

Article

Human Factor and Usability Testing of a Binocular Optical Coherence Tomography System

Reena Chopra¹, Pádraig J. Mulholland^{1,2}, Adam M. Dubis¹, Roger S. Anderson^{1,2}, and Pearse A. Keane¹

¹ NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK

² Optometry and Vision Science Research Group, School of Biomedical Sciences, Ulster University, Coleraine, UK

Correspondence: Pearse A. Keane, MD FRCOphth, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom. e-mail: p.keane@ucl.ac.uk

Received: 04 April 2017

Accepted: 07 July 2017

Published: 15 August 2017

Keywords: binocular; optical coherence tomography; automated; diagnostics; usability; human factors

Citation: Chopra R, Mulholland PJ, Dubis AM, Anderson RS, Keane PA. Human factor and usability testing of a binocular optical coherence tomography system. *Trans Vis Sci Tech.* 2017;6(4):16, doi:10.1167/tvst.6.4.16

Copyright 2017 The Authors

Purpose: To perform usability testing of a binocular optical coherence tomography (OCT) prototype to predict its function in a clinical setting, and to identify any potential user errors, especially in an elderly and visually impaired population.

Methods: Forty-five participants with chronic eye disease (mean age 62.7 years) and 15 healthy controls (mean age 53 years) underwent automated eye examination using the prototype. Examination included ‘whole-eye’ OCT, ocular motility, visual acuity measurement, perimetry, and pupillometry. Interviews were conducted to assess the subjective appeal and ease of use for this cohort of first-time users.

Results: All participants completed the full suite of tests. Eighty-one percent of the chronic eye disease group, and 79% of healthy controls, found the prototype easier to use than common technologies, such as smartphones. Overall, 86% described the device to be appealing for use in a clinical setting. There was no statistically significant difference in the total time taken to complete the examination between participants with chronic eye disease (median 702 seconds) and healthy volunteers (median 637 seconds) ($P = 0.81$).

Conclusion: On their first use, elderly and visually impaired users completed the automated examination without assistance. Binocular OCT has the potential to perform a comprehensive eye examination in an automated manner, and thus improve the efficiency and quality of eye care.

Translational Relevance: A usable binocular OCT system has been developed that can be administered in an automated manner. We have identified areas that would benefit from further development to guide the translation of this technology into clinical practice.

Introduction

Optical coherence tomography (OCT) imaging has transformed ophthalmology.^{1,2} In 2011, it is estimated that 20 million OCT examinations were performed worldwide, more than the sum of all other ophthalmic imaging combined.³ Although reasonably quick and safe to perform, the costs involved with operating OCT devices are not trivial.^{4,5} Commercial devices are expensive to purchase, and the costs of associated labor to capture the OCT scans are even greater.⁴ The requirement for OCT and other imaging also places increasing demand on ophthalmology clinics, with fragmented patient pathways and often extended

waiting times.^{6,7} Clinic efficiency has recently been reported to improve by stationing all necessary tests for the visit in one location, minimizing the time patients spend moving around and waiting.⁸

In this report, we explore a novel “binocular OCT” system (Envision Diagnostics Inc., El Segundo, CA) that incorporates many aspects of the eye examination into one single patient-operated instrument.⁹ This device aims to improve both the efficiency and quality of eye care while reducing the overall labor and equipment costs. Automatic alignment of the oculars to the user’s eyes allows patients to undertake the full suite of tests, including ‘whole-eye’ OCT imaging, without operator assistance. Paired with ‘smart technology’, such as customizable display

screens to present information to the user, and voice recognition to register user responses, the device also performs a range of other ophthalmic diagnostic tests, from visual acuity measurement to perimetry. The binocular design of the device permits simultaneous OCT image capture from both eyes, allowing OCT-derived assessment of binocular functions, such as pupillometry and ocular motility. With such a device, it may thus be possible in the future to perform a comprehensive, objective, quantitative ocular examination using a single instrument, and, additionally, may have a role in telemedicine to transfer generated data from remote locations.

While many such devices are capable of completing specific tasks, lack of “usability” prevents their widespread adoption (i.e., device operations are not easy to learn and remember, or are not efficient or user-friendly).^{10,11} Moreover, devices that are difficult to use or understand expose patients to clinical risk as a result of human error during usage. Structured, patient-centered, usability testing is essential to the design, clinical validation, regulatory approval, and widespread implementation of new medical devices.¹² This is particularly the case for a putative binocular OCT system – an automated device that will primarily be used in visually impaired, often elderly, populations.

In this study, we perform prospective usability testing of a binocular OCT prototype in a population with chronic eye disease, and in healthy volunteers, with a view to predicting likely function in a clinical setting and identifying any potential user errors that may expose the patient to clinical risk (EASE study – ClinicalTrials.gov Identifier: NCT02822612). The results of the study will facilitate an iterative process of operating software and workflow modifications, and thus aid the translational of this technology into clinical practice.

Methods

Study Population

Forty-five participants with chronic eye disease were prospectively recruited from glaucoma, retinal disease, and strabismus clinics at Moorfields Eye Hospital, London, UK. In addition, 15 healthy volunteers with no self-reported history of ocular disease were recruited as a control group. The sample size was based on usability literature,¹³ and draft guidance from the Food and Drug Administration “Human Factors” program.¹⁴ All participants were

required to have no significant hearing impairment that would affect their ability to respond to instructions delivered by the device. A conversational level of English was required for users to understand the instructions, and to be able to communicate with the device via an English language voice recognition system (VRS). No participants were excluded based on disease status to ensure our cohort consisted of everyday users of eye care services. Therefore, ocular comorbidities were permitted. Approval for data collection and analysis was obtained from a UK National Health Service Research Ethics Committee (REC) (London-Central) and the study adhered to the tenets of the Declaration of Helsinki.

