BMJ Open

Functional vision tests as clinical trial outcome measures in ophthalmology: a scoping review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-097970
Article Type:	Original research
Date Submitted by the Author:	13-Dec-2024
Complete List of Authors:	Raji, Shabnam; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences Thirunavukarasu, Arun; University of Oxford, Oxford University Clinical Academic Graduate School; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences Taylor, Laura; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences MacLaren, Robert; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital
Keywords:	OPHTHALMOLOGY, Clinical Trial, GENETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Functional vision tests as clinical trial outcome measures in ophthalmology: a scoping review

Authors: Shabnam Raji[1,2*], Arun J Thirunavukarasu[2,3], Laura J Taylor[1,2], Robert E MacLaren[1,2,]

Affiliations:

[1.] Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[2.] Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[3.] Oxford University Clinical Academic Graduate School, University of Oxford, Oxford, United Kingdom

***Corresponding author:** Shabnam Raji; Oxford Eye Hospital, Lower Ground 1 West Wing, John Radcliffe Hospital, Headley Way, Oxford, OX3 9DU, United Kingdom

Email: enquiries@eye.ox.ac.uk Phone: 01865231159

Word count: 4536

Abstract

Objectives: To identify currently available functional vision tests and evaluate their use as clinical trial outcome measures in ophthalmology.

Design: Scoping review using the Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews (PRISMA-ScR) guidelines.

Methods: A literature search was conducted in MEDLINE and Embase (via Ovid) for articles published between 1st January 2003 to 1st August 2024. Additional grey literature was sourced from institutional repositories, conference proceedings and a manual citation search. Article screening was conducted against a pre-defined inclusion criteria by two independent, masked reviewers, with a third reviewer acting as arbiter. The inclusion criteria were English language articles which feature a test assessing functional vision in patients with an ophthalmological disease. Details of source characteristics, test methodology and accessibility and evidence of test validation were collected.

Results: Of 2,995 articles returned by the search, 73 were included and 45 unique tests of functional vision were identified. Diseases affecting the peripheral retina were mainly affected, accounting for 77% (56 out of 73) of the diseases featured in all included studies. Overall, 82% (37 out of 45) functional vision tests reported evidence of statistical validation with varying robustness. Functional vision tests were mapped to domains of orientation and mobility, facial recognition, observer-rated task performance, visual search and driving. Obstacle courses assess vision-guided orientation and mobility, correlate highly with clinical measures of visual function in severe peripheral retinal disease and have been validated for use in clinical trials. Their requirement of physical space and time limits utility in multi-centre trials; equivalent tests leveraging virtual reality and eye tracking technologies are in development. Early iterations of visual search tests to simulated realistic scenes have demonstrated discriminative ability, even in paediatric patients.

Conclusions: Functional vision tests can facilitate research into future novel ophthalmological treatments that prioritises patients in terms of how clinical benefit is defined. The principal barriers to the uptake of these tests are lack of accessibility, low quality validation and that many tests remain early in their development stage. This review captures the current landscape of functional vision tests and serves as a reference for investigators and regulatory bodies to evaluate the suitability of these tests for ophthalmic clinical trials.

Keywords: functional vision, performance-based assessment, outcome measure, mobility, task performance

Strengths and limitations of this study

1. This review provides the first evaluation of functional vision tests in ophthalmology, focusing on their potential as clinical trial outcome measures.
2. A comprehensive grey literature search was performed to minimise the risk of bias.
3. Due to heterogeneity in reported test validation, only a qualitative synthesis of validation data was possible.
4. Incomplete or insufficiently detailed data in the included studies limited the scope of the analysis.

Acknowledgments: None

Author statement: S.R was responsible for the design of the scoping review, conducting the search, screening eligible sources, extracting and analysing data, and the writing and preparation of the manuscript. A.J.T. contributed to conducting the search, screening eligible sources, extracting and analysing data, and the writing and preparation of the manuscript. L.J.T. contributed to screening eligible sources and writing and preparation of the manuscript. R.E.M. contributed to writing and preparation of the manuscript.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: This work was supported by the NIHR Oxford Biomedical Research Centre.

Data Sharing Statement: As no original data was generated as part of this review, there are no datasets available for sharing. All relevant data sources are publicly available and cited within the manuscript.

Introduction

Functional vision tests measure how well individuals can interact with their visual environment (1), and these tests may characterise certain eye diseases better than standard clinical measures of visual function and patient reported outcome measures (2). Functional vision is distinct from visual function which describes the physiological function of the eye and associated visual system, often through contrived clinical tests such as perimetry or visual acuity. Functional vision tests are based on activities of daily living in several domains: mobility, object identification, facial recognition and reading, among others. They output objective scores and can conflate aspects of visual acuity, spatial vision, cognition, colour vision, light sensitivity and adaptation to assess overall function (3). They also consist of relatively complex tasks that assess higher-order visual processing which may offer a more holistic understanding of visual impairment. In this way, they are highly pertinent measures of a patient's overall quality of life and have broad potential application as clinically meaningful outcome measures in ophthalmology clinical trials.

Currently accepted visual function outcome measures in ophthalmology include best-corrected visual acuity, perimetry, full-field stimulus testing, microperimetry and mobility testing (4,5). Despite standardisation, visual acuity remains a gross characterisation of overall vision, insensitive to changes in retinal function away from the fovea and displays poor reliability in patients with visual impairment (6). Standard automated perimetry has been the gold standard for detecting optic nerve damage and has been used effectively as an outcome measure in glaucoma trials (7). However, perimetry is limited by low test-retest reliability, particularly in those with poor steady, central fixation in macular disease and certain oculomotor abnormalities, such as nystagmus (6). Fundus-controlled perimetry, or microperimetry, has gained favour in this regard and has become a key endpoint in several clinical trials (8).

Structural outcome measures in ophthalmology can offer precise, highly reproducible assessments of disease progression and can delineate anatomical biomarkers. However,

these measures may not be applicable if structure and function do not reliably correlate, for instance, where there is amblyopia or a gene defect affecting enzymes of the visual cycle. In these cases, it is unclear how anatomical changes in the eye translate to patient benefit (6).

In other medical specialties, functional tests have already been established as key clinical trial endpoints, such as in stroke medicine and multiple sclerosis (9,10). The US Food and Drug Administration (FDA) have published specific guidelines on patient-centred drug development (11) to prioritise the impact of novel treatments on patients. Similarly, the World Health Organisation's International Classification of Functioning, Disability and Health framework classifies health in terms of functioning and disability in daily life (12). It provides the basis for a more integrated understanding of health, with emphasis on practical function rather than solely biomedical variables. Research is ongoing in ophthalmology clinical trials to align with this framework.

Here, a review was undertaken to identify currently available functional vision tests and evaluate their application as clinical trial outcome measures in ophthalmology.

Methods

A scoping review was selected due to the heterogeneity of articles found in the preliminary literature search, and to allow for more exploratory analysis of functional vision tests as an outcome measure. The review was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (13). A literature search was conducted in MEDLINE and Embase (both via Ovid). Publication dates were restricted from 1st January 2003 to 1st August 2024. A grey literature search was conducted to minimise publication bias and maximise the scope of the review. Grey literature sources included a manual citation search, Google scholar, conference proceedings and the British Library Electronic Theses Online Service (EThOS). The full Boolean search string with combined index and free text terms is detailed in Appendix A.

Duplicates were manually removed by two reviewers. Title and abstract screening, and full text screening was conducted against a pre-defined inclusion criteria by two independent, masked reviewers, with a third reviewer acting as arbiter to resolve disagreement by casting a deciding vote. The inclusion criteria were as follows: 1. Written in the English language; 2. Is a primary research article; 3. Is not a retracted article; 4. Features a test designed for human patients; 5. Test assesses functional vision. Included tests were restricted to those used in patients with an ophthalmological disease. Psychophysical, visual function tests and patient reported outcome measures (PROMs) were excluded. Although an important domain of functional vision, reading tests were excluded in this search as they have been subject to extensive literature review (14).

Key features of the included texts were charted by two independent, masked reviewers with results synthesised by one reviewer. Data on study design, patient characteristics, test methodology, visual function correlates, validity and repeatability evidence and accessibility were extracted. Specifically, articles were searched for evidence of the following: test responsiveness, inter- and intra-rater reliability, test-retest reliability, content, construct and criterion validity. Repeatability and validity data were abstracted to only include statistical values of significance and correlation; purely qualitative statements were excluded. Data visualisation was performed with Microsoft Excel 2024 (Microsoft Corporation, USA) and Inkscape (version 0.92).

Patient and public involvement

There was no direct patient or public involvement in this review.

Results

The initial search yielded 2,665 articles. After screening, a total of 73 texts were included: 67 peer reviewed publications and six conference abstracts. The full search and screening process is shown in Figure 1. Source characteristics of all included studies are summarised

in Table 1. Forty-five unique functional vision tests were identified and listed in Table 2. All functional vision tests were grouped into thematic categories for further analysis. and are illustrated in Figure 2 along a continuum based on their reported ability to measure central or peripheral vision loss. The number of included articles contributing to each category of functional vision test is also shown in Figure 2. Orientation and mobility and observer-rated performance tasks accounted for the highest number of articles found with 25 and 22 respectively. Virtual reality was the least represented with four articles, although all were published within the last five years which predicts an expanding area of research, in line with the growth of new technologies. Figure 3 illustrates the disease of the patient population in the included articles categorised by structure of the eye affected, clinical phenotype and genotype. Functional vision tests were mainly investigated in diseases affecting the peripheral retina which accounted for 77% (56 out of 73) of the diseases featured in all included studies. Rod-cone dystrophies and optic nerve diseases were common, appearing in 37 and 19 articles respectively. Cone-rod dystrophies and macular disease (both inherited and acquired) featured in fewer studies; 6 and 9 respectively. The number of patients within studies ranged from 4 to 192 and the distribution of reported patient age across all studies is displayed in Figure 4. Only 14 out of 73 articles included a paediatric cohort of patient.

A clinical reference standard was identified in 29 out of the 45 functional vision tests. Overall, 37 out of 45 functional vision tests reported evidence of statistical validation, but these were of varying robustness. To date, 7 functional vision tests have been used as outcome measures in 10 separate clinical trials for retinal disease as outlined in Table 3.

Orientation and mobility tests

The most common format of functional vision test was obstacle course, assessing orientation and mobility. Performance on obstacle courses was generally assessed by speed and accuracy, which were often combined to produce an overall score. Metrics of speed include preferred walking speed, percentage of preferred walking speed (PPWS) and course completion time. Accuracy metrics include error number, number of collisions or incidents or

path departure. One study provided more detailed metrics on trajectory analyses and walking initiation time aided by measurement tools such as motion capture systems and inertial sensors (15). Some tests involved videotaped performances which were sent to reading centres for grading to reduce the risk of grader bias (16).

Courses ranged in size from 2.1 x 3.6m to 68 x 1.3m, and were located in purpose-built facilities, hospitals and real indoor rooms (e.g. a cafeteria). All tests identified in this review were performed indoors, although outdoor mobility tests have been described in the literature (17,18). Some tests were performed under multiple luminance levels, ranging from 0.2 to 500 lux, tested in stages to be sensitive to different levels of nyctalopia. No orientation and mobility test exposed patients to acute changes in illumination to test rapid light or dark adaptation, a common difficulty reported in retinitis pigmentosa, perhaps due to safety concerns. Better designed obstacle courses incorporated changes in floor elevation to assess depth perception. If featured in the course, obstacles were commonly made of cardboard or foam and were suspended at various heights. Some tests reported the Weber contrast values and chromaticity coordinates of the obstacles.

Orientation and mobility tests were predominately used on patients with rod-cone dystrophy or glaucoma. As such, the test is suitable for patients with low vision and defects of peripheral vision. The Multi Luminance Mobility Test (MLMT) was used as a primary outcome measure in the landmark clinical trial of voretigene neparvovec (Luxturna) for *RPE65*-related Leber's congenital amaurosis, the first approved gene therapy in ophthalmology (19). The MLMT adopts a binary instead of a continuous scoring system, is performed under seven different luminance levels and demonstrates ceiling effects (20). The low luminance conditions allowed the test to demonstrate sensitivity to changes in disease state; *RPE65* is an enzyme which facilitates dark adaptation of viable rod photoreceptors. It follows that a drug capable of rescuing the function of defective *RPE65* would result in enhanced scotopic vision (19). The success of the MLMT has subsequently inspired the development of several commercial, academic and dedicated facilities offering functional vision testing, to include Streetlab and

Ora (15,21–24). It should however be noted that MLMT is primarily an assessment of scotopic vision augmented by dark adaptation of rods and not necessarily the best method to assess cone function.

Applications of virtual reality technology

Virtual reality can overcome many limitations of orientation and mobility tests. Virtual reality may absolve the need for a physical, homogenously lit room whilst still maintaining a degree of realism (25). As such, it is more accessible for use in multi-centre clinical trials and can overcome the scaling challenges of physical obstacle courses. However, virtual reality-related motion sickness has been reported and as a result, patients may still instructed to walk in physical space to avoid this (26). Commonly used virtual reality headsets include the HTC Vive Pro Eye, Fove 0 and Oculus Rift, which are consumer devices commercially available at a relatively low cost. Proprietary, custom-made software was used on this hardware. Some studies included trackers mounted to patients' head, hands and feet to generate kinematic data (27,28). The technical specifications of VR devices were as follows: display screens were LED or AMOLED, panel sizes ranged from 18.5" to 80", resolution ranged from 1280 × 1440 to 4K, and the horizontal field of view ranged from 89 to 150 degrees. If reported, the display refresh rate was 90Hz. VR tests were conducted binocularly, although recent iterations enable monocular testing (28,29).

Visual search tests

Visual search tasks relate to several domains of functional vision including social interaction, reading, driving and mobility, and have been used to assess patients with various forms of visual impairment (30,31). Visual search may be performed binocularly in front of a display monitor with free head movements or using virtual reality headsets with in-built eye-tracking. Display screen sizes generally range from 17" to 27", although a hemispheric, panoramic screen covering 180 degrees of horizontal visual field has been reported (32). Eye tracking devices included the Tobii EyeX, Tobii 4C, Tobii Pro X3-120, Tobii AB (Tobii technology,

Stockholm, Sweden), HTC Vive trackers (HTC Corp., New Taipei, Taiwan), Oculus Quest Pro (Meta, Burlingame, CA) and the Eyelink II system, Eyelink 1000 system (SR Research Ltd., Ontario, Canada). Proprietary, custom-made software was used on this hardware. Task performance metrics were search time and correct responses.

Visual scenes included geometric shapes hidden in a computer-generated room and everyday objects hidden in photographs of real-world scenes. Psychophysical targets such as optotypes or geometric shapes are not intuitively reflective of real life and studies have shown that a Landolt C search task, compared to object identification in a real photograph, did not differentiate patients from visually healthy controls (33). All scenes found in visual search tasks were two-dimensional and static, and therefore not reflective of dynamic scenes of the real world. The realism and context provided by real world scenes is important as the role of global features and semantic guidance in object search has been well evidenced to influence visual behaviour (34,35). Early iterations of visual search tests in simulated realistic scenes have demonstrated discriminative ability, even in paediatric patients (36,37). One portable tablet-based visual search test was able to discriminate patients with severe diabetic macular oedema from an established normative database (38).

Driving simulator tests

Driving simulator tests have previously been used to evaluate safety, for example, in glaucoma and in the development of new multifocal intraocular lenses, but not treatment effectiveness in clinical trials (39,40). Driving simulator tests have been described in many forms. Moving base driving simulators exist that benefit from a realistic car body and wide-field scene projection but lack the accessibility of other portable simulators (41). Desktop-based driving simulators are low fidelity tests and the lack of real-world consequences from patient error has been reported to influence behaviour by overstating true driving performance (39). The artificial driving scenes in these desktop-based simulators can also cause the patient to subtend a smaller visual angle compared to real life which inadvertently affects the amplitude of saccadic eye moments – a common measure of performance in driving simulator tests.

Observer-rated visual performance tests

Observer-rated visual performance tests are simulated activities of daily living performed in a controlled environment and assessed by an observer. These tests have been shown to correlate with similar tasks performed at home (42). Tested activities include dialling a phone number, reading in reduced illumination or opening a lock with a key. The original Assessment of Function Related to Vision (AFREV) was limited by ceiling effects and was superseded by the Assessment of Disability Related to Vision (ADREV). The Compressed Assessment of Ability Related to Vision (CAARV) is a compressed version of this test requiring only 14 minutes to complete. In 2014, the Functional Low-Vision Observer Rated Assessment (FLORA) was developed as an untimed, home-based test for ultra-low vision patients in the context of a clinical trial for the Argus II retinal prosthesis; a validation study is ongoing (43). A validation study for the more recently developed Instrumental Activities of Daily Living Tools in Very-Low Vision (IADL-VLV) underscores the tests' potential as an outcome measure in vision restoration trials. It was developed using a Delphi consensus procedure, with input from occupational therapists and low-vision experts, maintaining high levels of content validity (44). Novel observer rated performance tests are in development with good repeatability and monocular testing (45). Limitations of potential observer bias were reported, although newer test iterations have incorporated automated scoring methods using sensors attached to objects to detect object displacement (46,47). The tests were also subject to floor and ceiling effects (48) and could place infeasible cognitive and motor demands on patients in line with the activities assessed, limiting their use to a select subset of suitable patients.

Facial recognition tests

The Cambridge Face Memory Test is a validated, computer-based, alternative forced choice task where a target face must be distinguished from two additional unfamiliar faces. The test is freely available online, performed binocularly and has an established normative reference score. The test demonstrates variable discriminative ability when applied to different disease cohorts. In patients with dry AMD, the test was not found to be sensitive to early or

intermediate stages of dry AMD but was able to discriminate individuals with features of late-stage disease such as geographic atrophy (49). Moreover, one study showed no significant correlation between facial discrimination performance and severity of diabetic macular oedema (38). Co-occurring psychiatric illness, neurological damage or neurodevelopmental disorders such as autism affect facial recognition (50) and facial recognition tests are used cautiously in these populations.

Discussion

A functional vision test has been used as a primary outcome measure in a landmark gene therapy clinical trial in ophthalmology. This has set the stage for the development of more unconventional assessments of vision which will be evaluated herein.

Existing functional vision tests in ophthalmology

Orientation and mobility tests were originally used in early clinical trials of retinal prosthesis implants in blind or ultra-low vision patients (51–53). They were favoured as these patients often had remnants of useful vision and light perception that were not captured in standard clinical tests of visual function. As such, these functional tests have relevance in end-stage disease than in early-stage disease where structural changes remain sensitive markers of clinical progression (54). They are useful in measuring low luminance mobility and peripheral vision loss although individuals with localised degeneration may employ head and eye movements to project the visual environment onto islands of functioning retina. In a study with choroideremia patients, no deficit in Multi Luminance Mobility Testing (MLMT) performance was observed due to preserved macular function even in the presence of advanced peripheral disease (55).

Orientation and mobility tests are constrained by several limitations and performance scores can be marred by many sources of error. Firstly, the tests are inherently influenced by patients' confidence and psychological state. For example, a distinguishing feature of orientation and

mobility tests is that an error committed results in an immediate physical response, such as colliding with an obstacle or wall. How individuals negotiate these physical responses varies widely, in terms of risk management or aversion. Furthermore, if patients are aware of being observed or recorded, then the results may be additionally confounded by the Hawthorne effect. The time taken to complete the course is likely to be affected by patient confidence which will improve if a patient has knowledge that they have just received a potentially sight-saving treatment, and thereby conferring a placebo effect. Performance scores may also be confounded by a learning effect and repeated testing is necessary to overcome this which can prove laborious for patients – if patients are instructed to repeatedly walk as fast as possible in multiple course runs to determine maximum performance speed, they may be limited by physical stamina rather than their vision.

Practically, the resources required to develop, conduct and maintain these tests limit their scalability and may preclude their continued use in multi-centre clinical trials. Several orientation and mobility VR tests have been described that offer easy manipulation of the digital visual environment and potentially unlimited course configurations. These tests provide greater optionality in assessing a range of diseases and control of experimental conditions, therefore improving test reproducibility. The automated scoring performance in VR can also reduce assessor bias. Moreover, VR can make an orientation and mobility test into a game by introducing interactive scoring, for example, tests exist that instruct patients to 'tag' obstacles with a controller (28). However, certain limitations arise from the use of VR. The physical VR headset detaches the user from reality and introduces a degree of abstraction to a task. Discrepancies in resolution between the retina and a VR display screen can affect true perception, particularly if the pixel density and resolution is considerably below human acuity (56). VR tests remain in their infancy and require validation in relevant patient populations to ascertain their usability as outcome measures.

VR has also been applied to visual search tests which have demonstrated discriminative ability, even in paediatric patients (36,37). The increased accessibility of eye tracking

technology as consumer devices, evidenced by the 2024 release of the Apple Vision Pro, assures further development of virtual reality and visual search tests. An avenue of future development may be wearable technologies that can monitor real-time visual search in daily life over extended periods of time. A similar application is the EMA approved endpoint of wearable sensors that quantify movement in muscular dystrophy trials (57).

Driving simulator tests have been described in several formats although if patients have been banned from driving due to deteriorating vision, then the psychological impact of being subjected to a driving test should be considered. Not all patients, particularly those with early onset inherited retinal diseases, ever learn to drive, limiting the accessibility of the test.

Inherited retinal diseases: a use case for functional vision tests

Well-designed tests of functional vision relate closely to the prevailing symptoms throughout the natural history of an ophthalmological disease. The symptoms of the disease guide test development to ensure that highly relevant concepts of interest are assessed, and that outcomes remain patient-relevant and pertinent to quality of life. Development and validation is challenging in diseases with variable phenotypes or low prevalence, both exhibited within inherited retinal diseases which collectively represent the leading cause of blindness among working age adults in England and Wales (58). Pathogenic mutations in over 280 genes have been identified as causing inherited retinal disease; each mutation is associated with its own phenotypic characteristics and so patient symptoms can be highly nuanced (59). Selected outcome measures will depend on the underlying disease mechanism and whether a gene-specific or gene-agnostic therapy is developed. The growth of research and development into therapies for these inherited retinal diseases calls for agile innovation in clinical trial outcomes measures to facilitate the arrival of novel gene therapies to market.

