Can Automated Imaging for Optic Disc and Retinal Nerve Fiber Layer Analysis Aid Glaucoma Detection?

Katie Banister, PhD,[1] Charles Boachie, BSc (Hons),[2] Rupert Bourne, FRCOphth,[3] Jonathan Cook, PhD,[4] Jennifer M. Burr, MD,[5] Craig Ramsay, PhD,[1] David Garway-Heath, MD,[6] Joanne Gray, MSc,[7] Peter McMeekin, PhD,[7] Rodolfo Hernández, MSc,[8] Augusto Azuara-Blanco, FRCOphth[9]

Purpose: To compare the diagnostic performance of automated imaging for glaucoma.

Design: Prospective, direct comparison study.

Participants: Adults with suspected glaucoma or ocular hypertension referred to hospital eye services in the United Kingdom.

Methods: We evaluated 4 automated imaging test algorithms: the Heidelberg Retinal Tomography (HRT; Heidelberg Engineering, Heidelberg, Germany) glaucoma probability score (GPS), the HRT Moorfields regression analysis (MRA), scanning laser polarimetry (GDx enhanced corneal compensation; Glaucoma Diagnostics (GDx), Carl Zeiss Meditec, Dublin, CA) nerve fiber indicator (NFI), and Spectralis optical coherence tomography (OCT; Heidelberg Engineering) retinal nerve fiber layer (RNFL) classification. We defined abnormal tests as an automated classification of outside normal limits for HRT and OCT or NFI $\geq$ 56 (GDx). We conducted a sensitivity analysis, using borderline abnormal image classifications. The reference standard was clinical diagnosis by a masked glaucoma expert including standardized clinical assessment and automated perimetry. We analyzed 1 eye per patient (the one with more advanced disease). We also evaluated the performance according to severity and using a combination of 2 technologies.

Main Outcome Measures: Sensitivity and specificity, likelihood ratios, diagnostic, odds ratio, and proportion of indeterminate tests.

Results: We recruited 955 participants, and 943 were included in the analysis. The average age was 60.5 years (standard deviation, 13.8 years); 51.1% were women. Glaucoma was diagnosed in at least 1 eye in 16.8%; 32% of participants had no glaucoma-related findings. The HRT MRA had the highest sensitivity (87.0%; 95% confidence interval [CI], 80.2%−92.1%), but lowest specificity (63.9%; 95% CI, 60.2%−67.4%); GDx had the lowest sensitivity (35.1%; 95% CI, 27.0%−43.8%), but the highest specificity (97.2%; 95% CI, 95.6%−98.3%). The HRT GPS sensitivity was 81.5% (95% CI, 73.9%−87.6%), and specificity was 67.7% (95% CI, 64.2%−71.2%); OCT sensitivity was 76.9% (95% CI, 69.2%−83.4%), and specificity was 78.5% (95% CI, 75.4%−81.4%). Including only eyes with severe glaucoma, sensitivity increased: HRT MRA, HRT GPS, and OCT would miss 5% of eyes, and GDx would miss 21% of eyes. A combination of 2 different tests did not improve the accuracy substantially.

Conclusions: Automated imaging technologies can aid clinicians in diagnosing glaucoma, but may not replace current strategies because they can miss some cases of severe glaucoma. *Ophthalmology 2016;123:930-938 Crown Copyright © 2016 Published by ELSEVIER Inc. on behalf of American Academy of Ophthalmology. This is an open access article under the Open Government Licence (OGL) (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/).*

Diagnosis of glaucoma by an experienced ophthalmologist remains the best reference standard.[1] However, diagnosis can be challenging, especially in people with early glaucoma. Accurate clinical diagnosis of glaucoma is limited by subjectivity, reliance on the examiner's experience, and a wide variation of optic disc structure among the population.[1,2]

Automated imaging of the optic nerve head or retinal nerve fiber layer (RNFL) increasingly is being introduced into practice for diagnosis and monitoring.[3] Interpretation of some of the outputs may require expertise, but classification of results as normal or abnormal also can be generated by automatic comparison with a normative database.

Several imaging technologies that quantify the structure of the retina and optic nerve head can be used in glaucoma.[4] Confocal scanning laser ophthalmoscopy is commercially available as Heidelberg retina tomograph (HRT; Heidelberg Retina Tomograph III [Heidelberg Engineering, Heidelberg, Germany]). It includes 2 classification algorithms, the

Moorfields regression analysis (MRA)[5] and the glaucoma probability score (GPS).[6,7] The RNFL can be assessed using either scanning laser polarimetry, currently available as the GDx-PRO (Glaucoma Diagnostics [GDx] Carl Zeiss Meditec, Dublin, CA)[8] or spectral-domain optical coherence tomography (OCT), with several commercial devices available.[9] These imaging tests are user friendly and provide automated quantitative classifications.[10]

Although many published data describe the diagnostic performance of imaging techniques in cohorts of retrospectively selected glaucoma patients or glaucoma-free normal subjects, there is no high-quality evidence of the comparative accuracy of current imaging techniques for identifying glaucoma in consecutive patients with unknown status screened for possible glaucoma.[4,11] Existing data from case-control studies may not be applicable to the clinically relevant population who undergo assessment and diagnosis.[12] We aimed to assess and compare the performance of these commercially available technologies to detect glaucoma in a prospective cohort. This work was conducted as part of a wider publicly funded study (National Institute for Health Research Health Technology Assessment [HTA], 09/22/111) that also evaluated cost-effectiveness of these imaging technologies in a triage setting in the United Kingdom.

