MRA. They found that the performance of MRA was similar to that of specialists and general ophthalmologists, but better than that of ophthalmology residents.

The purpose of the current study was to compare the diagnostic accuracy of HRT, GPS and MRA with that of assessment conducted by physicians with different experience of glaucoma.

**Subjects and methods**

Ophthalmologists and residents in ophthalmology were asked to grade disc photographs. Those who agreed to participate were asked to report their level of clinical experience by classifying themselves as glaucoma expert, general ophthalmologist, other subspecialist or ophthalmology resident, and they were subsequently divided into subgroups accordingly.

The research followed the tenets of the Declaration of Helsinki and was approved by the Regional Ethical Review Board in Lund Sweden, vetting the ethics of research involving humans.

**Subjects**

All disc photographs and HRT images were retrieved from an existing database of healthy and glaucoma subjects. The database has been described in detail elsewhere,\textsuperscript{25} and is here only briefly described.

The glaucoma subjects were patients managed at the Department of Ophthalmology, Malmö University Hospital, Malmö, Sweden. All had a confirmed diagnosis of glaucoma with reproducible glaucomatous visual field defects with standard automated perimetry conducted using the 30-2 full threshold program of the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA, USA). Patients had at least two consecutive visual field examinations classified as outside normal limits (ONL) by the glaucoma hemifield test with depressed points appearing in the same area of the visual field.\textsuperscript{25} Glaucma hemifield test has been validated as a reliable diagnostic visual field interpretation tool.\textsuperscript{27,28} Ninety-six glaucoma subjects met the definition of glaucoma by analysis of cluster with depressed visual field points as described by Katz et al\textsuperscript{27} and Anderson and Patella,\textsuperscript{29} there was only one subject with paracentral visual defect who did not meet this criteria. Photographs with obvious artifacts (eg, prominent reflections or the shutter half way down) were excluded. One eye per patient was selected. In patients with both eyes eligible (ie, with a diagnosis of glaucoma with reproducible visual field defects), the eye deemed best by the perimetric mean deviation (MD) value was selected. A total of 97 disc photographs and HRT images from 97 glaucoma patients were included. The mean age of the patients was 71 years (range 49–87 years). The average MD was −7.2 dB, (range −23.21 to 2.14 dB).

Healthy subjects were randomly selected among presumably healthy persons living in Malmö, Sweden.\textsuperscript{30} They all underwent a thorough ophthalmic examination including HRT imaging and disc photography. Inclusion and exclusion criteria were as follows: corrected visual acuity better than 0.8, intraocular pressure below 20 mm Hg, no history of serious eye trauma or eye surgery except uncomplicated cataract surgery, and no previous or current serious eye disease or neurological disease. As for the glaucoma subjects, all photographs with suboptimal quality or obvious artifacts were excluded. For the purpose of this study, healthy subjects younger than 50 years of age at the time of the data collection were excluded in order to better match the age of the glaucoma patients. Photographs and images from 138 healthy subjects were included. The mean age of the healthy subjects was 66 years (range 51–79 years).

**Photographs**

All disc photographs had been taken by the same experienced technician, using a Carl Zeiss fundus camera (Model 60306, Oberkochen, West Germany) with standard settings (aperture 5.5, flash strength
The HRT images were then assessed by the MRA 10 and GPS.11 Sensitivity and specificity were also calculated for eyes with advanced glaucoma, defined as an MD worse than −18 dB. The Marascuilo procedure was performed using Microsoft Excel for Macintosh (version 12.2.6, Microsoft Corp.), and descriptive statistics were derived using SPSS for Macintosh (version 16.0.0, SPSS Inc., Chicago, IL, USA).

Results

The grading of all disc photographs was completed by 45 physicians, who reported themselves to be the following: 10 glaucoma experts, 13 general ophthalmologists, 11 other subspecialists, and 11 ophthalmology residents. Regarding their knowledge and skills related to glaucoma, almost half of the 45 physicians (44%) indicated that they were experienced, 80% of the experts considered themselves to be very experienced and 10 of the 11 residents felt they were less experienced.