Tests that are selected as clinical trial outcome measures should also relate to the region of therapy delivery. For example, in a rod-dominated photoreceptor degeneration the main

symptom may be reduced peripheral vision, but if a drug is administered to rescue remaining photoreceptors at the macula, it is logical to preclude the use of a mobility test that may be insensitive to ultimately measure therapy efficacy. This emphasises the importance in judicious selection of appropriate and effective outcomes measures. Additionally, functional vision tests that are performed binocularly have limited utility in clinical trials featuring monocular interventions, particularly where therapy is delivered to the worse seeing eye—as is common practice—as the better seeing eye tends to predict visual functional ability (60). Ideally, both monocular and binocular assessments should be performed. Assessments of binocular function can provide understanding of overall function, leading to interpretations of quality of life and subsequent health economic analyses.

Several inherited retinal diseases are syndromic with systemic abnormalities that may additionally impair a patient's ability to perform a functional vision test, for reasons other than reduced vision due to retinal degeneration. An example of this is in Joubert's syndrome, whereby mutations in *CEP290* concurrently cause Leber's congenital amaurosis and psychomotor delay with cerebellar malformations, among other ciliopathy-associated abnormalities (61). Performing a functional vision test in these patients with cognitive and physical impairment would be unreliable in measuring changes in retinal function.

Challenges in paediatric validation of functional vision tests

There is a dearth of validated functional vision tests for use in paediatric patients. This is of particular relevance if novel therapies, that are proven to be efficacious in adults, are offered to patients at an earlier age, and in the case of diseases which typically have an early onset of presentation. Examples include Luxturna for *RPE65*-LCA, which used the MLMT in a trial involving adult patients, but for which treatment may be initiated in younger patients as index presentations are frequently early in life. Tests should be optimised for use in children with appropriate modifications to enable clinical trials and post-trial monitoring to capture the

benefit conferred by new treatments. Few functional vision tests identified in this review have been used in children (15,23,27,28,36,37,62–69).

Validation of novel functional vision tests

Treatments such as visual prostheses, stem cell transplantation, gene augmentation and editing therapies, antisense oligonucleotide therapy and optogenetic therapies are being developed at pace for previously untreatable ocular conditions. Progress in the development of these treatments requires validated outcomes. The paucity of validation in functional vision tests is evidenced in Table 2. Few articles reported a full description of test methodology to allow replication, and validation evidence was either absent or fragmented. The absence of an established gold standard test for the measurement of functional vision meant no studies were found to report concurrent validity. Clinically adjudicated reference standards to validate novel tests have been reported in other fields of medicine such as infectious disease diagnostics, and may be useful in the absence of an existing gold standard test (70).

The functional vision tests in this review correlate with clinical measures of visual function to varying degrees of significance and construct validity. The appropriateness of this correlation may be questioned, as functional vision tests measure a distinct aspect of vision rather than acting as surrogate indicators of visual function, raising the issue of whether full validation is required in all cases of test development. It can be said that drawing on the experience of clinicians and patients' perspectives should provide more weight in determining whether test measurements provide useful and clinically meaningful information.

Most current clinical trials adopt a monocular study design to benefit from the contralateral eye as a control but the need for standardised, precise and reliable outcome measures will become critical once treatments are delivered bilaterally (71). Standardised validation of functional vision tests can improve evidence synthesis, the inferential quality of results and enhances comparability of data between clinical trials with treatments for the same disease. It is

reasonable to suggest that functional vision tests should still be validated against standard clinical measures of visual function, but the strength of its validation, or lack thereof, should not solely dictate inclusion as an outcome measure in clinical trials.

In the 1990's, the increase in visual prosthesis development for vision restoration trials led to a greater need for clinically meaningful endpoints. The various centres that developed visual prosthesis used different efficacy measurements, making cross-comparison challenging. This led to the International Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials (HOVER) taskforce where experts from around the world collaboratively formed guidance to measure visual function in vision restoration clinical trials (72). Most functional vision tests found in this review have been applied to inherited retinal diseases, as shown in Table 3, yet there is currently no such directive for inherited retinal disease. Novel clinical trial outcome measures would benefit from being guided by consensus-building to retain standardisation. Stakeholders involved in such consensus-building should include patients, advocacy groups, clinical trial sponsors, disease experts, regulatory agencies and experts in the functional vision construct being measured.

Limitations

This study has limitations. Functional vision tests are in development globally and the regional cultural differences in activities of daily living were not explored in this review, nor were the sources of funding for centres developing functional vision tests.

Conclusion

Functional vision tests can facilitate research into future novel ophthalmological treatments that prioritises patients in terms of how clinical benefit is defined. The principal barriers to the uptake of these tests are lack of accessibility, low quality validation and that many tests remain early in their development stage. This review captures the current landscape of functional

vision tests and serves as a reference for investigators and regulatory bodies to evaluate the suitability of these tests for ophthalmic clinical trials.

References

1. Bennett CR, Bex PJ, Bauer CM, Merabet LB. The Assessment of Visual Function and Functional Vision. Semin Pediatr Neurol. 2019 Oct;31:30–40.

2. Warrian KJ, Altangerel U, Spaeth GL. Performance-based Measures of Visual Function. Survey of Ophthalmology. 2010 Mar 1;55(2):146–61.

3. High-Level Visual Processing: Cognitive Influences. In: Principles of Neural Science, Fifth Edition [Internet]. New York, NY: McGraw-Hill Education; 2014 [cited 2024 Jan 17]. Available from: neurology.mhmedical.com/content.aspx?aid=1101680178

4. Thompson DA, Iannaccone A, Ali RR, Arshavsky VY, Audo I, Bainbridge JWB, et al. Advancing Clinical Trials for Inherited Retinal Diseases: Recommendations from the Second Monaciano Symposium. Trans Vis Sci Tech. 2020 Jun 3;9(7):2.

5. Shi LF, Hall AJ, Thompson DA. Full-field stimulus threshold testing: a scoping review of current practice. Eye [Internet]. 2023 Jul 13 [cited 2023 Jul 18]; Available from: https://www.nature.com/articles/s41433-023-02636-3

6. Schmetterer L, Scholl H, Garhöfer G, Janeschitz-Kriegl L, Corvi F, Sadda SR, et al. Endpoints for clinical trials in ophthalmology. Progress in Retinal and Eye Research. 2023 Nov;97:101160.

7. Weinreb RN, Kaufman PL. The Glaucoma Research Community and FDA Look to the Future: A Report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium *. Invest Ophthalmol Vis Sci. 2009 Apr 1;50(4):1497.

8. Taylor LJ, Josan AS, Jolly JK, MacLaren RE. Microperimetry as an Outcome Measure in *RPGR*- associated Retinitis Pigmentosa Clinical Trials. Trans Vis Sci Tech. 2023 Jun 9;12(6):4.

9. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke. 2007 Mar;38(3):1091–6.

10. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Mult Scler. 2017 Apr;23(5):704–10.

11. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making | FDA [Internet]. [cited 2024 Jan 10]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory

12. World Health Organization. International classification of functioning, disability and health: children and youth version: ICF-CY. 2007;322.

13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467–73.

14. Rubin GS. Measuring reading performance. Vision Research. 2013 Sep;90:43–51.

15. Sahel JA, Grieve K, Pagot C, Authié C, Mohand-Said S, Paques M, et al. Assessing Photoreceptor Status in Retinal Dystrophies: From High-Resolution Imaging to Functional Vision. American Journal of Ophthalmology. 2021 Oct;230:12–47.

16. Chung DC, McCague S, Yu ZF, Thill S, DiStefano-Pappas J, Bennett J, et al. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. Clin Experiment Ophthalmol. 2018;46(3):247–59.

17. Marron JA, Bailey IL. Visual factors and orientation-mobility performance. Am J Optom Physiol Opt. 1982 May;59(5):413–26.

18. Kuyk T, Elliott JL, Fuhr PS. Visual correlates of mobility in real world settings in older adults with low vision. Optom Vis Sci. 1998 Jul;75(7):538–47.

19. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. The Lancet. 2017 Aug 26;390(10097):849–60.

20. Maguire A.M., Russell S., Wellman J.A., Chung D.C., Yu Z.-F., Tillman A., et al. Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation-Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. Ophthalmology. 2019;126(9):1273–85.

21. Azoulay-Sebban L, Zhao Z, Zenouda A, Lombardi M, Gutman E, Brasnu E, et al. Correlations Between Subjective Evaluation of Quality of Life, Visual Field Loss, and Performance in Simulated Activities of Daily Living in Glaucoma Patients. Journal of Glaucoma. 2020 Oct;29(10):970–4.

22. Adrian J, Authié C, Lebrun J, Lombardi M, Zenouda A, Gutman E, et al. Driving behaviour and visual compensation in glaucoma patients: Evaluation on a driving simulator. Clinical & Experimental Ophthalmology. 2022;50(4):420–8.

23. Kumaran N, Ali RR, Tyler NA, Bainbridge JWB, Michaelides M, Rubin GS. Validation of a Vision-Guided Mobility Assessment for RPE65-Associated Retinal Dystrophy. Transl vis sci technol. 2020;9(10):5.

24. Shapiro A., Corcoran P., Sundstrom C., Angjeli E., Rodriguez J.D., Abelson M.B., et al. Development and validation of a portable visual navigation challenge for assessment of retinal disease in multi-centered clinical trials. Invest Ophthalmol Vis Sci [Internet]. 2017;58(8). Available from: http://iovs.arvojournals.org/article.aspx?articleid=2638048 PT - Conference Abstract

25. Li Y, Gunasekeran DV, RaviChandran N, Tan TF, Ong JCL, Thirunavukarasu AJ, et al. The next generation of healthcare ecosystem in the metaverse. Biomedical Journal. 2023 Dec;100679.

26. Lam AKN, To E, Weinreb RN, Yu M, Mak H, Lai G, et al. Use of Virtual Reality Simulation to Identify Vision-Related Disability in Patients With Glaucoma. JAMA Ophthalmol. 2020 May;138(5):490–8.

27. Aleman TS, Miller AJ, Maguire KH, Aleman EM, Serrano LW, O'Connor KB, et al. A Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations: Testing a Proof-of-Concept After Gene Therapy. OPTH. 2021 Mar;Volume 15:939–52.

28. Bennett J, Aleman EM, Maguire KH, Nadelmann J, Weber ML, Maguire WM, et al. Optimization and Validation of a Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations. Trans Vis Sci Tech. 2023 Jan 30;12(1):28.

29. Authie C., Poujade M., Talebi A., Defer A., Zenouda A., Coen C., et al. Development and validation of a novel mobility test for IRDs, from reality to virtual reality. medRxiv [Internet]. 2023; Available from: https://www.medrxiv.org/

30. Senger C, Margarido MRRA, De Moraes CG, De Fendi LI, Messias A, Paula JS. Visual Search Performance in Patients with Vision Impairment: A Systematic Review. Current Eye Research. 2017 Nov 2;42(11):1561–71.

31. Fuhr P.S., Liu L., Kuyk T.K. Relationships between feature search and mobility performance in persons with severe visual impairment. Optom Vis Sci. 2007;84(5):393–400.

32. Thibaut M, Tran THC, Szaffarczyk S, Boucart M. Impact of age-related macular degeneration on object searches in realistic panoramic scenes. Clin Exp Optom. 2018;101(3):372–9.

33. Smith ND, Crabb DP, Garway-Heath DF. An exploratory study of visual search performance in glaucoma. Ophthalmic Physiologic Optic. 2011 May;31(3):225–32.

34. Biederman I, Mezzanotte RJ, Rabinowitz JC. Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology. 1982 Apr;14(2):143–77.

35. Torralba A, Castelhano MS, Oliva A, Henderson JM. Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search.

36. Zhang X, Manley CE, Micheletti S, Tesic I, Bennett CR, Fazzi EM, et al. Assessing visuospatial processing in cerebral visual impairment using a novel and naturalistic static visual search task. Res Dev Disabil. 2022;131(8709782):104364.

37. Manley CE, Bennett CR, Merabet LB. Assessing Higher-Order Visual Processing in Cerebral Visual Impairment Using Naturalistic Virtual-Reality-Based Visual Search Tasks. Children (Basel). 2022;9(8).

38. Taylor DJ, Alquiza PJ, Jones PR, Wilson I, Bi W, Sim DA, et al. Tablet-based tests of everyday visual function in a diabetic macular oedema (DME) clinic waiting area: A feasibility study. Ophthalmic Physiologic Optic. 2023 Dec 22;opo.13261.

39. Csaky K, Ferris F III, Chew EY, Nair P, Cheetham JK, Duncan JL. Report From the NEI/FDA Endpoints Workshop on Age-Related Macular Degeneration and Inherited Retinal Diseases. Investigative Ophthalmology & Visual Science. 2017 Jul 21;58(9):3456–63.

40. Drum BA, Rorer EM, Calogero D. Night Driving Performance in the National Advanced Driving Simulator vs. Clinical Tests of Vision. Investigative Ophthalmology & Visual Science. 2007 May 10;48(13):1511.

41. Kubler TC, Kasneci E, Rosenstiel W, Heister M, Aehling K, Nagel K, et al. Driving with Glaucoma: Task Performance and Gaze Movements. Optom Vis Sci. 2015;92(11):1037–46.

42. West SK, Rubin GS, Munoz B, Abraham D, Fried LP. Assessing functional status: correlation between performance on tasks conducted in a clinic setting and performance on the same task conducted at home. The Salisbury Eye Evaluation Project Team. J Gerontol A Biol Sci Med Sci. 1997 Jul;52(4):M209-217.

43. Second Sight Medical Products. A Validation of the Functional Low-Vision Observer Rated Assessment (FLORA-20) for Profoundly Blind Individuals [Internet]. clinicaltrials.gov; 2023 Nov [cited 2024 Jan 1]. Report No.: NCT04312763. Available from: https://clinicaltrials.gov/study/NCT04312763

44. Finger RP, McSweeney SC, Deverell L, F. O'Hare, Bentley SA, Luu CD, et al. Developing an Instrumental Activities of Daily Living Tool as Part of the Low Vision Assessment of Daily Activities Protocol. Investigative Ophthalmology & Visual Science. 2014 Dec 29;55(12):8458–66.

45. Peterson C.L., Htoon H.M., Man R., Lamoureux E.L., Fenwick E., Cheung G.C.M., et al. Development and Validation of Performancebased Assessments of Daily Living Tasks for Age-related Macular Degeneration. Invest Ophthalmol Vis Sci. 2023;64(8):5042.

46. Carlson M, Kim S, Batabyal S, Aldape M, Mohanty S. Design and Development of a Multi-luminance Shape Discrimination Test for Assessing Object Recognition Ability of Visually Impaired Individuals. 2024;

47. Rubenstein LZ, Schairer C, Wieland GD, Kane R. Systematic biases in functional status assessment of elderly adults: effects of different data sources. J Gerontol. 1984 Nov;39(6):686–91.

48. Terheyden JH, Fink DJ, Pondorfer SG, Holz FG, Finger RP. Instrumental Activities of Daily Living Tools in Very-Low Vision: Ready for Use in Trials? Pharmaceutics. 2022 Nov;14(11):2435.

49. Higgins BE, Taylor DJ, Bi W, Binns AM, Crabb DP. Novel computer-based assessments of everyday visual function in people with age-related macular degeneration. Lewin AS, editor. PLoS ONE. 2020 Dec 7;15(12):e0243578.

50. O'Hearn K, Schroer E, Minshew N, Luna B. Lack of developmental improvement on a face memory task during adolescence in autism. Neuropsychologia. 2010 Nov 1;48(13):3955–60.

51. Geruschat D.R., Bittner A.K., Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. Optom Vis Sci. 2012;89(9):1308–15.

52. Xu H, Zhong X, Pang C, Zou J, Chen W, Wang X, et al. First Human Results With the 256 Channel Intelligent Micro Implant Eye (IMIE 256). Trans Vis Sci Tech. 2021 Oct 27;10(10):14.

53. Petoe MA, Titchener SA, Kolic M, Kentler WG, Abbott CJ, Nayagam DAX, et al. A Second-Generation (44-Channel) Suprachoroidal Retinal Prosthesis: Interim Clinical Trial Results. Transl vis sci technol. 2021;10(10):12.

54. Abdalla Elsayed MEA, Taylor LJ, Josan AS, Fischer MD, MacLaren RE. Choroideremia: The Endpoint Endgame. IJMS. 2023 Sep 20;24(18):14354.

55. Chung DC, McCague S, Yu ZF, Thill S, DiStefano-Pappas J, Bennett J, et al. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies: Multi-luminance mobility test. Clinical & Experimental Ophthalmology. 2018 Apr;46(3):247–59.

56. Drascic D, Milgram P. Perceptual issues in augmented reality. In: Bolas MT, Fisher SS, Merritt JO, editors. San Jose, CA; 1996 [cited 2024 Jan 16]. p. 123–34. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1014000

57. Servais L, Yen K, Guridi M, Lukawy J, Vissière D, Strijbos P. Stride Velocity 95th Centile: Insights into Gaining Regulatory Qualification of the First Wearable-Derived Digital Endpoint for use in Duchenne Muscular Dystrophy Trials. J Neuromuscul Dis. 9(2):335–46.

58. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ Open. 2014 Feb 13;4(2):e004015.

59. RetNet: Summaries of Genes and Loci Causing Retinal Diseases [Internet]. [cited 2024 Feb 18]. Available from: https://web.sph.uth.edu/RetNet/sum-dis.htm

60. Kulkarni KM, Mayer JR, Lorenzana LL, Myers JS, Spaeth GL. Visual Field Staging Systems in Glaucoma and the Activities of Daily Living. American Journal of Ophthalmology. 2012 Sep;154(3):445-451.e3.

61. Valente EM, Silhavy JL, Brancati F, Barrano G, Krishnaswami SR, Castori M, et al. Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome. Nat Genet. 2006 Jun;38(6):623–5.

62. Roman AJ, Cideciyan AV, Wu V, Mascio AA, Krishnan AK, Garafalo AV, et al. Mobility test to assess functional vision in dark-adapted patients with Leber congenital amaurosis. BMC Ophthalmol. 2022 Dec;22(1):266.

63. Maguire AM, Russell S, Chung DC, Yu ZF, Tillman A, Drack AV, et al. Durability of Voretigene Neparvovec for Biallelic RPE65-Mediated Inherited Retinal Disease. Ophthalmology. 2021 Oct;128(10):1460–8.

64. Jacobson SG, Cideciyan AV, Sumaroka A, Roman AJ, Charng J, Lu M, et al. Outcome Measures for Clinical Trials of Leber Congenital Amaurosis Caused by the Intronic Mutation in the CEP290 Gene. Investigative Ophthalmology & Visual Science. 2017 May 16;58(5):2609–22.

65. Velikay-Parel M, Ivastinovic D, Koch M, Hornig R, Dagnelie G, Richard G, et al. Repeated mobility testing for later artificial visual function evaluation. J Neural Eng. 2007;4(1):S102-7.

66. Geruschat DR, Flax M, Tanna N, Bianchi M, Fisher A, Goldschmidt M, et al. FLORA™: Phase I development of a functional vision assessment for prosthetic vision users. Clinical and Experimental Optometry. 2015 Jul 1;98(4):342–7.

67. Pierce EA, Aleman TS, Jayasundera KT, Ashimatey BS, Kim K, Rashid A, et al. Gene Editing for CEP290-Associated Retinal Degeneration. New England Journal of Medicine. 2024 Jun 5;390(21):1972–84.

68. Pierce EA, Ashimatey BS, Jayasundera T, Hoyng C, Lam BL, Lorenz B, et al. Twelve-month Natural History Study of Centrosomal Protein 290 (CEP290)-associated Inherited Retinal Degeneration. Ophthalmology Science. 2024 Sep 1;4(5):100483.

69. Russell SR, Drack AV, Cideciyan AV, Jacobson SG, Leroy BP, Van Cauwenbergh C, et al. Intravitreal antisense oligonucleotide sepofarsen in Leber congenital amaurosis type 10: a phase 1b/2 trial. Nat Med. 2022 May;28(5):1014–21.

70. Patel R, Tsalik EL, Evans S, Fowler VG, Doernberg SB, for The Antibacterial Resistance Leadership Group. Clinically Adjudicated Reference Standards for Evaluation of Infectious Diseases Diagnostics. Clinical Infectious Diseases. 2023 Mar 4;76(5):938–43.

71. MacLaren RE. Benefits of gene therapy for both eyes. The Lancet. 2016 Aug 13;388(10045):635–6.

72. Ayton LN, Joseph F. Rizzo III, Bailey IL, Colenbrander A, Dagnelie G, Geruschat DR, et al. Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials: Recommendations from the International HOVER Taskforce. Translational Vision Science & Technology [Internet]. 2020 Jul [cited 2023 Aug 11];9(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426586/

Titles and legends to figures and tables

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) flow diagram of the study selection process

Figure 2. Number of included articles (n=73) contributing to each category of functional vision test. Six categories of functional vision test ordered on a continuum based on reported ability to measure central or peripheral vision loss. Exemplar fundus autofluorescence images depicting severe peripheral retinal degeneration due to *RPE65*-associated Leber's Congenital Amaurosis (left) and discrete central atrophy within the macula due to *RPGR*-associated cone dystrophy (right). In some severe retinal degenerations, such as end-stage Leber's Congenital Amaurosis, extensive peripheral degeneration encroaches centrally leading to complete loss of light perception.

Figure 3. Disease of patient population in included articles (n = 73) categorised by the structure of the eye affected, clinical phenotype and, where reported, genotype.

Figure 4. Reported age of patient population assessed with functional vision tests. The dashed line demarcates age 18, below which signifies paediatric testing. Five articles were omitted as no age data was available. Note that there are few studies testing paediatric patient populations and even fewer suitable for pre-school age children.

