Cochrane Library

Cochrane Database of Systematic Reviews

Tests for detecting strabismus in children aged 1 to 6 years in the community (Review)

Hull S, Tailor V, Balduzzi S, Rahi J, Schmucker C, Virgili G, Dahlmann-Noor A

WILEY

TABLE OF CONTENTS

Tests for detecting strabismus in children aged 1 to 6 years in the community

Sarah Hull[1], Vijay Tailor[1], Sara Balduzzi[2], Jugnoo Rahi[3], Christine Schmucker[4], Gianni Virgili[5], Annegret Dahlmann-Noor[1]

[1]NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. [2]Cochrane Italy, Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy. [3]Department of Epidemiology, UCL Institute of Child Health and UCL Institute of Ophthalmology, London, UK. [4]Cochrane Germany, Medical Center - Univ. of Freiburg, Faculty of Medicine, Univ. of Freiburg, Freiburg, Germany. [5]Department of Translational Surgery and Medicine, Eye Clinic, University of Florence, Florence, Italy

Contact address: Annegret Dahlmann-Noor, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, 162 City Road, London, EC1V 2PD, UK. annegret.dahlmann-noor@moorfields.nhs.uk.

A B S T R A C T

Background

Strabismus (misalignment of the eyes) is a risk factor for impaired visual development both of visual acuity and of stereopsis. Detection of strabismus in the community by non-expert examiners may be performed using a number of different index tests that include direct measures of misalignment (corneal or fundus reflex tests), or indirect measures such as stereopsis and visual acuity. The reference test

to detect strabismus by trained professionals is the cover‐ uncover test.

Objectives

To assess and compare the accuracy of tests, alone or in combination, for detection of strabismus in children aged 1 to 6 years, in a community setting by non-expert screeners or primary care professionals to inform healthcare commissioners setting up childhood screening programmes.

Secondary objectives were to investigate sources of heterogeneity of diagnostic accuracy.

Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL; 2016, Issue 12) (which contains the Cochrane Eyes and Vision Trials Register) in the Cochrane Library, the Health Technology Assessment Database (HTAD) in the Cochrane Library (2016, Issue 4), MEDLINE Ovid (1946 to 5 January 2017), Embase Ovid (1947 to 5 January 2017), CINAHL (January 1937 to 5 January 2017), Web of Science Conference Proceedings Citation Index-Science (CPCI-S) (January 1990 to 5 January 2017), BIOSIS Previews (January 1969 to 5 January 2017), MEDION (to 18 August 2014), the Aggressive Research Intelligence Facility database (ARIF) (to 5 January 2017), the ISRCTN registry (www.isrctn.com/editAdvancedSearch); searched 5 January 2017, ClinicalTrials.gov (www.clinicaltrials.gov); searched 5 January 2017 and the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) (www.who.int/ictrp/search/en); searched 5 January 2017. We did not use any date or language restrictions in the electronic searches for trials. In addition, orthoptic journals and conference proceedings without electronic listings were searched.

Selection criteria

All prospective or retrospective population-based test accuracy studies of consecutive participants were included. Studies compared a single or combination of index tests with the reference test. Only those studies with sufficient data for analysis were included specifically to calculate sensitivity and specificity and determine diagnostic accuracy.

Participants were aged 1 to 6 years. Studies reporting participants outside this range were included if subgroup data were available.

Permitted settings included population-based vision screening programmes or opportunistic screening programmes, such as those performed in schools.

Data collection and analysis

We used standard methodological procedures expected by Cochrane. In brief, two review authors independently assessed titles and abstracts for eligibility and extracted the data, with a third senior author resolving any disagreement. We analysed data primarily for specificity and sensitivity.

Main results

One study from a total of 1236 papers, abstracts and trials was eligible for inclusion with a total number of participants of 335 of which 271 completed both the screening test and the gold standard test. The screening test using an automated photoscreener had a sensitivity of 0.46 (95% confidence interval (CI) 0.19 to 0.75) and specificity of 0.97 (CI 0.94 to 0.99). The overall number affected by strabismus was low at 13 (4.8%).

Authors' conclusions

There is very limited data in the literature to ascertain the accuracy of tests for detecting strabismus in the community as performed by non-expert screeners. A large prospective study to compare methods would be required to determine which tests have the greatest accuracy.

P L A I N   L A N G U A G E   S U M M A R Y

Tests for detecting strabismus in children aged one to six years in the community

Review aim

The aim of this Cochrane Review was to find out how well different tests work to detect strabismus in children aged 1 to 6 years old outside of eye departments. These tests were used in the community and were performed by screeners who were not eye specialists.

Background

Strabismus (also known as squint) occurs when the eyes are not aligned. It can lead to reduced vision and failure of the eyes to work properly together, including for 3D vision. A number of different tests can be used to screen for strabismus directly, by measuring the misalignment; or indirectly, by measuring the level of vision in each eye (visual acuity); or by measuring 3D vision (stereopsis). It is unknown which of these tests is the most accurate in correctly identifying children with strabismus.

Results and conclusion

Only one study was found that met the standards to be included in this review. This study used a photoscreener (a type of camera that measures refractive error and misalignment). Following screening, all children were offered an examination by an eye-care specialist to confirm which children did have strabismus. The photoscreener was very accurate in identifying those children without strabismus (highly specific) but not accurate in correctly identifying those children with strabismus (low sensitivity only).

As only one study could be included in this review, it was not possible to conclude which test is the most accurate for screening for strabismus. Further studies would be needed to determine this. However, they would need to include very large numbers of children to be able to make statistically valid conclusions.

How up to date is this review?

Cochrane researchers searched for studies that had been published up to 5 January 2017.

BACKGROUND

Target condition being diagnosed

Strabismus is a physical condition in which the eyes are not aligned. It is associated with deficient binocularity, the mechanism that integrates visual information from both eyes. Strabismus can be primary, or it can be a consequence of poor vision in one eye or of refractive errors. Less commonly, strabismus can be caused by lesions affecting the oculomotor, trochlear or abducens nerve, or higher neurological pathways. Strabismus is rarely caused by developmental or traumatic defects of the extraocular muscles. Strabismus is a risk factor for the development of amblyopia during the 'sensitive period' of vision development. During this period, neural plasticity is greatest, and it begins to decline around the age of 6 years; clinical interventions are typically offered to children up to the age of 10 years. Screening programmes therefore attempt to identify children with amblyogenic risk factors before the age of 6 years, to allow remedial treatment.

Prevalence figures for strabismus vary. The most recent screening study in Baltimore, USA, found a prevalence of manifest strabismus of 3.3% in Caucasian and 2.1% in African American children aged 6 to 71 months (Friedman 2009). Other population-based studies have reported a prevalence of childhood strabismus between 0.01% and 3.1%, indicating that prevalence may vary greatly by ethnicity, age, type of strabismus and definitions used (Graham 1974; Matsuo 2007a; Matsuo 2007b; Preslan 1996; Traboulsi 2008; Turacli 1995; Wedner 2000; Appendix 1).

Relevance of strabismus in children

There are many subtypes of strabismus. In the context of childhood vision screening programmes, the most relevant distinction is between manifest and latent strabismus. Manifest strabismus is a risk factor for the development of amblyopia, the commonest vision disorder in children (prevalence 1.6% to 3.6% in Western societies) (Simons 1996a).

Amblyopia is a developmental anomaly of spatial vision, usually associated with strabismus, anisometropia or from deprivation early in life (Ciuffreda 1991). Amblyopes have reduced visual acuity in one or both eyes, reduced contrast sensitivity and reduced contour integration. Clinical definitions of amblyopia are based on visual acuity only, taking into consideration the age of the child and progressive improvement of 'normal visual acuity' in the early years. Unilateral amblyopia is often defined as an interocular difference in best-corrected visual acuity (of 2 logMAR or Snellen chart lines) (Friedman 2009), or best-corrected visual acuity of 0.30 logMAR or worse in either eye (Rahi 2002; Traboulsi 2008). In 3-year-old children, bilateral amblyopia is suspected if best-corrected visual acuity is worse than 0.40 logMAR in one eye and worse than 0.3 logMAR in the other eye in the presence of a bilateral amblyogenic

risk factor. In 4-year-old children, the thresholds are 0.3 and 0.18 logMAR, respectively (Schmidt 2004).

Strabismus has a profound effect on stereopsis or perception of depth. Stereopsis normally develops within the first 3 to 4 months of age and reaches adult levels by the age of 24 to 36 months (Braddick 1980; Fawcett 2005; Fox 1980; Petrig 1981; Takai 2005). Two studies reported that stereoacuity continues to develop beyond the age of 3 years, and may not yet be fully mature at 5 years or 12 years of age, respectively (Simons 1981a; Walraven 1993). Normal adult stereopsis is 50 to 60 seconds of arc; some childhood vision screening programmes have used a threshold of 400 seconds of arc for "suspicion of amblyopia" (Traboulsi 2008). Reduced stereopsis adversely affects motor skills, particularly fine motor skills (Grant 2007; Hrisos 2006; O'Connor 2010; Webber 2008).

Significant misalignment can affect development (through unilateral reduced acuity, lack of depth perception and limitation of peripheral visual field), social interactions, and emotional well-being. In children with infantile esotropia, surgical correction of strabismus leads to improvement in general development as measured by the Bayley scale (Rogers 1982). Scores on anxiety and depression scales such as the National Eye Institute Visual Functioning Questionnaire and the Hospital Anxiety and Depression Scale are significantly different from non-strabismic children, and improve following surgical strabismus correction (Bernfeld 1982; Chai 2009). Children with strabismus may have significantly greater conduct and externalising problems (Koklanis 2006).

Strabismus can also be an indicator of severe eye and health problems. As it can indicate poor vision, it may in rare cases be the first sign of childhood cataract, glaucoma, or tumours of the eye, optic nerve, orbit or brain, such as retinoblastoma, glioma, or rhabdomyosarcoma.

Gross misalignment of the eyes is usually noticed by members of the family or carers. Small angles of deviation are not necessarily apparent. In young children, features such as a broad nasal bridge or certain lid positions and shape (epicanthus) can give rise to pseudostrabismus, i.e. a perception of strabismus when in fact the eyes are straight.