Methods

Study Design and Participants

We conducted a pragmatic multicenter, within-patient, comparative evaluation of the diagnostic accuracy of automated imaging techniques for diagnosis of glaucoma, the Glaucoma Automated Tests Evaluation (GATE). We selected participants prospectively, and they underwent imaging with all technologies under evaluation and then had the reference standard diagnosis (clinical assessment by a glaucoma expert, including examination of the fundus by biomicroscopy and visual field testing with Humphrey 24-2 Swedish interactive threshold algorithm testing, masked to the imaging test results). The study was approved by the North of Scotland Research Ethics Committee (reference, 10/S0801/58) and was conducted according to the tenets of the Declaration of Helsinki. The full study protocol is publicly available.[13] We sought patient views on the design, conduct, and analysis of the study through representatives from the International Glaucoma Society.

The study was coordinated from a central study office in the Health Services Research Unit, University of Aberdeen, and was conducted in 5 National Health Service hospital eye services in the United Kingdom (Aberdeen, Bedford, Hinchingbrooke, Liverpool, and Moorfields). We identified eligible patients from their referral letter as being adults (age ≥18 years) who were newly referred from primary care to the department of ophthalmology of the recruiting hospital with a possible glaucoma diagnosis or glaucoma-related finding. This included high intraocular pressure; possible abnormalities in the optic disc, visual field test results, or both; and possible narrow anterior chamber angle. Patients were ineligible if they had a previous diagnosis of glaucoma or had already been seen by an ophthalmologist.

Participant Recruitment Process

We sent information about the study to potential eligible patients at each recruiting hospital, before their first hospital appointment. At their first clinic appointment, we then approached patients, and those patients who agreed to participate and signed the consent

Table 1. Possible Diagnoses by the Clinician Performing the Reference Standard Measurement, Ranked in Order of Severity

Diagnosis	Definition
Glaucoma	
Severe	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss† with MD of −12.01 dB or worse
Moderate	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss† with MD between −6.01 dB and −12 dB
Mild	Evidence of glaucomatous optic neuropathy* and a characteristic visual field loss† with MD of −6 dB or better
Glaucoma suspect	
Disc suspect	Appearance suggestive of glaucomatous optic neuropathy, but also may represent a variation of normality, with normal visual fields (with or without high IOP)
Visual field suspect	Visual field loss suggestive of glaucoma, but also may represent a variation of normality, with normal appearance of the optic disc (with or without high IOP)
Visual field and disc suspect	Both the optic disc and visual field have some features that resemble glaucoma, but also may represent a variation of normal (with or without high IOP)
Ocular hypertension	Both the visual field and optic nerve appear normal in the presence of elevated pressure >21 mmHg
Primary angle closure	Closed anterior chamber angle (appositionally or synechial) in at least 270° and at least 1 of the following 2: IOP >21 mmHg and presence of peripheral anterior synechiae; both visual field and optic nerve appear normal
Primary angle-closure suspect	Closed anterior chamber angle (appositionally without any synechiae) in at least 270°, with IOP ≤21 mmHg; both visual field and optic nerve appear normal
No glaucoma-related findings	Absence of any of the above diagnoses

IOP = intraocular pressure; MD = mean deviation.

*Any of the following: optic disc or retinal nerve fiber layer structural abnormalities; diffuse thinning, focal narrowing, or notching of the optic disc rim, especially at the inferior or superior poles; documented, progressive thinning of the neuroretinal rim with an associated increase in cupping of the optic disc; diffuse or localized abnormalities of the peripapillary retinal nerve fiber layer, especially at the inferior or superior poles; disc rim or peripapillary retinal nerve fiber layer hemorrhages; and optic disc neural rim asymmetry of the 2 eyes consistent with loss of neural tissue.