Table 1. Summary source characteristics of all included studies

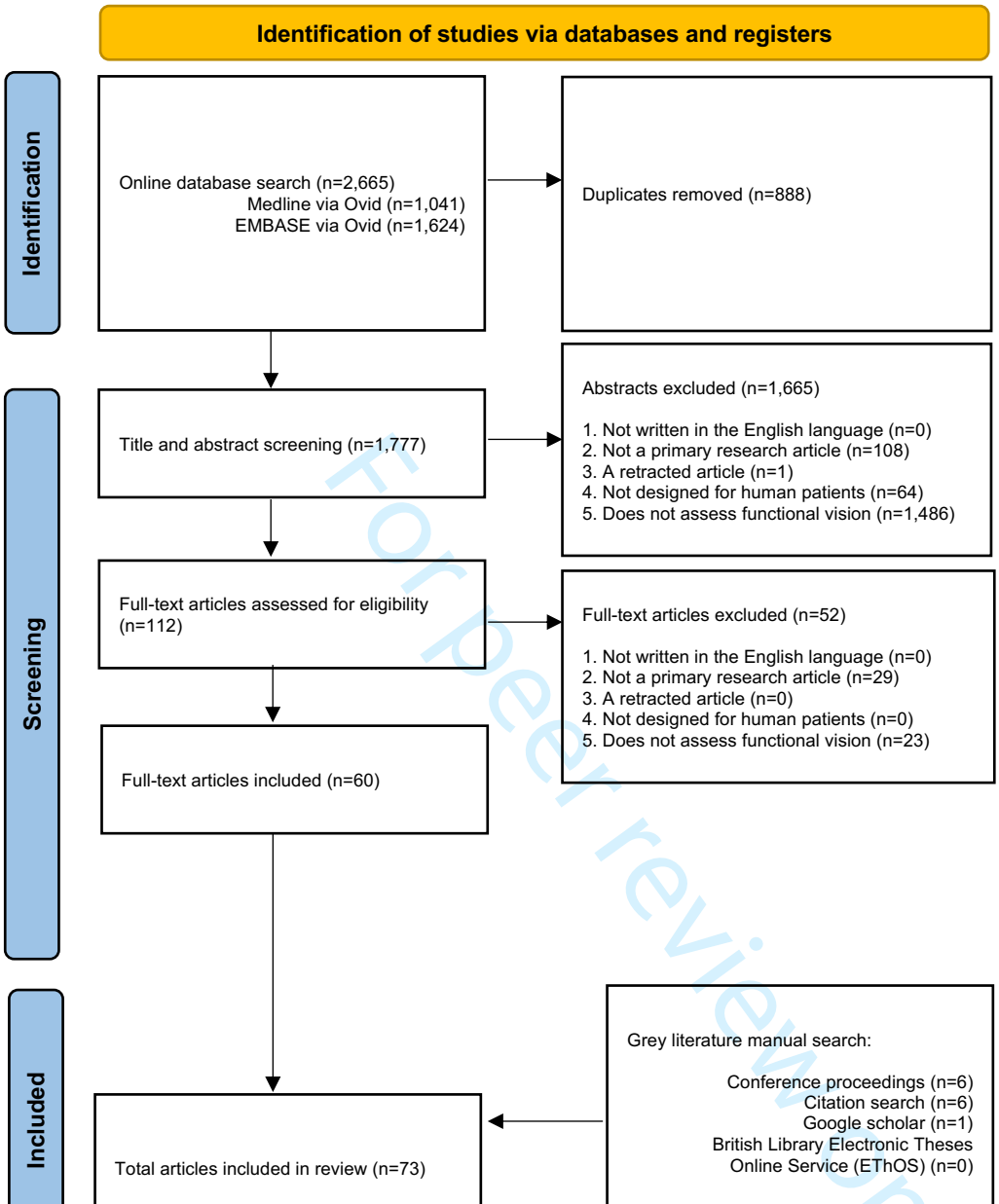
Publication year	**Number of studies**
2005-2010	8
2011-2015	15
2016-2020	24
2021-2024	26
Study design	
Interventional study	
Phase I/II randomised controlled trial	3
Phase III randomised controlled trial	1
Pilot/Feasibility	1
Observational studies	
Cross-sectional	49
Case series	10
Case-control	2
Cohort	1
Conference proceedings	
Abstract	6
Country of institutional affiliation [a]	
North America	38
Europe	24
Asia	4
Oceania	4
Middle East	2
South America	1
Africa	0

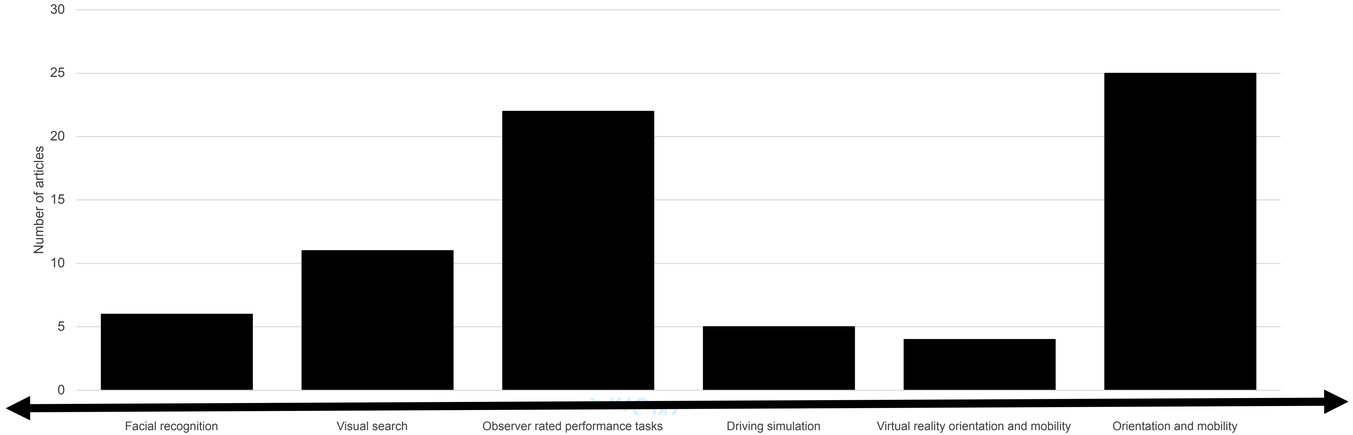
Table 2. Patient population, reference standard, test outcomes, and repeatability and validity data of all included studies featuring a functional vision test

(Uploaded as a separate document due to landscape format)

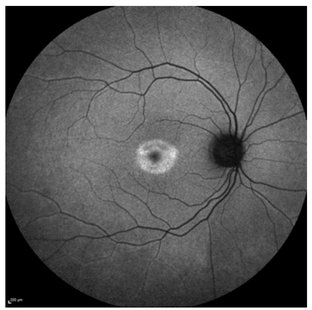
Table 3. Functional vision tests used as clinical trial outcome measures

Name of functional vision test	Disease population	ClinicalTrials.gov Identifier	Type of outcome measure
Multi Luminance Mobility Test (MLMT)	*RPE65*-related Leber's congenital amaurosis	NCT00999609	Primary
	NR2E3 and RHO-related retinitis pigmentosa	NCT05203939	Efficacy
The Functional Low-Vision Observer Rated Assessment (FLORA for Argus II prosthesis)	End-stage retinitis pigmentosa	NCT02303288; NCT03406416	Primary; Secondary
Low Luminance Mobility Testing (LLMT)	Retinitis pigmentosa	NCT03073733	Secondary
Visual Navigation Challenge (Ora-VNC)	*CEP290*-related Leber's congenital amaurosis	NCT03140969; NCT03872479	Secondary
Multi-Luminance Y-Mobility Test (MLYMT)	Retinitis pigmentosa	NCT04945772	Secondary
Vision-guided mobility assessment	*RPE65*-related retinal dystrophy	NCT02781480	Secondary
Orientation and mobility for Argus II prosthesis	End-stage retinitis pigmentosa	NCT00407602	Secondary

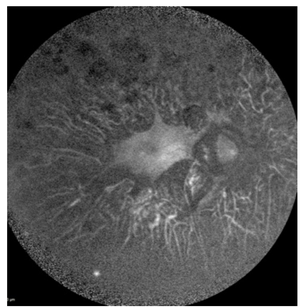




Central vision loss



Peripheral vision loss



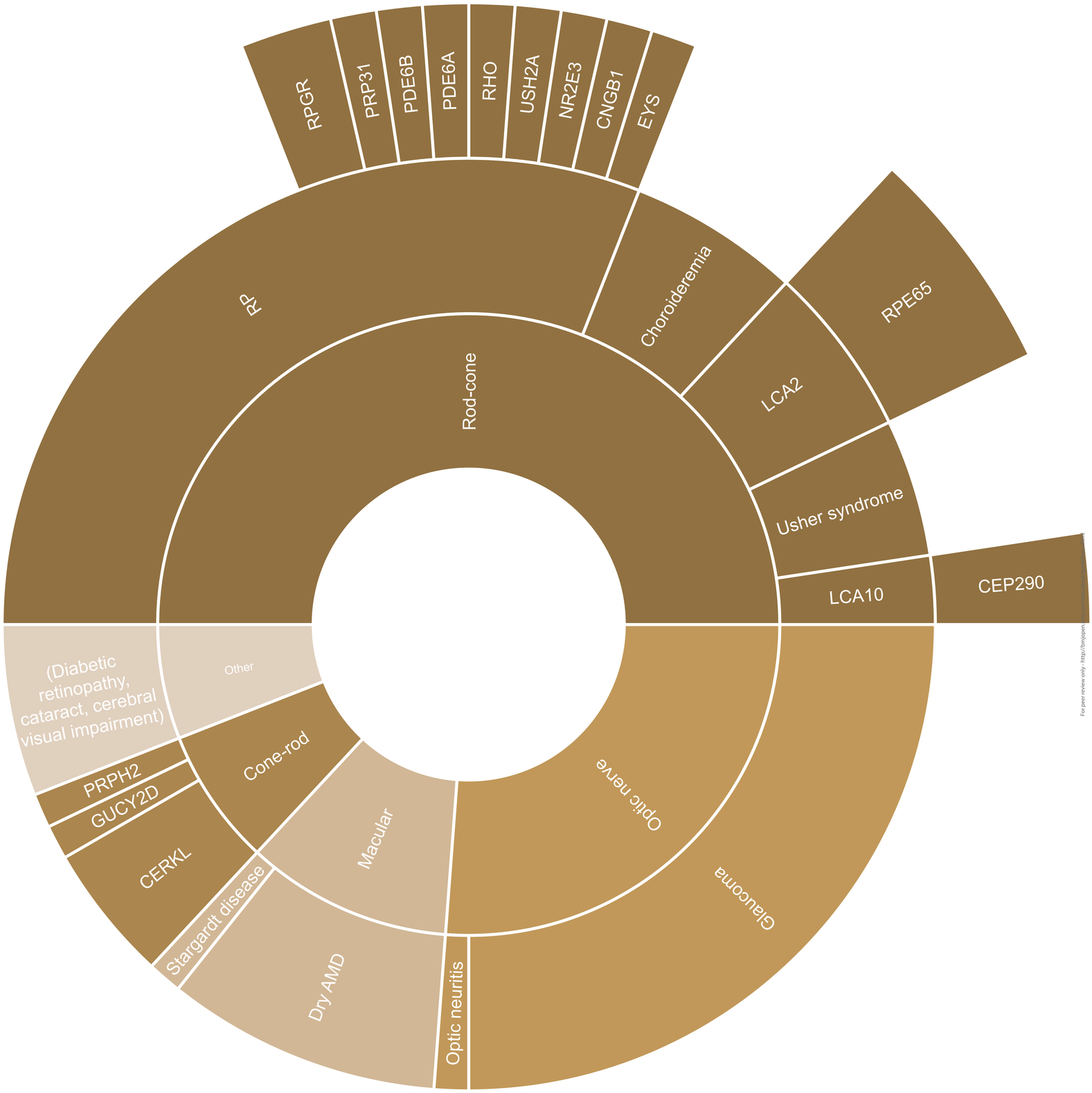


Table 2. Patient population, reference standard, test outcomes, and repeatability and validity data of all included studies featuring a functional vision test

Citation	Patient population	Functional vision test	Reference standard(s)	Test outcome(s)	Reported repeatability and validity data
Orientation and mobility (O&M)					
Roman et al., 2022	10 patients with *GUCY2D*- and *CEP290*-associated Leber's congenital amaurosis	**Mobility test for rod-mediated vision**	VA; FST	Navigation success over a fixed number of trials; Travel duration	**Content validity** - Mobility demonstrated a linear relationship with FST. No correlation between VA and mobility **Construct validity** - No significant difference between controls and patients in suprathreshold transit time (p=0.63). At threshold and dimmer luminance levels transit times increased for both patients and normal subjects.
Sahel et al., 2021	25 patients with retinitis pigmentosa and *RPE65*-associated Leber's congenital amaurosis	**StreetLab mobility course**	VA; VF; CS; Dark adaptation	Course completion time; PWS; PPWS; Number of collisions; Walking initiation time; trajectory analyses/segments; Distance travelled	**Construct validity** – Patients performed worse than controls for PWS, PPWS, number of collisions and walking initiation time under both low and high illumination
Bertaud et al., 2021	22 patients with glaucoma				**Construct validity** – No difference in mobility performance between patients and controls under photopic luminance. Under glare conditions, PWS and PPWS were significantly lower in patients than controls (p=0.049 and p=0.038 respectively). Mobility time was significantly longer in patients than controls (p=0.046). Distance travelled, mobility incidents, and trajectory segmentations not significantly different between patients and controls.
Chung et al., 2018; Maguire et al., 2019 Maguire et al., 2021	19 patients with *RPE65*-associated Leber's congenital amaurosis 19 patients with *RPE65*-associated Leber's congenital amaurosis	**Multi-Luminance Mobility Test (MLMT)**	VA; VF; FST (white light)	MLMT binocular change score (number of collisions and time to navigate course)	**Content validity** - Variable correlation of accuracy score with quality-of-life questionnaire (r=−0.54 to −0.7). Correlation of mean accuracy score with VA ranged from −0.75 to 0.86. Correlation between mean accuracy score and total degrees of visual field ranged from −0.53 to −0.53. **Construct validity** - Able to distinguish controls from patients. **Repeatability-** High inter-grader agreement for scoring (Cohen's kappa=97.9%). High concordance between scores at baseline visits ranging from 86% to 92%. **Sensitivity to change** - Over 1-year observation period controls had an MLMT change score of 0, representing no change and 20 patients had an MLMT change score of 0. Few patients had an MLMT change score of −1 or −2 (i.e. a worsening).
Lam et al., 2024*	18 patients with *NR2E3 and RHO*-associated retinitis pigmentosa			MLMT monocular change score	**Construct validity –** 6 out of 7 *RHO* patients had stable or improved MLMT scores, including 2 patients that demonstrated a 3-luminance level improvement. All autosomal dominant-*NR2E3* patients had no improvement
Kammer et al., 2021*	20 patients with retinitis pigmentosa	**Low Luminance Mobility Test (LLMT)**	VA; CS; VF; VA LV VFQ-48	Critical Illumination Level; Maximum Step Speed score	**Content validity -** All visual function measures significantly related to Critical Illumination Level in a multiple regression model, R^2 =0.75 (p=0.004) **Construct validity** - Able to distinguish controls from patients. **Repeatability** - No change in Critical Illumination Level between test sessions for 75% of patients. Inter-rater and intra-rater grading biases close to zero and no significant differences between graders (p>0.05).
Xu et al., 2021	5 patients with retinitis pigmentosa	**Orientation and mobility test (256 Channel Intelligent Micro Implant Eye implant)**		Effort; Average completion time	Not reported
Boyer et al., 2023*	27 patients with advanced retinitis pigmentosa	**Multi-Luminance Y-Mobility Test (MLYMT)**			Not reported

Kumaran et al., 2020	19 patients with *RPE65*-related retinal dystrophy	**Vision-guided mobility assessment**	VA; CS; VF; FST; Impact of Vision Impairment Questionnaire	Completion time; error number; walking speed; PPWS	**Repeatability** – Large repeatability coefficient of 1.30 m/s **Content validity -** Mean retinal sensitivity (p=0.022) and total hill of vision (p=0.022) predicted walking speed with significance. No correlation between walking speed and VA (p=0.340) or CS (p=0.433) **Criterion validity** - Walking speed approached significance (p=0.052) and was positively associated with affected subjects' perceived difficulties with mobility
Jacobson et al., 2017	22 patients with *CEP290*-associated Leber's congenital amaurosis	**Mobility performance task**	FST	Number of patient incidents (obstacles/wall bumps or reorientations)	**Content validity** – Correlation between mobility score and VA (p =0.002).
Alshaghthrah et al., 2014; Al Saqr et al., 2017	20 patients with retinitis pigmentosa	**Portable mobility course**	VA; CS	PPWS; Collision score	**Content validity** - Significant correlation between CS and collision incidences (p=0.03). No significant correlation between CS and mobility scores (p > 0.05). **Repeatability** - PPWS scores not significantly different (p>0.05) on repeat testing. Collision incidences significantly lower at the second visit (p=0.012). Agreement in collision incidences between the two visits suggestive of no learning effect.
Shapiro et al., 2017*;	Inherited retinal disease	**Ora-VNC (Visual Navigation Challenge)**		Navigation time; Composite score	**Construct validity** - Navigation times for controls, mild and severe retinitis pigmentosa were significantly different across all light levels (p<0.05) and between groups (p < 0.05).
Pierce et al., 2024; Pierce et al., 2024	26 patients with *CEP290*-associated retinal dystrophy				**Content validity** – Composite score was correlated with LCVA, white light FST and red light FST in both eyes, and blue light FST in the better eye (p < 0.05). **Construct validity** – Nine participants (64%) showed a meaningful improvement from baseline. **Repeatability –** Mean test-retest variability from baseline to retest in the worse eye was 0.6 for VNC composite score (95% confidence interval = −0.1, 1.3). **Sensitivity to change –** Mean change from baseline to 6 months test in the worse eye was -0.1 (-1.2, 1.0).
Russell et al., 2022	11 patients with *CEP290*-associated Leber congenital amaurosis				**Construct validity** - Mean (±standard deviation) improvement in composite score was +2.50±3.118 in treated eyes compared to +1.75±2.383 in untreated eyes (p=0.10). A greater improvement in the composite score from baseline to month 12 was seen in the low dose group (+4.00±3.114 and +2.67±2.714 for treated and untreated eyes, respectively) compared to the higher dose group (+0.25±1.323 and +0.38±0.750, respectively).
Ivanov et al., 2016	25 patients with retinitis pigmentosa	**Natural environment walking task with eye tracking**		PPWS; Number of obstacle collisions; Eye position variability	**Construct validity** - Average PPWS for controls (97%) was higher than all other patient groups.
Ikeda et al., 2015	8 patients with retinitis pigmentosa	**Walking test**		Number of trial failures; Time taken to reach goal	Not reported
Nau et al., 2014	36 patients with low vision	**Obstacle course for BrainPort device**		PPWS; Percentage of obstacle collisions	Not reported
Geruschat et al., 2012	8 patients with advanced retinitis pigmentosa	**Orientation and mobility assessment in retinal prosthesis**	VA; VF	Course completion time; Obstacle contacts	**Construct validity** – Significantly increased obstacle contacts between subjects with worse and those with better VA and VF. No significant difference in course completion time
Kiser et al., 2008	22 patients with age-related macular degeneration	**Mobility obstacle course**		Course completion time; Obstacle contacts	Not reported
Fuhr et al., 2007	44 patients with severe visual impairment	**High density obstacle course**		Course completion time; Obstacle contacts	**Construct validity** – Longer course completion time in patients than age matched controls with significant group effect (p<0.0005). Patients made more obstacle contacts than controls. Analyses of mean number of obstacle contacts showed a significant group effect (p=0.001).
Velikay Parel et al., 2007	10 patients with retinitis pigmentosa, Usher syndrome and optic nerve atrophy	**Mobility assessment**	VA; VF	Average speed; Obstacle contacts	**Content validity** - VA and VF had no significant effect on passing time (p=0.08 and p=0.23 respectively) **Construct validity** - Average passing times between the groups were significantly different (p=0.03). No significant difference in the average number of contacts between groups (p=0.15)
Virtual reality O&M					

bmjopen-2024-097970 on 27 May 2025. Downloaded from

Authie et al., 2023	30 patients with retinitis pigmentosa	**MObility Standardised Test (MOST)**	VA; CS; VF; Dark adaptation	Trial duration; Number of collisions; Number of steps and flags touched; Entries in the dead end; Course redirections	**Construct validity** - Demonstrates discrimination between patients and controls (accuracy larger than 95% in all conditions) and between early and late stages of the disease (mean accuracy of 82.3%). **Content validity** - Average performance score strongly correlated with VA, CS and VF. **Reliability** - Highly reproducible (intraclass correlation coefficient>0.98) and reliable (VR and real-life correlation r=0.98)
Aleman et al., 2021; Bennett et al., 2023	29 patients with choroideremia, *RPE65*-associated Leber's congenital amaurosis, *EYS-, CNGB1-, NR2E3-, RPGR-, CRKL-, PRPH2-, USH2A-, PRPF31*-associated retinitis pigmentosa	**Virtual reality orientation and mobility**	VF; FST; VA	Speed; Accuracy (obstacle identification, departures from the path, direction of movement, collisions, and whether the subject missed any arrows or repeated them)	**Content validity** – Better performance in patients with better VA and larger VF extents **Construct validity** – Significant difference in the time to complete obstacle testing between patients and controls (p=0.0027). Controls identified approximately 50% of the obstacles at the dimmest course luminance. All but two patients were able to complete the test, although they required higher luminance levels (by >2 log units) to identify 50% of the obstacles. **Repeatability –** Small improvement in object detection on the second test leading to positive test-retest differences. Greater test-retest values at the dimmest obstacle course luminance level suggestive of a minor learning effect.
Daga et al., 2017	31 patients with glaucoma	**Virtual Environment Human Navigation Task (VEHuNT)**	VF	Time to complete task	**Construct validity** - Significant difference on average time to complete task between patients and controls for room A (p=0.001). No significant difference on average time to complete the task between patients and controls for room B (p=0.514). Significant relationship between time to complete the task and visual field loss for room A but not for room B (p=0.001).
Facial recognition					
Hirji et al., 2020; Hirji et al., 2021	72 patients with primary open angle glaucoma with glaucomatous macular damage	**The Cambridge Face Memory test**	VF; CS	Percentage of correctly identified faces	**Content validity** - Significant correlation between facial recognition and VF mean deviation (p<0.0001)
Glen et al., 2012; Glen et al., 2013	54 patients with glaucoma				**Construct validity** - Patients with advanced VF defects identified fewer faces on average than those with early and moderate defects and controls (p<0.05).
Mazzoli et al., 2019	64 patients with age-related macular degeneration and 48 patients with primary open angle glaucoma				**Construct validity** – Test scores were lower in patients compared to controls (p<0.001).
Taylor et al., 2018	30 patients with non-neovascular age-related macular degeneration				**Construct validity** - Geographic atrophy patients identified significantly fewer faces on average than early and intermediate AMD patients and controls (p=0.04).
Observer-rated performance tests					
Delyfer et al., 2021 Karapanos et al., 2021, Petoe et al., 2021	18 patients with retinitis pigmentosa 4 patients with retinitis pigmentosa	**Functional Low-Vision Observer Rated Assessment (FLORA)**		Final impact rating; Task performance score	Not reported
Greenberg et al. 2015	30 patients with retinitis pigmentosa				
Yoon et al., 2021	5 patients with retinitis pigmentosa				
Geruschat et al., 2015	26 patients with retinitis pigmentosa				
Altangerel et al., 2006	43 patients with primary open angle glaucoma	**Assessment of Function Related to Vision (AFREV)**	VF; VA; CS	AFREV score	**Content validity** - AFREV scores highly correlated with CS (r = 0.772), binocular VA (r=−0.768), better-eye VA (r =−0.737), worse-eye VA (r =−0.675), and VF scores (r = 0.606) and NEI-VFQ scores (r = 0.70). **Construct validity** – Distinguishes between mild, moderate and severe binocular VF loss.