Diagnosis of strabismus: the cover test

The cover test is based on the observation of a refixation movement of a deviated eye when the fixing eye is covered (Gamble 1950; McKean 1976; Romano 1971; Scott 1973). The basic form of

the cover test, the cover-uncover test, establishes the diagnosis of manifest strabismus. An occluder is introduced in front of one eye, then removed, re-establishing binocular viewing. If an eye moves when the other is covered, this indicates that this eye was not fixing before the cover was introduced. Any eye movement is interpreted as 'test positive' and 'manifest strabismus present'; the magnitude of the movement is often categorised as small, moderate or large. This test is used in some screening programmes to detect strabismus

(Fogt 2000). The accuracy of the cover test in detecting strabismus may be affected by the child's age at screening, with better test performance in children over the age of 3 years (Williams 2001). Variations of the cover test are used to diagnose latent strabismus, and to measure the magnitude of both manifest and latent strabismus. The presence of latent strabismus is assessed by using the alternate cover test. The occluder is moved from eye to eye, allowing viewing of the target with one eye only, without permitting binocular viewing. The observer notes refixation movements of either eye as the cover is removed.

Quantitative measurements are obtained by neutralising the strabismus with prisms held in front of one eye whilst performing the cover⁻ uncover test (simultaneous prism cover test) or the alternate cover test (prism alternate cover test); the endpoint of measurements is the prism with which no refixation movement is observed when the cover is removed. To trained professionals (orthoptists) refixation on cover test can indicate strabismus; however, this has not been used in published screening studies.

All cover tests are carried out with the participant fixing on a target presented at distances of 6, 4 or 3 metres, and then at near distances (33 cm or 40 cm). In children, the distance target is often presented at 3 metres. In very young children the test is often only carried out at near fixation.

The cover⁻ uncover test aims to detect strabismus, but not refractive errors, the other significant group of amblyogenic risk factors. Its accuracy as a standalone amblyopia screening test is therefore limited (Schmidt 2004). Conversely, addition of the cover⁻ uncover test to vision screening tests increases the detection rate of strabismus (VIP 2007). Vision screening programmes for children between 4 and 6 years traditionally use optotype testing to determine visual acuity (matching or naming letters or pictures), with or without a cover test to detect strabismus. In an effort to screen younger children to identify and treat problems early, these 'manual' screening programmes are increasingly supplemented or even replaced by the use of devices such as photorefractors, which also aim to provide information about refractive amblyogenic risk factors. The American Association of Pediatric Ophthalmology and Strabismus (AAPOS) recently published updated recommendations for automated screening programmes (Donahue 2013). Screening methods were categorised into refractive and non-refractive screening instruments. With regards to detection of strabismus, the AAPOS recommends that non-refractive screening devices should detect manifest strabismus greater than 8 prism dioptres (PD) in primary position (Donahue 2013). UK recommendations suggest that screening at age 4 to 5 years old provides the most accuracy and allows adequate time to treat (Solebo 2015).

Index test(s)

Different tests are in use to detect strabismus in a community or primary care setting by non-expert screeners or primary care professionals.

- Type 1 tests directly identify ocular misalignment, for example corneal (Hirschberg) or fundus reflections tests (Brückner).
- Type 2 tests assess binocular function such as stereoacuity, e.g. contour and random dot stereotests from which presence of strabismus is deduced.
- Type 3 tests are designed to detect reduced central vision/ visual acuity which may in turn be associated with strabismus, e.g. HOTV, LEA symbols, Keeler (previously Glasgow) crowded logMAR, Sonksen crowded logMAR, crowded Kay picture test, displayed as paper- or computer-based charts or conventional retroilluminated charts.
- Type 4 tests are those automated refraction devices which are also designed to report ocular misalignment.

We planned to include studies that report combinations of several index tests.

Other tests for strabismus, such as controlled binocular acuity test, suppression tests, blur test, and tests designed to detect reduced fusional reserve (prism reflex test, prism fusion range) are not used by lay screeners, but only by trained professionals (orthoptists). Orthoptist-delivered screening is not within the remit of this review.

Principles underlying each type of index test

Detailed information about each index test is given in Appendix 2.

Type 1: tests which directly identify ocular misalignment

In manifest strabismus with childhood onset, information from the deviating eye is suppressed, so that a person does not perceive double vision. The principle of the cover test is that when the fixating eye is covered, the deviating eye will move to primary position (looking straight ahead position) to take up fixation, as long as it has some vision and does not have eccentric fixation or severe eye movement deficit. Presence, speed and magnitude of this refixation movement are the outcomes of the cover test.

Type 2: tests of binocular function ⁻ stereopsis

The visual axes need to be within a certain angle to each other in order to detect information that is presented in stereotests. Strabismus may be associated with reduced stereopsis.

Type 3: tests designed to detect reduced central vision/visual acuity

Though not a specific indicator, visual acuity tests may indicate the presence of strabismus-induced amblyopia.

Type 4: automated refraction devices designed to report ocular misalignment

Some autorefractors indicate asymmetry of corneal reflections.

Clinical pathway

Childhood vision screening programmes vary around the world, and may differ between high- and low- to middle-income countries.

World Health Organization (WHO) Member States are grouped into low- and middle-income countries (LMIC) by WHO region, separating out high-income countries within each of these regions, based on the World Bank's gross national income per capita (World Bank; World Health Organization 2014). Low- and middle-income states have a gross national income per capita of less than USD 12,276 whereas high-income states have a gross national income per capita of USD 12,276 or more.

High-income countries often have national guidelines for screening, though these are not necessarily matched by an established national screening programme. For example in Israel and Sweden, there are established national screening programmes (Schmucker 2009). In the UK, the National Screening Committee recommends vision screening of children during the school year of their fifth birthday, to be delivered under the supervision of orthoptists with the focus on screening for visual impairment as the target condition, not other risk factors for amblyopia (Hall 2003; UK National Screening Committee). Abnormal screening results trigger referral for a comprehensive eye examination. Despite this recommendation, implementation of childhood vision screening continues to show regional variation. In the USA, Canada, Belgium and Germany there are no national screening programmes although regional programmes do exist. There are, however, guidelines in the USA which recommend vision and alignment screening between the age of 3 and 3.5 years by a suitably trained individual (American Academy of Ophthalmology 2012). In Canada, there are national guidelines for screening visual acuity and ocular alignment at 3 to 5 years of age, but no established screening programme (Canadian Paediatric Society 2009). In many countries office-based paediatricians, ophthalmologists and optometrists offer annual 'child health' or 'eye health' checks, respectively, but these occur outside national programmes.

In low- and middle-income countries, little information on national screening programmes is available in the literature or online (World Health Organization 2014). One exception is Iran, where a national screening programme of 3 to 6 year olds performed by kindergarten teachers assessing visual acuity with illiterate 'E' Snellen charts has been in place since 1996 with an estimated up-

take of 67% of eligible children in 2005 (Khandekar 2009). In India there is no national screening (Jose 2009).

There are efforts to find cost-effective strategies for screening in developing countries such as a remote photoscreener system piloted in Brazil and China, and a home-based screening programme in China performed by parents (Donahue 2008; Lan 2012).

Rationale

Strabismus is a risk factor for the development of amblyopia. Whilst large deviations may be detected by family, friends or lay screeners, small deviations may go unnoticed, leading to suppression of visual information from the deviated eye. Childhood vision screening programmes use varying combinations of tests, depending on the age at which children are tested, and the type of professionals carrying out tests. Strabismus tests as part of a combination of tests may increase the precision of childhood amblyopia screening (VIP 2007). Published vision screening studies often lack specific information on strabismus detection, instead reporting overall precision in detecting amblyopia.

This review does not propose screening for strabismus that is of no aesthetic concern or visual consequence, but it is important to summarise current evidence on accuracy of tests in detecting strabismus in a screening setting, in order to enable healthcare commissioners to implement the most effective screening programme.

O B J E C T I V E S

To assess and compare the accuracy of tests, alone or in combination, for detection of strabismus in children aged 1 to 6 years, in a community setting by non-expert screeners or primary care professionals to inform healthcare commissioners setting up childhood screening programmes.

Secondary objectives

Other objectives were to investigate sources of heterogeneity of diagnostic accuracy, including:

- age;
- setting;
- type of professionals performing the test;
- study design;
- study size (< 100 vs. ≥ 100 participants, which may reflect the adoption of different sampling strategies);
- variation in the way a test is carried out;
- type of strabismus (convergent vs. divergent, horizontal/ vertical);
- severity of strabismus (amount of misalignment, constant/ intermittent/latent);

METHODS

Criteria for considering studies for this review

Types of studies

We have included all prospective or retrospective population-based test accuracy studies of consecutive participants. By 'population-based' we mean not only screening studies, implying sampling based on census, but also studies recruiting from community services such as schools or paediatric health districts. Hospital cohorts were excluded, unless the sampling from a community service was clearly described.

We have included studies that compare a single index test, or a combination of index tests, with a reference standard (cover test, performed as a standalone test or as part of a comprehensive eye examination). Case-control studies, in which children are selected based on their disease status, have been excluded unless they are nested in large prospective consecutive studies. Studies had to provide sufficient data to calculate diagnostic accuracy (sensitivity, specificity). We planned to include studies in which only a subgroup of participants had undergone the reference test; the result from these to be considered by subgroup analysis.

Participants

We included children aged 1 to 6 years old. Strabismus is a risk factor for the development of amblyopia during the 'sensitive period' of vision development. During this period, neural plasticity is greatest, and it begins to decline around the age of 6 years; clinical interventions are typically offered to children up to the age of 10 years. Screening programmes therefore attempt to identify children with amblyogenic risk factors before the age of 6 years, to allow remedial treatment. We set the lower limit of the age range at 1 year to avoid overlap with early postnatal eye screening programmes.

In countries where children start school in the academic year of their fifth birthday, screening programmes aim to capture children aged 4 to 5 years, i.e. during their first year at school. In other countries, the year of school entry can be earlier or later, and vision screening programmes may be carried out in the first year of school, or independent from schools. An age range of 1 to 6 years allows inclusion of all population-based studies in children at risk of developing amblyopia from strabismus.