Clinical Data Collection

Best-corrected visual acuity was initially measured monocularly using Early Treatment Diabetic Retinopathy Study (ETDRS) charts. For participants with glaucoma, visual field mean deviation scores were recorded from their most recent (<6 months) SITA standard 24-2 examination on the Humphrey Visual Field Analyzer (Carl Zeiss Meditec, Dublin, CA). If worn, participants’ habitual refractive error correction was measured using an automatic lensmeter (Grafton Optical, Berkhamsted, UK). Both acuity and habitual refractive error were inputted into custom software connected to the binocular OCT device.

Prototype Binocular OCT Technical Specifications

The binocular OCT prototype was a similar size as other commercial OCT systems in use today. It was mounted on a motorized base that allows users to adjust the instrument height. It was a Class 1 laser system that has two internal full color displays and one swept-source OCT laser (Axsun Technologies, Billerica, MA) centered at 1060 nm that enables OCT imaging of both eyes. The laser power was limited to the lowest power allowed by Class 1 limits for single pulse, pulse train, and average power across 30,000 seconds of use for a conservative duty cycle estimate of 66% and subtended beam angle of 1.5 mrad. Custom optics on independent linear motion stages were used to direct and focus light into the subject’s eyes. A hardware VRS (Sensory, Inc., Santa Clara, CA) and a text-to-speech module (TextSpeak, Inc., Westport, CO) were used for communications with the participant. The prototype was connected to an external central computer system that, along with

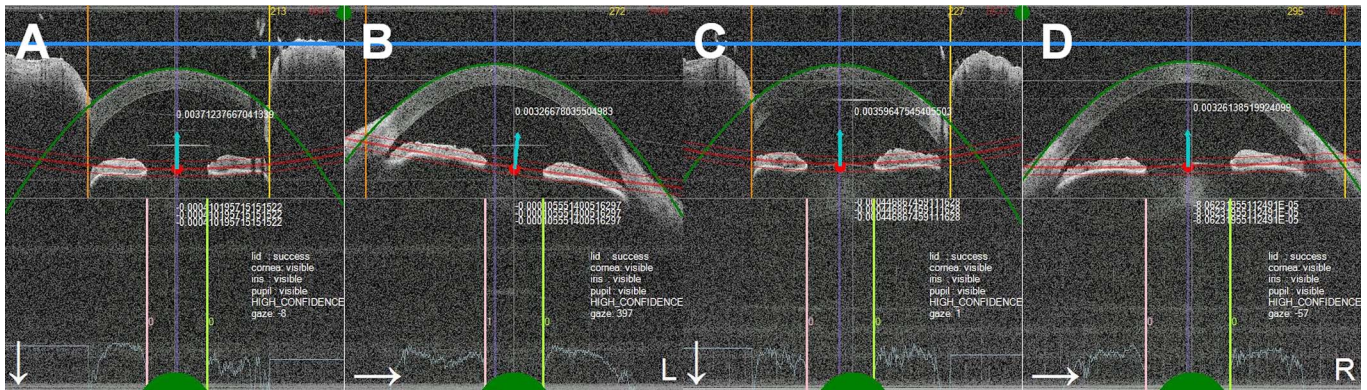


Figure 1. Real-time segmentation of optical coherence tomography (OCT) images of the anterior cornea and pupil center. Images are captured in the vertical (A, C) and horizontal (B, D) meridians of both eyes. Segmentation aids accurate alignment of the optics within the device to the user's eyes.

internal custom electronics, handles data processing and output.

Binocular OCT Examination

Prior to examination a spherical equivalent of the participant's habitual refractive correction was remotely placed within the prototype device. Without prior training, participants then underwent automated binocular OCT examination, under direct supervision by a study investigator (RC). Instructions for the examination and individual tests were delivered in an automated manner using TextSpeak and a speaker system built into the prototype. All instructions were spoken in a female British English voice. Participants were asked to listen carefully to the instructions and respond verbally when asked to do so by the device. Participants were advised that the examination would consist of several tests and that the device will inform them when the examination has concluded. All examinations were video recorded with participant consent.

The order of testing is listed below. While 1 and 2 were performed simultaneously on both eyes, tests 3, 4, 5, and 6 were performed on the left eye first followed by the right eye. This order was arbitrary and preset for this device.

Test 1: Introduction to the Examination and Initial Ocular Alignment

The device instructed the user to place their head in the mask, and respond "ready" to begin examination. The mask incorporated a nose rest and a disposable forehead rest. Disposable safety goggles made of American National Standards Institute (ANSI) Z97.1 compliant polycarbonate was placed on the interface for protection from the moving optics within the

device. Once the user was comfortable and ready to begin, the device presented a circular fixation target covering a 2.5° field on the retina to eye with better acuity. The device then proceeded by moving the optics of the device to align with the user's eyes. Real-time segmentation of the cornea and iris planes provided the device with simultaneous feedback on accurate alignment (Fig. 1). Ocular alignment was automatically reassessed prior to each individual test to ensure the user was still in the correct position. The device was able to recognize when the user was leaning away from the mask or not fixating – in these cases, the machine provided the user with additional instructions to adjust their head position, or to remind them to look at the fixation target. The optics within the device were simultaneously adjusted to regain alignment before proceeding with the next test.