Kulkarni et al., 2012; Warrian et al., 2010; Warrian et al., 2009 Richman et al., 2010, Richman et al., 2010	192 patients with glaucoma 91 patients with diabetic retinopathy 112 patients with age-related macular degeneration 192 patients with glaucoma	**Assessment of Disability Related to Vision (ADREV)**	VF VA; CS; VF; VFQ-25 VA; CS; VF; VFQ-25 VA; CS; VF; Stereopsis	ADREV score	**Content validity** - Highest correlation with the total ADREV score was the integrated VF score (p=-0.49). **Content validity** – All of the ADREV's scales were correlated with one or more clinical measures of visual function except the Ambulation test. **Content validity** – 66% of correlations made between clinical ophthalmic measurements and ADREV scores were significant to P<0.0007. 55% of correlations made between the ADREV and the VFQ total and subscale scores were significant to P< 0.0004. **Content validity** – ADREV performance was strongly associated with binocular VA (P<0.001) and binocular CS (P<0.001). Monocular and binocular VF results had a weaker correlation with the ability to perform the ADREV tasks (P<0.05).
Edwards et al., 2018	6 patients with advanced retinitis pigmentosa implanted with Retina Implant Alpha AMS - *USH2A, PDE6B, RPE65, RPGR, CERKL*	**Tabletop object and clock face recognition**	No. of correctly location and named items	Not reported	
Azoulay-Sebban et al., 2020; Lombardi et al., 2018	32 patients with glaucoma	**Homelab at StreetLab**	VA; CS; VF; NEI VFQ-25	Path travel time; Mobility incidents; Movement onset; movement initiation time and duration; Localisation of people time; Face orientation recognition time	**Construct validity** - No significant difference in path travel time between patients and controls. Number of mobility incidents was higher in advanced glaucoma group than in other 2 groups (p=0.0126 and 0.0281, for controls and early glaucoma respectively). **Content validity** – Integrated binocular field and VA demonstrated significant correlation with test outcomes. Overall movement duration for small objects in reaching and grasping tasks was significantly longer in glaucoma patients compared with controls. Mobility incidents and the reaching and grasping task parameters were not significantly correlated with quality-of-life questionnaire scores.
Wei et al., 2012 Sun et al., 2016 Waisbourd et al., 2019	9 patients with glaucoma 161 patients with glaucoma 153 patients with glaucoma	**CAARV (Compressed Assessment of Ability Related to Vision)**	VA; CS; VF VF VA; CS; VF; VFQ-25	Total CAARV score	**Content validity** – Strongest correlation was between the central VF cluster and total CAARV score (P<0.001). Central VF cluster in the better eye positively correlated with the majority of CAARV and NEI VFQ-25 subscales. **Construct validity** – Compared to non-rapid progressors, patients who had rapidly progressing glaucoma presented with lower baseline CAARV scores for reading street signs (p=0.01), facial recognition (p=0.01), and total score (p<0.001).
Reighard et al., 2019	145 patients with glaucoma	**I-CAARV (Indian - Compressed Assessment of Ability Related to Vision)**	VA; VF; CS; Indian-VFQ	I-CAARV score	**Content validity -** I-CAARV scores and the Indian VFQ were significantly correlated (P<0.01). Rasch-calibrated scores on the I-CAARV were also significantly correlated with VF MD, presenting VA, best-corrected VA, and CS in both the better-seeing eye (p=0.60, p=-0.91, p=-0.53, p=0.76 respectively) and worse-seeing eye (p=0.48, p=-0.61, p=-0.53, p=0.69). **Repeatability –** Rasch analysis found that the I-CAARV had moderate reliability (0.74) and measurement precision was fair (person separation 1.67 logits). Rasch analysis found good **construct validity** (infit range 0.66-1.13; outfit range 0.65-1.21)
Peterson et al., 2023*	36 patients with age-related macular degeneration	**Performance-based activities of daily living task tests (ADLTT)**	VA; CS; MP	Task completion time	**Construct validity –** Longer task completion time in patients than controls for money counting task using worse eye vision and binocular vision (both p<0.001) and object drink making task using monocular worse eye vision (p=0.033). **Content validity –** Only the money counting task demonstrated moderate to strong correlations with VA, CS, and MP. **Divergent validity** was demonstrated when correlated with race and gender in most ADLTTs except for facial expression task.

					Repeatability - Moderate to good test-retest reliability for money counting and drink making tasks only using monocular worse eye vision.
Ni et al., 2012	64 patients with age-related cataract	**Real-Life Vision Test (RLVT)**	VA; CS; Stereopsis; Colour perception; VFQ-25	Time taken to complete task	**Construct validity –** Controls performed significantly better than patients (P<0.01). Significant difference reported between patients with different cataract severity **Content validity** - All RLVT subscales remained highly associated with most clinical measures, after controlling for age, years of education, Mini Mental State Examination scores, self-rating depression scores, and reaction time.
Finger et al., 2014	40 patients with rod-cone dystrophy	**Very Low Vision Instrumental Activities of Daily Living (IADL-VLV)**	VA; VF	Completion and accuracy score	**Content validity –** VA and VF were associated with IADL performance. **Construct validity –** Patients with worse VA or VF scored lower (p<0.00 and p=0.001 respectively)
Visual search					
Higgins et al., 2020	38 patients with non-neovascular age related macular degeneration	**Computer based assessment (Visual search task and simulated dynamic driving scene)**	VA; CS; MP; EuroQol-5D questionnaire	Total correct responses; Median response time	**Construct validity** - Slower performance in visual search tasks associated with more severe disease. No significant difference between groups for total correct responses (p=0.342). Significant difference in median response time between the groups (p=0.007). Early and intermediate group's median response time were not significantly slower than the controls. **Content validity** - Response time was associated with measures of VA and CS.
Taylor et al., 2017	31 patients with dry age-related macular degeneration		VA; CS	Median search time; Fixation duration; Saccadic amplitude; Saccades per second	**Content validity** – Significant associations between average search time and VA (p<0.001) and CS (p<0.001) **Construct validity –** 61% of patients exceeded the normative limits for average search time; this was statistically significant (p<0.0001). No differences between groups in fixation duration or saccades per second. Yet saccadic amplitude remained significantly smaller for patients compared to controls (p<0.001).
Thibaut et al., 2018	21 patients with age related macular degeneration	**Object search in realistic panoramic scenes**		Percentage of correct target detection; percentage of false positives; scene views explored; search time	**Construct validity -** No significant differences in performance between patients and age-matched controls.
Wan et al., 2020	30 patients with age-related cataract	**Visual search and facial recognition task**		Fixation count and total duration; total visit duration; Forward and backward saccade count per line; percentage of regressive saccades; percentage of correctly identified faces	**Construct validity** – Significant difference before and after surgery for the percentage of correctly identified objects and faces (p=0.049 and p=0.004 respectively), average search time (p<0.001), fixation count (p<0.001), total fixation duration (p= 0.039) and total visit duration (p=0.008). No significant change was in mean fixation duration. **Repeatability** - No significant difference between baseline and follow-up assessment (all parameters p<0.05)
Kartha et al., 2023	37 patients with ultra-low vision	**Virtual reality visual performance test**	Berkeley Rudimentary Vision Test	Item measure; Person measure	**Content validity –** Negative correlation between patients with poorer visual acuity having lower person measures (p=0.002, r^2=0.2, mean absolute error=0.43). **Construct validity** – Items measures ranged between −0.09 to 0.39 in relative d′ units. Person measures ranged between -0.74 and 2.2 relative d' units.
Martínez-Almeida et al., 2021	33 patients with glaucoma	**Virtual reality system with gaze monitoring**		Fixation number and duration; Saccadic amplitude and velocity; Fixation/saccade ratio; Total search and execution time; Number of collisions	**Construct validity** – Significant differences between controls and patients for the static task in terms of number of fixations (p=0.012), mean saccadic velocity (p=0.02 and 0.017), fixations/saccades ratio (p=0.035 and 0.04), and the search and total execution times during visual search exercise (p=0.004 and 0.027, respectively). For the dynamic task, Significant differences were found on average saccades amplitude (p=0.02), average saccades velocity (p=0.03) and the number of collisions (p=0.02).
Kurek et al., 2023*	30 patients with retinitis pigmentosa	**Virtual reality visual search task with natural scenes**	CS	Performance score (encompassing search duration and rate of performance success)	**Construct validity –** Able to discriminate between patients and controls (Accuracy >86%) **Repeatability** – Good agreement of performance score between sessions (Intraclass correlation coefficient>0.89) **Content validity** - Correlation with CS was p=0.76. 83% of RP participants indicated that the virtual reality test was representative of their difficulties in daily li

Zhang et al., 2022; Manley et al., 2022	63 patients with cerebral visual impairment	**Virtual toybox and virtual hallway**		Success rate; Reaction time; Gaze error; Visual search area; Off-screen percent (an index of task compliance	**Construct validity** – For the virtual toybox task, mean success rate for patients was significantly lower compared to controls (p<0.001). Significant difference with respect to mean reaction time with patients taking longer to find the target compared to controls (p < 0.001). For the virtual hallway task, mean success rate for patients was significantly lower compared to controls (p<0.001). Mean reaction time was significantly greater in patients compared to controls (p<0.001)
Roux-Sibilon et al., 2018	22 patients with glaucoma	**Scene and face recognition**	VF	Participant's response; Reaction time for response	**Construct validity** - Patients demonstrated deficit in both detection and categorization of all low-contrast images compared to controls.
Smith et al., 2012	40 patients with glaucoma	**Visual search task with eye tracking**	VF; CS	Average number of saccades per second; average saccade amplitude; Average search duration	**Construct validity** - Average rate of saccades by patients was significantly smaller than controls during the visual search task (p=0.02). No difference in average saccade amplitude between the patients and controls (p=0.09). **Content validity** - Average number of saccades was negatively correlated with CS (p=0.006) and more severe VF defects (p=0.037).
Driving simulators					
Adrian et al., 2022	14 patients with glaucoma	**Fixed base driving simulator at StreetLab**		Reaction times; Longitudinal regulation; lateral control; eye and head movements; Fixation duration and number per second; Fixation duration; horizontal and vertical gaze direction; head yaw	**Construct validity -** Compared to controls, patients demonstrated a longer mean duration of lateral excursions (p=0.045), and more lane excursions in a wide left lane (p=0.045). Patients demonstrated a larger standard deviation of horizontal gaze (p=0.034). No significant difference was established for the other measured outcomes.
Kübler et al., 2015	6 patients with glaucoma	**Simulated driving test**		Driving lane positions; time to line crossing (indicates steering stability); driving speed; head and eye tracking	Not reported
Lee et al., 2019	31 patients with glaucoma	**DriveSafe (slide recognition test)**	VA; VF; CS; UFOV® test	Total number of correctly identified road user features (DriveSafe score); number of fixations points; average fixation duration; average saccade amplitude; horizontal and vertical search variance	**Construct validity** - Patients had significantly worse DriveSafe scores (p=0.03), fixated on road users for shorter durations (p<0.001), exhibited smaller saccades (p=0.02), reduced fixation duration and saccadic amplitudes compared to controls (p<0.001 and p=0.02). No other significant group differences were found. **Content validity** - Significant relationship between clinical measures and DriveSafe scores: UFoV 2 (p=0.005), worse-eye VF mean deviation (p=0.003), CS (p=0.03) and UFoV 3 (p=0.05).
Devos et al., 2018	17 patients with glaucoma	**Performance based visual field test in a driving simulator**	VF; UFOV®	Total crashes; Speed exceedances; Correct stops at traffic lights; Centre line crossings; Road edge excursions; Complex response time; Target identification accuracy; Number of missed responses; Response time	**Construct validity** - Patients identified fewer VF symbols (p=0.047) and took longer (p=0.048) to detect the VF symbols compared to controls. No significant differences for the other driving performance measures. **Content validity** - Correlation between performance-based VF test scores and horizontal FOV of the Keystone vision screener and UFOV® divided attention subtest (p=0.02 and p=0.046 respectively). **Repeatability** – Intraclass correlation ranged between 0.17 for response time and 0.92 for correct responses.
Prado-Vega et al., 2013	23 patients with glaucoma	**Driving simulator with eye-scanning**	VF	Steering activity; Lane keeping; Longitudinal and lateral distance to obstacle; Collisions	**Construct validity** - No significant difference between patients and controls for lane keeping, obstacle avoidance, and eye-scanning behaviour. Steering activity was significantly higher for patients than for controls. **Content validity** – No significance correlation between the percentage of depressed IVF points and driving performance measures (p>0.2).

VA = visual acuity; BCVA = best corrected visual acuity; VF = visual field; CS = contrast sensitivity; MP = microperimetry; FST = Full-field stimulus testing; FLORA = functional low-vision observer rated assessment; PWS = preferred walking speed; PPWS = percentage preferred walking speed; O&M = orientation and mobility; POAG: primary open angle glaucoma; AMD: age-related macular degeneration; VFQ-25 = Visual Functioning Questionnaire-25; VA LV VFQ-48 = Veterans Affairs Low-Vision Visual Functioning Questionnaire; UFOV = useful-field-of-view. *Indicates a conference abstract. Where a genetic mutation was reported, this has been included in italics. If a form of validation evidence (e.g. construct validity) is absent from table, it was not reported in the original article.

Appendix A.

Search strategy performed in MEDLINE and Embase (via Ovid) on 1st August 2024
Functional vision.ti,ab.
Functional ability.ti,ab.
Functional disability.ti,ab.
Functional impairment.ti,ab.
Performance based.ti,ab.
Real world vision.ti,ab.
Real world task.ti,ab.
Daily living task*.ti,ab.
Mobility.ti,ab.
Vis* task.ti,ab.
Visual search.ti,ab.
Eye-Tracking Technology/
Fac* recognition.ti,ab.
1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
Eye Diseases/
Visual* impair*.ti,ab.
Vision impaired.ti,ab.
Glaucoma/
Inherited retinal disease.ti,ab.
Achromatopsia.ti,ab.
Choroideremia.ti,ab.
Stargardt Disease/
Usher Syndromes/
Leber Congenital Amaurosis/

Optic Atrophy, Hereditary, Leber/
Retinitis Pigmentosa/
Macular Degeneration/
15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
14 and 28
limit 29 to (english language and yr="2003 -Current")

Table A1. Full Boolean search strategy divided into two concepts: functional vision and eye disease

BMJ Open

Functional vision tests as clinical trial outcome measures in ophthalmology: a scoping review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-097970.R1
Article Type:	Original research
Date Submitted by the Author:	28-Apr-2025
Complete List of Authors:	Raji, Shabnam; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital Thirunavukarasu, Arun; University of Oxford, Oxford University Clinical Academic Graduate School; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences Taylor, Laura; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital MacLaren, Robert; University of Oxford, Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences; Oxford University Hospitals NHS Foundation Trust, Oxford Eye Hospital
Primary Subject Heading:	Ophthalmology
Secondary Subject Heading:	Research methods, Genetics and genomics
Keywords:	OPHTHALMOLOGY, Clinical Trial, GENETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Functional vision tests as clinical trial outcome measures in ophthalmology: a scoping review

Authors: Shabnam Raji[1,2]*, Arun J. Thirunavukarasu[2,3], Laura J. Taylor[1,2], Robert E. MacLaren[1,2,]

Affiliations:

[1.] Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[2.] Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[3.] Oxford University Clinical Academic Graduate School, University of Oxford, Oxford, United Kingdom

***Corresponding author:** Shabnam Raji; Oxford Eye Hospital, Lower Ground 1 West Wing, John Radcliffe Hospital, Headley Way, Oxford, OX3 9DU, United Kingdom

Email: enquiries@eye.ox.ac.uk Phone: 01865231159

Word count: 4536

ABSTRACT

Objectives: To identify currently available functional vision tests and evaluate their use as clinical trial outcome measures in ophthalmology.

Design: Scoping review using the Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews (PRISMA-ScR) guidelines.

Methods: A literature search was conducted in MEDLINE and Embase (via Ovid) for articles published between 1st January 2003 to 1st August 2024. Additional grey literature was sourced from institutional repositories, conference proceedings and a manual citation search. Article screening was conducted against a pre-defined inclusion criteria by two independent, masked reviewers, with a third reviewer acting as arbiter. The inclusion criteria were English language articles which feature a test assessing functional vision in patients with an ophthalmological disease. Details of source characteristics, test methodology and accessibility and evidence of test validation were collected.

Results: Of 2,665 articles returned by the search, 73 were included and 45 unique tests of functional vision were identified. Diseases affecting the peripheral retina were mainly affected, accounting for 77% (56 out of 73) of the diseases featured in all included studies. Overall, 82% (37 out of 45) functional vision tests reported evidence of statistical validation with varying robustness. Functional vision tests were mapped to domains of orientation and mobility, facial recognition, observer-rated task performance, visual search and driving. Obstacle courses assess vision-guided orientation and mobility, correlate highly with clinical measures of visual function in severe peripheral retinal disease and have been validated for use in clinical trials. Their requirement of physical space and time limits utility in multi-centre trials; equivalent tests leveraging virtual reality and eye tracking technologies are in development. Early iterations of visual search tests to simulated realistic scenes have demonstrated discriminative ability, even in paediatric patients.

Conclusions: Functional vision tests can facilitate research into future novel ophthalmological treatments that prioritises patients in terms of how clinical benefit is defined. The principal barriers to the uptake of these tests are lack of accessibility, low quality validation and that many tests remain early in their development stage. This review captures the current landscape of functional vision tests and serves as a reference for investigators and regulatory bodies to evaluate the suitability of these tests for ophthalmic clinical trials.

Keywords: functional vision, performance-based assessment, outcome measure, mobility, task performance

STRENGTHS AND LIMITATIONS OF THIS STUDY

1. This review provides the first evaluation of functional vision tests in ophthalmology, focusing on their potential as clinical trial outcome measures.
2. A comprehensive grey literature search was performed to minimise the risk of bias.
3. Due to heterogeneity in reported test validation, in-depth statistical analysis of validation data was not undertaken.
4. Incomplete or insufficiently detailed data in the included studies limited the scope of the analysis.

Acknowledgments: None

Author statement: S.R was responsible for the design of the scoping review, conducting the search, screening eligible sources, extracting and analysing data, and the writing and preparation of the manuscript. A.J.T. contributed to conducting the search, screening eligible sources, extracting and analysing data, and the writing and preparation of the manuscript. L.J.T. contributed to screening eligible sources and writing and preparation of the manuscript. R.E.M. contributed to writing and preparation of the manuscript. R.E.M is the guarantor of the manuscript. All authors approved the final version of the manuscript.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: This work was supported by the NIHR Oxford Biomedical Research Centre.

Ethics Approval: This study did not involve human participants or animals, and therefore ethics approval was not required.

Data Sharing Statement: As no original data was generated as part of this review, there are no datasets available for sharing. All relevant data sources are publicly available and cited within the manuscript.

INTRODUCTION

Functional vision tests measure how well individuals can interact with their visual environment [1], and these tests may characterise certain eye diseases better than standard clinical measures of visual function and patient reported outcome measures [2]. Functional vision is distinct from visual function which describes the physiological function of the eye and associated visual system, often through contrived clinical tests such as perimetry or visual acuity. Functional vision tests are based on activities of daily living in several domains: mobility, object identification, facial recognition and reading, among others. They output objective scores and can conflate aspects of visual acuity, spatial vision, cognition, colour vision, light sensitivity and adaptation to assess overall function [3]. They also consist of relatively complex tasks that assess higher-order visual processing which may offer a more holistic understanding of visual impairment. In this way, they are highly pertinent measures of a patient's overall quality of life and have broad potential application as clinically meaningful outcome measures in ophthalmology clinical trials.

Currently accepted visual function outcome measures in ophthalmology include best-corrected visual acuity, perimetry, full-field stimulus testing, microperimetry and mobility testing [4,5]. Despite standardisation, visual acuity remains a gross characterisation of overall vision, insensitive to changes in retinal function away from the fovea and displays poor reliability in patients with visual impairment [6]. Standard automated perimetry has been the gold standard for detecting optic nerve damage and has been used effectively as an outcome measure in glaucoma trials [7]. However, perimetry is limited by low test-retest reliability, particularly in those with poor steady, central fixation in macular disease and certain oculomotor abnormalities, such as nystagmus [6]. Fundus-controlled perimetry, or microperimetry, has gained favour in this regard and has become a key endpoint in several clinical trials [8].

Structural outcome measures in ophthalmology can offer precise, highly reproducible assessments of disease progression and can delineate anatomical biomarkers. However,

these measures may not be applicable if structure and function do not reliably correlate, for instance, where there is amblyopia or a gene defect affecting enzymes of the visual cycle. In these cases, it is unclear how anatomical changes in the eye translate to patient benefit [6].