When studies included children outside the range of 1 to 6 years, we tried to obtain subgroup data. If we did not obtain subgroup data we excluded these studies. We intended to include these studies if the proportion of children beyond age 6 is less than an agreed threshold, e.g. 20%, and we would have conducted sensitivity or subgroup analyses as appropriate.

We considered children attending population-based vision screening programmes. We included opportunistic screening programmes, such as including children attending schools. We excluded orthoptist-delivered programmes, as these include the reference standard.

Index tests

We included any test used by lay screeners to detect strabismus, either directly by identifying misalignment, or indirectly by identifying a consequence of strabismus such as loss of stereovision. The participant age range means that different tests may be used, as appropriate for the age of participants in each particular study. We described the index test by test type rather than enumerated.

- Type 1: tests which directly identify ocular misalignment:
  - corneal reflections tests (Hirschberg);
  - fundus reflections test (Brückner).
- Type 2: tests of binocular function: stereoacuity:
  - e.g. contour and random dot stereotests.
- Type 3: tests designed to detect reduced central vision/ visual acuity:
  - e.g. HOTV, LEA symbols, Keeler (previously Glasgow) crowded logMAR, Sonksen crowded logMAR, crowded Kay picture test, displayed as paper- or computer-based charts or conventional retroilluminated charts.
- Type 4: automated refraction devices designed to report ocular misalignment.

If the search had revealed several high-quality studies for each test type for inclusion in this review, we would have considered splitting the review by test type group.

Finally, we did not consider tests that require specialist skills, such as the 4-dioptre prism test, since we are concerned with population screening which is typically carried out by non-expert professionals, not by orthoptists, optometrists or ophthalmologists, who would directly use our reference standard, the cover‾ uncover test.

Target conditions

The target condition is constant or intermittent manifest strabismus of any magnitude and type (esotropia, microtropia, exotropia, hyper/hypotropia).

Reference standards

The reference standard considered in this review was the cover‾ uncover test, whether used alone or within a comprehensive ophthalmic examination, or in combination with other tests, by trained personnel.

We included studies that use the cover‾ uncover test, regardless of the type of professional performing the test. Type of professional (ophthalmologist, orthoptist, optometrist, trained technician, non-expert screener) will be noted and analysed as subgroups.

We included studies in which the cover- uncover or alternate cover test is used as part of a comprehensive eye examination, which often also includes visual acuity, biomicroscopy and refraction. For the latter scenario there is a risk of incorporation bias. This bias can be avoided by ensuring that the tests that are part of the reference standard 'comprehensive eye examination' do not belong to the same type of test as the index test included in that analysis. We excluded the whole study if a single test is assessed and there is incorporation bias, and we excluded part of the study data if a study comparing several index tests suffered from incorporation bias regarding a specific test.

Search methods for identification of studies

Electronic searches

The Cochrane Eyes and Vision Information Specialist searched the following electronic databases. There were no language or publication year restrictions.

- Cochrane Central Register of Controlled Trials (CENTRAL; 2016, Issue 12) (which contains the Cochrane Eyes and Vision Trials Register) in the Cochrane Library (searched 5 January 2017) (Appendix 3);
- Health Technology Assessment Database (HTAD) in the Cochrane Library (2016, Issue 4, searched 5 January 2017) (Appendix 4);
- MEDLINE Ovid (1946 to 5 January 2017) (Appendix 5);
- Embase Ovid (1980 to 5 January 2017) (Appendix 5);
- CINAHL (Cumulative Index to Nursing and Allied Health Literature) (1982 to 5 January 2017) (Appendix 6);
- Web of Science Conference Proceedings Citation Index-Science (CPCI-S) (January 1990 to 5 January 2017) (Appendix 7);
- BIOSIS (January 1969 to 5 January 2017) (Appendix 8);
- MEDION (searched 18 August 2014) (Appendix 9);
- Aggressive Research Intelligence Facility database (ARIF) (www.birmingham.ac.uk/research/activity/mds/projects/HaPS/ PHEB/ARIF/index.aspx: searched 5 January 2017) (Appendix 10);
- ISRCTN registry (www.isrctn.com/editAdvancedSearch: searched 5 January 2017) (Appendix 11);
- US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (www.clinicaltrials.gov: searched 5 January 2017) (Appendix 12);
- World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) (www.who.int/ictrp: searched 5 January 2017) (Appendix 9).

Searching other resources

We used the weblink pcwww.liv.ac.uk/˜rowef/index_files/ Page646.htm to search the following orthoptic journals and conference proceedings which are not electronically listed: British and Irish Orthoptic Journal, American Orthoptic Journal, Australian Orthoptic Journal, European Strabismus Association, International Strabismus Association and the International Orthoptic Congress. We contacted study authors for further clarification when required.

Data collection and analysis

Selection of studies

Two review authors (SH, VT) independently assessed the titles and abstracts for eligibility. We sorted the abstracts into 'definitely exclude' and 'possibly include' categories, recognising that sometimes it is not possible to judge from the abstract whether a reference fulfils the criteria or not. We placed all abstracts selected by at least one review author in the 'possibly include' category. We resolved disagreements at each step by discussion between the two review authors and a third senior author (AD-N).

Data extraction and management

We extracted the number of:

- true positives (TP), i.e. participants categorised as strabismic by both the reference and index test;
- false negatives (FN), i.e. participants categorised as strabismic by the reference test, but as non-strabismic by the index test;
- true negatives (TN), i.e. participants categorised as non-strabismic by both the reference and index tests;
- false positives (FP), i.e. participants categorised as non-strabismic by the reference test, but as strabismic by the index test;
- participants with uninterpretable index test results;
- missing data, i.e. participants included in the study, but not in the analyses, by causes of exclusion.

Uninterpretable test results at individual participant level were recorded in primary publications when a child did not comply with a test, i.e. refused to give an answer during assessment of visual acuity or stereovision, or did not fixate on targets for automated devices, or in case of ocular abnormalities affecting the clarity of cornea, lens or vitreous, or a combination of the three. For each study, we recorded how such cases were treated in the analyses.

Two review authors independently extracted the data to ensure consistency and entered these into Cochrane's statistical software, Review Manager 5 (RevMan 5) (Review Manager 2014). We have extracted the data shown in Table 1, which we have displayed in the Characteristics of included studies tables.

Assessment of methodological quality

We used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool to evaluate the risk of bias and applicability of primary studies (www.bris.ac.uk/quadas/quadas-2). QUADAS-2 consists of four key domains: patient selection, index test, reference standard, and flow and timing. The tool is completed in four phases.

1. The review question is stated.
2. Development of review-specific guidance.
3. Review of the published flow diagram for the primary study or construction of a flow diagram if none is reported.
4. Judgement of bias and applicability.

Each domain is assessed in terms of the risk of bias and the first three are also assessed in terms of concerns regarding applicability. To help reach a judgement on the risk of bias, signalling questions are included. These flag aspects of study design related to the potential for bias and aim to help review authors make risk of bias judgements. Two review authors independently assessed the methodological quality of the included studies. A third senior author resolved disagreements on study quality. Table 2 shows the guidance the review authors used when judging the methodological quality of studies.

We scored the risk of bias signalling questions as 'yes/no/unclear' as detailed in Table 2. Risk of bias was judged as 'low', 'high' or 'unclear'. When we answered 'yes' for all signalling questions for a domain then we could judge the risk of bias 'low'. If we answered any question as 'no', this flagged the potential for bias. When this occurred, we followed the guidelines developed in phase 2 of the quality assessment process to judge the risk of bias. We used the 'unclear' category only when insufficient data were reported to permit a judgment.

We judged applicability of primary studies to the review question in a similar manner.

We also recorded study sponsorship.

Statistical analysis and data synthesis

We used two-by-two data of index and reference test results to calculate the sensitivities and specificities, with their 95% confidence intervals. We used the RevMan 5 software for descriptive analyses, and plotted individual studies in forest plots.

Considering test threshold across different test types was the most important analytic issue in this review. We planned to use a continuous output measure for most tests: ocular misalignment as prism dioptres (PD) or degrees for test type 1; stereoscopic acuity as seconds of arc for test type 2; visual acuity in logMAR for test type 3 (acknowledging that comparison of values may be hampered by use of charts with different optotype size steps, and that simple mathematical conversion from Snellen to logMAR may be inaccurate); and millimetres or a ratio for test type 4. Other tests listed in Appendix 2 and used in the diagnosis of, but not in screening for, strabismus are not based on an explicit common measure.

However, in practice the heterogeneous execution and technical characteristics of the tests made it difficult to consider using an explicit threshold in statistical analyses, and implicit threshold effects are more likely.

Analyses within each test type

We intended to analyse different tests within each test type group, using the following strategy. For each study, we intended to extract data at specific thresholds if available. We attempted to extract cut-offs of 8 PD for horizontal and 1 PD for vertical deviations in test type 1; 400 arc seconds for test type 2; and visual acuity 0.2 logMAR for test type 3. UK screening recommendations specify "less than 0.2 logMAR" as referral threshold (UK National Screening Committee); guidelines from the AAPOS specify that optotype-based screening (which covers test type 3) should detect visual acuity of less than 0.176 logMAR (Snellen 20/30) at all ages. Threshold values for test type 4 have not been published; we therefore used "any asymmetry, in millimetres or as ratio" as threshold. Thresholds are summarised in Table 3

Investigations of heterogeneity

The framework for likely sources of heterogeneity was described previously and mainly includes setting and study population, particularly regarding referral method and inclusion criteria; type of professional executing the reference standard; and study quality assessment.

We planned to investigate heterogeneity in the first instance through visual examination of forest plots of sensitivities and specificities and through visual examination of the receiver operating characteristic (ROC) plot of the raw data. However, we had insufficient data to investigate these secondary objectives.

Sensitivity analyses

Where appropriate (i.e. if not already explored in our analyses of heterogeneity) and if sufficient data were available, we planned to explore the sensitivity of any summary accuracy estimates to aspects of study quality such as nature of masking and type of reference standard, guided by the anchoring statements developed in our QUADAS-2 exercise.

Assessment of reporting bias

We did not assess publication bias since there is no standard method to achieve this in diagnostic test accuracy reviews (Deeks 2005). For selective outcome reporting issues, such as the use of a specific cut-off of ocular misalignment, we did not search for a protocol to assess within-study reporting bias, since protocols of diagnostic accuracy studies are not routinely reported.