†Reliable visual field abnormality considered a valid representation of the subject's functional status. Visual field damage consistent with retinal nerve fiber layer damage (e.g., nasal step, arcuate field defect, or paracentral depression in clusters of test sites). Visual field loss in 1 hemifield that is different from the other hemifield, that is, across the horizontal midline (in early or moderate cases). Absence of other known explanations.
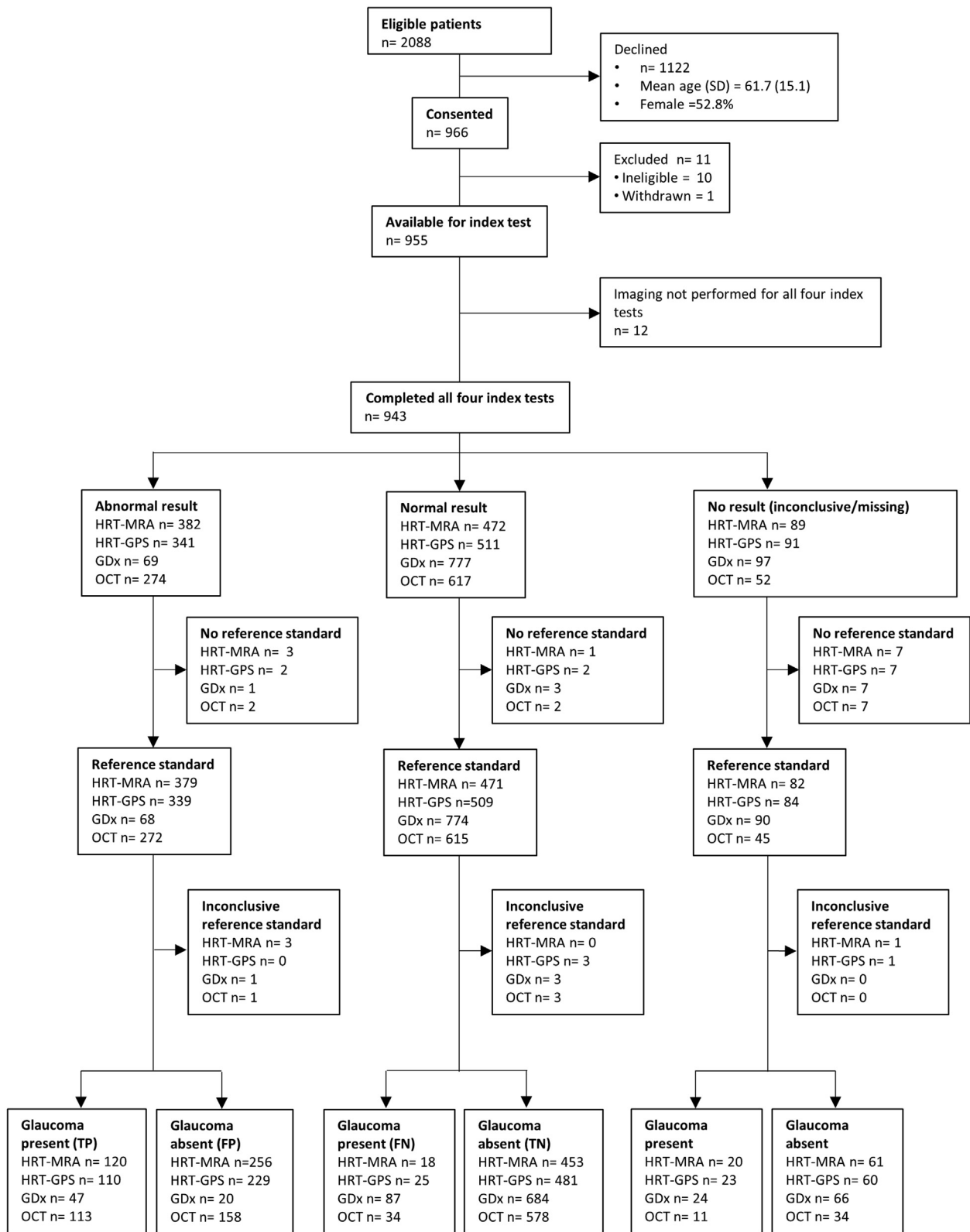
Figure 1. Flow diagram showing diagnosis of the cohort. FN = false negative; FP = false positive; GPS = glaucoma probability score; HRT = Heidelberg retinal tomography; MRA = Moorfields regression analysis; OCT = optical coherence tomography; TN = true negative; TP = true positive.

form were enrolled. We recorded the demographics (age, gender) of those patients who declined to take part.

Testing Regimen

Before seeing the ophthalmologist, participants were imaged using all imaging tests in a random order followed by visual field measurement with standard automated perimetry with the Humphrey 24-2 Swedish interactive threshold algorithm strategy. An ophthalmologist with glaucoma expertise (reference standard) who was masked to the imaging results then assessed participants. We conducted all tests on the same day in 2 centers, and in 3 centers, the clinician assessment was on a separate date, within 2 weeks.

Index Tests

The following imaging technologies were compared.

Heidelberg Retina Tomograph III. A confocal laser scanning ophthalmoscope (Heidelberg Retina Tomograph III; Heidelberg Engineering) measures quantitative structural information of the optic disc anatomy and creates a topographic map of the retinal surface. Images are given a quality index (the mean topography standard deviation), which the manufacturer recommends should be less than 40 μm. We assessed 2 classification tools: the MRA and the GPS. Both methods produced an overall classification of within normal limits, borderline, or outside normal limits by comparison with normative data. Abnormal index test results for MRA and GPS were defined as an overall classification of outside normal limits.

Scanning Laser Polarimetry. Scanning laser polarimetry (the GDx with enhanced corneal compensation or the GDx PRO; Carl Zeiss Meditec) analyzes polarized light reflected from the fundus to measure the RNFL thickness. Images are given a quality index Q, which the manufacturer recommends should be 7 or more. The software generated an automated discriminating classifier of glaucoma, the nerve fiber indicator (NFI) value. The manufacturers report that 95% of the normal population has an NFI value of 35 or less and that 99% of the normal population has an NFI value of 55 or less.[14] An abnormal index test result for the GDx was taken as an NFI value of 56 or more. An NFI value between 36 and 55 was considered to be similar to the borderline category of the other imaging tests. Measurements were obtained using either the GDx PRO or GDx variable corneal compensation (VCC) with an updated enhanced corneal compensation module.

Optical Coherence Tomography. Spectral-domain OCT was performed using the Spectralis (Heidelberg Engineering), which produces a detailed cross-sectional image of the retina using reflected light in a similar manner to B-mode ultrasound. Images are given a quality index Q, which the manufacturer recommends should be more than 15. The software (RNFL software version 5) automatically compared an average RNFL thickness with a normative database and produced an overall classification of within normal limits, borderline, or outside normal limits. Abnormal index test results for OCT were taken as a classification of outside normal limits.

We classed images as good quality if they met the manufacturer quality threshold as described above. Pupils were not dilated routinely unless image quality was inadequate. Between 1 and 3 experienced technicians performed the imaging tests at each center. We did not have any restriction on the same technician performing all imaging tests on an individual. The imaging technologies automatically generated a classification measurement without user input, except the HRT MRA, which required manual identification of the margin of the optic disc.