Test 2: Whole-Eye Swept-Source OCT Imaging

'Whole-eye' imaging, as defined by recent literature,^{15,16} was the first diagnostic test to be performed. The user was instructed to look at a central fixation target, presented to the eye with better acuity. The following OCT images were captured from both eyes simultaneously: (1) anterior segment OCT imaging – 128 raster scans consisting of 1350 A-scans each covering an area of 16×16 mm of the anterior segment, and one horizontal and one vertical averaged scan, (2) posterior segment OCT imaging – 128 raster scans consisting of 1350 A-scans each covering an area of 14×14 mm of the retina centered on the macula, and one horizontal and one vertical averaged scan, and (3) vitreous OCT imaging – raster scan of 128 slices imaging up to an inner depth of 8 to 10 mm from the retina, and one horizontal and one vertical averaged scan. All averaged scans were generated

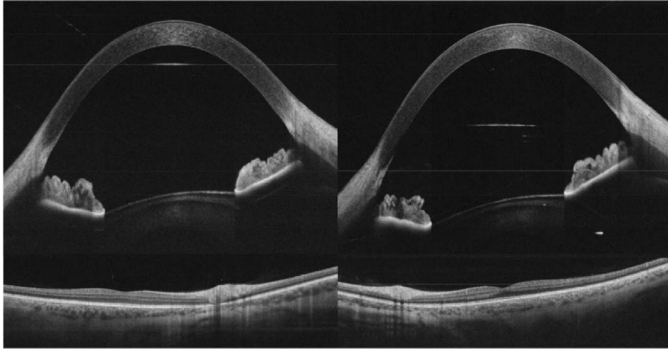


Figure 2. Averaged OCT B-scan images acquired using the binocular OCT prototype in a healthy volunteer. Anterior and posterior segment images are captured of both eyes simultaneously.

from 16 B-scans through the central meridians (Fig. 2).

Test 3: Ocular Motility

This examination was performed with one eye fixating at a time, while capturing simultaneous OCT images of the vertical and horizontal planes of the anterior segments of both eyes. A fixation target was presented in primary position (center) and eight positions of gaze 4° from center (west, northwest, north, northeast, east, southeast, south, southwest). The coordinates of the position of the anterior segment of both the fixating and nonfixating eyes could subsequently be mapped to detect the presence of heterophorias, heterotropias, and abnormal eye movements.

Test 4: Visual Acuity

Visual acuity was assessed monocularly. Each eye was presented with single ETDRS targets on the display screen. The device instructed the user to verbally report the letter they can see or respond “I don’t know” if they were unable to discern it. Responses were registered using a VRS with an automated threshold algorithm¹⁷ used to determine final acuity. For each presented letter, the user was given a window of 10 seconds to respond. The device presents the next largest visual acuity target if the response is incorrect or not heard. In the current system, the size of the largest visual acuity targets range from 20/16 to 20/800. (See Supplementary Video S1).

Test 5: Suprathreshold Perimetry

A 2° × 2° high-contrast, square-wave grating stimulus was randomly presented in the same eight peripheral subfields as tested in the motility exam.

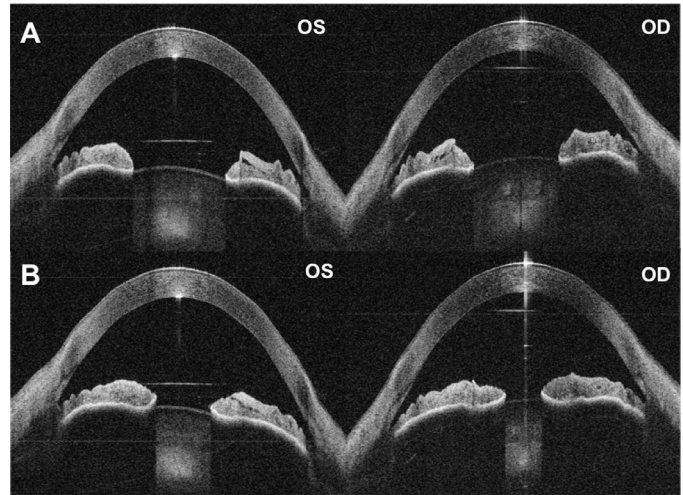


Figure 3. Automated quantitative pupillometry using the binocular OCT prototype. (A) Pupils of both eyes dilated immediately prior to stimulation. (B) Pupils at maximum constriction after controlled flash of light presented to the right eye. Note the anisocoria – the left pupil does not constrict equally to the right pupil. In post-processing, the presence and extent of pupillary defects can be calculated using OCT-derived measurements of the pupil circumference.

Participants were instructed to focus on the central fixation target and respond “yes” when the stimulus was visible. Users have a timeframe of 2 seconds to respond on top of a random time delay of up to 3 seconds before the next stimulus is presented.

Test 6: Pupillometry

Pupil reactions were assessed using simultaneous OCT capture of the anterior segments including the iris plane. Each eye was stimulated independently and sequentially with a single, bright, 250-ms flash of white light. B-scan recordings are captured at regular intervals of 350 ms prior to stimulation and 4000 ms post-stimulation. Measurements of the pupil circumference could subsequently be calculated to identify pupil abnormalities and relative afferent pupillary defects (Fig. 3).