RESULTS

Results of the search

The searches yielded a total of 2327 records (Figure 1). After de-duplication we screened 1236 studies/papers which, following independent screening by two authors (SH, VT), led to the exclusion of 1129 studies not meeting the inclusion criteria (including age range, examiner type and primary use of cover test), lack of relevance or lack of results. The remaining 107 studies underwent full text review, with disagreements resolved by a third author (AD)

(Figure 1). The authors of six posters were contacted to ascertain relevant publications and data for those posters; three replied with one publication identified that did not meet the inclusion criteria as no strabismus outcomes were reported (Shallo-Hoffman 2004). In addition authors for three published studies were contacted where full analysis of the results required additional data (Enzenauer 2000; Robinson 1999; Tung 2006); two of these authors responded but further data were unavailable (summarised in the Characteristics of excluded studies table). One study met the inclusion criteria and had sufficient data for analysis (Arthur 2009).

Figure 1. Study flow diagram.
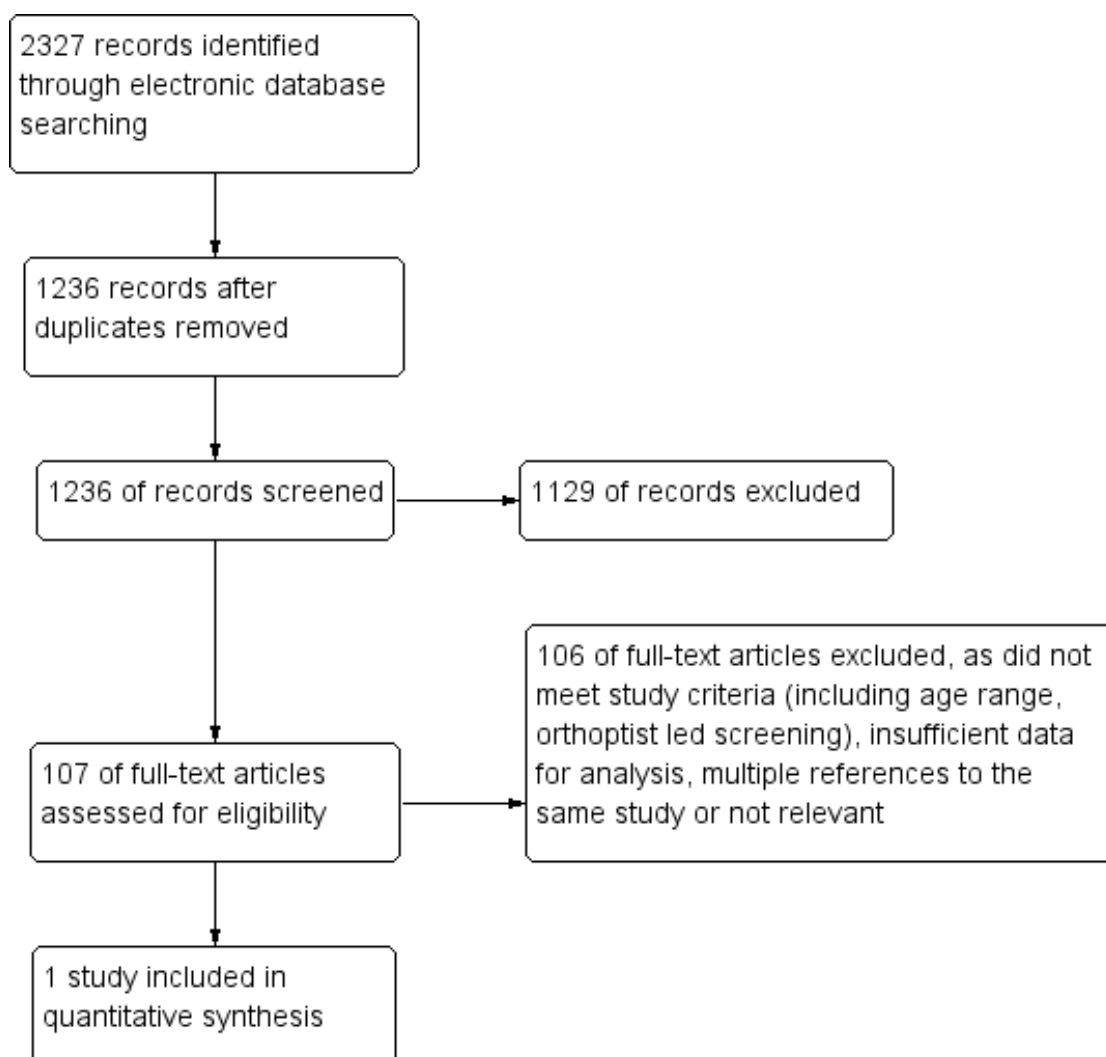
Methodological quality of included studies

Arthur 2009 was a prospective study performed in a community setting with all eligible children invited for screening and all screened participants offered a gold standard examination (Summary of findings 1). Eligible children for the study were all junior kindergarten students in a specific school district of Ontario, Canada; and 98% of those enrolled were 4 or 5 years of age. The screening was conducted by certified dental assistants conjointly with an existing dental screening programme. The dental assistants underwent training on the plusoptiX S04 photoscreener (Plusoptix GmbH) with defined criteria for failing the test of a corneal reflex more than 10 degrees from the centre. Bias assessment indicated an unclear risk of bias for the patient selection domain but a low risk of bias for all other QUADAS-2 domains (Figure 2; Figure 3). This was due to a relatively low uptake of screening at 25% with included children volunteering and not sampled randomly or consecutively. There was no available data on the prevalence of strabismus in non-responders compared to responders. There were no adverse outcomes reported.

Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study
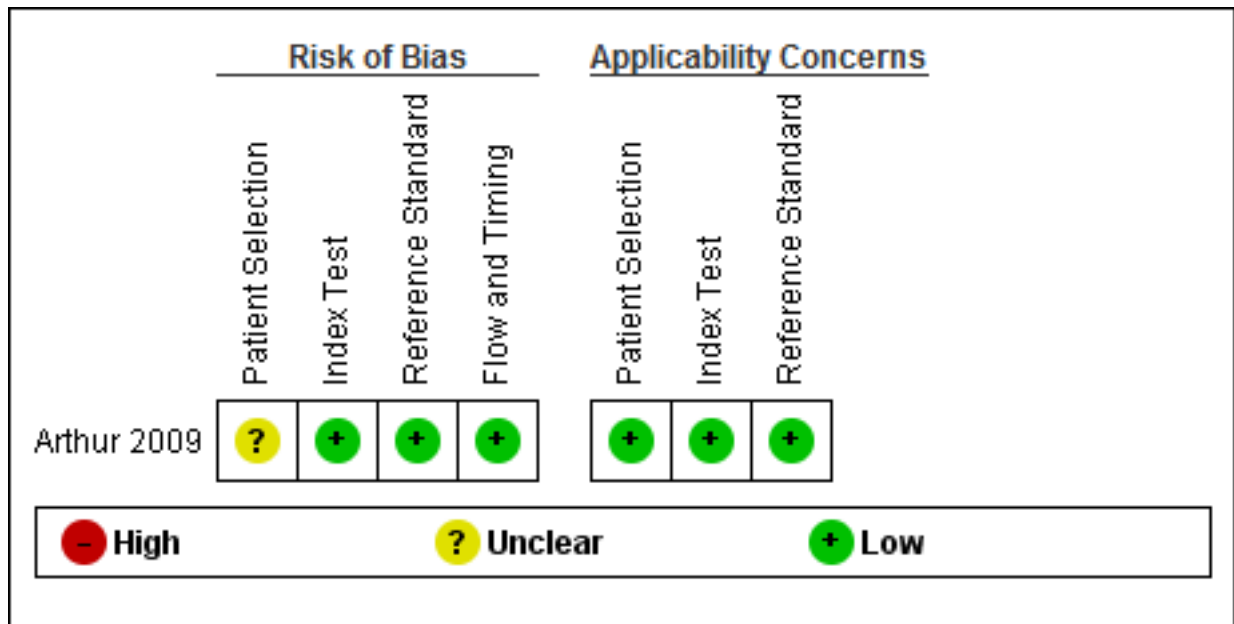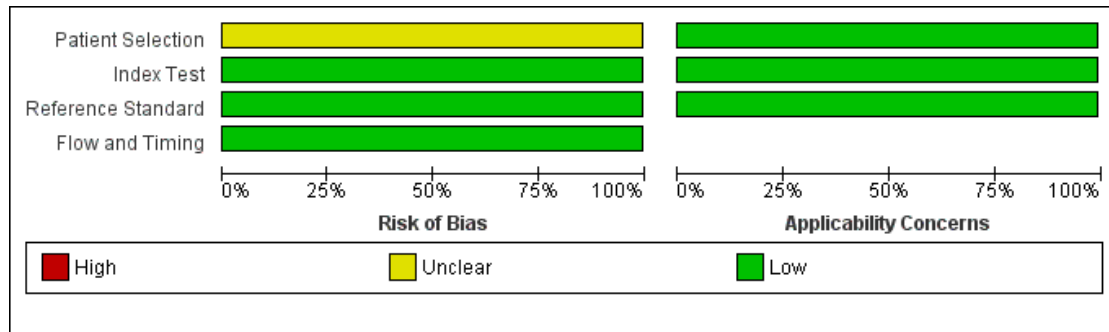
Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies

Findings

Three hundred and six children were screened by the photoscreener. Two hundred and seventy-one had both interpretable screening photographs and completed the gold standard examination, the others having declined (n = 14), being unable to attend within the study timeframe (n = 11), become uncontactable (n = 6), having had uninterpretable photographs (n = 3) or incomplete examination (n = 1). The photoscreener was used to ascertain refractive error, anisocoria and ocular misalignment with 14 children referred specifically for ocular misalignment. A total of 13 children were identified to have strabismus on gold standard examination of which six had been referred for ocular misalignment, two had been referred for refractive error and five had passed the screening test. The two participants referred for refractive error not ocular misalignment and found to have strabismus were considered as false negatives for calculating accuracy. The main outcomes were a sensitivity of 0.46 (95% confidence interval (CI) 0.19 to 0.75), and a specificity of 0.97 (CI 0.94 to 0.99) (Figure 4). The estimated prevalence of strabismus in the screened population was 4.8%. The types of strabismus identified were intermittent exotropia (n = 3 well-controlled, n = 2 poorly-controlled), esotropia (n = 4), hypertropia (n = 2) and exotropia (n = 2).