MRA was compatible with all optic discs from healthy individuals and glaucoma patients, and thus able to classify all 235 images. GPS was compatible with all the glaucoma eyes, but eight healthy eyes were nevertheless incompatible with the GPS database, and thus not classified. Those eight were removed from the denominator in the specificity calculation for GPS. The same eight healthy discs were all correctly classified as ‘WNL’ by the MRA. The relative proportion of optic discs classified as BL was 13% with MRA and 17% with GPS. By comparison, the physicians classified an average of 17% of the disc photographs as uncertain.

When the more specific classification approach was used, both MRA and GPS showed significantly better sensitivity ($P < 0.05$) than the average physician, when the more sensitive approach was applied, only MRA yielded significantly better sensitivity ($P < 0.05$; Table 1). None of the HRT methods yielded better sensitivity compared with the glaucoma experts. With the more sensitive approach specificity was slightly, but not significantly, better for the average physician as compared with the HRT methods (Table 1). The average optic disc size was larger in the glaucoma patients than in the healthy subjects: 2.25 and 1.96 mm$^2$, respectively. Large discs were observed in 28% of the glaucoma patients and 6% of the healthy subjects, and the corresponding proportions of small discs were 8 and 17%. GPS and MRA offered perfect sensitivity (100%) in eyes with large discs, as determined by both the more sensitive and the more specific approach. Using the more specific approach indicated sensitivities of 64 and 66% for assessments of eyes with large discs by the average physician and the experts, respectively. The specificity

Heidelberg Retina Tomograph

HRT images (Heidelberg Engineering, software 1.11, standard reference plane) were obtained within ±6 months of the disc photographs, and they were all of good quality (pixel standard deviation $\leq 40 \mu m$).31 To be able to evaluate the HRT images by use of MRA and GPS, the HRT data were manually retrieved from archive discs and upgraded to the newer software (software 3.1, Heidelberg Engineering), and new topographies were calculated. All optic disc margins were outlined by one of the authors (SA) with the help of the disc photographs; this procedure has been reported to improve the definition of the margins.32 The HRT images were then assessed by the MRA$^{10}$ and GPS.11

A total of 235 HRT images and disc photographs were graded. The overall MRA and GPS results classify images into one of the three categories: within normal limits (WNL), borderline (BL), or ONL. In a similar manner, the physicians classified disc photographs as healthy, uncertain, or glaucomatous.

Analyses

Sensitivity and specificity of the classification performed by the physicians were calculated in two ways: by a more specific approach treating ‘uncertain’ as healthy and a more sensitive approach treating ‘uncertain’ as glaucomatous. The same approaches were applied to compute the sensitivity and specificity of MRA and GPS. Sensitivity and specificity were determined for the overall average physicians and for each subgroup. MRA and GPS results were compared with the average for all physicians and with the subgroup averages using the Marascuilo procedure for multiple proportions.33 The overall level of significance was set to 0.05 and this was used for all calculations with the Marascuilo procedure. Statistical comparisons of the subgroups were not done. Sensitivity and specificity were also calculated for eyes with different disc sizes according to HRT measurements31 (ie, small < 1.6 mm$^2$, medium 1.6–2.5 mm$^2$, or large > 2.5 mm$^2$), but no comparisons were performed because of low statistical power. Sensitivity was also calculated separately for eyes with advanced glaucoma, defined as an MD worse than −18 dB. The Marascuilo procedure was performed using Microsoft Excel for Macintosh (version 12.2.6, Microsoft Corp.), and descriptive statistics were derived using SPSS for Macintosh (version 16.0.0, SPSS Inc., Chicago, IL, USA).
related to large discs was low 38% with MRA and 50% with GPS, even when the more specific approach was used. Considering eyes with small discs, the more sensitive approach showed better sensitivity for MRA than for GPS (88% and 50%, respectively), and somewhat better sensitivity for the glaucoma experts compared with the average physician (85% and 70%, respectively; Figure 1).

Discussion

Overall, sensitivity was higher for both MRA and GPS compared with the average physician, but not all differences reached statistically significance (Table 1).