Test Duration and Completion Rates

The following information was collected during and after the examination:

- Overall examination time;
- Time needed to complete individual diagnostic tests;
- Examination completion rates for the whole examination and for each individual diagnostic test; and

Table 1. Clinical and Demographic Characteristics

	Age, Mean \pm SD (Range) (y)	Visual Acuity Better Eye, Mean \pm SD (Range) (logMAR)	Visual Acuity Worse Eye, Mean \pm SD (Range) (logMAR)	Visual Fields Mean Deviation Better Eye, Mean \pm SD (dB)	Visual Fields Mean Deviation Worse Eye, Mean \pm SD (dB)	Refractive Error Spherical Equivalent, Mean \pm SD (DS)
Chronic eye disease (<i>n</i> = 45)						
Glaucoma (<i>n</i> = 13)	64.1 \pm 14.7 (27–83)	0.10 \pm 0.12 (–0.08 to 0.24)	0.20 \pm 0.12 (0.00–0.42)	–6.31 \pm 8.17	–11.75 \pm 10.41	–0.54 \pm 0.76
Retinal disease (<i>n</i> = 14)	71.0 \pm 11.4 (50–88)	0.35 \pm 0.24 (0.02–0.84)	0.68 \pm 0.34 (0.24–1.30)			0.89 \pm 2.21
Strabismus (<i>n</i> = 14)	50.8 \pm 19.1 (23–74)	–0.04 \pm 0.08 (–0.20 to 0.10)	0.24 \pm 0.27 (–0.10 to 0.78)			0.54 \pm 2.19
Ocular comorbidities (<i>n</i> = 4)	70.5 \pm 8.0 (61–81)	0.24 \pm 0.42 (–0.10 to 0.78)	0.63 \pm 0.53 (–0.1 to 1.18)	–7.59 \pm 6.86	–11.28 \pm 5.81	1.11 \pm 1.80
Healthy controls (<i>n</i> = 15)	53.1 \pm 11.2 (30–67)	0.01 \pm 0.14 (–0.16 to 0.30)	0.02 \pm 0.15 (–0.16 to 0.30)			–0.29 \pm 1.26

- Observed user and device errors that may lead to the generation of erroneous examination findings.

Interview and Questionnaire

A pretest interview was conducted prior to binocular OCT examination to gauge participants' levels of experience with common technologies such as computers, smartphones, the Internet, and email. Immediately after the examination, a short debriefing discussion was conducted. Participants were asked to rate the ease of the examination in comparison to the technologies they commonly use. Subjective ease of use, duration, and appeal were rated on a 5-point Likert scale. Verbatim comments were also recorded.

Patient and Public Involvement (PPI)

Prior to study setup, a patient focus group of 10 contributors was convened to assist with informing the design and implementation of the study. A second group of nine contributors (including 4 people from the first focus group) was convened after study completion to advise on dissemination of preliminary results, and hence provide discourse for recommendations to improve the device to aid its translation into a clinical setting. The events lasted for approximately 3 hours. The PPI team from the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital facilitated the discussions.

Engaging end-users is essential for all stages of the clinical validation of health technologies.^{18,19} The inclusion of PPI was particularly relevant for a self-operated binocular OCT device intended for use in elderly and visually impaired populations to identify user requirements while the device was in early-stage development. The recommendations of participants, patients, and the public will guide the modification of the interface and workflow of the binocular OCT system.

Results

Patient Demographics

Thirteen participants had glaucoma only (12 with primary open angle glaucoma [POAG], 1 with glaucoma secondary to hypertensive uveitis). Fourteen had strabismus only (7 with esotropia, 6 with exotropia, and 1 with hypertropia). Fourteen had bilateral retinal disease only (including 8 with age-related macular degeneration [AMD], 4 with diabetic macular edema, 1 with central serous retinopathy, and 1 with retinal vein occlusion with cystoid macular edema). Four participants had ocular comorbidities: two with bilateral POAG and AMD; one had unilateral POAG and a symptomatic epiretinal membrane in the fellow eye; and one had bilateral POAG and congenital convergent strabismus. [Table 1](#) presents their clinical and demographic characteristics.

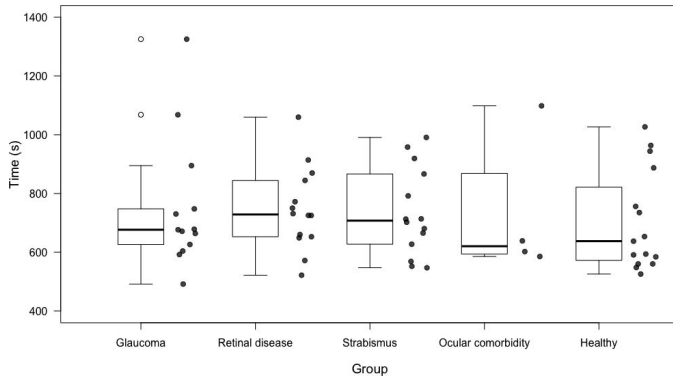


Figure 4. Box plots showing total examination time for each group. The *horizontal lines* within each box represent the median for each group; the ends of the boxes are the upper and lower quartiles, and the *whiskers* represent minimum and maximum values. The data for each individual participant is included as peripheral scatter plots.

Binocular OCT Examination

All participants completed the full suite of tests using the binocular OCT device without assistance. The Shapiro-Wilk test revealed that overall examination times were not normally distributed. The median time to complete the full suite of diagnostic tests on the binocular OCT prototype was 702 seconds (interquartile range [IQR]: 627–845 seconds) in the group with chronic eye disease, and 638 seconds (IQR: 572–821 seconds) in the healthy control group. There was no significant difference in the time taken to complete the full suite of tests on the binocular OCT system between the two groups ($P = 0.30$, Mann-Whitney U test). Similarly, there was no statistical difference in the overall examination time between participants with glaucoma, strabismus,

retinal disease, ocular comorbidities, and healthy volunteers ($P = 0.81$, Kruskal-Wallis; Fig. 4).