Figure 4. Forest plot of 1 Photoscreener.

Summary of findings

Accuracy of a photoscreener to detect strabismus in the community						
Patient/population: children aged 1 to 6 years old **Setting:** school **Index test:** plusoptix S04 photoscreener **Target condition:** constant and intermittent manifest strabismus **Reference standard:** cover test at distance and near						
Number of studies	Number of participants	Number affected by target condition	Sensitivity of test (95% CI)	Specificity of test (95% CI)	Risk of bias based on QUADAS-2 domains	Comments
1 Arthur 2009	271	13	0.46 (0.19 to 0.75)	0.97 (0.94 to 0.99)	Unclear risk	Low participation rate of 25%

CI: confidence intervals

Study ID	Arthur 2009
Clinical features and settings	Previous testing and results: unknown. Setting: elementary school. Referral route/selection: all who were screened offered gold standard examination
Participants	Sample size: 306 screened (1343 invited to study (consents sent: this may have introduced selection bias from concerned parents being more likely to return consent forms), 387 returned, 45 excluded as consents too late, 7 excluded for document errors, 28 absent on day of screening, 1 uncooperative), 275 gold standard exam (14 declined, 11 unable to attend within time frame, 6 uncontactable) of which 271 data interpretable for both index and reference (3 photographs unusable, 1 did not complete exam) Socio-demographic items: 98% 4 to 5 years of age, gender and ethnicity not given, no ocular abnormalities (i.e. media opacities, which would affect test results/technical failure rates). Geographic region: Limestone school district, Ontario, Canada
Study design	Selection: all patients with data available on both index and reference tests as single group. Enrolment: consecutive series, enrolled by post in combination with dental screening programme. Identification: prospective. If more than one test: one test.
Target condition	Constant and intermittent manifest strabismus (esotropia, exotropia, vertical tropia, microtropia), prevalence of the target condition in the sample: 13 (of 271)
Reference standard	Test definition and description: monocular visual acuity with occlusion glasses (crowded Keeler logMAR letter matching test/Crowded Kay pictures/Cardiff cards), cover test at distance and near, ocular movements and convergence, binocular single vision assessment (20D base-out prism test and/or stereopsis) and red reflex test Standards: discharged if VA 0.2 logMAR or better, binocular single vision at distance and near and no suspected ocular pathology, 6- to 12-week review and re-check if borderline, cycloplegic refraction/dilated examination all others. Test operator(s): optometrist or orthoptist or ophthalmologist. Timing of reference standard: separate visit to hospital but timing unknown
Index tests	plusoptiX S04 photoscreener, co-axial camera, handheld at 1 m. Criteria for positive test result: eye alignment > 10 degrees from centre (manually flagged as abnormal) anisometropia > 1D, astigmatism > 1.25D, myopia > 3D, hyperopia > 3.5D, anisocoria > 1 mm. Details of test operators: certified dental assistants after 3 hours of training. Timing: 5 to 10 seconds image acquisition time repeated if necessary. Manufacturer: Plusoptix GmbH. Technical characteristics: 3rd generation, infrared, coaxial video camera, portable, handheld, non-contact
Follow-up	How many participants were lost to follow-up: 31. How many have missing or uninterpretable test results: 4. Adverse events noted that could be caused by the test: 0.

| Notes | Sources of funding: none declared. |
| | Anything else of relevance: low participation rate (25%). |

DISCUSSION

Screening for strabismus in children in the community may be achieved by tests that directly ascertain misalignment of the eyes (corneal or fundus reflections) or indirectly detect associated reduced vision or stereopsis. Small deviations may not be noticed by family but may have significant impact on visual development hence the rationale for screening.

Summary of main results

There is limited available data on strabismus screening in the community as performed by lay examiners with the majority of published screening studies predominantly focusing on amblyopia screening. One study was identified that met the full inclusion criteria for this review in which all children screened with a photo-screener were offered a gold standard examination (Arthur 2009). There was an unclear risk of bias. The results indicated high specificity but low sensitivity implying the potential for significant false negatives. Absolute numbers found to have strabismus by gold standard examination were small at 13 in total out of 271 children.

Strengths and weaknesses of the review

Only one study was analysed in this review, prohibiting any conclusion on the accuracy of screening tests. It remains unclear whether other screening modalities would have significant accuracy for screening in this context.

Applicability of findings to the review question

The findings have limited applicability to the review question in which the assessment and comparison of the accuracy of multiple tests in screening for strabismus was to be ascertained. The single study included suggests that the plusoptiX S04 photoscreener for detecting ocular misalignment could provide a specific but not sensitive test, and this single study is not sufficient for robust conclusion (95% CI about 20% to 75%). Further studies are needed. Due to the lack of relevant studies, the secondary objectives to investigate sources of heterogeneity of diagnostic accuracy could not be assessed.

Through review of the literature, other studies were identified with relevant results that did not meet the inclusion criteria for this review and as such could not be included. This included the Vision in Preschoolers (VIP) study, a large multicentre trial performed in two phases to ascertain the accuracy of various screening tools for children aged 3 to 4 years old (VIP 2007). In phase I, trained eye care professionals performed the screening assessment but in phase II, trained nurses and lay screeners performed the screening tests. The population screened were enriched from a preceding generalised screening programme, with all those who failed screening included in the VIP study as well as a proportion of those who did

not. The aim was to enrich for ocular pathology within the study to better ascertain accuracy of screening methods. As such, this study could not be included in this review but still has relevant conclusions.

VIP 2007 specifically assessed methods for screening for strabismus; and for the lay and nurse screeners it included four tests; Retinomax autorefraction, SureSight Vision Screener autorefractor, LEA symbols visual acuity testing and Stereo Smile Test II stereoacuity testing. Of 4040 children screened, 157 (3.9%) were found to have strabismus. For lay screeners the combination of both the Stereo Smile test and the SureSight autorefractor and the Stereo Smile test and the Retinomax autorefractor were associated with a statistically significant increase in sensitivity of strabismus detection for a 90% specificity but no such increase was observed for other test combinations or for the nurse screeners. The study concluded that the addition of tests for eye alignment to acuity or refraction tests alone would depend on a screening programme's goals and resources. It also indicates that tests of visual acuity alone would be insufficient for identifying all cases of strabismus.

A large prospective, consecutively enrolled study of all 3 to 6 year olds in an eastern province of Taiwan used two different index tests for screening with all children offered a gold standard examination by a single ophthalmologist (Tung 2006). Screening for strabismus performed in 2003 on 2868 children was conducted by trained kindergarten teachers using both a National Taiwan University (NTU) random dot stereogram to detect stereopsis less than 300 seconds of arc and Hirschberg corneal reflexes at 1 metre, with any displacement of the light reflexes considered abnormal. The number screened and then unavailable for the gold standard assessment was not disclosed. Detailed outcome numbers were not provided and as such this study could not be included in this review. However, the overall sensitivity and specificity for the NTU random dot stereogram were 38.9% and 90.4% respectively and for the Hirschberg light reflex were 75% and 98.9% respectively suggesting good efficacy for the Hirschberg light reflexes as a screening modality.

In summary the applicability of available studies to primary screening programmes is limited. Future screening studies should also consider the optimum screening age, which for optotype-based tests is around 4 to 5 years (Solebo 2015). Lastly, we would recommend further research into long-term visual and psychosocial outcomes of childhood strabismus, to explore the benefits of early detection.

AUTHORS' CONCLUSIONS

Implications for practice

Identifying strabismus as part of a screening programme is most important if it impacts on visual acuity (leading to amblyopia) or stereopsis. Therefore screening in the community does not need to

directly test for strabismus by ocular misalignment although there is the suggestion from other studies that sensitivity is increased by doing so. There is a lack of evidence for which tests are most accurate in detecting strabismus specifically in a normal population being screened by non-expert screeners.

Implications for research

Cochrane Reviews of the accuracy of screening tests to detect anisometropia and amblyopia would complement the evidence review on screening strategies. Given the prevalence of amblyopia and amblyogenic risk factors, primary vision and strabismus screening studies would require large numbers of children to be screened. Such studies may be cost-effective if run alongside existing vision screening programmes. As visual acuity alone may not be sufficiently sensitive to detect strabismus, addition of autorefractor,

stereoacuity, corneal light reflex testing or novel devices should be considered (VIP 2007). Although sensitivity of screening tests should be around 80% specificity may not need to be, as the further assessment for amblyopia is non-invasive and does not carry a risk of harm.

REFERENCES

References to studies included in this review

Arthur 2009 *{published data only}*

Arthur BW, Riyaz R, Rodriguez S, Wong J. Field testing of the plusoptiX S04 photoscreener. *Journal of AAPOS* 2009; **13**(1):51–7.

References to studies excluded from this review

Enzenauer 2000 *{published and unpublished data}*

Enzenauer RW, Freeman HL, Larson MR, Williams TL. Photoscreening for amblyogenic factors by public health personnel: the Eyecor Camera System. *Ophthalmic Epidemiology* 2000;**7**(1):1–12.

Robinson 1999 *{published and unpublished data}*

Robinson B, Bobier WR, Martin E, Bryant L. Measurement of the validity of a preschool vision screening program. *American Journal of Public Health* 1999;**89**(2):193–8.

Shallo-Hoffman 2004 *{published data only}*

Shallo-Hoffmann J, Coulter RA, Oliver P, Hardigan P, Blavo C. A study of pre-school screening tests' testability, validity and duration: do group differences matter?. *Strabismus* 2004;**12**(2):115–23.

Tung 2006 *{published and unpublished data}*

Tung IC, Tsai RK, Chang CH, Sheu MM. Comparison of trained kindergarten teachers and public health nurses in the administration of preschool amblyopia and strabismus screening tests. *Tzu Chi Medical Journal* 2006;**18**(1):29–33.