In other medical specialties, functional tests have already been established as key clinical trial endpoints, such as in stroke medicine and multiple sclerosis [9,10]. The US Food and Drug Administration (FDA) have published specific guidelines on patient-centred drug development [11] to prioritise the impact of novel treatments on patients. Similarly, the World Health Organisation's International Classification of Functioning, Disability and Health framework classifies health in terms of functioning and disability in daily life [12]. It provides the basis for a more integrated understanding of health, with emphasis on practical function rather than solely biomedical variables. Research is ongoing in ophthalmology clinical trials to align with this framework.

Here, a review was undertaken to identify currently available functional vision tests and evaluate their application as clinical trial outcome measures in ophthalmology.

METHODS

Search strategy

A scoping review was selected due to the heterogeneity of articles found in the preliminary literature search, and to allow for more exploratory analysis of functional vision tests as an outcome measure. The review was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [13]. A literature search was conducted in MEDLINE and Embase (both via Ovid). Publication dates were restricted from 1st January 2003 to 1st August 2024. A grey literature search was conducted to minimise publication bias and maximise the scope of the review. Grey literature sources included a manual citation search, Google scholar, conference

proceedings and the British Library Electronic Theses Online Service (EThOS). The full Boolean search string with combined index and free text terms is detailed in Table S1.

Duplicates were manually removed by two reviewers. Title and abstract screening, and full text screening was conducted against a pre-defined inclusion criteria by two independent, masked reviewers, with a third reviewer acting as arbiter to resolve disagreement by casting a deciding vote.

Inclusion and exclusion criteria

The inclusion criteria were as follows: 1. Written in the English language; 2. Is a primary research article; 3. Is not a retracted article; 4. Features a test designed for human patients; 5. Test assesses functional vision. Included tests were restricted to those used in patients with an ophthalmological disease. Psychophysical, visual function tests and patient reported outcome measures (PROMs) were excluded. Although an important domain of functional vision, reading tests were excluded in this search as they have been subject to extensive literature review [14].

Data extraction and analysis

Key features of the included texts were charted by two independent, masked reviewers with results synthesised by one reviewer. Data on study design, patient characteristics, test methodology, visual function correlates, validity and repeatability evidence and accessibility were extracted. Specifically, articles were searched for evidence of the following: test responsiveness, inter- and intra-rater reliability, test-retest reliability, content, construct and criterion validity. Repeatability and validity data were abstracted to only include statistical values of significance and correlation; purely qualitative statements were excluded. Data visualisation was performed with Microsoft Excel 2024 (Microsoft Corporation, USA) and Inkscape (version 0.92).

Patient and public involvement

There was no direct patient or public involvement in this review.

RESULTS

The initial search yielded 2,665 articles. After screening, a total of 73 texts were included: 67 peer reviewed publications and six conference abstracts. The full search and screening process is shown in Figure 1. Source characteristics of all included studies are summarised in Table 1. Forty-five unique functional vision tests were identified and listed in Table S2. An abridged list of functional vision tests is listed in Table 2. All functional vision tests were grouped into thematic categories for further analysis. and are illustrated in Figure 2 along a continuum based on their reported ability to measure central or peripheral vision loss. The number of included articles contributing to each category of functional vision test is also shown in Figure 2. Orientation and mobility and observer-rated performance tasks accounted for the highest number of articles found with 25 and 22 respectively. Virtual reality was the least represented with four articles, although all were published within the last five years which predicts an expanding area of research, in line with the growth of new technologies. Figure 3 illustrates the disease of the patient population in the included articles categorised by structure of the eye affected, clinical phenotype and genotype. Functional vision tests were mainly investigated in diseases affecting the peripheral retina which accounted for 77% (56 out of 73) of the diseases featured in all included studies. Rod-cone dystrophies and optic nerve diseases were common, appearing in 37 and 19 articles respectively. Cone-rod dystrophies and macular disease (both inherited and acquired) featured in fewer studies; 6 and 9 respectively. The number of patients within studies ranged from 4 to 192 and the distribution of reported patient age across all studies is displayed in Figure 4. Only 14 out of 73 articles included a paediatric cohort of patient.

A clinical reference standard was identified in 29 out of the 45 functional vision tests. Overall, 37 out of 45 functional vision tests reported evidence of statistical validation, but these were

of varying robustness. To date, 7 functional vision tests have been used as outcome measures in 10 separate clinical trials for retinal disease as outlined in Table 3.

Orientation and mobility tests

The most common format of functional vision test was obstacle course, assessing orientation and mobility. Performance on obstacle courses was generally assessed by speed and accuracy, which were often combined to produce an overall score. Metrics of speed include preferred walking speed, percentage of preferred walking speed (PPWS) and course completion time. Accuracy metrics include error number, number of collisions or incidents or path departure. One study provided more detailed metrics on trajectory analyses and walking initiation time aided by measurement tools such as motion capture systems and inertial sensors [15]. Some tests involved videotaped performances which were sent to reading centres for grading to reduce the risk of grader bias [16].

Courses ranged in size from 2.1 x 3.6m to 68 x 1.3m, and were located in purpose-built facilities, hospitals and real indoor rooms (e.g. a cafeteria). All tests identified in this review were performed indoors, although outdoor mobility tests have been described in the literature [17,18]. Some tests were performed under multiple luminance levels, ranging from 0.2 to 500 lux, tested in stages to be sensitive to different levels of nyctalopia. No orientation and mobility test exposed patients to acute changes in illumination to test rapid light or dark adaptation, a common difficulty reported in retinitis pigmentosa, perhaps due to safety concerns. Better designed obstacle courses incorporated changes in floor elevation to assess depth perception. If featured in the course, obstacles were commonly made of cardboard or foam and were suspended at various heights. Some tests reported the Weber contrast values and chromaticity coordinates of the obstacles.

Orientation and mobility tests were predominately used on patients with rod-cone dystrophy or glaucoma. As such, the test is suitable for patients with low vision and defects of peripheral vision. The Multi Luminance Mobility Test (MLMT) was used as a primary outcome measure

in the landmark clinical trial of voretigene neparvovec (Luxturna) for *RPE65*-related Leber's congenital amaurosis, the first approved gene therapy in ophthalmology [19]. The MLMT adopts a binary instead of a continuous scoring system, is performed under seven different luminance levels and demonstrates ceiling effects [20]. The low luminance conditions allowed the test to demonstrate sensitivity to changes in disease state; *RPE65* is an enzyme which facilitates dark adaptation of viable rod photoreceptors. It follows that a drug capable of rescuing the function of defective *RPE65* would result in enhanced scotopic vision [19]. The success of the MLMT has subsequently inspired the development of several commercial, academic and dedicated facilities offering functional vision testing, to include Streetlab and Ora [15,21–24]. It should however be noted that MLMT is primarily an assessment of scotopic vision augmented by dark adaptation of rods and not necessarily the best method to assess cone function.

Applications of virtual reality technology

Virtual reality can overcome many limitations of orientation and mobility tests. Virtual reality may absolve the need for a physical, homogenously lit room whilst still maintaining a degree of realism [25]. As such, it is more accessible for use in multi-centre clinical trials and can overcome the scaling challenges of physical obstacle courses. However, virtual reality-related motion sickness has been reported and as a result, patients may still instructed to walk in physical space to avoid this [26]. Commonly used virtual reality headsets include the HTC Vive Pro Eye, Fove 0 and Oculus Rift, which are consumer devices commercially available at a relatively low cost. Proprietary, custom-made software was used on this hardware. Some studies included trackers mounted to patients' head, hands and feet to generate kinematic data [27,28]. The technical specifications of VR devices were as follows: display screens were LED or AMOLED, panel sizes ranged from 18.5" to 80", resolution ranged from 1280 × 1440 to 4K, and the horizontal field of view ranged from 89 to 150 degrees. If reported, the display refresh rate was 90Hz. VR tests were conducted binocularly, although recent iterations enable monocular testing [28,29].

Visual search tests

Visual search tasks relate to several domains of functional vision including social interaction, reading, driving and mobility, and have been used to assess patients with various forms of visual impairment [30,31]. Visual search may be performed binocularly in front of a display monitor with free head movements or using virtual reality headsets with in-built eye-tracking. Display screen sizes generally range from 17" to 27", although a hemispheric, panoramic screen covering 180 degrees of horizontal visual field has been reported [32]. Eye tracking devices included the Tobii EyeX, Tobii 4C, Tobii Pro X3-120, Tobii AB (Tobii technology, Stockholm, Sweden), HTC Vive trackers (HTC Corp., New Taipei, Taiwan), Oculus Quest Pro (Meta, Burlingame, CA) and the Eyelink II system, Eyelink 1000 system (SR Research Ltd., Ontario, Canada). Proprietary, custom-made software was used on this hardware. Task performance metrics were search time and correct responses.

Visual scenes included geometric shapes hidden in a computer-generated room and everyday objects hidden in photographs of real-world scenes. Psychophysical targets such as optotypes or geometric shapes are not intuitively reflective of real life and studies have shown that a Landolt C search task, compared to object identification in a real photograph, did not differentiate patients from visually healthy controls [33]. All scenes found in visual search tasks were two-dimensional and static, and therefore not reflective of dynamic scenes of the real world. The realism and context provided by real world scenes is important as the role of global features and semantic guidance in object search has been well evidenced to influence visual behaviour [34,35]. Early iterations of visual search tests in simulated realistic scenes have demonstrated discriminative ability, even in paediatric patients [36,37]. One portable tablet-based visual search test was able to discriminate patients with severe diabetic macular oedema from an established normative database [38].

Driving simulator tests

Driving simulator tests have previously been used to evaluate safety, for example, in glaucoma and in the development of new multifocal intraocular lenses, but not treatment effectiveness in clinical trials [39,40]. Driving simulator tests have been described in many forms. Moving base

driving simulators exist that benefit from a realistic car body and wide-field scene projection but lack the accessibility of other portable simulators [41]. Desktop-based driving simulators are low fidelity tests and the lack of real-world consequences from patient error has been reported to influence behaviour by overstating true driving performance [39]. The artificial driving scenes in these desktop-based simulators can also cause the patient to subtend a smaller visual angle compared to real life which inadvertently affects the amplitude of saccadic eye moments – a common measure of performance in driving simulator tests.

Observer-rated visual performance tests

Observer-rated visual performance tests are simulated activities of daily living performed in a controlled environment and assessed by an observer. These tests have been shown to correlate with similar tasks performed at home [42]. Tested activities include dialling a phone number, reading in reduced illumination or opening a lock with a key. The original Assessment of Function Related to Vision (AFREV) was limited by ceiling effects and was superseded by the Assessment of Disability Related to Vision (ADREV). The Compressed Assessment of Ability Related to Vision (CAARV) is a compressed version of this test requiring only 14 minutes to complete. In 2014, the Functional Low-Vision Observer Rated Assessment (FLORA) was developed as an untimed, home-based test for ultra-low vision patients in the context of a clinical trial for the Argus II retinal prosthesis; a validation study is ongoing [43]. A validation study for the more recently developed Instrumental Activities of Daily Living Tools in Very-Low Vision (IADL-VLV) underscores the tests' potential as an outcome measure in vision restoration trials. It was developed using a Delphi consensus procedure, with input from occupational therapists and low-vision experts, maintaining high levels of content validity [44]. Novel observer rated performance tests are in development with good repeatability and monocular testing [45]. Limitations of potential observer bias were reported, although newer test iterations have incorporated automated scoring methods using sensors attached to objects to detect object displacement [46,47]. The tests were also subject to floor and ceiling effects [48] and

could place infeasible cognitive and motor demands on patients in line with the activities assessed, limiting their use to a select subset of suitable patients.

Facial recognition tests

The Cambridge Face Memory Test is a validated, computer-based, alternative forced choice task where a target face must be distinguished from two additional unfamiliar faces. The test is freely available online, performed binocularly and has an established normative reference score. The test demonstrates variable discriminative ability when applied to different disease cohorts. In patients with dry AMD, the test was not found to be sensitive to early or intermediate stages of dry AMD but was able to discriminate individuals with features of late-stage disease such as geographic atrophy [49]. Moreover, one study showed no significant correlation between facial discrimination performance and severity of diabetic macular oedema [38]. Co-occurring psychiatric illness, neurological damage or neurodevelopmental disorders such as autism affect facial recognition [50] and facial recognition tests are used cautiously in these populations.

DISCUSSION

A functional vision test has been used as a primary outcome measure in a landmark gene therapy clinical trial in ophthalmology. This has set the stage for the development of more unconventional assessments of vision which will be evaluated herein.

Existing functional vision tests in ophthalmology

Orientation and mobility tests were originally used in early clinical trials of retinal prosthesis implants in blind or ultra-low vision patients [51–53]. They were favoured as these patients often had remnants of useful vision and light perception that were not captured in standard clinical tests of visual function. As such, these functional tests have relevance in end-stage disease than in early-stage disease where structural changes remain sensitive markers of clinical

progression [54]. They are useful in measuring low luminance mobility and peripheral vision loss although individuals with localised degeneration may employ head and eye movements to project the visual environment onto islands of functioning retina. In a study with choroideremia patients, no deficit in Multi Luminance Mobility Testing (MLMT) performance was observed due to preserved macular function even in the presence of advanced peripheral disease [55].

Orientation and mobility tests are constrained by several limitations and performance scores can be marred by many sources of error. Firstly, the tests are inherently influenced by patients' confidence and psychological state. For example, a distinguishing feature of orientation and mobility tests is that an error committed results in an immediate physical response, such as colliding with an obstacle or wall. How individuals negotiate these physical responses varies widely, in terms of risk management or aversion. Furthermore, if patients are aware of being observed or recorded, then the results may be additionally confounded by the Hawthorne effect. The time taken to complete the course is likely influenced by patient confidence which may improve if a patient is aware that they have received a potentially sight-saving treatment, thereby conferring a placebo effect. Performance scores may also be confounded by a learning effect and repeated testing is necessary to overcome this which can prove laborious for patients – if patients are instructed to repeatedly walk as fast as possible in multiple course runs to determine maximum performance speed, they may be limited by physical stamina rather than their vision.

Practically, the resources required to develop, conduct and maintain these tests limit their scalability and may preclude their continued use in multi-centre clinical trials. Several orientation and mobility VR tests have been described that offer easy manipulation of the digital visual environment and potentially unlimited course configurations. These tests provide greater optionality in assessing a range of diseases and control of experimental conditions, therefore improving test reproducibility. The automated scoring performance in VR can also reduce assessor bias. Moreover, VR can make an orientation and mobility test into a game by introducing interactive scoring, for example, tests exist that instruct patients to 'tag'

obstacles with a controller [28]. However, certain limitations arise from the use of VR. The physical VR headset detaches the user from reality and introduces a degree of abstraction to a task. Discrepancies in resolution between the retina and a VR display screen can affect true perception, particularly if the pixel density and resolution is considerably below human acuity [56]. VR tests remain in their infancy and require validation in relevant patient populations to ascertain their usability as outcome measures.

VR has also been applied to visual search tests which have demonstrated discriminative ability, even in paediatric patients [36,37]. The increased accessibility of eye tracking technology as consumer devices, evidenced by the 2024 release of the Apple Vision Pro, assures further development of virtual reality and visual search tests. An avenue of future development may be wearable technologies that can monitor real-time visual search in daily life over extended periods of time. A similar application is the EMA approved endpoint of wearable sensors that quantify movement in muscular dystrophy trials [57].

Driving simulator tests have been described in several formats although if patients have been banned from driving due to deteriorating vision, then the psychological impact of being subjected to a driving test should be considered. Not all patients, particularly those with early onset inherited retinal diseases, ever learn to drive, limiting the accessibility of the test.

Inherited retinal diseases: a use case for functional vision tests

Well-designed tests of functional vision relate closely to the prevailing symptoms throughout the natural history of an ophthalmological disease. The symptoms of the disease guide test development to ensure that highly relevant concepts of interest are assessed, and that outcomes remain patient-relevant and pertinent to quality of life. Development and validation is challenging in diseases with variable phenotypes or low prevalence, both exhibited within inherited retinal diseases which collectively represent the leading cause of blindness among working age adults in England and Wales [58]. Pathogenic mutations in over 280 genes have

been identified as causing inherited retinal disease; each mutation is associated with its own phenotypic characteristics and so patient symptoms can be highly nuanced [59]. Selected outcome measures will depend on the underlying disease mechanism and whether a gene-specific or gene-agnostic therapy is developed. The growth of research and development into therapies for these inherited retinal diseases calls for agile innovation in clinical trial outcomes measures to facilitate the arrival of novel gene therapies to market.

Tests that are selected as clinical trial outcome measures should also relate to the region of therapy delivery. For example, in a rod-dominated photoreceptor degeneration the main symptom may be reduced peripheral vision, but if a drug is administered to rescue remaining photoreceptors at the macula, it is logical to preclude the use of a mobility test that may be insensitive to ultimately measure therapy efficacy. This emphasises the importance of judiciously selecting appropriate and effective outcome measures. Additionally, functional vision tests that are performed binocularly have limited utility in clinical trials featuring monocular interventions, particularly where therapy is delivered to the worse seeing eye – as is common practice – as the better seeing eye tends to predict visual functional ability [60]. Ideally, both monocular and binocular assessments should be performed. Assessments of binocular function can provide understanding of overall function, leading to interpretations of quality of life and subsequent health economic analyses.

Several inherited retinal diseases are syndromic with systemic abnormalities that may additionally impair a patient's ability to perform a functional vision test, for reasons other than reduced vision due to retinal degeneration. An example of this is in Joubert's syndrome, whereby mutations in *CEP290* concurrently cause Leber's congenital amaurosis and psychomotor delay with cerebellar malformations, among other ciliopathy-associated abnormalities [61]. Performing a functional vision test in these patients with cognitive and physical impairment would be unreliable in measuring changes in retinal function, and it may be difficult to isolate the true measurement of retinal disease due to the confounding effect of systemic abnormalities.

Challenges in the paediatric validation of functional vision tests

There is a dearth of validated functional vision tests for use in paediatric patients. This is of particular relevance if novel therapies, that are proven to be efficacious in adults, are offered to patients at an earlier age, and in the case of diseases which typically have an early onset of presentation. Examples include Luxturna for *RPE65*-LCA, which used the MLMT in a trial involving adult patients, but for which treatment may be initiated in younger patients as index presentations are frequently early in life. Tests should be optimised for use in children with appropriate modifications to enable clinical trials and post-trial monitoring to capture the benefit conferred by new treatments. Few functional vision tests identified in this review have been used in children [15,23,27,28,36,37,62–69].

Validation of novel functional vision tests

Treatments such as visual prostheses, stem cell transplantation, gene augmentation and editing therapies, antisense oligonucleotide therapy and optogenetic therapies are being developed at pace for previously untreatable ocular conditions [70]. Progress in the development of these treatments requires validated outcomes. The paucity of validation in functional vision tests is evidenced in Table 2 and S2. Few articles reported a full description of test methodology to allow replication, and validation evidence was either absent or fragmented. The absence of an established gold standard test for the measurement of functional vision meant no studies were found to report concurrent validity. Clinically adjudicated reference standards to validate novel tests have been reported in other fields of medicine such as infectious disease diagnostics, and may be useful in the absence of an existing gold standard test [71].

The functional vision tests in this review correlate with clinical measures of visual function to varying degrees of significance and construct validity. The appropriateness of this correlation

may be questioned, as functional vision tests measure a distinct aspect of vision rather than acting as surrogate indicators of visual function, raising the issue of whether full validation is required in all cases of test development. It can be said that drawing on the experience of clinicians and patients' perspectives should provide more weight in determining whether test measurements provide useful and clinically meaningful information.

Most current clinical trials adopt a monocular study design to benefit from the contralateral eye as a control but the need for standardised, precise and reliable outcome measures will become critical once treatments are delivered bilaterally [72]. Standardised validation of functional vision tests can improve evidence synthesis, the inferential quality of results and enhances comparability of data between clinical trials with treatments for the same disease. It is reasonable to suggest that functional vision tests should still be validated against standard clinical measures of visual function, but the strength of its validation, or lack thereof, should not solely dictate inclusion as an outcome measure in clinical trials.

In the 1990's, the increase in visual prosthesis development for vision restoration trials led to a greater need for clinically meaningful endpoints. The various centres that developed visual prosthesis used different efficacy measurements, making cross-comparison challenging. This led to the International Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials (HOVER) taskforce where experts from around the world collaboratively formed guidance to measure visual function in vision restoration clinical trials [73]. Most functional vision tests found in this review have been applied to inherited retinal diseases, as shown in Table 3, yet there is currently no such directive for inherited retinal disease. Novel clinical trial outcome measures would benefit from being guided by consensus-building to retain standardisation. Stakeholders involved in such consensus-building should include patients, advocacy groups, clinical trial sponsors, disease experts, regulatory agencies and experts in the functional vision construct being measured.

Limitations

The limitations of this review and directions of future research should be considered. A scoping review was selected because of the heterogeneity of the articles identified in the literature search, and it can serve as a foundation for a systematic review or meta-analysis. Test validation in the included studies was reported with varying levels of detail and as such, in-depth statistical analysis of validation data was not undertaken. Incomplete or insufficiently reported descriptions of tests and data limited the scope of the analysis in some cases. This review aimed to address these limitations by critically evaluating their implications and providing evidence-based recommendations to guide future reporting practices.

Functional vision tests are in development globally and the regional cultural differences in activities of daily living were not explored in this review, nor were the sources of funding for centres developing functional vision tests. Furthermore, given that functional vision tests assess aspects of higher-order visual processing [3], exploring correlations of functional vision performance scores with primary visual cortex activity may also be an avenue for future research [37].

CONCLUSION

Functional vision tests can facilitate research into future novel ophthalmological treatments that prioritises patients in terms of how clinical benefit is defined. The principal barriers to the uptake of these tests are lack of accessibility, low quality validation and that many tests remain early in their development stage. This review captures the current landscape of functional vision tests and serves as a reference for investigators and regulatory bodies to evaluate the suitability of these tests for ophthalmic clinical trials.