Individual Diagnostic Test Times

Test times for each of the diagnostic tests are presented in Table 2. These values include the time spent reassessing and realigning the optics of the device with the participant's eyes prior to each test (if required). There was no statistically significant difference between the chronic eye disease group and the healthy volunteers in the time taken to complete the individual tests, except in the case of visual acuity measurement ($P = 0.004$), where diseased eyes took longer (median 168 seconds, compared with 133 seconds for healthy volunteers).

Major Examination Components

To determine how much time the prototype spent on various activities, the examination times were dissected further into four major exam components as indicated in Figure 5 including (1) audio instructions (the machine speaking to the user), (2) voice responses (the machine waiting for and interpreting voice input from the user), (3) OCT imaging (amount of time spent gathering OCT data for the tests), and (4) component movements (including repositioning of the motors and other optical components within the machine to maintain optical alignment with the user's eyes).

The device spent a median time of 178 seconds providing instructions to the participant (range 130–367 seconds). The time taken to provide the standard instructions was fixed at 130 seconds for each exam. The remainder of the time was spent providing additional instructions when the device recognized

Table 2. Time Taken to Complete Individual Diagnostic Tests

	Chronic Eye Disease ($n = 45$) Median Time (s) and IQR	Healthy Controls ($n = 15$) Median Time (s) and IQR	Mann-Whitney U Test P Value
Introduction and ocular alignment	94 (55–125)	80 (64–101)	0.500
OCT imaging	159 (149–179)	155 (151–169)	0.932
Ocular motility	78 (66–96)	72 (64–85)	0.591
Visual acuity	168 (138–196)	133 (110–152)	0.004*
Suprathreshold perimetry	122 (111–137)	113 (106–170)	0.620
Pupillometry	69 (54–89)	67 (62–106)	0.394
Overall	702 (627–845)	637 (572–821)	0.302

There was no statistically significant difference between the two groups, except for the visual acuity test.

* Significant at $P < 0.05$.

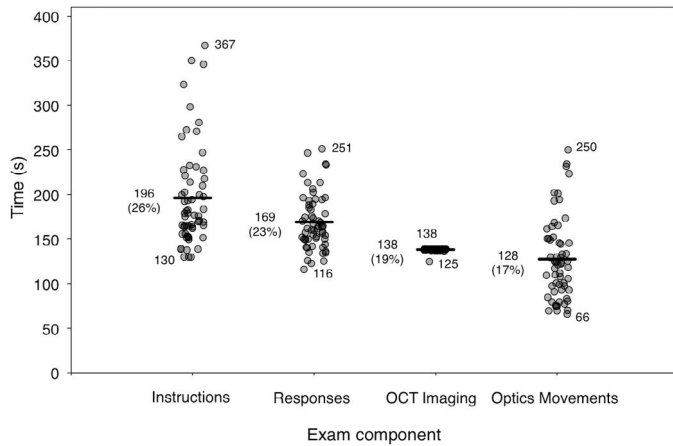


Figure 5. Scatter plots illustrating the time spent on the major exam components for all participants. The *horizontal lines* represent the median time, Q2, for each component.

that the user was not fixating or keeping their eyes open. In these cases, the device would automatically attempt to find their eyes and would give them instructions to look at the fixation target.

A median time of 163 seconds was spent on voice recognition (range, 116–251 seconds). Voice recognition was only required for visual acuity and perimetry at present, therefore examinations that spent longer on voice recognition indicate that a longer time was spent on these two sections of the examination.

The device spent an average of 139 seconds performing OCT imaging (range, 125–139 seconds) during the entire exam. This was a fairly constant amount of time required to perform ‘whole-eye’ imaging.

An average of 122 seconds was spent moving optics within the machine per exam (range, 66–250 seconds). This was strongly correlated with the amount of time spent providing instructions to the participant (Pearson’s correlation $r = 0.82$, $P < 0.001$). Examinations where additional instructions were provided also required simultaneous repositioning of the optics.

Observed Errors

As presented in [Table 3](#), the majority of examinations generated usable data. Both device and user errors affected the quality of data produced. OCT imaging was classified as ‘ungradable’ if the OCT scans were poor quality (i.e., if there were severe artifacts or generalized reductions in signal strength to the extent that major interfaces could not be identified). Poor quality posterior segment and vitreous images were captured if the participant blinked during image capture, or if the position of

Table 3. Number of Examinations that Generated Gradable Data

	Chronic Eye Disease ($n = 45$) N (%)	Healthy Controls ($n = 15$) N (%)
Anterior segment imaging		
OS	45 (100)	14 (93.3)
OD	45 (100)	14 (93.3)
Posterior segment imaging		
OS	38 (84.4)	13 (86.7)
OD	42 (93.3)	13 (86.7)
Vitreous imaging		
OS	38 (84.4)	13 (86.7)
OD	42 (93.3)	13 (86.7)
Motility		
OS fixating	21 (60.0)	14 (93.3)
OD fixating	26 (74.2)	14 (93.3)
Visual acuity		
OS	42 (93.3)	15 (100)
OD	45 (100)	15 (100)
Suprathreshold perimetry		
OS	43 (95.6)	15 (100)
OD	44 (97.8)	15 (100)
Pupillometry		
OS flash	39 (86.7)	13 (86.7)
OD flash	41 (91.1)	13 (86.7)

the eye was incorrect. One healthy volunteer had their eyes closed throughout the imaging test and no anterior segment imaging data were acquired. Good quality anterior segment images were obtained for all other participants.