VIP 2007 *{published data only}*

Vision in Preschoolers Study Group. Does assessing eye alignment along with refractive error or visual acuity increase sensitivity for detection of strabismus in preschool vision screening?. *Investigative Ophthalmology & Visual Science* 2007;**48**(7):3115–25.

Additional references

Adams 2005

Adams WE, Hrisos S, Richardson S, Davis H, Frisby JP, Clarke MP. Frisby Davis distance stereoacuity values in visually normal children. *British Journal of Ophthalmology* 2005;**89**(11):1438–41.

Almeida 2012

Almeida JD, Silva AC, Paiva AC, Teixeira JA. Computational methodology for automatic detection of strabismus in digital images through Hirschberg test. *Computers in Biology and Medicine* 2012;**42**(1):135–46.

American Academy of Ophthalmology 2012

American Academy of Ophthalmology. Preferred practice pattern - Amblyopia. www.aao.org/ppp (accessed 30 April 2013).

Amitava 2012

Amitava AK, Kewlani D, Khan Z, Razzak A. Assessment of a modification of Brückner's test as a screening modality for anisometropia and strabismus. *Oman Journal of Ophthalmology* 2012;**3**(3):131–5.

Anonymous 2004

Vision in Preschoolers Study Group. Preschool visual acuity screening with HOTV and Lea symbols: testability and between-test agreement. *Optometry and Vision Science* 2004; **81**(9):678–83.

Arnold 2000

Arnold RW, Gionet EG, Jastrzebski AI, Kovtoun TA, Machida CJ, Armitage MD, et al. The Alaska Blind Child Discovery project: rationale, methods and results of 4000 screenings. *Alaska Medicine* 2000;**42**(3):58–72.

Arnold 2013

Arnold RW, Arnold AW, Armitage MD, Shen JM, Hepler TE, Woodard TL. Pediatric Photoscreeners in High Risk

Patients 2012: A Comparison Study of Plusoptix, iScreen and SPOT. *Binocular Vision and Strabology Quarterly, Simms-Romano's* 2013;**28**(1):20–8.

Barry 1997
Barry JC, Backes A. Limbus versus pupil center for ocular alignment measurement with corneal reflexes. *Investigative Ophthalmology & Visual Science* 1997;**38**(12):2597–607.

Basmak 2007
Basmak H, Sahin A, Yildirim N, Saricicek T, Yurdakul S. The angle kappa in strabismic individuals. *Strabismus* 2007;**15**(4):193–6.

Bernfeld 1982
Bernfeld A. Psychological repercussions of strabismus in children. *Journal Francais d'Ophtalmologie* 1982;**5**(8-9):523–30.

Braddick 1980
Braddick O, Atkinson J, Julesz B, Kropfl W, Bodis-Wollner I, Raab E. Cortical binocularity in infants. *Nature* 1980;**288**(5789):363–5.

Broadbent 1990
Broadbent H, Westall C. An evaluation of techniques for measuring stereopsis in infants and young children. *Ophthalmic and Physiological Optics* 1990;**10**(1):3–7.

Brodie 1987
Brodie SE. Photographic calibration of the Hirschberg test. *Investigative Ophthalmology & Visual Science* 1987;**28**(4):736–42.

Brückner 1965
Brückner R. Practical use of the illumination test in the early diagnosis of strabismus. *Ophthalmologica* 1965;**149**(6):497–503.

Canadian Paediatric Society 2009
Amit M, Canadian Paediatric Society. Vision screening in infants, children and youth. *Paediatric Child Health* 2009;**14**(4):246–51.

Carrera 1993
Carrera A, Saornil MA, Zamora MI, Maderuelo A, Canamares S, Pastor JC. Detecting amblyogenic diseases with the photographic Bruckner. *Strabismus* 1993;**1**(1):3–9.

Carter 1978
Carter AJ, Roth N. Axial length and the Hirschberg test. *American Journal of Optometry and Physiological Optics* 1978;**55**(6):361–4.

Chai 2009
Chai Y, Shao Y, Lin S, Xiong KY, Chen WS, Li YY, et al. Vision-related quality of life and emotional impact in children with strabismus: a prospective study. *Journal of International Medical Research* 2009;**37**(4):1108–14.

Choi 1998
Choi RY, Kushner BJ. The accuracy of experienced strabismologists using the Hirschberg and Krimsky tests. *Ophthalmology* 1998;**105**(7):1301–6.

Cibis 1994
Cibis GW. Video vision development assessment (VVDA): combining the Bruckner test with eccentric photorefraction for dynamic identification of amblyogenic factors in infants and children. *Transactions of the American Ophthalmological Society* 1994;**92**:643–85.

Ciuffreda 1991
Ciuffreda KJ, Levi DM, Selenow A. *Amblyopia: Basic and Clinical Aspects.* Boston: Butterworth Heinemann, 1991.

Dahlmann-Noor 2009a
Dahlmann-Noor AH, Vrotsou K, Kostakis V, Brown J, Heath J, Iron A, et al. Vision screening in children by Plusoptix Vision Screener compared with gold-standard orthoptic assessment. *British Journal of Ophthalmology* 2009;**93**(3):342–5.

Dahlmann-Noor 2009b
Dahlmann-Noor AH, Comyn O, Kostakis V, Misra A, Gupta N, Heath J, et al. Plusoptix Vision Screener: the accuracy and repeatability of refractive measurements using a new autorefractor. *British Journal of Ophthalmology* 2009;**93**(3):346–9.

daSilva 1991
daSilva OA, Henriques J, Pinto F, Neves C. Visual screening in children. *Acta Medica Portuguesa* 1991;**4**(4):183–7.

Deeks 2005
Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**(9):882–93.

DeRespinis 1989
DeRespinis PA, Naidu E, Brodie SE. Calibration of Hirschberg test photographs under clinical conditions. *Ophthalmology* 1989;**96**(7):944–9.

Dobson 1978
Dobson V, Teller DY, Lee CP, Wade B. A behavioral method for efficient screening of visual acuity in young infants. I. Preliminary laboratory development. *Investigative Ophthalmology & Visual Science* 1978;**17**(12):1142–50.

Donahue 2008
Donahue SP, Lorenz S, Johnson T. Photo screening around the world: Lions Club International Foundation experience. *Seminars in Ophthalmology* 2008;**23**(5):294–7.

Donahue 2013
Donahue SP, Arthur B, Neely DE, Arnold RW, Silbert D, Ruben JB, et al. Guidelines for automated preschool vision screening: a 10-year, evidence-based update. *Journal of AAPOS* 2013;**17**(1):4–8.

Emsley 1948
Emsley HH. *Visual Optics.* London: Hatton Press, 1948.

Eskridge 1988
Eskridge JB, Wick B, Perrigin D. The Hirschberg test: a double-masked clinical evaluation. *American Journal of Optometry and Physiological Optics* 1988;**65**(9):745–50.

Fawcett 2005

Fawcett SL, Wang YZ, Birch EE. The critical period for susceptibility of human stereopsis. *Investigative Ophthalmology & Visual Science* 2005;**46**(2):521–5.

Fogt 2000

Fogt N, Baughman BJ, Good G. The effect of experience on the detection of small eye movements. *Optometry and Vision Science* 2000;**77**(12):670–4.

Fox 1980

Fox R, Aslin RN, Shea SL, Dumais ST. Stereopsis in human infants. *Science* 1980;**207**(4428):323–4.

Friedman 2009

Friedman DS, Repka MX, Katz J, Giordano L, Ibironke J, Hawse P, et al. Prevalence of amblyopia and strabismus in white and African American children aged 6 through 71 months the Baltimore Pediatric Eye Disease Study. *Ophthalmology* 2009;**116**(11):2128–34.

Fu 2006

Fu VL, Birch EE, Holmes JM. Assessment of a new Distance Randot stereoacuity test. *Journal of AAPOS* 2006;**10**(5): 419–23.

Gamble 1950

Gamble JD. Identifying deviations by the cover test. *Optical Journal and Review of Optometry* 1950;**87**(16):31.

Graf 2012

Graf M, Alhammouri Q, Vieregge C, Lorenz, B. The Bruckner transillumination test: limited detection of small-angle esotropia. *Ophthalmology* 2012;**118**(12):2504–9.

Graham 1974

Graham PA. Epidemiology of strabismus. *British Journal of Ophthalmology* 1974;**58**(3):224–31.

Grant 2007

Grant S, Melmoth DR, Morgan MJ, Finlay AL. Prehension deficits in amblyopia. *Investigative Ophthalmology & Visual Science* 2007;**48**(3):1139–48.

Griffin 1986

Griffin JR, Cotter SA. The Bruckner test: evaluation of clinical usefulness. *American Journal of Optometry and Physiological Optics* 1986;**63**(12):957–61.

Griffin 1989

Griffin JR, McLin LN, Schor CM. Photographic method for Bruckner and Hirschberg testing. *Optometry and Vision Science* 1989;**66**(7):474–9.

Hall 2003

Hall DMB, Elliman D. *Health For All Children*. Oxford: OUP, 2003.

Hasebe 1995

Hasebe S, Ohtsuki H, Tadokoro Y, Okano M, Furuse T. The reliability of a video-enhanced Hirschberg test under clinical conditions. *Investigative Ophthalmology & Visual Science* 1995;**36**(13):2678–85.

Hasebe 1998

Hasebe S, Ohtsuki H, Kono R, Nakahira Y. Biometric confirmation of the Hirschberg ratio in strabismic children. *Investigative Ophthalmology & Visual Science* 1998;**39**(13): 2782–5.

Hatt 2008

Hatt SR, Mohney BG, Leske DA, Holmes JM. Variability of stereoacuity in intermittent exotropia. *American Journal of Ophthalmology* 2008;**145**(3):556–61.

Hered 1997

Hered RW, Murphy S, Clancy M. Comparison of the HOTV and Lea Symbols charts for preschool vision screening. *Journal of Pediatric Ophthalmology and Strabismus* 1997;**34**(1):24–8.

Hirschberg 1881

Hirschberg J. The quantitative analysis of diplopic strabismus. *British Medical Journal* 1881;**1**(1044):5–9.

Holmes 2005

Holmes JM, Fawcett SL. Testing distance stereoacuity with the Frisby-Davis 2 (FD2) test. *American Journal of Ophthalmology* 2005;**139**(1):193–5.