Reference Standard

An ophthalmologist with glaucoma expertise performed a clinical assessment that formed the reference standard (Table 1). All glaucoma experts participating in the study were trained in the procedures and used a preagreed definition of glaucoma (Table 1). Clinical examination included Goldmann applanation tonometry, gonioscopy, biomicroscopic examination of the optic disc (pupils dilated unless contraindicated), and evaluation of the visual field test with Humphrey Swedish interactive threshold algorithm 24-2. This represents currently recommended clinical practice in the United Kingdom.[15] Imaging test results were not available to the ophthalmologist. If a clinical diagnosis could not be established, (e.g., unreliable visual field measurement), an inconclusive diagnosis was recorded. To ensure a valid and consistent application of the agreed reference standard, a limited number of consultant ophthalmologists provided the assessment (n = 11) according to the diagnosis chart shown in Table 1.

Outcomes

Primary outcomes were sensitivity and specificity. Secondary outcomes were diagnostic odds ratio, likelihood ratio, and proportion of indeterminate test results.

Sample Size

We calculated the sample size using a 5% significance level based on a 2-sided test using standard diagnostic accuracy study methods.[16] A study of 897 individuals would have 90% power to detect a 9% difference in accuracy for the primary outcome of diagnosis of glaucoma based on conservative assumptions (a probability of disagreement of 0.18, a glaucoma rate of 25%, and a sensitivity of 86%).[11] This sample size also would yield 80% power for detecting a 6% difference in accuracy should the sensitivity be 93% (the current best estimate from meta-analyses of high-quality diagnostic studies). For specificity, there would be more than 90% power to detect a 5% difference. We assumed that 6% of test results would be indeterminate, and this increased the total sample size to 954.

Statistical Analysis

We included a single eye per patient in the analysis. We ranked the clinical diagnosis in order of decreasing severity according to Table 1. We used the worst eye based on this ranking for our analyses. When both eyes had the same diagnosis, we chose an eye at random. We calculated diagnostic measures (sensitivity, specificity, likelihood ratios, and diagnostic odds ratio) for each test for detection of glaucoma. We compared the diagnostic performance (sensitivity and specificity) of the imaging tests using McNemar's test[16] and generated corresponding 95% paired confidence intervals (CIs).[17] We did not impute any missing data. We reported an indeterminate result when no automatic classification was generated, imaging quality was low, or an artifact was present. We explored the following using sensitivity analyses: performance of the tests across the spectrum of disease, classifying borderline test results as abnormal, combining results of more than one imaging test, and considering glaucoma suspects as glaucoma. In combining test results, if either test gave abnormal results, we classified this as abnormal. We only classified combined results as normal if results from both tests were normal. We calculated the area under the receiver operating characteristic curve (ROC) for each test using the global parameter for HRT or OCT and the NFI parameter for GDx, and we compared these using the method of DeLong et al.[18] We

have reported the results according to the Standards for Reporting Diagnostic accuracy studies (STARD) guidelines.[19] We conducted a post hoc analysis, suggested by a peer reviewer, to explore informally the difference in test performance between tests by calculating individual recruiting center sensitivity and specificity (results not reported here).

Results

We recruited 966 participants (48% of those approached) between April 2011 and July 2013. After obtaining informed consent, we excluded 11 participants: 10 were ineligible (preexisting glaucoma, n = 4; referred from other ophthalmologist, n = 4; or not referred for glaucoma, n = 2), and 1 person withdrew from the study. Therefore, 955 participants were available for the index test comparison. Imaging was not implemented for all index tests in 12

Table 2. Participant Demographics and Ocular Characteristics

Characteristic	Value		
	All Participants	*Glaucoma*	*Nonglaucoma*
No.	943	158	770
Mean age (SD), yrs	60.5 (13.8)	67.4 (12.7)	59.2 (13.6)
Female gender, no. (%)	482 (51.1)	74 (46.8)	401 (52.1)
Ethnicity, no. (%)*			
Black or black Caribbean	25 (2.7)	4 (2.5)	21 (2.7)
Black or black British-African	20 (2.1)	6 (3.8)	14 (1.8)
Asian or Asian British-Indian	18 (1.9)	5 (3.2)	13 (1.7)
Asian or Asian British-Pakistani	4 (0.4)	0 (0)	4 (0.5)
Chinese	1 (0.1)	1 (0.6)	0 (0)
Other Asian background	4 (0.4)	1 (0.6)	3 (0.4)
Mixed white and black African	1 (0.1)	1 (0.6)	0 (0)
White British	826 (89.2)	140 (88.6)	686 (89.1)
Other	29 (3.1)	0 (0)	29 (3.8)
	Right Eye	Left Eye	
Refraction			
Mean sphere D, Mean (SD), n	0.4 (3.3), 571	1.0 (3.6), 561	
Myopia greater than −5 D, N/n (%)	37/943 (3.9)	36/943 (3.8)	
Hyperopia greater than +5 D, N/n (%)	38/943 (4.0)	51/943 (5.4)	
Astigmatism greater than 3 D, N/n (%)	16/943 (1.7)	16/943 (1.7)	
Visual acuity logMAR, Mean SD, n	0.0 (0.30), 925	0.0 (0.3), 926	

D = diopter; logMAR = logarithm of the minimum angle of resolution; SD = standard deviation.
*Self-reported ethnicity; no ethnicity was recorded for 15 participants.