Motility exams were classified as ungradable if the user did not fixate on the target displayed in the various positions of gaze. The motility test currently consists of a fixation target that appears in different locations, rather than requiring the user to ‘fix and follow’. Unsurprisingly, we found that participants with advanced POAG were not able to detect the target when presented in a scotoma. Similarly, participants with poor visual acuity were unable to see the fixation target due to its low contrast. Other observed errors included one participant with POAG misinterpreting the motility test for a visual field test and therefore not following the motility target.

Data for visual acuity and perimetry exams were classified as ungradable if the participant did not respond, and thus a measurement could not be generated. Three participants did not verbally respond when required during the visual acuity test.

Table 4. Subjective Ratings for Binocular OCT Examination, Rated on a 5-Point Likert Scale

	Chronic Eye Disease (<i>n</i> = 45)		Healthy Controls (<i>n</i> = 15)	
	<i>N</i> (%)	Mean Age and SD (y)	<i>N</i> (%)	Mean Age and SD (y)
Ease of use				
5 (very easy)	20 (44.4)	63.80 ± 12.63	6 (40.0)	52.33 ± 12.29
4	15 (33.3)	58.07 ± 21.75	7 (46.7)	52.43 ± 11.77
3	8 (17.8)	65.75 ± 17.07	1 (6.7)	50
2	1 (2.2)	67	1 (6.7)	65
1 (very difficult)	1 (2.2)	80	0 (0.0)	-
Duration				
5 (very short time)	15 (33.3)	66.13 ± 10.93	2 (13.3)	55.00 ± 7.07
4	11 (24.4)	63.27 ± 17.78	6 (40.0)	53.83 ± 13.48
3	17 (37.8)	60.29 ± 20.13	7 (46.7)	51.86 ± 11.29
2	2 (4.4)	53.50 ± 31.82	0 (0.0)	-
1 (very long time)	0 (0.0)	-	0 (0.0)	-
Appeal				
5 (very appealing)	21 (46.7)	66.29 ± 11.52	8 (53.3)	50.75 ± 14.08
4	18 (40.0)	59.55 ± 18.23	5 (33.3)	55.00 ± 6.00
3	3 (6.7)	40.33 ± 26.63	2 (13.3)	57.50 ± 10.61
2	2 (4.4)	77.50 ± 4.94	0 (0.0)	-
1 (very unappealing)	1 (2.2)	80	0 (0.0)	-

Similarly, only one participant did not respond during perimetry. We also observed errors in the accuracy of the voice recognition system. The sensitivity of the VRS was calculated by comparing the participant's verbal response to the response interpreted by the device. Average sensitivity measured as 64% overall for all 60 participants (range, 12.5%–100%). This appeared to be related to the system misinterpreting the user (e.g., “A” heard when the user responds “K” in the acuity test), or if the user responded with multiple answers (e.g., “Y or V”), or attempted to change their answer.

Pupillometry values were classified as ungradable if the pupil response data could not be generated from the examination. Errors were observed if the user blinked during the test or looked away from the central fixation target so that the pupil was occluded.

Interview

Eighty-two percent of our cohort with chronic eye disease and 93% of healthy volunteers used common technologies, such as computers and smartphones at least a few times per week. All of these participants regularly used the Internet and email communication. Four participants, including one healthy volunteer, never used these technologies.

Eighty-one percent of participants with chronic eye

disease and 79% of healthy volunteers subjectively found the binocular OCT system easier to use than a computer or smartphone. Table 4 presents subjective ratings for ease of use, appeal, and duration. Seventy-eight percent of participants with chronic eye disease and 87% of healthy volunteers rated the binocular OCT 4 or 5 on ease of use (1 = very difficult, 5 = very easy).

Fifty-eight percent of chronic eye disease participants and 54% of healthy volunteers rated 4 or 5 on duration of test (1 = very long time, 5 = very short time). Overall, 86% found the device appealing to use in a clinical setting. There was no significant correlation between age and subjective ratings for ease of use, duration, or appeal. There was however a significant correlation between subjective ratings for ease of use and duration ($P < 0.001$, Spearman's rank) and ease of use and appeal ($P < 0.001$) in the overall cohort. Interestingly, we found no significant correlation between overall examination time and subjective ratings for ease of use, appeal, or duration. All four subjects that were unfamiliar with common technologies rated the examination 4 or 5 on ease, appeal, and duration.

Verbatim comments revealed that participants felt the device delivered clear instructions and was easy to use. Of participants, 37% commented that the device

headrest and mask interface was physically uncomfortable. Participants with poor acuity found the fixation target to be unclear.

PPI Recommendations

Generally, contributors to the focus group felt that performing their own eye examination facilitated more control over their care. They felt confident that the generation of robust, standardized data from smart devices, such as the binocular OCT system would aid monitoring of eye disease. The group welcomed the potential significant reduction in waiting time, and felt that the automated eye examination would benefit both patients and clinicians. They highlighted the importance of feedback to patients during automated testing to reassure that the tests were being correctly completed. Concerns were also raised regarding the concept of automation and whether patients would still have an opportunity to interact with their clinician if the device was implemented in clinics.

Discussion

In this study, we performed prospective usability testing of an early binocular OCT prototype in a population of study participants with chronic eye disease, as well as in healthy volunteers. To our knowledge, this is the first system that can perform a comprehensive eye examination – with functional and diagnostic testing in addition to conventional OCT structural imaging – in an end-to-end, automated manner.

Historically, eye examinations in hospital eye clinics have been fragmented, inefficient, and costly.^{4-6,8} The binocular OCT prototype combines many routine tests into one single instrument, with the aim of improving the speed and efficiency of patient flow, in addition to providing reproducible and quantitative data for several aspects of the eye examination. Designed as an automated, patient-operated device, usability testing is indispensable to predict the likelihood of future successful implementation in eye clinics. Moreover, usability testing can identify potential user and device errors, and thus facilitate continued improvement of the device in an iterative process.