Howland 1974

Howland HC, Howland B. Photorefraction: a technique for study of refractive state at a distance. *Journal of the Optical Society of America* 1974;**64**(2):240–9.

Howland 2009

Howland HC. Photorefraction of eyes: history and future prospects. *Optometry and Vision Science* 2009;**86**(6):603–6.

Hrisos 2006

Hrisos S, Clarke MP, Kelly T, Henderson J, Wright CM. Unilateral visual impairment and neurodevelopmental performance in preschool children. *British Journal of Ophthalmology* 2006;**90**(7):836–8.

Huynh 2005

Huynh SC, Ojaimi E, Robaei D, Rose K, Mitchell P. Accuracy of the Lang II stereotest in screening for binocular disorders in 6-year-old children. *American Journal of Ophthalmology* 2005;**140**(6):1130–2.

Jose 2009

Jose R, Sachdeva S. School eye screening and the National Program for Control of Blindness. *Indian Pediatrics* 2009; **46**(3):205–8.

Kaakinen 1979

Kaakinen K. A simple method for screening of children with strabismus, anisometropia or ametropia by simultaneous photography of the corneal and the fundus reflexes. *Acta Ophthalmologica* 1979;**57**(2):161–71.

Kay 1983

Kay H. New method of assessing visual acuity with pictures. *British Journal of Ophthalmology* 1983;**67**(2):131–3.

Kaye 2005

Kaye SB. Testing distance stereoacuity with the Frisby-Davis 2 (FD2) test. *American Journal of Ophthalmology* 2005;**140** (2):346–7.

Khandekar 2009

Khandekar R, Parast N, Arabi A. Evaluation of 'vision screening' program for three to six-year-old children in the

Republic of Iran. *Indian Journal of Ophthalmology* 2009;**57**(6):437–42.

Koklanis 2006

Koklanis K, Abel LA, Aroni R. Psychosocial impact of amblyopia and its treatment: a multidisciplinary study. *Clinical and Experimental Ophthalmology* 2006;**34**(8): 743–50.

Kothari 2007

Kothari MT. Can the Bruckner test be used as a rapid screening test to detect significant refractive errors in children?. *Indian Journal of Ophthalmology* 2007;**55**(3): 213–5.

Krimsky 1951

Krimsky E. Method for objective investigation of strabismus. *Journal of the American Medical Association* 1951;**145**(8): 539–44.

Lan 2012

Lan W, Zhao F, Li Z, Zeng J, Liu W, Lu J, et al. Validation and cost-effectiveness of a home-based screening system for amblyopia. *Ophthalmology* 2012;**119**(6):1265–71.

Lang 1983

Lang J. Microtropia. *International Ophthalmology* 1983;**6**(1):33–6.

LeGrand 1980

LeGrand Y, ElHage SG. *Physiological Optics*. Berlin Heidelberg: Springer-Verlag, 1980.

Matsuo 2007a

Matsuo T, Matsuo C. Comparison of prevalence rates of strabismus and amblyopia in Japanese elementary school children between the years 2003 and 2005. *Acta Medica Okayama* 2007;**61**(6):329–34.

Matsuo 2007b

Matsuo T, Matsuo C, Matsuoka H, Kio K. Detection of strabismus and amblyopia in 1.5- and 3-year-old children by a preschool vision-screening program in japan. *Acta Medica Okayama* 2007;**61**(1):9–16.

McCormick 2002

McCormick A, Bhola R, Brown L, Squirrel D, Giles J, Pepper I. Quantifying relative afferent pupillary defects using a Sbisa bar. *British Journal of Ophthalmology* 2002;**86**(9):985–7.

McKean 1976

McKean HE, Wirtschafter JD, Marx D. Bias of the cover test in the diagnosis of alternating tropia. *Annals of Ophthalmology* 1976;**8**(4):435–7.

Miller 1993

Miller JM, Mellinger M, Greivenkemp J, Simons K. Videographic Hirschberg measurement of simulated strabismic deviations. *Investigative Ophthalmology & Visual Science* 1993;**34**(11):3220–9.

Miller 1995

Miller JM, Hall HL, Greivenkamp JE, Guyton DL. Quantification of the Bruckner test for strabismus. *Investigative Ophthalmology & Visual Science* 1995;**36**(5): 897–905.

Model 2012

Model D, Eizenman M. An automated Hirschberg test for infants. *IEEE Transactions on Bio-Medical Engineering* 2012; **58**(1):103–9.

Moghaddam 2012

Moghaddam AA, Kargozar A, Zarei-Ghanavati M, Najjaran M, Nozari V, Shakeri MT. Screening for amblyopia risk factors in pre-verbal children using the Plusoptix photoscreener: a cross-sectional population-based study. *British Journal of Ophthalmology* 2012;**96**(1):83–6.

Nuzzi 1986

Nuzzi G. Binocular Polaroid test. *Journal of Pediatric Ophthalmology and Strabismus* 1986;**23**(1):31–3.

O'Connor 2010

O'Connor AR, Birch EE, Anderson S, Draper H. The functional significance of stereopsis. *Investigative Ophthalmology & Visual Science* 2010;**51**(4):2019–23.

Ohlsson 2002

Ohlsson J, Villarreal G, Sjostrom A, Abrahamsson M, Sjostrand J. Screening for amblyopia and strabismus with the Lang II stereo card. *Acta Ophthalmologica Scandinavica* 2002;**80**(2):163–6.

Pai 2012

Pai AS, Rose KA, Samarawickrama C, Fotedar R, Burlutsky G, Varma R, et al. Testability of refraction, stereopsis, and other ocular measures in preschool children: the Sydney Paediatric Eye Disease Study. *Journal of AAPOS* 2012;**16**(2):185–92.

Paysse 2001

Paysse EA, Williams GC, Coats DK, Williams EA. Detection of red reflex asymmetry by pediatric residents using the Bruckner reflex versus the MTI photoscreener. *Pediatrics* 2001;**108**(4):E74.

Petrig 1981

Petrig B, Julesz B, Kropfl W, Baumgartner G, Anliker M. Development of stereopsis and cortical binocularity in human infants: electrophysiological evidence. *Science* 1981; **213**(4514):1402–5.

Pott 1998

Pott JW, Oosterveen DK, Van Hof-van Duin J. Screening for suppression in young children: the polaroid suppression test. *Journal of Pediatric Ophthalmology and Strabismus* 1998;**35**(4):216–22.

Pott 2003

Pott JW, Kingma C, Verhoeff K, Grootendorst RJ, de Faber JT. The polaroid suppression test in a pediatric population with ophthalmologic disorders. *Journal of AAPOS* 2003;**7**(2):137–41.

Prakash 1996

Prakash P, Sharma P, Rao VM, Shastry P, Menon V. Polaroid scotometer: a new device to chart suppression scotomata. *Journal of Pediatric Ophthalmology and Strabismus* 1996;**33**(3):181–4.

Preslan 1996

Preslan MW, Novak A. Baltimore Vision Screening Project. *Ophthalmology* 1996;**103**(1):105–9.

Rahi 2002

Rahi J, Logan S, Timms C, Russell-Eggitt I, Taylor D. Risk, causes, and outcomes of visual impairment after loss of vision in the non-amblyopic eye: a population-based study. *Lancet* 2002;**360**(9333):597–602.

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riddell 1994

Riddell PM, Hainline L, Abramov I. Calibration of the Hirschberg test in human infants. *Investigative Ophthalmology & Visual Science* 1994;**35**(2):538–43.

Rogers 1982

Rogers GL, Chazan S, Fellows R, Tsou BH. Strabismus surgery and its effect upon infant development in congenital esotropia. *Ophthalmology* 1982;**89**(5):479–83.

Romano 1971

Romano PE, von Noorden GK. Limitations of cover test in detecting strabismus. *American Journal of Ophthalmology* 1971;**72**(1):10–2.

Romano 2006

Romano PE. Individual case photogrammetric calibration of the Hirschberg Ratio (HR) for corneal light reflection test strabometry. *Binocular Vision and Strabismus Quarterly* 2006;**21**(1):45–6.

Rosner 1984

Rosner J, Clift GD. The validity of the Frisby stereotest as a measure of precise stereoacuity. *Journal of the American Optometric Association* 1984;**55**(7):505–6.

Rutstein 2000

Rutstein RP, Corliss DA. Distance stereopsis as a screening device. *Optometry and Vision Science* 2000;**77**(3):135–9.

Sansonetti 2004

Sansonetti A, Perisset J, Reinhardt M. Screening of vision disorders in young children. *Revue Medicale de la Suisse Romande* 2004;**124**(8):514–6.

Schaeffel 2002

Schaeffel F. Kappa and Hirschberg ratio measured with an automated video gaze tracker. *Optometry and Vision Science* 2002;**79**(5):329–34.

Schmidt 2003

Schmidt PP, Maguire MG, Moore B, Cyert L. Testability of preschoolers on stereotests used to screen vision disorders. *Optometry and Vision Science* 2003;**80**(11):753–7.

Schmidt 2004

Schmidt P, Maguire M, Dobson V, Quinn G, Ciner E, Cyert L, et al. Comparison of preschool vision screening tests as administered by licensed eye care professionals in the Vision In Preschoolers Study. *Ophthalmology* 2004;**111**(4): 637–50.

Schmucker 2009

Schmucker C, Grosselfinger R, Riemsma R, Antes G, Lange S, Lagreze W, et al. Effectiveness of screening preschool children for amblyopia: a systematic review. *BMC Ophthalmology* 2009;**9**:3.

Scott 1973

Scott AB. Editorial: Strabismus--beyond the cover test. *Investigative Ophthalmology* 1973;**12**(10):719–20.

Silbert 2013

Silbert DI, Arnold RW, Matta NS. Comparison of the iScreen and the MTI photoscreeners for the detection of amblyopia risk factors in children. *Journal of AAPOS* 2013; **17**(1):34–7.

Simons 1981a

Simons K. Stereoacuity norms in young children. *Archives of Ophthalmology* 1981;**99**(3):439–45.

Simons 1981b

Simons K. A comparison of the Frisby, Random-Dot E, TNO, and Randot circles stereotests in screening and office use. *Archives of Ophthalmology* 1981;**99**(3):446–52.