REFERENCES

1. Bennett CR, Bex PJ, Bauer CM, Merabet LB. The Assessment of Visual Function and Functional Vision. Semin Pediatr Neurol. 2019 Oct;31:30–40.

2. Warrian KJ, Altangerel U, Spaeth GL. Performance-based Measures of Visual Function. Survey of Ophthalmology. 2010 Mar 1;55(2):146–61.

3. High-Level Visual Processing: Cognitive Influences. In: Principles of Neural Science, Fifth Edition [Internet]. New York, NY: McGraw-Hill Education; 2014 [cited 2024 Jan 17]. Available from: neurology.mhmedical.com/content.aspx?aid=1101680178

4. Thompson DA, Iannaccone A, Ali RR, Arshavsky VY, Audo I, Bainbridge JWB, et al. Advancing Clinical Trials for Inherited Retinal Diseases: Recommendations from the Second Monaciano Symposium. Trans Vis Sci Tech. 2020 Jun 3;9(7):2.

5. Shi LF, Hall AJ, Thompson DA. Full-field stimulus threshold testing: a scoping review of current practice. Eye [Internet]. 2023 Jul 13 [cited 2023 Jul 18]; Available from: https://www.nature.com/articles/s41433-023-02636-3

6. Schmetterer L, Scholl H, Garhöfer G, Janeschitz-Kriegl L, Corvi F, Sadda SR, et al. Endpoints for clinical trials in ophthalmology. Progress in Retinal and Eye Research. 2023 Nov;97:101160.

7. Weinreb RN, Kaufman PL. The Glaucoma Research Community and FDA Look to the Future: A Report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium *. Invest Ophthalmol Vis Sci. 2009 Apr 1;50(4):1497.

8. Taylor LJ, Josan AS, Jolly JK, MacLaren RE. Microperimetry as an Outcome Measure in *RPGR*- associated Retinitis Pigmentosa Clinical Trials. Trans Vis Sci Tech. 2023 Jun 9;12(6):4.

9. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke. 2007 Mar;38(3):1091–6.

10. Motl RW, Cohen JA, Benedict R, Phillips G, LaRocca N, Hudson LD, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Mult Scler. 2017 Apr;23(5):704–10.

11. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making | FDA [Internet]. [cited 2024 Jan 10]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory

12. World Health Organization. International classification of functioning, disability and health: children and youth version: ICF-CY. 2007;322.

13. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467–73.

14. Rubin GS. Measuring reading performance. Vision Research. 2013 Sep;90:43–51.

15. Sahel JA, Grieve K, Pagot C, Authié C, Mohand-Said S, Paques M, et al. Assessing Photoreceptor Status in Retinal Dystrophies: From High-Resolution Imaging to Functional Vision. American Journal of Ophthalmology. 2021 Oct;230:12–47.

16. Chung DC, McCague S, Yu ZF, Thill S, DiStefano-Pappas J, Bennett J, et al. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. Clin Experiment Ophthalmol. 2018;46(3):247–59.

17. Marron JA, Bailey IL. Visual factors and orientation-mobility performance. Am J Optom Physiol Opt. 1982 May;59(5):413–26.

18. Kuyk T, Elliott JL, Fuhr PS. Visual correlates of mobility in real world settings in older adults with low vision. Optom Vis Sci. 1998 Jul;75(7):538–47.

19. Russell S, Bennett J, Wellman JA, Chung DC, Yu ZF, Tillman A, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. The Lancet. 2017 Aug 26;390(10097):849–60.

20. Maguire A.M., Russell S., Wellman J.A., Chung D.C., Yu Z.-F., Tillman A., et al. Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation-Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. Ophthalmology. 2019;126(9):1273–85.

21. Azoulay-Sebban L, Zhao Z, Zenouda A, Lombardi M, Gutman E, Brasnu E, et al. Correlations Between Subjective Evaluation of Quality of Life, Visual Field Loss, and Performance in Simulated Activities of Daily Living in Glaucoma Patients. Journal of Glaucoma. 2020 Oct;29(10):970–4.

22. Adrian J, Authié C, Lebrun J, Lombardi M, Zenouda A, Gutman E, et al. Driving behaviour and visual compensation in glaucoma patients: Evaluation on a driving simulator. Clinical & Experimental Ophthalmology. 2022;50(4):420–8.

23. Kumaran N, Ali RR, Tyler NA, Bainbridge JWB, Michaelides M, Rubin GS. Validation of a Vision-Guided Mobility Assessment for RPE65-Associated Retinal Dystrophy. Transl vis sci technol. 2020;9(10):5.

24. Shapiro A., Corcoran P., Sundstrom C., Angjeli E., Rodriguez J.D., Abelson M.B., et al. Development and validation of a portable visual navigation challenge for assessment of retinal disease in multi-centered clinical trials. Invest Ophthalmol Vis Sci [Internet]. 2017;58(8). Available from: http://iovs.arvojournals.org/article.aspx?articleid=2638048 PT - Conference Abstract

25. Li Y, Gunasekeran DV, RaviChandran N, Tan TF, Ong JCL, Thirunavukarasu AJ, et al. The next generation of healthcare ecosystem in the metaverse. Biomedical Journal. 2023 Dec;100679.

26. Lam AKN, To E, Weinreb RN, Yu M, Mak H, Lai G, et al. Use of Virtual Reality Simulation to Identify Vision-Related Disability in Patients With Glaucoma. JAMA Ophthalmol. 2020 May;138(5):490–8.

27. Aleman TS, Miller AJ, Maguire KH, Aleman EM, Serrano LW, O'Connor KB, et al. A Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations: Testing a Proof-of-Concept After Gene Therapy. OPTH. 2021 Mar;Volume 15:939–52.

28. Bennett J, Aleman EM, Maguire KH, Nadelmann J, Weber ML, Maguire WM, et al. Optimization and Validation of a Virtual Reality Orientation and Mobility Test for Inherited Retinal Degenerations. Trans Vis Sci Tech. 2023 Jan 30;12(1):28.

29. Authie C., Poujade M., Talebi A., Defer A., Zenouda A., Coen C., et al. Development and validation of a novel mobility test for IRDs, from reality to virtual reality. medRxiv [Internet]. 2023; Available from: https://www.medrxiv.org/

30. Senger C, Margarido MRRA, De Moraes CG, De Fendi LI, Messias A, Paula JS. Visual Search Performance in Patients with Vision Impairment: A Systematic Review. Current Eye Research. 2017 Nov 2;42(11):1561–71.

31. Fuhr P.S., Liu L., Kuyk T.K. Relationships between feature search and mobility performance in persons with severe visual impairment. Optom Vis Sci. 2007;84(5):393–400.

32. Thibaut M, Tran THC, Szaffarczyk S, Boucart M. Impact of age-related macular degeneration on object searches in realistic panoramic scenes. Clin Exp Optom. 2018;101(3):372–9.

33. Smith ND, Crabb DP, Garway-Heath DF. An exploratory study of visual search performance in glaucoma. Ophthalmic Physiologic Optic. 2011 May;31(3):225–32.

34. Biederman I, Mezzanotte RJ, Rabinowitz JC. Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology. 1982 Apr;14(2):143–77.

35. Torralba A, Castelhano MS, Oliva A, Henderson JM. Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search.

36. Zhang X, Manley CE, Micheletti S, Tesic I, Bennett CR, Fazzi EM, et al. Assessing visuospatial processing in cerebral visual impairment using a novel and naturalistic static visual search task. Res Dev Disabil. 2022;131(8709782):104364.

37. Manley CE, Bennett CR, Merabet LB. Assessing Higher-Order Visual Processing in Cerebral Visual Impairment Using Naturalistic Virtual-Reality-Based Visual Search Tasks. Children (Basel). 2022;9(8).

38. Taylor DJ, Alquiza PJ, Jones PR, Wilson I, Bi W, Sim DA, et al. Tablet-based tests of everyday visual function in a diabetic macular oedema (DME) clinic waiting area: A feasibility study. Ophthalmic Physiologic Optic. 2023 Dec 22;opo.13261.

39. Csaky K, Ferris F III, Chew EY, Nair P, Cheetham JK, Duncan JL. Report From the NEI/FDA Endpoints Workshop on Age-Related Macular Degeneration and Inherited Retinal Diseases. Investigative Ophthalmology & Visual Science. 2017 Jul 21;58(9):3456–63.

40. Drum BA, Rorer EM, Calogero D. Night Driving Performance in the National Advanced Driving Simulator vs. Clinical Tests of Vision. Investigative Ophthalmology & Visual Science. 2007 May 10;48(13):1511.

41. Kubler TC, Kasneci E, Rosenstiel W, Heister M, Aehling K, Nagel K, et al. Driving with Glaucoma: Task Performance and Gaze Movements. Optom Vis Sci. 2015;92(11):1037–46.

42. West SK, Rubin GS, Munoz B, Abraham D, Fried LP. Assessing functional status: correlation between performance on tasks conducted in a clinic setting and performance on the same task conducted at home. The Salisbury Eye Evaluation Project Team. J Gerontol A Biol Sci Med Sci. 1997 Jul;52(4):M209-217.

43. Second Sight Medical Products. A Validation of the Functional Low-Vision Observer Rated Assessment (FLORA-20) for Profoundly Blind Individuals [Internet]. clinicaltrials.gov; 2023 Nov [cited 2024 Jan 1]. Report No.: NCT04312763. Available from: https://clinicaltrials.gov/study/NCT04312763

44. Finger RP, McSweeney SC, Deverell L, F. O'Hare, Bentley SA, Luu CD, et al. Developing an Instrumental Activities of Daily Living Tool as Part of the Low Vision Assessment of Daily Activities Protocol. Investigative Ophthalmology & Visual Science. 2014 Dec 29;55(12):8458–66.

45. Peterson C.L., Htoon H.M., Man R., Lamoureux E.L., Fenwick E., Cheung G.C.M., et al. Development and Validation of Performancebased Assessments of Daily Living Tasks for Age-related Macular Degeneration. Invest Ophthalmol Vis Sci. 2023;64(8):5042.

46. Carlson M, Kim S, Batabyal S, Aldape M, Mohanty S. Design and Development of a Multi-luminance Shape Discrimination Test for Assessing Object Recognition Ability of Visually Impaired Individuals. 2024;

47. Rubenstein LZ, Schairer C, Wieland GD, Kane R. Systematic biases in functional status assessment of elderly adults: effects of different data sources. J Gerontol. 1984 Nov;39(6):686–91.

48. Terheyden JH, Fink DJ, Pondorfer SG, Holz FG, Finger RP. Instrumental Activities of Daily Living Tools in Very-Low Vision: Ready for Use in Trials? Pharmaceutics. 2022 Nov;14(11):2435.

49. Higgins BE, Taylor DJ, Bi W, Binns AM, Crabb DP. Novel computer-based assessments of everyday visual function in people with age-related macular degeneration. Lewin AS, editor. PLoS ONE. 2020 Dec 7;15(12):e0243578.

50. O'Hearn K, Schroer E, Minshew N, Luna B. Lack of developmental improvement on a face memory task during adolescence in autism. Neuropsychologia. 2010 Nov 1;48(13):3955–60.

51. Geruschat D.R., Bittner A.K., Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. Optom Vis Sci. 2012;89(9):1308–15.

52. Xu H, Zhong X, Pang C, Zou J, Chen W, Wang X, et al. First Human Results With the 256 Channel Intelligent Micro Implant Eye (IMIE 256). Trans Vis Sci Tech. 2021 Oct 27;10(10):14.

53. Petoe MA, Titchener SA, Kolic M, Kentler WG, Abbott CJ, Nayagam DAX, et al. A Second-Generation (44-Channel) Suprachoroidal Retinal Prosthesis: Interim Clinical Trial Results. Transl vis sci technol. 2021;10(10):12.

54. Abdalla Elsayed MEA, Taylor LJ, Josan AS, Fischer MD, MacLaren RE. Choroideremia: The Endpoint Endgame. IJMS. 2023 Sep 20;24(18):14354.

55. Chung DC, McCague S, Yu ZF, Thill S, DiStefano-Pappas J, Bennett J, et al. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies: Multi-luminance mobility test. Clinical & Experimental Ophthalmology. 2018 Apr;46(3):247–59.

56. Drascic D, Milgram P. Perceptual issues in augmented reality. In: Bolas MT, Fisher SS, Merritt JO, editors. San Jose, CA; 1996 [cited 2024 Jan 16]. p. 123–34. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1014000

57. Servais L, Yen K, Guridi M, Lukawy J, Vissière D, Strijbos P. Stride Velocity 95th Centile: Insights into Gaining Regulatory Qualification of the First Wearable-Derived Digital Endpoint for use in Duchenne Muscular Dystrophy Trials. J Neuromuscul Dis. 9(2):335–46.

58. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ Open. 2014 Feb 13;4(2):e004015.

59. RetNet: Summaries of Genes and Loci Causing Retinal Diseases [Internet]. [cited 2024 Feb 18]. Available from: https://web.sph.uth.edu/RetNet/sum-dis.htm

60. Kulkarni KM, Mayer JR, Lorenzana LL, Myers JS, Spaeth GL. Visual Field Staging Systems in Glaucoma and the Activities of Daily Living. American Journal of Ophthalmology. 2012 Sep;154(3):445-451.e3.

61. Valente EM, Silhavy JL, Brancati F, Barrano G, Krishnaswami SR, Castori M, et al. Mutations in CEP290, which encodes a centrosomal protein, cause pleiotropic forms of Joubert syndrome. Nat Genet. 2006 Jun;38(6):623–5.

62. Roman AJ, Cideciyan AV, Wu V, Mascio AA, Krishnan AK, Garafalo AV, et al. Mobility test to assess functional vision in dark-adapted patients with Leber congenital amaurosis. BMC Ophthalmol. 2022 Dec;22(1):266.

63. Maguire AM, Russell S, Chung DC, Yu ZF, Tillman A, Drack AV, et al. Durability of Voretigene Neparvovec for Biallelic RPE65-Mediated Inherited Retinal Disease. Ophthalmology. 2021 Oct;128(10):1460–8.

64. Jacobson SG, Cideciyan AV, Sumaroka A, Roman AJ, Charng J, Lu M, et al. Outcome Measures for Clinical Trials of Leber Congenital Amaurosis Caused by the Intronic Mutation in the CEP290 Gene. Investigative Ophthalmology & Visual Science. 2017 May 16;58(5):2609–22.

65. Velikay-Parel M, Ivastinovic D, Koch M, Hornig R, Dagnelie G, Richard G, et al. Repeated mobility testing for later artificial visual function evaluation. J Neural Eng. 2007;4(1):S102-7.

66. Geruschat DR, Flax M, Tanna N, Bianchi M, Fisher A, Goldschmidt M, et al. FLORA™: Phase I development of a functional vision assessment for prosthetic vision users. Clinical and Experimental Optometry. 2015 Jul 1;98(4):342–7.

67. Pierce EA, Aleman TS, Jayasundera KT, Ashimatey BS, Kim K, Rashid A, et al. Gene Editing for CEP290-Associated Retinal Degeneration. New England Journal of Medicine. 2024 Jun 5;390(21):1972–84.

68. Pierce EA, Ashimatey BS, Jayasundera T, Hoyng C, Lam BL, Lorenz B, et al. Twelve-month Natural History Study of Centrosomal Protein 290 (CEP290)-associated Inherited Retinal Degeneration. Ophthalmology Science. 2024 Sep 1;4(5):100483.

69. Russell SR, Drack AV, Cideciyan AV, Jacobson SG, Leroy BP, Van Cauwenbergh C, et al. Intravitreal antisense oligonucleotide sepofarsen in Leber congenital amaurosis type 10: a phase 1b/2 trial. Nat Med. 2022 May;28(5):1014–21.

70. STEM CELL THERAPIES, GENE-BASED THERAPIES, OPTOGENETICS, AND RETINAL PROSTHETICS: CURRENT STATE AND IMPLICATIONS FOR THE FUTURE - PMC [Internet]. [cited 2025 Apr 25]. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6492547/

71. Patel R, Tsalik EL, Evans S, Fowler VG, Doernberg SB, for The Antibacterial Resistance Leadership Group. Clinically Adjudicated Reference Standards for Evaluation of Infectious Diseases Diagnostics. Clinical Infectious Diseases. 2023 Mar 4;76(5):938–43.

72. MacLaren RE. Benefits of gene therapy for both eyes. The Lancet. 2016 Aug 13;388(10045):635–6.

73. Ayton LN, Joseph F. Rizzo III, Bailey IL, Colenbrander A, Dagnelie G, Geruschat DR, et al. Harmonization of Outcomes and Vision Endpoints in Vision Restoration Trials: Recommendations from the International HOVER Taskforce. Translational Vision Science & Technology [Internet]. 2020 Jul [cited 2023 Aug 11];9(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426586/

Titles and legends to figures and tables

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) flow diagram of the study selection process

Figure 2. Number of included articles (n=73) contributing to each category of functional vision test. Six categories of functional vision test ordered on a continuum based on reported ability to measure central or peripheral vision loss. Exemplar fundus autofluorescence images depicting severe peripheral retinal degeneration due to *RPE65*-associated Leber's Congenital Amaurosis (left) and discrete central atrophy within the macula due to *RPGR*-associated cone dystrophy (right). In some severe retinal degenerations, such as end-stage Leber's Congenital Amaurosis, extensive peripheral degeneration encroaches centrally leading to complete loss of light perception.

Figure 3. Disease of patient population in included articles (n = 73) categorised by the structure of the eye affected, clinical phenotype and, where reported, genotype.

Figure 4. Reported age of patient population assessed with functional vision tests. The dashed line demarcates age 18, below which signifies paediatric testing. Five articles were omitted as no age data was available. Note that there are few studies testing paediatric patient populations and even fewer suitable for pre-school age children.

Table 1. Summary source characteristics of all included studies

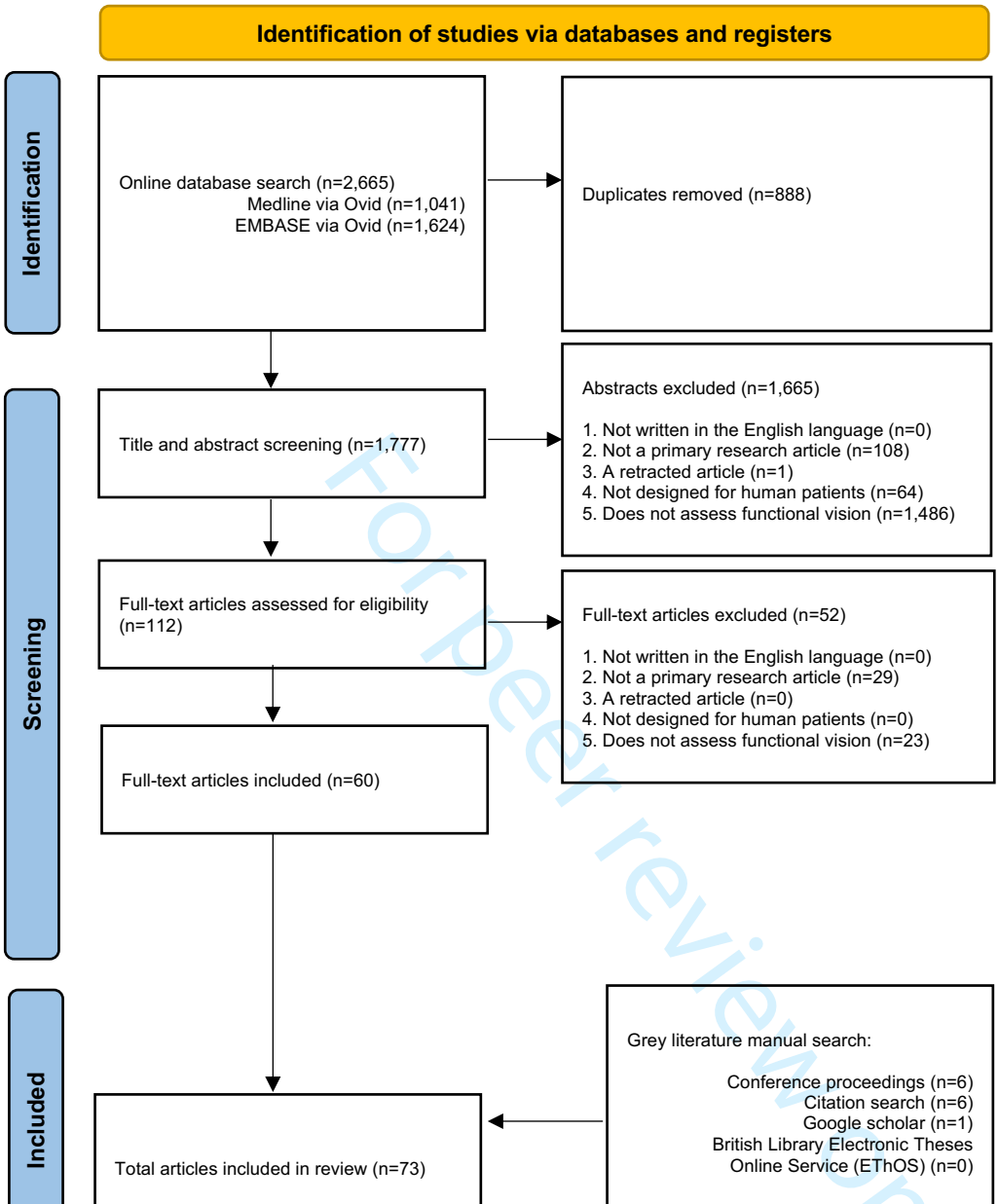
Publication year	**Number of studies**
2005-2010	8
2011-2015	15
2016-2020	24
2021-2024	26
Study design	
Interventional study	
Phase I/II randomised controlled trial	3
Phase III randomised controlled trial	1
Pilot/Feasibility	1
Observational studies	
Cross-sectional	49
Case series	10
Case-control	2
Cohort	1
Conference proceedings	
Abstract	6
Country of institutional affiliation [a]	
North America	38
Europe	24
Asia	4
Oceania	4
Middle East	2
South America	1
Africa	0