Our cohort of first time users was able to complete the full suite of tests using the binocular OCT prototype without any previous training and without any assistance during the examination. Participants commented that the device provided “clear instruc-

tions” and was “easy to use.” The majority of our cohort was familiar with operating common technologies, such as computers and smartphones, and found the prototype to be easier to use in comparison. However, those unfamiliar with technology also rated the device highly on ease of use.

We found that the subjective ratings for ease of use and appeal of the device correlated with ratings for test duration. Participants who perceived the examination took a short time, rated the ease and appeal more positively. However, subjective ratings for duration did not correlate with the total examination time. We observed no significant difference in the total examination time between participants with chronic eye disease and healthy controls. Overall, the median time taken to complete the examination was 702 seconds (11.7 minutes) for participants with chronic eye disease and 638 seconds (10.6 minutes) for healthy controls. By comparison, Callaway et al.⁸ reported a mean clinic time of 28.8 minutes for patients to undergo technician work-up (including history-taking and visual acuity measurement), and acquisition of retinal OCT in the photography suite. In many real-world settings, mean diagnostic testing time is likely to greatly exceed this, particularly in public healthcare settings, which are often overburdened and under resourced. Thus, binocular OCT examination is likely to be more efficacious than current workflows as patients undergo all tests in one location in an automated manner, reducing the time patients spend travelling through the eye clinic. Nonetheless, an important aspect of iterative usability testing is to try to identify examination components that can be further improved in terms of speed.

A small proportion of examinations generated ungradable data. In the case of visual acuity and perimetry measurements, this was a consequence of the user not responding verbally when required. This occurred more frequently for the visual acuity exam in the left eye – the first eye to be tested that required a verbal response. A more complete set of results was obtained for the right eye in these users, and for the subsequent perimetry exam. This may be explained by a learning effect, where the user subsequently understood that the task required a verbal response. As similarly reported in perimetry literature, increased exposure to the device on repeated testing is likely to yield more reliable results and improve all aspects of usability.²⁰ Similarly, we found that the motility exam generated more reliable data when the second eye was fixating. Ungradable data for motility exams was more prevalent in users with chronic eye disease. This was

likely related to poor visual acuity or reduced visual fields, affecting the ability to discern the motility target. In future iterations, this test could be improved by using a high luminance ‘fix and follow’ target.

We obtained good quality automated OCT images of the anterior segment in all except one participant, and good quality posterior segment and vitreous images in the majority of participants. Given our aging populations, and the increasing prevalence of ocular comorbidities, the collection of objective, quantitative OCT data for the whole-eye is likely to be valuable for longitudinal monitoring and detection of eye disease. In addition, the ability to derive OCT measurements of vitreous activity could have applications in monitoring vitritis.²¹ The quality of OCT data generated for both whole-eye imaging and pupillometry was affected by ocular misalignment – a consequence of the user moving their eyes, poor fixation, or blinking during the examination. Our prototype is susceptible to these errors due to the relatively long image acquisition time to complete simultaneous whole-eye OCT (mean 139 seconds). With advances in swept-source OCT laser technology, image acquisition speed is likely to improve in further iterations of the device, rendering the device less vulnerable to artifacts. The quality of OCT imaging, in particular, posterior segment and vitreous imaging also appeared to be affected by large angle strabismus. In these cases, the images for the fixating eye were acceptable, whereas the OCT lasers were unable to image directly through the pupil in the nonfixating heterotropic eye due to the large angle between the pupil plane and the direction of the laser.

To be functional as an automated and interactive device, it is essential that the system is responsive to the user. In the current binocular OCT prototype, this encompasses elements such as voice recognition. In our study, the sensitivity of the VRS was only 64%. This is likely related to the wide variation in articulation, volume, and regional accents of our cohort. Although the examination was undertaken in a quiet room, background noise from the machine itself may have impacted the response heard by the system. As voice recognition technology becomes more sophisticated, the error rate is likely to reduce, but may not be eliminated. Other interactive features, such as registering responses via buttons, in a similar way to many current visual field tests, may be more appropriate for some functional tests. In some populations (e.g., pediatrics), where responses may be unreliable, objective tests using OCT imaging to track responses may be more suitable. For example,

visual acuity could be measured by presenting optokinetic stimuli to the user while simultaneously tracking the movement of the fovea on OCT.

Dissecting the examinations further into the time taken to complete major exam components, we found that delivering instructions took the greatest amount of time, especially in exams where the user was repeatedly reminded to fixate. Participants commented that the central target was unclear – improving its intensity is likely to improve fixation. Although participants found the instructions to be “clear,” articulating instructions in a more succinct manner will reduce the overall examination time – this is one aspect that is particularly likely to benefit from patient and public input.

The main advantage of human operators is the immediate recognition of the discussed errors and artifacts, whereas fully automated devices will require an inherent feedback mechanism to assess the quality of the generated data. This is important for determining whether tests need to be repeated, or if the user requires further or specialist examination beyond the scope of the device. This was also one concern highlighted at our PPI event. Reassurance could be provided to the patient via a visual or audio notification, or indirectly through feedback from a technician working in the clinic; however, this would be most beneficial in parallel with an in-built tool for simultaneous quality control.

User requirements encompass more than clinical effectiveness, and the ergonomics of the device must also be considered. Many of our participants commented that the device interface was physically uncomfortable. Assessing the needs of multiple types of users is essential to encourage continued use of medical devices.¹⁰ For late-stage prototypes, extensive ergonomic testing will be essential prior to commercial release.