Table 3. Diagnosis of Participants' Worse Eye by Clinician in Secondary Care (Reference Standard)

Diagnosis by Clinician	Worse Eye	
	No.	%
Number of eyes	**932**	
Glaucoma (reference standard positive test results)	158	17.0
Mild	78	8.4
Moderate	45	4.8
Severe	26	2.8
Severity not recorded	9	1.0
Disc suspect	170	18.2
VF suspect	36	3.9
VF + disc suspect	36	3.9
OHT	115	12.3
PAC	31	3.3
PAC suspect	83	8.9
No glaucoma-related findings	299	32.1
Inconclusive diagnosis	4	0.4

OHT = ocular hypertension; PAC = primary angle closure; VF = visual field.

participants, and these were excluded from all analyses. There were no adverse events reported during the study. Figure 1 shows a flow diagram of participant flow and classification of results by test and reference standard. The baseline demographics and ocular characteristics of the 943 participants who underwent all index tests are shown in Table 2. Age and gender of nonparticipants were similar to those of participants (see Fig 1).

In 11 of the 943 participants who underwent all the imaging tests, no reference standard was collected (Fig 1). Reference standard results (diagnosis by clinician) of the remaining 932 participants' worse eyes (as described in "Methods") are shown in Table 3. Table 4 (available at www.aaojournal.org) shows intraocular pressure and visual field mean deviation according to diagnosis. The most common diagnosis was no glaucoma-related findings (32% of participants had no glaucoma-related findings in either eye). Glaucoma was diagnosed in at least 1 eye in 17% of the cohort, and 6.6% had glaucoma in both eyes at referral. Most eyes that were diagnosed with glaucoma had primary open-angle glaucoma (78%).

The distribution of test results, including indeterminate rates, is shown in Table 5 (available at www.aaojournal.org). Indeterminate rates were low (<10%) for all index tests; OCT had the lowest rate of indeterminate test results (4.2%). Indeterminate test results (including those with low quality) were not included in the primary analysis.

Figure 1 shows the categorization of the test results according to the reference standard finding (true and false positive, true and false negative). Of the 943 participants for whom all 4 tests were performed, 158 were classified by the clinician as glaucoma and 770 as not glaucoma. No conclusive diagnosis could be made for 4 participants. Table 6 (available at www.aaojournal.org) shows the diagnostic performance of the 4 tests, including CIs. The HRT MRA had the highest sensitivity (87.0%) but the lowest specificity (63.9%); GDx had the lowest sensitivity (35.1%) but the highest specificity (97.2%); and the other 2 tests provided intermediate results (HRT GPS values were very similar to the HRT MRA results, and OCT had very similar sensitivity and specificity values). The diagnostic odds ratio ranged from 9.24 for HRT GPS to 18.48 for GDx. Table 7 (available at www.aaojournal.org) shows the paired comparison of the 4 imaging tests.
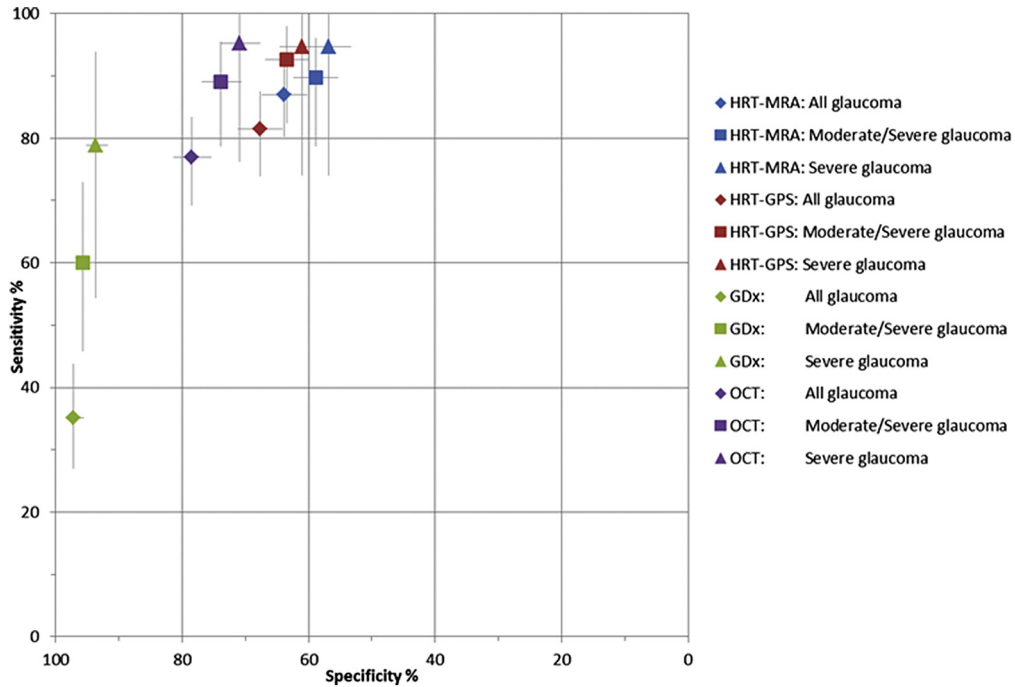
Figure 2. Graph showing the sensitivity and specificity of the tests across the spectrum of disease. GPS = glaucoma probability score; HRT = Heidelberg retinal tomography; MRA = Moorfields regression analysis; OCT = optical coherence tomography.

We explored with sensitivity analyses the performance of the tests across the spectrum of disease by excluding mild or moderate glaucoma diagnoses, or both, as shown in Figure 2. Restricting the reference standard to a diagnosis of severe glaucoma yielded very high sensitivity (95%) for HRT MRA, HRT GPS, and OCT; of these 3, OCT demonstrated the highest specificity (71%).