Also, considering the subgroups of graders, there was a tendency towards the best sensitivity being achieved by the experts and the poorest by those designated other subspecialists (Table 1). Regarding specificity, we found that the physicians tended to be better than both the MRA and GPS, but only with the more sensitive approach, which could be expected since higher sensitivities generally are accompanied by lower specificity and the opposite. The general ophthalmologists tended to be better than the other subgroups, residents were often associated with the lowest specificity, and glaucoma experts were only slightly better than residents. The differences concerning specificity and sensitivity among the subgroups were not tested for significance because the number of graders was rather limited.

When the more specific approach was applied, the specificity ranged from 86% with MRA to 94% with GPS; with the more sensitive approach, it ranged from 69% for MRA to 79% for other subspecialists (Table 1). Several studies have demonstrated better specificity with MRA than with GPS.16,17,19,23,24 We observed a similar trend in our results, but the fact that GPS was unable to classify images from eight healthy subjects might have contributed to better specificity values for GPS.

It has previously been reported that disc size affects the diagnostic accuracy of subjective assessment,34 as well as MRA and particularly GPS classifications.12,19,21–23 Disc size also proved to be an important factor influencing diagnostic accuracy in our study. The sensitivity of GPS was only 25% in patients with small discs when the specific approach was applied, and the value increased to 50% when the more sensitive approach was used. The corresponding figures for MRA were 75% and 88%, respectively. For the glaucoma experts, the mean sensitivity was 93% in eyes with advanced disease.

| Table 1 | Sensitivity and specificity of the Heidelberg Retinal Tomograph (HRT) algorithms and subjective optic discs classification by physicians |
|-----------------|-------------------------------------------------|-----------------|-----------------|-----------------|
|                | Borderline/uncertain analysed as healthy         | Borderline/uncertain analysed as glaucoma |
|                | Sensitivity | Specificity | Sensitivity | Specificity |
| HRT Moorfields regression analysis | 87%abc | 86% | 94%a | 69% |
| HRT Glaucoma probability score | 79%c | 94%d | 93% | 72%d |
| All physicians (n = 45) | 62% | 92% | 82% | 76% |
| Mean (min-max) | 4–97 | 63–100 | 58–99 | 32–99 |
| Glaucan experts (n = 10) | 72% | 90% | 88% | 75% |
| Mean (min-max) | 45–97 | 63–100 | 70–99 | 32–96 |
| General ophthalmologists (n = 13) | 59% | 97% | 83% | 80% |
| Mean (min-max) | 24–87 | 84–100 | 59–97 | 57–99 |
| Other subspecialists (n = 11) | 53% | 93% | 77% | 79% |
| Mean (min-max) | 4–81 | 79–100 | 58–95 | 58–99 |
| Residents (n = 11) | 66% | 87% | 81% | 71% |
| Mean (min-max) | 35–86 | 77–96 | 58–96 | 49–94 |

*aSignificant difference (P < 0.05) when compared with all physicians.
*bSignificant difference (P < 0.05) when compared with general ophthalmologists.
*cSignificant difference (P < 0.05) when compared with other subspecialists.
*dEight optic discs in healthy subjects were incompatible with the GPS database and were therefore excluded from the calculation of specificity.
Optic disc appearance was not used as selection criterion of either healthy individuals or glaucoma subjects to avoid any bias affecting the subjective optic disc assessment.

By MRA alone, 100% of the discs in eyes with advanced visual field defects were classified as being ONL. GPS correctly classified 67%, glaucoma experts 93%, and all physicians 84%. The cutoff at MD worse than \(-18\) dB for advanced glaucoma was arbitrary set with the purpose of analysing the diagnostic accuracy for those with more advanced glaucomatous damage, since it is a great disadvantage to misclassify these subjects regardless if assessment is made in a screening setting or a clinical environment. If instead using the criteria for advanced glaucoma as defined by Hodapp, that is, MD worse than \(-12\) dB, MRA correctly classifies 84% as ONL, GPS 74%, glaucoma experts 83% and all physicians 72%. Using a definition of MD worse than \(-15\) dB for advanced glaucoma gives results of 80%, 60%, 82% and 67%, respectively. Regardless of cutoff criterion chosen for MD, our results show that MRA classifies more glaucoma subjects correctly. Reddy et al.\(^1\) have previously found that MRA provided 89% sensitivity in eyes with advanced glaucomatous visual field loss, when eyes with MD values worse than \(-15\) dB were included in the assessment.