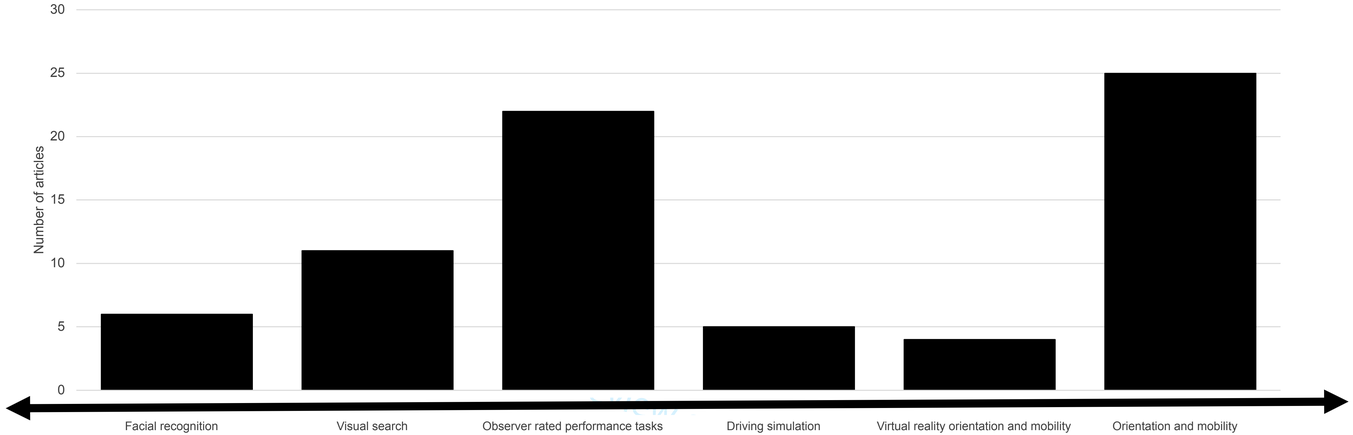
Table 2. Patient population, reference standard, test outcomes, and repeatability and validity data of all included studies featuring a functional vision test

(Uploaded as a separate document due to landscape format)

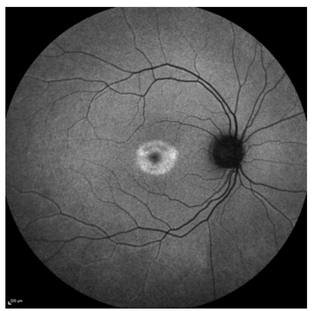
Table 3. Functional vision tests used as clinical trial outcome measures

Name of functional vision test	Disease population	ClinicalTrials.gov Identifier	Type of outcome measure
Multi Luminance Mobility Test (MLMT)	*RPE65*-related Leber's congenital amaurosis	NCT00999609	Primary
	NR2E3 and RHO-related retinitis pigmentosa	NCT05203939	Efficacy
The Functional Low-Vision Observer Rated Assessment (FLORA for Argus II prosthesis)	End-stage retinitis pigmentosa	NCT02303288; NCT03406416	Primary; Secondary
Low Luminance Mobility Testing (LLMT)	Retinitis pigmentosa	NCT03073733	Secondary
Visual Navigation Challenge (Ora-VNC)	*CEP290*-related Leber's congenital amaurosis	NCT03140969; NCT03872479	Secondary
Multi-Luminance Y-Mobility Test (MLYMT)	Retinitis pigmentosa	NCT04945772	Secondary
Vision-guided mobility assessment	*RPE65*-related retinal dystrophy	NCT02781480	Secondary
Orientation and mobility for Argus II prosthesis	End-stage retinitis pigmentosa	NCT00407602	Secondary

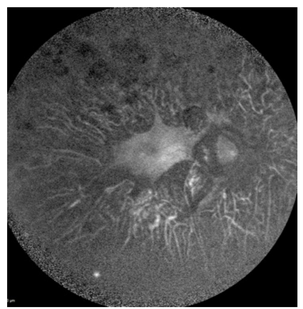




Central vision loss



Peripheral vision loss



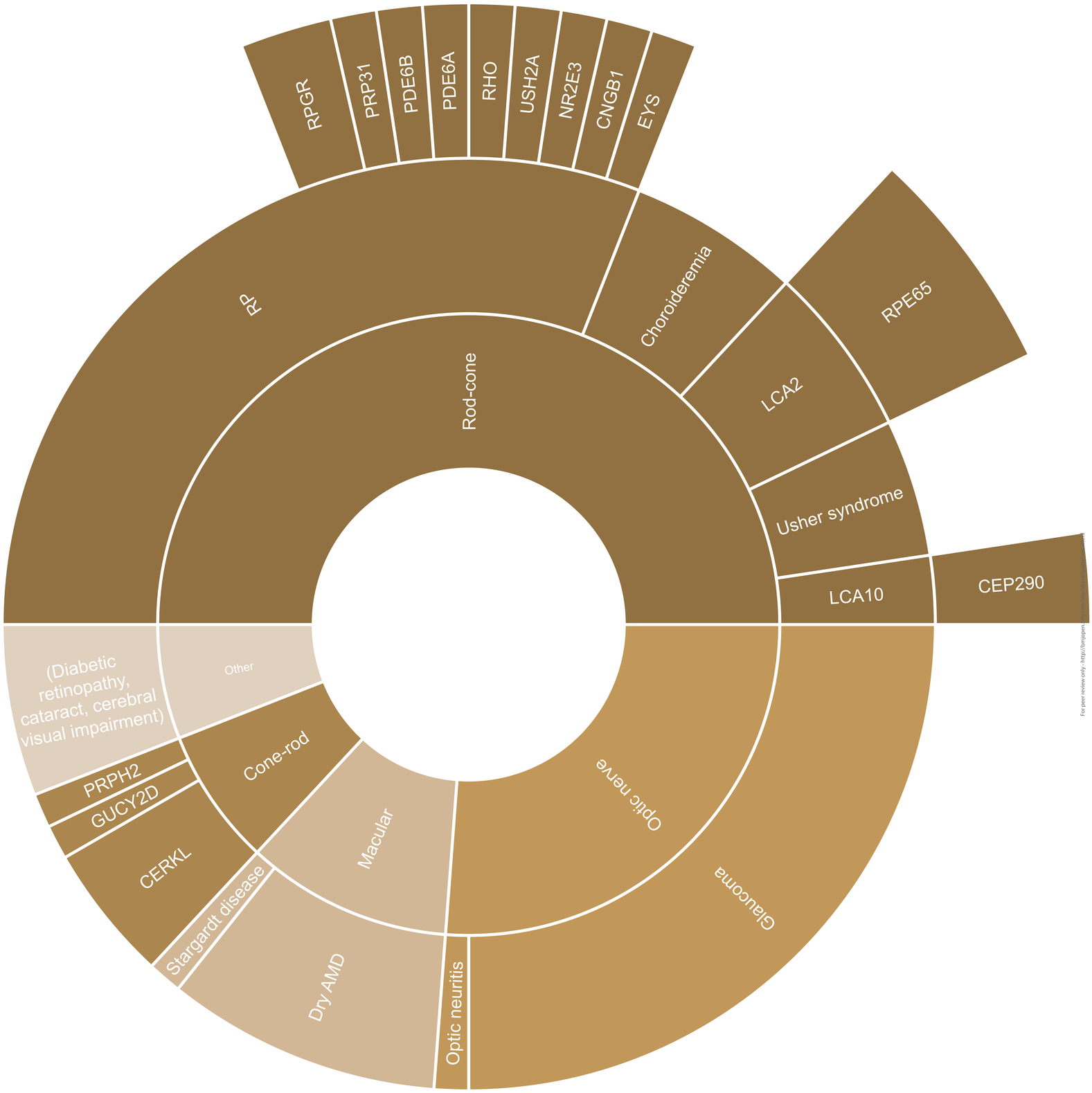


Table 2. Patient population, reference standard, test outcomes, and repeatability and validity data of included studies featuring a functional vision test (abridged; full table available in Supplementary Table S2)

Citation	Patient population	Functional vision test	Reference standard(s)	Test outcome(s)	Reported repeatability and validity data
Orientation and mobility (O&M)					
Roman et al., 2022	10 patients with *GUCY2D*- and *CEP290*-associated Leber's congenital amaurosis	**Mobility test for rod-mediated vision**	VA; FST	Navigation success over a fixed number of trials; Travel duration	**Content validity** - Mobility demonstrated a linear relationship with FST. No correlation between VA and mobility **Construct validity** - No significant difference between controls and patients in suprathreshold transit time (p=0.63). At threshold and dimmer luminance levels transit times increased for both patients and normal subjects.
Sahel et al., 2021	25 patients with retinitis pigmentosa and *RPE65*-associated Leber's congenital amaurosis	**StreetLab mobility course**	VA; VF; CS; Dark adaptation	Course completion time; PWS; PPWS; Number of collisions; Walking initiation time; trajectory analyses/segments; Distance travelled	**Construct validity** – Patients performed worse than controls for PWS, PPWS, number of collisions and walking initiation time under both low and high illumination
Bertaud et al., 2021	22 patients with glaucoma				**Construct validity** – No difference in mobility performance between patients and controls under photopic luminance. Under glare conditions, PWS and PPWS were significantly lower in patients than controls (p=0.049 and p=0.038 respectively). Mobility time was significantly longer in patients than controls (p=0.046). Distance travelled, mobility incidents, and trajectory segmentations not significantly different between patients and controls.
Chung et al., 2018; Maguire et al., 2019	19 patients with *RPE65*-associated Leber's congenital amaurosis	**Multi-Luminance Mobility Test (MLMT)**	VA; VF; FST (white light)	MLMT binocular change score (number of collisions and time to navigate course)	**Content validity** - Variable correlation of accuracy score with quality-of-life questionnaire (r=−0.54 to −0.7). Correlation of mean accuracy score with VA ranged from −0.75 to 0.86. Correlation between mean accuracy score and total degrees of visual field ranged from −0.53 to −0.53. **Construct validity** - Able to distinguish controls from patients. **Repeatability-** High inter-grader agreement for scoring (Cohen's kappa=97.9%). High concordance between scores at baseline visits ranging from 86% to 92%. **Sensitivity to change** - Over 1-year observation period controls had an MLMT change score of 0, representing no change and 20 patients had an MLMT change score of 0. Few patients had an MLMT change score of −1 or −2 (i.e. a worsening).
Lam et al., 2024*	18 patients with *NR2E3 and RHO*-associated retinitis pigmentosa			MLMT monocular change score	**Construct validity –** 6 out of 7 *RHO* patients had stable or improved MLMT scores, including 2 patients that demonstrated a 3-luminance level improvement. All autosomal dominant-*NR2E3* patients had no improvement
Kumaran et al., 2020	19 patients with *RPE65*-related retinal dystrophy	**Vision-guided mobility assessment**	VA; CS; VF; FST; Impact of Vision Impairment Questionnaire	Completion time; error number; walking speed; PPWS	**Repeatability** – Large repeatability coefficient of 1.70 m/s **Content validity -** Mean retinal sensitivity (p=0.022) and total hill of vision (p=0.022) predicted walking speed with significance. No correlation between walking speed and VA (p=0.340) or CS (p=0.433) **Criterion validity** - Walking speed approached significance (p=0.052) and was positively associated with affected subjects' perceived difficulties with mobility
Pierce et al., 2024; Pierce et al., 2024	26 patients with *CEP290*-associated retinal dystrophy	**Ora-VNC (Visual Navigation Challenge)**		Navigation time; Composite score	**Content validity** – Composite score was correlated with BCVA, white light FST and red light FST in both eyes, and blue light FST in the better eye (p < 0.05). **Construct validity** – Nine participants (64%) showed a meaningful improvement from baseline. **Repeatability –** Mean test-retest variability from baseline to retest in the worse eye was 0.6 for VNC composite score (95% confidence interval = −0.1, 1.3). **Sensitivity to change –** Mean change from baseline to 12 months test in the worse eye was -0.1 (-1.2, 1.0).
Virtual reality O&M					
Authie et al., 2023	30 patients with retinitis pigmentosa	**MObility Standardised Test (MOST)**	VA; CS; VF; Dark adaptation	Trial duration; Number of collisions; Number of steps and flags touched; Entries in the dead end; Course redirections	**Construct validity** - Demonstrates discrimination between patients and controls (accuracy larger than 95% in all conditions) and between early and late stages of the disease (mean accuracy of 82.3%). **Content validity** - Average performance score strongly correlated with VA, CS and VF. **Reliability** - Highly reproducible (intraclass correlation coefficient>0.98) and reliable (VR and real-life correlation r=0.98)
Aleman et al., 2021; Bennett et al., 2023	29 patients with choroideremia, *RPE65*-associated Leber's congenital amaurosis, *EYS-, CNGB1-, NR2E3-,*	**Virtual reality orientation and mobility**	VF; FST; VA	Speed; Accuracy (obstacle identification, departures from the path, direction of movement, collisions, and whether the subject missed any arrows or repeated them)	**Content validity** – Better performance in patients with better VA and larger VF extents **Construct validity** – Significant difference in the time to complete obstacle testing between patients and controls (p=0.0027). Controls identified approximately 50% of the obstacles at the dimmest course luminance. All but two patients were able to complete the test although they required higher luminance levels (by >2 log units) to identify 50% of the obstacles.

	RPGR-, CRKL-, PRPH2-, USH2A-, PRPF31-associated retinitis pigmentosa				**Repeatability –** Small improvement in object detection on the second test leading to positive test-retest differences. Greater test-retest values at the dimmest obstacle course luminance level suggestive of a minor learning effect.
Facial recognition					
Hirji et al., 2020; Hirji et al., 2021	72 patients with primary open angle glaucoma with glaucomatous macular damage	**The Cambridge Face Memory test**	VF; CS	Percentage of correctly identified faces	**Content validity** - Significant correlation between facial recognition and VF mean deviation (p<0.0001)
Observer-rated performance tests					
Azoulay-Sebban et al., 2020; Lombardi et al., 2018	32 patients with glaucoma	**Homelab at StreetLab**	VA; CS; VF; NEI VFQ-25	Path travel time; Mobility incidents; Movement onset; movement initiation time and duration; Localisation of people time; Face orientation recognition time	**Construct validity** - No significant difference in path travel time between patients and controls. Number of mobility incidents was higher in advanced glaucoma group than in other 2 groups (p=0.0126 and 0.0281, for controls and early glaucoma respectively). **Content validity** – Integrated binocular field and VA demonstrated significant correlation with test outcomes. Overall movement duration for small objects in reaching and grasping tasks was significantly longer in glaucoma patients compared with controls. Mobility incidents and the reaching and grasping task parameters were not significantly correlated with quality-of-life questionnaire scores.
Visual search					
Higgins et al., 2020	38 patients with non-neovascular age related macular degeneration	**Computer based assessment (Visual search task and simulated dynamic driving scene)**	VA; CS; MP; EuroQol-5D questionnaire	Total correct responses; Median response time	**Construct validity** - Slower performance in visual search tasks associated with more severe disease. No significant difference between groups for total correct responses (p=0.342). Significant difference in median response time between the groups (p=0.007). Late and intermediate group's median response time were not significantly slower than the controls. **Content validity** - Response time was associated with measures of VA and CS.
Kartha et al., 2023	37 patients with ultra-low vision	**Virtual reality visual performance test**	Berkeley Rudimentary Vision Test	Item measure; Person measure	**Content validity –** Negative correlation between patients with poorer visual acuity having lower person measures (p=0.002, r^2=0.2, mean absolute error=0.43). **Construct validity** – Items measures ranged between −.09 to 0.39 in relative d′ units. Person measures ranged between -0.74 and 2.2 relative d' units.
Zhang et al., 2022; Manley et al., 2022	63 patients with cerebral visual impairment	**Virtual toybox and virtual hallway**		Success rate; Reaction time; Gaze error; Visual search area; Off-screen percent (an index of task compliance	**Construct validity** – For the virtual toybox task, mean success rate for patients was significantly lower compared to controls (p<0.001). Significant difference with respect to mean reaction time with patients taking longer to find the target compared to controls (p < 0.001). For the virtual hallway task, mean success rate for patients was significantly lower compared to controls (p<0.001). Mean reaction time was significantly greater in patients compared to controls (p<0.001)
Driving simulators					
Adrian et al., 2022	14 patients with glaucoma	**Fixed base driving simulator at StreetLab**		Reaction times; Longitudinal regulation; lateral control; eye and head movements; Fixation duration and number per second; Fixation duration; horizontal and vertical gaze direction; head yaw	**Construct validity -** Compared to controls, patients demonstrated a longer mean duration of lateral excursions (p=0.045), and more lane excursions in a wide left curve (p=0.045). Patients demonstrated a larger standard deviation of horizontal gaze (p=0.034). No significant difference was established for the other measured outcomes.
Lee et al., 2019	31 patients with glaucoma	**DriveSafe (slide recognition test)**	VA; VF; CS; UFOV® test	Total number of correctly identified road user features (DriveSafe score); number of fixations points; average fixation duration; average saccade amplitude; horizontal and vertical search variance	**Construct validity** - Patients had significantly worse DriveSafe scores (p=0.03), fixated on road users for shorter durations (p<0.001), exhibited smaller saccades (p=0.02), reduced fixation duration and saccadic amplitudes compared to controls (p<0.001 and p=0.02). No other significant group differences were found. **Content validity** - Significant relationship between clinical measures and DriveSafe scores: UFoV 2 (p=0.005), worse-eye VF mean deviation (p=0.003), CS (p=0.03) and UFoV 3 (p=0.05).

VA = visual acuity; BCVA = best corrected visual acuity; VF = visual field; CS = contrast sensitivity; MP = microperimetry; FST = Full-field stimulus testing; FLORA = functional low-vision observer rated assessment; PWS = preferred walking speed; PPWS = percentage preferred walking speed; O&M = orientation and mobility; POAG: primary open angle glaucoma; AMD: age-related macular degeneration; VFQ-25 = Visual Functioning Questionnaire-25; VA LV VFQ-48 = Veterans Affairs Low-Vision Visual Functioning Questionnaire; UFOV = useful-field-of-view. *Indicates a conference abstract. Where a genetic mutation was reported, this has been included in italics. If a form of validation evidence (e.g. construct validity) is absent from table, it was not reported in the original article.

bmjopen-2024-097970 on 27 May 2025. Downloaded from

Supplementary Material

Table S1. Full Boolean search strategy divided into two concepts: functional vision and eye disease

Search strategy performed in MEDLINE and Embase (via Ovid) on 1st August 2024
Functional vision.ti,ab.
Functional ability.ti,ab.
Functional disability.ti,ab.
Functional impairment.ti,ab.
Performance based.ti,ab.
Real world vision.ti,ab.
Real world task.ti,ab.
Daily living task*.ti,ab.
Mobility.ti,ab.
Vis* task.ti,ab.
Visual search.ti,ab.
Eye-Tracking Technology/
Fac* recognition.ti,ab.
1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
Eye Diseases/
Visual* impair*.ti,ab.
Vision impaired.ti,ab.
Glaucoma/
Inherited retinal disease.ti,ab.
Achromatopsia.ti,ab.
Choroideremia.ti,ab.
Stargardt Disease/
Usher Syndromes/
Leber Congenital Amaurosis/
Optic Atrophy, Hereditary, Leber/
Retinitis Pigmentosa/
Macular Degeneration/
15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
14 and 28
limit 29 to (english language and yr="2003 -Current")

Table S2. Patient population, reference standard, test outcomes, and repeatability and validity data of all included studies featuring a functional vision test

Citation	Patient population	Functional vision test	Reference standard(s)	Test outcome(s)	Reported repeatability and validity data
Orientation and mobility (O&M)					
Roman et al., 2022	10 patients with *GUCY2D*- and *CEP290*-associated Leber's congenital amaurosis	**Mobility test for rod-mediated vision**	VA; FST	Navigation success over a fixed number of trials; Travel duration	**Content validity** - Mobility demonstrated a linear relationship with FST. No correlation between VA and mobility **Construct validity** - No significant difference between controls and patients in suprathreshold transit time (p=0.63). At threshold and dimmer luminance levels, transit times increased for both patients and normal subjects.
Sahel et al., 2021	25 patients with retinitis pigmentosa and *RPE65*-associated Leber's congenital amaurosis	**StreetLab mobility course**	VA; VF; CS; Dark adaptation	Course completion time; PWS; PPWS; Number of collisions; Walking initiation time; trajectory analyses/segments; Distance travelled	**Construct validity** – Patients performed worse than controls for PWS, PPWS, number of collisions and walking initiation time under both low and high illumination
Bertaud et al., 2021	22 patients with glaucoma				**Construct validity** – No difference in mobility performance between patients and controls under photopic luminance. Under glare conditions, PWS and PPWS were significantly lower in patients than controls (p=0.049 and p=0.038 respectively). Mobility time was significantly longer in patients than controls (p=0.046). Distance travelled, mobility incidents, and trajectory segmentations not significantly different between patients and controls.
Chung et al., 2018; Maguire et al., 2019 Maguire et al., 2021	19 patients with *RPE65*-associated Leber's congenital amaurosis	**Multi-Luminance Mobility Test (MLMT)**	VA; VF; FST (white light)	MLMT binocular change score (number of collisions and time to navigate course)	**Content validity** - Variable correlation of accuracy score with quality-of-life questionnaire (r=−0.54 to −0.7). Correlation of mean accuracy score with VA ranged from −0.75 to 0.86. Correlation between mean accuracy score and total degrees of visual field ranged from −0.57 to −0.53. **Construct validity** - Able to distinguish controls from patients. **Repeatability-** High inter-grader agreement for scoring (Cohen's kappa=97.9%). High concordance between scores at baseline visits ranging from 86% to 94%. **Sensitivity to change** - Over 1-year observation period controls had an MLMT change score of 0, representing no change and 20 patients had an MLMT change score of 0. Few patients had an MLMT change score of −1 or −2 (i.e. a worsening).
Lam et al., 2024*	18 patients with *NR2E3 and RHO*-associated retinitis pigmentosa			MLMT monocular change score	**Construct validity –** 6 out of 7 *RHO* patients had stable or improved MLMT scores, including 2 patients that demonstrated a 3-luminance level improvement. Autosomal dominant-*NR2E3* patients had no improvement
Kammer et al., 2021*	20 patients with retinitis pigmentosa	**Low Luminance Mobility Test (LLMT)**	VA; CS; VF; VA LV VFQ-48	Critical Illumination Level; Maximum Step Speed score	**Content validity -** All visual function measures significantly related to Critical Illumination Level in a multiple regression model, R^2 =0.75 (p=0.004) **Construct validity** - Able to distinguish controls from patients. **Repeatability** - No change in Critical Illumination Level between test sessions for 75% of patients. Inter-rater and intra-rater grading biases close to zero and no significant differences between graders (p>0.05).
Xu et al., 2021	5 patients with retinitis pigmentosa	**Orientation and mobility test (256 Channel Intelligent Micro Implant Eye implant)**		Effort; Average completion time	Not reported
Boyer et al., 2023*	27 patients with advanced retinitis pigmentosa	**Multi-Luminance Y-Mobility Test (MLYMT)**			Not reported