We also explored the effect of including a borderline classification as an abnormal index test in addition to outside normal limits results, shown in Figure 3. Figure 4 (available at www.aaojournal.org) shows the categorization flow diagram for the analysis in Figure 3. Sensitivity was higher for all tests than under the default analysis, but with correspondingly lower
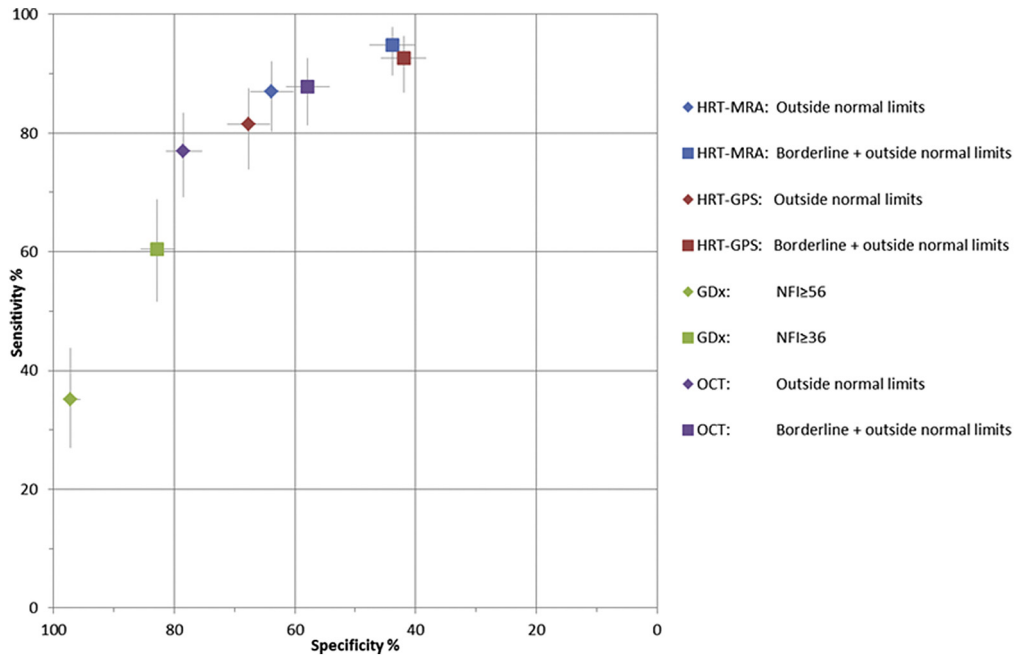
Figure 3. Graph showing comparison of the sensitivity and specificity of the tests when test results classified as borderline or outside normal limits were included as abnormal test results, compared with test results of outside normal limits only. GPS = glaucoma probability score; HRT = Heidelberg retinal tomography; MRA = Moorfields regression analysis; NFI = nerve fiber indicator; OCT = optical coherence tomography.

specificity. Likelihood ratios (and 95% CIs) showed evidence of both being able to rule in and out the presence of glaucoma for all 4 imaging tests (CIs did not contain 1.0). The positive likelihood ratio for GDx was substantially higher than for the other tests (less than 5). This was because of the very high specificity that more than offset the lower relative sensitivity performance. The diagnostic odds ratios ranged from 7.36 for GDx to 14.62 for HRT MRA. Table 8 (available at www.aaojournal.org) shows a further analysis of the diagnostic performance of the imaging tests when glaucoma suspect cases as well as glaucoma cases were classified as cases of disease for the reference standard.

Figure 5 (available at www.aaojournal.org) shows the receiver operating characteristics (ROC) curves for each test under the default analysis. The areas under the ROC curve were similar, with OCT having the highest (0.84) and MRA the lowest (0.79). There was no statistical evidence of a difference between them ($P = 0.077$).

We further explored the diagnostic accuracy of using combinations of tests (Table 9; available at www.aaojournal.org). The HRT MRA was combined with the other 3 tests to form a combined test. Combining tests in this way increased the sensitivity as expected, but only marginally, and at the expense of a greater decrease in specificity. Finally, the post hoc informal exploration of center differences showed some indication of MRA specificity varying between centers.

Discussion

The GATE study was a large, prospective, within-patient, comparative diagnostic accuracy study that provided the sensitivity and specificity data of 4 diagnostic imaging tests for glaucoma. In our study, most participants had successful testing and all 4 imaging tests had some value in terms of diagnosing or ruling out glaucoma. The HRT MRA had the highest sensitivity, but lower specificity, compared with other tests. By contrast, GDx had the best specificity, although the lowest sensitivity. The HRT GPS results were similar to those of the HRT MRA, as might be anticipated given that their analysis is based on imaging the same structure (i.e., the optic disc), and this finding has been reported previously in different populations.[6] Sensitivity for OCT was of very similar magnitude to its specificity. Most tests conducted yielded good-quality automated test results, between 92% (GDx) and 97% (OCT). Optical coherence tomography had the lowest percentage of low-quality imaging results and GDx had the highest, according to the image quality classification provided in the device software. There also was a small percentage of cases (<5%) in which an automated imaging test result could not be produced.

A number of sensitivity analyses were carried out to assess the robustness of the findings of the primary analysis. The HRT MRA had consistently the highest sensitivity across most analyses, but at a cost of lower specificity compared with other tests. When excluding cases of mild glaucoma, HRT GPS had higher sensitivity than HRT MRA. For severe glaucoma, OCT had the highest sensitivity and specificity. By contrast, GDx had the highest specificity, but the lowest sensitivity, across all analyses. None of the imaging tests had a sensitivity of 100% in identifying severe glaucoma.