In summary, the results of our usability study, and related focus group testing, make it clear that patients are receptive to the concept of an automated eye examination. To be attractive to users, easy to use, and effective at performing automated eye examinations, the system will need to be quick, responsive, and comfortable. For a system that aspires to be fully automated (i.e., operated by the patient without assistance), ongoing patient and public input will be essential to guide the design of the device. Further studies will be required to validate the diagnostic accuracy of each of the tests offered by the system. Once established, binocular OCT will offer objective, quantifiable information about almost every aspect of the eye examination and has the potential to

supersede many traditional but flawed testing methods. It is unlikely that the automated eye examination will be suitable for use in all patients. However, if such a system can replace some aspects of the eye examination, workflows and waiting times are likely to improve, costs are likely to reduce, and clinicians will be able to devote more time to patient care. Ultimately, this will improve the overall experience for both the patient and the clinician, and improve the overall quality of patient care.

Acknowledgments

The authors would like to thank Andi Skilton (PPI Senior Research Associate Lead) and Richard Cable (PPI Research Assistant) from the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital for organizing and facilitating the PPI events.

Dr. Keane has received speaker fees from Heidelberg Engineering, Topcon, Haag-Streit, Allergan, Novartis, and Bayer. He has served on advisory boards for Novartis and Bayer, and is an external consultant for DeepMind and Optos. Dr Mulholland and Prof. Anderson have received travel support from Heidelberg Engineering. Ms Chopra receives studentship support from the College of Optometrists, UK. This study was supported by the Miss Barbara Mary Wilmot Deceased Discretionary Trust.

This report is independent research arising from a Clinician Scientist award (CS-2014-14-023) supported by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Disclosure: **R. Chopra**, None; **P.J. Mulholland**, Heidelberg (R); **A.M. Dubis**, None; **R.S. Anderson**, Heidelberg (R); **P.A. Keane**, Allergan (F), Bayer (F), DeepMind (C), Haag-Streit (F), Heidelberg (F), Novartis (F), Optos (C), Topcon (F)

References

1. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*. 1991;254:1178–1181.
2. Keane PA, Sadda SR. Retinal imaging in the twenty-first century: state of the art and future directions. *Ophthalmology*. 2014;121:2489–2500.
3. News. OCT Estimates of Ophthalmic OCT Market Size and the Dramatic Reduction in Reimbursement Payments. 2012. Available at: www.octnews.org. Accessed January 26, 2017.
4. Murray TG, Tornambe P, Dugel P, Tong KB. Evaluation of economic efficiencies in clinical retina practice: activity-based cost analysis and modeling to determine impacts of changes in patient management. *Clin Ophthalmol*. 2011;5: 913–925.
5. Dugel PU, Tong KB. Development of an activity-based costing model to evaluate physician office practice profitability. *Ophthalmology*. 2011;118: 203–208, e201–e203.
6. Van Vliet EJ, Bredenhoff E, Sermeus W, Kop LM, Sol JC, Van Harten WH. Exploring the relation between process design and efficiency in high-volume cataract pathways from a lean thinking perspective. *Int J Qual Health Care*. 2011;23:83–93.
7. McMullen M, Netland PA. Wait time as a driver of overall patient satisfaction in an ophthalmology clinic. *Clin Ophthalmol (Auckland, NZ)*. 2013;7:1655.
8. Callaway NF, Park JH, Maya-Silva J, Leng T. Thinking lean: improving vitreoretinal clinic efficiency by decentralizing optical coherence tomography. *Retina*. 2016;36:335–341.
9. Walsh AC. Binocular optical coherence tomography. *Ophthalmic Surg Lasers Imaging*. 2011; 42(Suppl):S95–S105.
10. Martin JL, Norris BJ, Murphy E, Crowe JA. Medical device development: the challenge for ergonomics. *Appl Ergon*. 2008;39:271–283.
11. Lee C, Coughlin JF. Perspective: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. *J Prod Innov Manage*. 2015;32:747–759.
12. Wiklund ME, Kendler J, Strohlic AY. *Usability testing of medical devices*. 2nd ed. Boca Raton, FL: CRC Press; 2016.
13. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behav Res Methods Instrum Comput*. 2005;35:379–383.
14. US Food and Drug Administration. Applying human factors and usability engineering to optimize medical device design. Rockville, MA: FDA Center for Devices and Radiological Health; 2011:1–45.

15. Grulkowski I, Liu JJ, Potsaid B, et al. Retinal, anterior segment and full eye imaging using ultrahigh speed swept source OCT with vertical-cavity surface emitting lasers. *Biomed Opt Express*. 2012;3:2733–2751.
16. Dai C, Zhou C, Fan S, et al. Optical coherence tomography for whole eye segment imaging. *Opt Express*. 2012;20:6109–6115.
17. Beck RW, Moke PS, Turpin AH, et al. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol*. 2003;135:194–205.
18. Menon D, Stafinski T. Role of patient and public participation in health technology assessment and coverage decisions. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11:75–89.
19. Shah SG, Robinson I. Benefits of and barriers to involving users in medical device technology development and evaluation. *Int J Technol Assess Health Care*. 2007;23:131–137.
20. Wild JM, Dengler-Harles M, Searle AET, O’Neill EC, Crews SJ. The influence of the learning effect on automated perimetry in patients with suspected glaucoma. *Acta Ophthalmol*. 1989;67:537–545.
21. Keane PA, Balaskas K, Sim DA, et al. Automated analysis of vitreous inflammation using spectral-domain optical coherence tomography. *Transl Vis Sci Technol*. 2015;4(5):4.