Kumaran et al., 2020	19 patients with *RPE65*-related retinal dystrophy	**Vision-guided mobility assessment**	VA; CS; VF; FST; Impact of Vision Impairment Questionnaire	Completion time; error number; walking speed; PPWS	**Repeatability** – Large repeatability coefficient of 1.30 m/s **Content validity -** Mean retinal sensitivity (p=0.022) and total hill of vision (p=0.022) predicted walking speed with significance. No correlation between walking speed and VA (p=0.340) or CS (p=0.433) **Criterion validity** - Walking speed approached significance (p=0.052) and was positively associated with affected subjects' perceived difficulties with mobility
Jacobson et al., 2017	22 patients with *CEP290*-associated Leber's congenital amaurosis	**Mobility performance task**	FST	Number of patient incidents (obstacles/wall bumps or reorientations)	**Content validity** – Correlation between mobility score and VA (p =0.002).
Alshaghthrah et al., 2014; Al Saqr et al., 2017	20 patients with retinitis pigmentosa	**Portable mobility course**	VA; CS	PPWS; Collision score	**Content validity** - Significant correlation between CS and collision incidences (p=0.03). No significant correlation between CS and mobility scores (p > 0.05). **Repeatability** - PPWS scores not significantly different (p>0.05) on repeat testing. Collision incidences significantly lower at the second visit (p=0.012). Agreement in collision incidences between the two visits suggestive of no learning effect.
Shapiro et al., 2017*; Pierce et al., 2024; Pierce et al., 2024	Inherited retinal disease 26 patients with *CEP290*-associated retinal dystrophy	**Ora-VNC (Visual Navigation Challenge)**		Navigation time; Composite score	**Construct validity** - Navigation times for controls, mild and severe retinitis pigmentosa were significantly different across all light levels (p<0.05) and between groups (p < 0.05). **Content validity** – Composite score was correlated with LCVA, white light FST and red light FST in both eyes, and blue light FST in the better eye (p < 0.05). **Construct validity** – Nine participants (64%) showed a meaningful improvement from baseline. **Repeatability –** Mean test-retest variability from baseline to retest in the worse eye was 0.6 for VNC composite score (95% confidence interval = −0.1, 1.3). **Sensitivity to change –** Mean change from baseline to 6 months test in the worse eye was -0.1 (-1.2, 1.0).
Russell et al., 2022	11 patients with *CEP290*-associated Leber congenital amaurosis				**Construct validity** - Mean (±standard deviation) improvement in composite score was +2.50±3.118 in treated eyes compared to +1.75±2.383 in untreated eyes (p=0.10). A greater improvement in the composite score from baseline to month 12 was seen in the low dose group (+4.00±3.114 and +2.67±2.714 for treated and untreated eyes, respectively) compared to the higher dose group (+0.25±1.323 and +0.38±0.750, respectively).
Ivanov et al., 2016	25 patients with retinitis pigmentosa	**Natural environment walking task with eye tracking**		PPWS; Number of obstacle collisions; Eye position variability	**Construct validity** - Average PPWS for controls (97%) was higher than all other patient groups.
Ikeda et al., 2015	8 patients with retinitis pigmentosa	**Walking test**		Number of trial failures; Time taken to reach goal	Not reported
Nau et al., 2014	36 patients with low vision	**Obstacle course for BrainPort device**		PPWS; Percentage of obstacle collisions	Not reported
Geruschat et al., 2012	8 patients with advanced retinitis pigmentosa	**Orientation and mobility assessment in retinal prosthesis**	VA; VF	Course completion time; Obstacle contacts	**Construct validity** – Significantly increased obstacle contacts between subjects with worse and those with better VA and VF. No significant difference in course completion time
Kiser et al., 2008	22 patients with age-related macular degeneration	**Mobility obstacle course**		Course completion time; Obstacle contacts	Not reported
Fuhr et al., 2007	44 patients with severe visual impairment	**High density obstacle course**		Course completion time; Obstacle contacts	**Construct validity** – Longer course completion time in patients than age matched controls with significant group effect (p<0.0005). Patients made more obstacle contacts than controls. Analyses of mean number of obstacle contacts showed a significant group effect (p=0.001).
Velikay Parel et al., 2007	10 patients with retinitis pigmentosa, Usher syndrome and optic nerve atrophy	**Mobility assessment**	VA; VF	Average speed; Obstacle contacts	**Content validity** - VA and VF had no significant effect on passing time (p=0.08 and p=0.23 respectively) **Construct validity** - Average passing times between the groups were significantly different (p=0.03). No significant difference in the average number of contacts between groups (p=0.15)
Virtual reality O&M					

Authie et al., 2023	30 patients with retinitis pigmentosa	**MObility Standardised Test (MOST)**	VA; CS; VF; Dark adaptation	Trial duration; Number of collisions; Number of steps and flags touched; Entries in the dead end; Course redirections	**Construct validity** - Demonstrates discrimination between patients and controls (accuracy larger than 95% in all conditions) and between early and late stages of the disease (mean accuracy of 82.3%). **Content validity** - Average performance score strongly correlated with VA, CS and VF. **Reliability** - Highly reproducible (intraclass correlation coefficient>0.98) and reliable (VR and real-life correlation r=0.98)
Aleman et al., 2021; Bennett et al., 2023	29 patients with choroideremia, *RPE65*-associated Leber's congenital amaurosis, *EYS-, CNGB1-, NR2E3-, RPGR-, CRKL-, PRPH2-, USH2A-, PRPF31*-associated retinitis pigmentosa	**Virtual reality orientation and mobility**	VF; FST; VA	Speed; Accuracy (obstacle identification, departures from the path, direction of movement, collisions, and whether the subject missed any arrows or repeated them)	**Content validity** – Better performance in patients with better VA and larger VF extents **Construct validity** – Significant difference in the time to complete obstacle testing between patients and controls (p=0.0027). Controls identified approximately 50% of the obstacles at the dimmest course luminance. All but two patients were able to complete the test, although they required higher luminance levels (by >2 log units) to identify 50% of the obstacles. **Repeatability –** Small improvement in object detection on the second test leading to positive test-retest differences. Greater test-retest values at the dimmest obstacle course luminance level suggestive of a minor learning effect.
Daga et al., 2017	31 patients with glaucoma	**Virtual Environment Human Navigation Task (VEHuNT)**	VF	Time to complete task	**Construct validity** - Significant difference on average time to complete task between patients and controls for room A (p=0.001). No significant difference on average time to complete the task between patients and controls for room B (p=0.514). Significant relationship between time to complete the task and visual field loss for room A but not for room B (p=0.001).
Facial recognition					
Hirji et al., 2020; Hirji et al., 2021	72 patients with primary open angle glaucoma with glaucomatous macular damage	**The Cambridge Face Memory test**	VF; CS	Percentage of correctly identified faces	**Content validity** - Significant correlation between facial recognition and VF mean deviation (p<0.0001)
Glen et al., 2012; Glen et al., 2013	54 patients with glaucoma				**Construct validity** - Patients with advanced VF defects identified fewer faces on average than those with early and moderate defects and controls (p<0.05).
Mazzoli et al., 2019	64 patients with age-related macular degeneration and 48 patients with primary open angle glaucoma				**Construct validity** – Test scores were lower in patients compared to controls (p<0.001).
Taylor et al., 2018	30 patients with non-neovascular age-related macular degeneration				**Construct validity** - Geographic atrophy patients identified significantly fewer faces on average than early and intermediate AMD patients and controls (p=0.04).
Observer-rated performance tests					
Delyfer et al., 2021 Karapanos et al., 2021, Petoe et al., 2021	18 patients with retinitis pigmentosa 4 patients with retinitis pigmentosa	**Functional Low-Vision Observer Rated Assessment (FLORA)**		Final impact rating; Task performance score	Not reported
Greenberg et al. 2015	30 patients with retinitis pigmentosa				
Yoon et al., 2021	5 patients with retinitis pigmentosa				
Geruschat et al., 2015	26 patients with retinitis pigmentosa				
Altangerel et al., 2006	43 patients with primary open angle glaucoma	**Assessment of Function Related to Vision (AFREV)**	VF; VA; CS	AFREV score	**Content validity** - AFREV scores highly correlated with CS (r = 0.772), binocular VA (r=−0.768), better-eye VA (r =−0.737), worse-eye VA (r =−0.675), and VF scores (r = 0.606) and NEI-VFQ scores (r = 0.70). **Construct validity** – Distinguishes between mild, moderate and severe binocular VF loss.

Kulkarni et al., 2012; Warrian et al., 2010; Warrian et al., 2009 Richman et al., 2010, Richman et al., 2010	192 patients with glaucoma 91 patients with diabetic retinopathy 112 patients with age-related macular degeneration 192 patients with glaucoma	**Assessment of Disability Related to Vision (ADREV)**	VF VA; CS; VF; VFQ-25 VA; CS; VF; VFQ-25 VA; CS; VF; Stereopsis	ADREV score	**Content validity** - Highest correlation with the total ADREV score was the integrated VF score (p=-0.49). **Content validity** – All of the ADREV's scales were correlated with one or more clinical measures of visual function except the Ambulation test. **Content validity** – 66% of correlations made between clinical ophthalmic measurements and ADREV scores were significant to P<0.0007. 55% of correlations made between the ADREV and the VFQ total and subscale scores were significant to P< 0.0004. **Content validity** – ADREV performance was strongly associated with binocular VA (P<0.001) and binocular CS (P<0.001). Monocular and binocular VF results had a weaker correlation with the ability to perform the ADREV tasks (P<0.05).
Edwards et al., 2018	6 patients with advanced retinitis pigmentosa implanted with Retina Implant Alpha AMS - *USH2A, PDE6B, RPE65, RPGR, CERKL*	**Tabletop object and clock face recognition**	No. of correctly location and named items	Not reported	
Azoulay-Sebban et al., 2020; Lombardi et al., 2018	32 patients with glaucoma	**Homelab at StreetLab**	VA; CS; VF; NEI VFQ-25	Path travel time; Mobility incidents; Movement onset; movement initiation time and duration; Localisation of people time; Face orientation recognition time	**Construct validity** - No significant difference in path travel time between patients and controls. Number of mobility incidents was higher in advanced glaucoma group than in other 2 groups (p=0.0126 and 0.0281, for controls and early glaucoma respectively). **Content validity** – Integrated binocular field and VA demonstrated significant correlation with test outcomes. Overall movement duration for small objects in reaching and grasping tasks was significantly longer in glaucoma patients compared with controls. Mobility incidents and the reaching and grasping task parameters were not significantly correlated with quality-of-life questionnaire scores.
Wei et al., 2012 Sun et al., 2016 Waisbourd et al., 2019	9 patients with glaucoma 161 patients with glaucoma 153 patients with glaucoma	**CAARV (Compressed Assessment of Ability Related to Vision)**	VA; CS; VF VF VA; CS; VF; VFQ-25	Total CAARV score	**Content validity** – Strongest correlation was between the central VF cluster and total CAARV score (P<0.001). Central VF cluster in the better eye positively correlated with the majority of CAARV and NEI VFQ-25 subscales. **Construct validity** – Compared to non-rapid progressors, patients who had rapidly progressing glaucoma presented with lower baseline CAARV scores for reading street signs (p=0.01), facial recognition (p=0.01), and total score (p<0.001).
Reighard et al., 2019	145 patients with glaucoma	**I-CAARV (Indian - Compressed Assessment of Ability Related to Vision)**	VA; VF; CS; Indian-VFQ	I-CAARV score	**Content validity -** I-CAARV scores and the Indian VFQ were significantly correlated (P<0.01). Rasch-calibrated scores on the I-CAARV were also significantly correlated with VF MD, presenting VA, best-corrected VA, and CS in both the better-seeing eye (p=0.60, p=-0.91, p=-0.53, p=0.76 respectively) and worse-seeing eye (p=0.48, p=-0.61, p=-0.53, p=0.69). **Repeatability –** Rasch analysis found that the I-CAARV had moderate reliability (0.74) and measurement precision was fair (person separation 1.67 logits). Rasch analysis found good **construct validity** (infit range 0.66-1.13; outfit range 0.65-1.21)
Peterson et al., 2023*	36 patients with age-related macular degeneration	**Performance-based activities of daily living task tests (ADLTT)**	VA; CS; MP	Task completion time	**Construct validity –** Longer task completion time in patients than controls for money counting task using worse eye vision and binocular vision (both p<0.001) and object drink making task using monocular worse eye vision (p=0.033). **Content validity –** Only the money counting task demonstrated moderate to strong correlations with VA, CS, and MP. **Divergent validity** was demonstrated when correlated with race and gender in most ADLTTs except for facial expression task.

					Repeatability - Moderate to good test-retest reliability for money counting and drink making tasks only using monocular worse eye vision.
Ni et al., 2012	64 patients with age-related cataract	**Real-Life Vision Test (RLVT)**	VA; CS; Stereopsis; Colour perception; VFQ-25	Time taken to complete task	**Construct validity –** Controls performed significantly better than patients (P<0.01). Significant difference reported between patients with different cataract severity **Content validity** - All RLVT subscales remained highly associated with most clinical measures, after controlling for age, years of education, Mini Mental State Examination scores, self-rating depression scores, and reaction time.
Finger et al., 2014	40 patients with rod-cone dystrophy	**Very Low Vision Instrumental Activities of Daily Living (IADL-VLV)**	VA; VF	Completion and accuracy score	**Content validity –** VA and VF were associated with IADL performance. **Construct validity –** Patients with worse VA or VF scored slower (p<0.00 and p=0.001 respectively)
Visual search					
Higgins et al., 2020	38 patients with non-neovascular age related macular degeneration	**Computer based assessment (Visual search task and simulated dynamic driving scene)**	VA; CS; MP; EuroQol-5D questionnaire	Total correct responses; Median response time	**Construct validity** - Slower performance in visual search tasks associated with more severe disease. No significant difference between groups for total correct responses (p=0.342). Significant difference in median response time between the groups (p=0.007). Early and intermediate group's median response time were not significantly slower than the controls. **Content validity** - Response time was associated with measures of VA and CS.
Taylor et al., 2017	31 patients with dry age-related macular degeneration		VA; CS	Median search time; Fixation duration; Saccadic amplitude; Saccades per second	**Content validity** – Significant associations between median search time and VA (p<0.001) and CS (p<0.001) **Construct validity –** 61% of patients exceeded the normative limits for average search time; this was statistically significant (p<0.0001). No differences between groups in fixation duration or saccades per second. Yet saccadic amplitude remained significantly smaller for patients compared to controls (p<0.001).
Thibaut et al., 2018	21 patients with age related macular degeneration	**Object search in realistic panoramic scenes**		Percentage of correct target detection; percentage of false positives; scene views explored; search time	**Construct validity -** No significant differences in performance between patients and age-matched controls.
Wan et al., 2020	30 patients with age-related cataract	**Visual search and facial recognition task**		Fixation count and total duration; total visit duration; Forward and backward saccade count per line; percentage of regressive saccades; percentage of correctly identified faces	**Construct validity** – Significant difference before and after surgery for the percentage of correctly identified objects and faces (p=0.049 and p=0.004 respectively), average search time (p<0.001), fixation count (p<0.001), total fixation duration (p= 0.039) and total visit duration (p=0.008). No significant change was in mean fixation duration. **Repeatability** - No significant difference between baseline and follow-up assessment (all parameters p<0.05)
Kartha et al., 2023	37 patients with ultra-low vision	**Virtual reality visual performance test**	Berkeley Rudimentary Vision Test	Item measure; Person measure	**Content validity –** Negative correlation between patients with poorer visual acuity having lower person measures (p=0.002, r^2=0.2, mean absolute error=0.43). **Construct validity** – Items measures ranged between −0.09 to 0.39 in relative d′ units. Person measures ranged between -0.74 and 2.2 relative d' units.
Martínez-Almeida et al., 2021	33 patients with glaucoma	**Virtual reality system with gaze monitoring**		Fixation number and duration; Saccadic amplitude and velocity; Fixation/saccade ratio; Total search and execution time; Number of collisions	**Construct validity** – Significant differences between controls and patients for the static task in terms of number of fixations (p=0.012), mean saccadic velocity (p=0.02 and 0.017), fixations/saccades ratio (p=0.035 and 0.04), and the search and total execution times during visual search exercise (p=0.004 and 0.027, respectively). For the dynamic task, Significant differences were found on average saccades amplitude (p=0.02), average saccades velocity (p=0.03) and the number of collisions (p=0.02).
Kurek et al., 2023*	30 patients with retinitis pigmentosa	**Virtual reality visual search task with natural scenes**	CS	Performance score (encompassing search duration and rate of performance success)	**Construct validity –** Able to discriminate between patients and controls (Accuracy >86%) **Repeatability** – Good agreement of performance score between sessions (Intraclass correlation coefficient>0.89) **Content validity** - Correlation with CS was p=0.76. 83% of RP participants indicated that the virtual reality test was representative of their difficulties in daily li

Zhang et al., 2022; Manley et al., 2022	63 patients with cerebral visual impairment	**Virtual toybox and virtual hallway**		Success rate; Reaction time; Gaze error; Visual search area; Off-screen percent (an index of task compliance	**Construct validity** – For the virtual toybox task, mean success rate for patients was significantly lower compared to controls (p<0.001). Significant difference with respect to mean reaction time with patients taking longer to find the target compared to controls (p < 0.001). For the virtual hallway task, mean success rate for patients was significantly lower compared to controls (p<0.001). Mean reaction time was significantly greater in patients compared to controls (p<0.001)
Roux-Sibilon et al., 2018	22 patients with glaucoma	**Scene and face recognition**	VF	Participant's response; Reaction time for response	**Construct validity** - Patients demonstrated deficit in both detection and categorization of all low-contrast images compared to controls.
Smith et al., 2012	40 patients with glaucoma	**Visual search task with eye tracking**	VF; CS	Average number of saccades per second; average saccade amplitude; Average search duration	**Construct validity** - Average rate of saccades by patients was significantly smaller than controls during the visual search task (p=0.02). No difference in average saccade amplitude between the patients and controls (p=0.09). **Content validity** - Average number of saccades was negatively correlated with CS (p=0.006) and more severe VF defects (p=0.037).
Driving simulators					
Adrian et al., 2022	14 patients with glaucoma	**Fixed base driving simulator at StreetLab**		Reaction times; Longitudinal regulation; lateral control; eye and head movements; Fixation duration and number per second; Fixation duration; horizontal and vertical gaze direction; head yaw	**Construct validity -** Compared to controls, patients demonstrated a longer mean duration of lateral excursions (p=0.045), and more lane excursions in a wide left lane (p=0.045). Patients demonstrated a larger standard deviation of horizontal gaze (p=0.034). No significant difference was established for the other measured outcomes.
Kübler et al., 2015	6 patients with glaucoma	**Simulated driving test**		Driving lane positions; time to line crossing (indicates steering stability); driving speed; head and eye tracking	Not reported
Lee et al., 2019	31 patients with glaucoma	**DriveSafe (slide recognition test)**	VA; VF; CS; UFOV® test	Total number of correctly identified road user features (DriveSafe score); number of fixations points; average fixation duration; average saccade amplitude; horizontal and vertical search variance	**Construct validity** - Patients had significantly worse DriveSafe scores (p=0.03), fixated on road users for shorter durations (p<0.001), exhibited smaller saccades (p=0.02), reduced fixation duration and saccadic amplitudes compared to controls (p<0.001 and p=0.02). No other significant group differences were found. **Content validity** - Significant relationship between clinical measures and DriveSafe scores: UFoV 2 (p=0.005), worse-eye VF mean deviation (p=0.003), CS (p=0.03) and UFoV 3 (p=0.05).
Devos et al., 2018	17 patients with glaucoma	**Performance based visual field test in a driving simulator**	VF; UFOV®	Total crashes; Speed exceedances; Correct stops at traffic lights; Centre line crossings; Road edge excursions; Complex response time; Target identification accuracy; Number of missed responses; Response time	**Construct validity** - Patients identified fewer VF symbols (p=0.047) and took longer (p=0.048) to detect the VF symbols compared to controls. No significant differences for the other driving performance measures. **Content validity** - Correlation between performance-based VF test scores and horizontal FOV of the Keystone vision screener and UFOV® divided attention subtest (p=0.02 and p=0.046 respectively). **Repeatability** – Intraclass correlation ranged between 0.17 for response time and 0.92 for correct responses.
Prado-Vega et al., 2013	23 patients with glaucoma	**Driving simulator with eye-scanning**	VF	Steering activity; Lane keeping; Longitudinal and lateral distance to obstacle; Collisions	**Construct validity** - No significant difference between patients and controls for lane keeping, obstacle avoidance, and eye-scanning behaviour. Steering activity was significantly higher for patients than for controls. **Content validity** – No significance correlation between the percentage of depressed IVF points and driving performance measures (p>0.2).

VA = visual acuity; BCVA = best corrected visual acuity; VF = visual field; CS = contrast sensitivity; MP = microperimetry; FST = Full-field stimulus testing; FLORA = functional low-vision observer rated assessment; PWS = preferred walking speed; PPWS = percentage preferred walking speed; O&M = orientation and mobility; POAG: primary open angle glaucoma; AMD: age-related macular degeneration; VFQ-25 = Visual Functioning Questionnaire-25; VA LV VFQ-48 = Veterans Affairs Low-Vision Visual Functioning Questionnaire; UFOV = useful-field-of-view. *Indicates a conference abstract. Where a genetic mutation was reported, this has been included in italics. If a form of validation evidence (e.g. construct validity) is absent from table, it was not reported in the original article.