Varying the test definition of abnormal imaging test results by including the borderline category as positive results

had the anticipated impact of improving the detection of glaucoma, although at the expense of more nonglaucoma cases being classified falsely as being cases of glaucoma. Using combinations of 2 imaging tests led to a marginal improvement in the detection of glaucoma and a reduction in the number of indeterminate test results. Because the HRT GPS and HRT MRA analyses consistently gave the best performing sensitivity and both methods are available on the same machine (HRT), an increase in HRT performance in terms of sensitivity could be obtained by assigning a diagnosis of glaucoma when either the HRT GPS or HRT MRA classifications were outside normal limits. However, the improvement was smaller than the loss in performance in terms of correctly identifying nondisease cases. Considering the additional cost and practical implications, including training and the purchase of additional equipment, suggests that the use of a single technology is to be preferred. The informal exploration of test performance by center suggested that HRT MRA may vary slightly more in this regard. However, there were likely to be some differences between centers in terms of population, which may explain this, or alternatively, it may be a chance finding.

Although the automated imaging classifications performed well at identifying disease (sensitivity), they did not perform as well in identifying normal cases (specificity). The exception to this was GDx, which had very poor sensitivity but moderate specificity. The choice of which imaging test is to be preferred to aid diagnosis reflects the inherent trade-off regarding diagnostic testing, where the desire not to miss glaucoma when present must be balanced against the desire to identify correctly those who are without disease.

Among the strengths of this study is the fact that the enrolled population included consecutive recruitment of subjects without a known history of disease, which reflects a potential clinical application of a diagnostic test. Patients were referred from community optometrists, and this gives a representative sample of current UK practice. The recruited cohort included a large percentage of patients without disease as well as those across the spectrum of glaucoma, including glaucoma suspects and those with ocular hypertension. Prevalence of glaucoma in this cohort was nearly 20%. Other reported studies evaluating the performance of these diagnostic technologies often have used patients already diagnosed with glaucoma, with patients already identified as normal, which has a risk of selection bias. A previous head-to-head comparison of the best-performing parameters of previous versions of the imaging devices the Heidelberg Retina Tomograph II, GDx VCC, and Stratus OCT found similar performance across techniques among a prospective sample of glaucoma patients compared with healthy volunteers.[20] However, the sample did not represent routine practice (i.e., when the disease status is unknown), and the technologies have been upgraded since.

The tests that we compared are all routinely available in clinical practice, and each provided an automated classification of glaucoma status. By using the automated outputs, there is no need to interpret the large number of images and parameters that are produced by the technologies, and this therefore may be attractive to clinicians who are not glaucoma experts. An additional strength of our study is that we

used a paired design in which participants underwent all 4 tests, allowing a head-to-head comparison. We chose as the reference standard the clinical assessment provided by different ophthalmologists with glaucoma expertise. All ophthalmologists were trained in the study protocols and agreed to a common set of criteria to define glaucoma and normality and were masked to the imaging test results when making their diagnosis. By using different ophthalmologists working at different units, the results of the study are more likely to be generalizable than results from studies performed in a single unit.

Among the limitations, we recognize that diagnosing glaucoma during the very early stage of disease is challenging. Although consensus was sought through structured discussions and agreement, some assessor differences may have remained between the centers. Ideally, a longitudinal follow-up would provide the best possible reference standard. This was proposed by Medeiros et al.[21] who used optic nerve head progression on stereophotographic examination as the criterion for glaucoma diagnosis. We used a single eye (worse diagnosis) per patient, and assessment of the presence of glaucoma in either eye may have a slightly different diagnostic performance. A limitation of the HRT MRA analysis is that it requires a trained technician to identify manually the margins of the optic disc on the acquired image, and this may affect the diagnostic performance if it is not carried out appropriately by a trained examiner. There are other OCT instruments in clinical use with glaucoma diagnostic capabilities, and the results of this study using the Spectralis device may not be applicable fully to other OCT technologies. In this study, we analyzed only a small amount of potential data generated. In addition, OCT hardware and software are evolving rapidly, and new developments may improve their diagnostic performance. Also, current OCT technology now is able to quantify macular parameters that may be useful to diagnose glaucoma, for example, in myopia.

The results of this study have implications for clinical practice. No test demonstrated or could have been expected to demonstrate 100% sensitivity and specificity, and less-than-perfect results reflect a combination of factors that include an imperfect reference standard and normative databases, at least in part because of the variability of normal anatomic features and instrument operator−dependent factors. Thus, relying solely on the automatic classification to diagnose glaucoma or using this as a replacement for visual field testing is not recommended. We suggest automated imaging technologies are best suited as an adjunct to the ophthalmologist's assessment. It should also be noted that the clinical biomicroscopic examination of the optic nerve and nerve fiber layer yields a true stereoscopic view of the structures that may reveal glaucomatous features that are not visible in the computer-generated images from the imaging technologies used in this study, such as optic disc hemorrhages. These imaging instruments probably have their greatest usefulness when assisting in the diagnosis of more subtle glaucomatous optic neuropathy, particularly in situations where the nerve fiber layer and optic disc are indistinct clinically. Additionally, there is a role to assist clinicians who may not be glaucoma specialists or in the case of trainees where the additional information provided by these devices can be a valuable component of training. From a UK perspective, using automated imaging classification of imaging technologies as a triage test may prove to be cost-effective.[22]

In conclusion, this study offers valuable insight into the comparative diagnostic performance of automated diagnostic tests using some of their outputs (disc classification for HRT and RNFL classification for GDx enhanced corneal compensation and Spectralis OCT) that are easy to read and interpret and are encountered in a real-life setting. Automated testing is helpful in and aids glaucoma diagnosis. However, reliance solely on imaging (as used in this study) as a means of diagnosis is not recommended because some patients with severe glaucoma may be missed.