In a smaller study conducted by Reus et al., the diagnostic accuracy of MRA was compared with that of classification performed by subjective graders. Four graders in each of four categories (glaucoma experts, general ophthalmologists, ophthalmology residents, and optometrists) assessed disc photographs from 40 healthy and 48 glaucoma subjects. The results showed that glaucoma specialists and general ophthalmologists performed just as well as MRA, and the residents were not as successful as the other grader subgroups. Our investigation was larger, including 235 eyes and slightly more graders per subgroup, and we found that only glaucoma experts achieved sensitivities comparable to that of MRA. Thus, the performance of MRA in our study was impressive in many respects: it showed better sensitivity than most graders, except for glaucoma experts; it was at least as efficient as glaucoma experts in analysis of small discs; it was the only method that could classify all eyes with severe field defects as being ONL. Although none of the diagnostic methods we investigated was perfect, we conclude that HRT MRA can perform at least as well as the best clinicians. The fact that no eyes with advanced glaucoma were missed by MRA indicates an extra advantage of this method, which, for example, would be particularly beneficial in glaucoma population screening by use of an imaging device. However, on a less encouraging note, the diagnostic approaches we studied provided relatively poor specificities in eyes with large discs and poor sensitivities in eyes with small discs. The method of diagnosing glaucoma by imaging of the optic disc is usually, but not always, correct; and of course, this conclusion applies to other diagnostic techniques as well.

![Graph showing sensitivity and false positives (1-specificity) obtained with MRA, the GPS, and subjective classification by the average physician.](image)

**Figure 1** Sensitivity and false positives (1-specificity) obtained with MRA, the GPS, and subjective classification by the average physician. The letters in the coloured circles represent disc size: 'L', large; 'M', medium; 'S', small. (a) Analysis using the more specific approach considering 'BL' as healthy. GPS had 100% sensitivity and low specificity in eyes with large discs, and low sensitivity but perfect specificity in eyes with small discs. For MRA, sensitivity was 100% in eye with large discs, and 75% in small discs; corresponding values for assessment by the average physician were 68% and 60%, respectively. In large discs, specificity was low for both MRA and GPS. (b) Using the more sensitive approach considering 'BL' as glaucoma. GPS had high sensitivity in large discs, but only 50% in eyes with small discs. MRA had similar sensitivity in large discs, but better sensitivity in small discs, as compared with GPS. The average doctor reached 86% sensitivity in large and 77% in small discs. Specificity was low for both MRA (25%) and GPS (38%) in large discs. * Eight optic discs in healthy subjects were incompatible with the GPS database (four small and four medium sized discs) and were thus excluded from the calculation of specificity.
Summary

What was known before
- Results of research regarding optic disc classification for glaucoma with the Heidelberg Retina Tomograph (HRT), as compared with subjective assessment, are to some extent conflicting. Some have shown similar performance, some have presented that either HRT or subjective assessment is superior. Few studies have compared the latest HRT instrument, which includes Moorfields regression analysis (MRA) and glaucoma probability score (GPS), with subjective assessments. There is a lack of knowledge of how physicians with different experiences in glaucoma perform compared with MRA and GPS.

What this study adds
- We conclude that the sensitivity of MRA is superior to that of the average ophthalmologist, but not to the performance of glaucoma experts. MRA also correctly classified all eyes with advanced glaucoma, and was comparable to glaucoma experts in assessing small optic discs. GPS showed a disadvantage since some of the optic discs from healthy subjects were incompatible with the HRT database.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to all the participating physicians for their interest and collaboration in this project. This study was supported by the Swedish Research Council grant K2005-74X-10426-13A, the Herman Järnhardt Foundation, the Foundation for Visually Impaired in the Former Malmöhus County, and Crown Princess Margareta’s Foundation for the Visually Impaired.

References


21 Yip LW, Mikelberg FS. A comparison of the glaucoma probability score to earlier Heidelberg retina tomograph data analysis tools in classifying normal and glaucoma patients. *J Glaucoma* 2008; **17**: 513–516.


34 Heijl A, Molder H. Optic disc diameter influences the ability to detect glaucomatous disc damage. *Acta Ophthalmol (Copenh)* 1993; **71**: 122–129.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/