References

1. Prum BE Jr, Rosenburg LF, Gedde SJ, et al. Primary Open-Angle Glaucoma Preferred Practice Pattern® Guidelines. Ophthalmology 2016;123:p41–p111.
2. Reus NJ, Lemij HG, Garway-Heath DF, et al. Clinical assessment of stereoscopic optic disc photographs for glaucoma: the European Optic Disc Assessment Trial. Ophthalmology 2010;117:717–23.
3. Stein JD, Talwar N, Laverne AM, et al. Trends in use of ancillary glaucoma tests for patients with open-angle glaucoma from 2001 to 2009. Ophthalmology 2012;119:748–58.
4. Lin SC, Singh K, Jampel HD, et al; American Academy of Ophthalmology; Ophthalmic Technology Assessment Committee Glaucoma Panel. Optic nerve head and retinal nerve fiber layer analysis: a report by the American Academy of Ophthalmology. Ophthalmology 2007;114:1937–49.
5. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. Ophthalmology 1998;105:1557–63.
6. Coops A, Henson DB, Kwartz AJ, Artes PH. Automated analysis of heidelberg retina tomograph optic disc images by glaucoma probability score. Invest Ophthalmol Vis Sci 2006;47:5348–55.
7. Burgansky-Eliash Z, Wollstein G, Bilonick RA, et al. Glaucoma detection with the Heidelberg retina tomograph 3. Ophthalmology 2007;114:466–71.
8. Garas A, Vargha P, Holló G. Comparison of diagnostic accuracy of the RTVue Fourier-domain OCT and the GDx-VCC/ECC polarimeter to detect glaucoma. Eur J Ophthalmol 2012;22:45–54.
9. Sehi M, Iverson SM. Glaucoma diagnosis and monitoring using advance imaging technologies. US Ophthalmic Rev 2013;6:15–25.

10. Tay E, Andreou P, Xing W, et al. A questionnaire survey of patient acceptability of optic disc imaging by HRT II and GDx. Br J Ophthalmol 2004;88:719–20.

11. Burr J, Mowatt G, Siddiqui MAR, et al. The clinical and cost effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. Health Technol Assess 2007;11:iii–iv. ix-x, 1–190.

12. Medeiros FA, Ng D, Zangwill LM, et al. The effects of study design and spectrum bias on the evaluation of diagnostic accuracy of confocal scanning laser ophthalmoscopy in glaucoma. Invest Ophthalmol Vis Sci 2007;48:214–22.

13. National Institute for Health Research. Evaluation, Trials and Studies. Available at: http://www.nets.nihr.ac.uk/projects/hta/0922111. Accessed February 1, 2015.

14. Carl Zeiss Meditec, Inc. GDxPRO Scanning Laser Polarimeter: User Manual. URL: www.amedeolucente.it/pdf/GDxPRO_User_Manual.pdf; Accessed September 2014.

15. National Institute for Health and Clinical Excellence. Glaucoma: Diagnosis and Management of Chronic Open Angle Glaucoma and Ocular Hypertension. Clinical Guidelines CG85. 2009. Available at, http://guidance.nice.org.uk/CG85. Accessed August 1, 2015.

16. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12:153–7.

17. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. Stat Med 1998;17:2635–50.

18. DeLong ER, DeLon DM, Clarke-Pearson DL. Comparing the area under two or more correlated receiver operating characteristic curve: a non-parametric approach. Biometrics 1988;44:837–45.

19. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for reporting of diagnostic accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1–12.

20. Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. Comparison of the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and Stratus OCT optical coherence tomograph for the detection of glaucoma. Arch Ophthalmol 2004;122:827–37.

21. Medeiros FA, Zangwill LM, Bowd C, et al. Use of progressive glaucoma or optic disc change as the reference standard for evaluation of diagnostic tests in glaucoma. Am J Ophthalmol 2005;139:1010–8.

22. Azuara-Blanco A, Banister K, Boachie C, et al. Automated imaging technologies for the diagnosis of glaucoma: a comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost-effectiveness (GATE study). Health Technol Assess 2016;20:1–168.

Footnotes and Financial Disclosures

[1] Health Services Research Unit, University of Aberdeen, Aberdeen, United Kingdom.

[2] Robertson Centre for Biostatistics, University of Glasgow, Glasgow, United Kingdom.

[3] Vision & Eye Research Unit, Postgraduate Institute, Anglia Ruskin University, Cambridge, United Kingdom.

[4] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom.

[5] School of Medicine, University of St. Andrews, St. Andrews, United Kingdom.

[6] National Institute for Health Research Biomedical Research Centre, Moorfields Eye Hospital and University College London Institute of Ophthalmology, London, United Kingdom.

[7] Faculty of Health and Life Sciences, Northumbria University, Newcastle-upon-Tyne, United Kingdom.

[8] Health Economics Research Unit, University of Aberdeen, Aberdeen, United Kingdom.

[9] Centre for Experimental Medicine, Queen's University, Belfast, United Kingdom.

Abbreviations and Acronyms:

CI = confidence interval; **GDx** = Glaucoma Diagnostics; **GPS** = glaucoma probability score; **HRT** = Heidelberg retina tomograph; **MRA** = Moorfields regression analysis; **NFI** = nerve fiber indicator; **OCT** = optical coherence tomography; **RNFL** = retinal nerve fiber layer; **ROC** = receiver operating characteristic.

Correspondence:

Augusto Azuara-Blanco, FRCOphth, Centre for Experimental Medicine, Institute of Clinical Sciences - Block A, Queen's University, Grosvenor Road, Belfast, BT12 6BA United Kingdom. E-mail: a.azuara-blanco@qub.ac.uk.