Simons 1996a

Simons K. Preschool vision screening: rationale, methodology and outcome. *Survey of Ophthalmology* 1996; **41**(1):3–30.

Simons 1996b

Simons K, Avery KE, Novak A. Small-target random dot stereogram and binocular suppression testing for preschool vision screening. *Journal of Pediatric Ophthalmology and Strabismus* 1996;**33**(2):104–13.

Smith 1985

Smith EL 3rd, Levi DM, Manny RE, Harwerth RS, White JM. The relationship between binocular rivalry and strabismic suppression. *Investigative Ophthalmology & Visual Science* 1985;**26**(1):80–7.

Solebo 2015

Solebo AL, Cumebrland PM, Rahi JS. Whole-population vision screening in children aged 4-5 years to detect amblyopia. *Lancet* 2015;**385**(9984):2308–19.

Takai 2005

Takai Y, Sato M, Tan R, Hirai T. Development of stereoscopic acuity: longitudinal study using a computer-based random-dot stereo test. *Japanese Journal of Ophthalmology* 2005;**49**(1):1–5.

Thomson 1999

Thomson WD, Evans B. A new approach to vision screening in schools. *Ophthalmic and Physiological Optics* 1999;**19**(3): 196–209.

Tongue 1981

Tongue AC, Cibis GW. Bruckner test. *Ophthalmology* 1981; **88**(10):1041–4.

Tongue 1987

Tongue AC. Refractive errors in children. *Pediatric Clinics of North America* 1987;**34**(6):1425–37.

Traboulsi 2008

Traboulsi EI, Cimino H, Mash C, Wilson R, Crowe S, Lewis H. Vision First, a program to detect and treat eye

diseases in young children: the first four years. *Transactions of the American Ophthalmological Society* 2008;**106**:179–85.

Turacli 1995
Turacli ME, Aktan SG, Duruk K. Ophthalmic screening of school children in Ankara. *European Journal of Ophthalmology* 1995;**5**(3):181–6.

UK National Screening Committee
UK National Screening Committee. The UK NSC policy on vision defects screening in children. legacy.screening.nhs.uk/vision-child (accessed December 2013).

VanEenwyk 2008
VanEenwyk J, Agah A, Giangiacomo J, Cibis G. Artificial intelligence techniques for automatic screening of amblyogenic factors. *Transactions of the American Ophthalmological Society* 2008;**106**:64–73.

VIP 2005
The Vision in Preschoolers Study Group. Preschool vision screening tests administered by nurse screeners compared with lay screeners in the vision in preschoolers study. *Investigative Ophthalmology & Visual Science* 2005;**46**(8): 2639–48.

Walraven 1993
Walraven J, Janzen P. TNO stereopsis test as an aid to the prevention of amblyopia. *Ophthalmic & Physiological Optics* 1993;**13**(4):350–6.

Walsh 2000
Walsh LA, Laroche GR, Tremblay F. The use of binocular visual acuity in the assessment of intermittent exotropia. *Journal of AAPOS* 2000;**4**(3):154–7.

Webber 2008
Webber AL, Wood JM, Gole GA, Brown B. The effect of amblyopia on fine motor skills in children. *Investigative Ophthalmology & Visual Science* 2008;**49**(2):594–603.

Wedner 2000
Wedner SH, Ross DA, Balira R, Kaji L, Foster A. Prevalence of eye diseases in primary school children in a rural area of Tanzania. *British Journal of Ophthalmology* 2000;**84**(11): 1291–7.

Weinand 1998
Weinand F, Graf M, Demming K. Sensitivity of the MTI photoscreener for amblyogenic factors in infancy and early childhood. *Graefe's Archive for Clinical and Experimental Ophthalmology* 1998;**236**(11):801–5.

Wick 1980
Wick B, London R. The Hirschberg test: analysis from birth to age 5. *Journal of the American Optometric Association* 1980;**51**(11):1009–10.

Williams 2001
Williams C, Harrad RA, Harvey I, Sparrow JM, ALSPAC Study Team. Screening for amblyopia in preschool children: results of a population-based, randomised controlled trial. ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. *Ophthalmic Epidemiology* 2001; **8**(5):279–95.

World Bank
World Bank. How we classify countries. data.worldbank.org/about/country-classifications (accessed 30 April 2013).

World Health Organization 2014
World Health Organization. Definition of region groupings. www.who.int/healthinfo/global˙burden˙disease/definition˙regions/en/ (accessed 9 June 2014).

Yang, 2012
Yang HK, Han SB, Hwang JM, Kim YJ, Jeong CB, Kim KG. Assessment of binocular alignment using the three-dimensional Strabismus Photo Analyzer. *British Journal of Ophthalmology* 2012;**96**(1):78–82.

References to other published versions of this review

Tailor 2014
Tailor V, Balduzzi S, Hull S, Rahi J, Schmucker C, Virgili G, Dahlmann-Noor A. Tests for detecting strabismus in children age 1 to 6 years in the community. *Cochrane Database of Systematic Reviews* 2014, Issue 7. [DOI: 10.1002/14651858.CD011221

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Arthur 2009

Study characteristics	
Patient sampling	Invitation to all children in kindergarten.
Patient characteristics and setting	No prior testing.
Index tests	plusoptix S04 screener, conducted by dental nurses.
Target condition and reference standard(s)	Amblyopia and strabismus target conditions with standards of visual acuity, pupils, motility, cover test, binocular sensory tests, ± cycloplegic refraction and dilated fundus examination
Flow and timing	Reference standard completed within 2 to 3 months of screening. 31 no reference test (14 declined, 11 unable to attend within time frame, 6 uncontactable)
Comparative	
Notes	A possible risk of bias could arise from concerned parents being more likely to return consent forms for the study. There was only a 25% participation rate

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Arthur 2009 *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Enzenauer 2000	Subgroup data for 1- to 6-year-old participants unavailable. Author contacted but further data not available
Robinson 1999	Information unavailable on the false positives and true negatives for the strabismus outcomes for each year (table 4 in paper) so full data analysis could not be done. Author contacted but data not available
Shallo-Hoffman 2004	No reported strabismus outcomes.
Tung 2006	Information unavailable on the exact numbers of true positives, true negatives, false positives and false negatives that were used to generate table 4 for both the Hirschberg corneal light reflex test and the NTU-random dot stereograms. Full data analysis could not be done. Senior author contacted

(Continued)

VIP 2007	The tests for this and all other published VIP studies were conducted on a population enriched for eye disease and so did not meet the study criteria for this review

D A T A

Presented below are all the data for all of the tests entered into the review.

Tests.   Data tables by test

Test	No. of studies	No. of participants
1 Photoscreener	1	271

Test 1.   Photoscreener.

Review: Tests for detecting strabismus in children aged 1 to 6 years in the community

Test: 1 Photoscreener

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Arthur 2009	6	8	7	250	0.46 [0.19, 0.75]	0.97 [0.94, 0.99]		

A D D I T I O N A L   T A B L E S

Table 1.   Data extraction from included studies

Study ID	First author, year of publication
Clinical features and settings	Previous testing and results. Setting: community/school/clinic (office) setting. Referral route/selection.
Participants	Sample size. Socio-demographic items: age, gender, ethnicity, frequency of ocular abnormalities (i.e. media opacities, which would affect test results/technical failure rates), geographic region
Study design	Selection: as single group/as separate group with/without target condition. Enrolment: consecutive series. Identification: prospective/retrospective. If more than one test: how were tests allocated to individuals, did each individual receive all tests?

Table 1. Data extraction from included studies (*Continued*)

Target condition	Constant and intermittent manifest strabismus (esotropia, exotropia, vertical tropia, microtropia), including the prevalence of the target condition in the sample
Reference standard	Test definition and description, i.e. cover test; 'comprehensive eye examination' (visual acuity, cover test, cycloplegic refraction). Test operator(s). Timing of reference standard.
Index tests	Test definition and description. Criteria for positive test result. Details of test operators. Timing. Manufacturer. Technical characteristics.
Follow-up	How many participants were lost to follow-up: unknown. How many have missing or uninterpretable test results: unknown. Adverse events noted that could be caused by the test: none reported
Notes	Sources of funding. Abbreviations. Anything else of relevance.

Table 2. QUADAS-2 assessment guidance

Domain	Yes	No	Unclear
PATIENT SELECTION **Describe methods of patient selection: Describe included patients (prior testing, presentation, intended use of index test and setting)**			
Was a consecutive or random sample of patients enrolled?	Consecutive sampling or random sampling of children according to inclusion criteria	Non-random sampling or sampling based on volunteering or referral	Unclear whether consecutive or random sampling used.
Was a case-control design avoided?	Yes for all studies since case-control studies are excluded unless nested in cohort studies	N/A	N/A
Did the study avoid inappropriate exclusions?	Exclusions are detailed and felt to be appropriate (systemic disease causing strabismus) Children with known strabismus can be excluded.	Inappropriate exclusions are reported e.g. of children in whom strabismus has been suspected in primary care but not confirmed by trained professionals	Exclusions are not detailed (pending contact with study authors)

Table 2. QUADAS-2 assessment guidance *(Continued)*

Risk of bias: could the selection of patients have introduced bias?	-	-	-
Concerns regarding applicability: are there concerns that the included patients do not match the review question?	Inclusion of children in community settings, such as school or screening settings, with no previous diagnosis of any eye disease	Inclusion of children over the age of 6 years, referred to clinical settings, referred to eye professionals for suspect eye disease, or assessed in commercial settings on a volunteer basis; or previous diagnosis of failed screening test or strabismus	Unclear inclusion criteria.

INDEX TEST
Describe the index test and how it was conducted and interpreted

Were the index test results interpreted without knowledge of the results of the reference standard?	Test performed "blinded" or "independently and without knowledge of" reference standard results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding	Reference standard results available to those who conducted or interpreted the index tests	Unclear whether results are interpreted independently.
If a threshold was used, was it pre-specified?	Many included index tests are based on continuous measures (e.g. eye deviation, stereopsis, refractive error, visual acuity); the study authors declare that the selected cut-off used to dichotomise data was specified a priori, or a protocol is available with this information	A study is classified at higher risk of bias if the authors define the optimal cut-off post hoc based on their own study data	No information on pre-selection of index test cut-off values
Risk of bias: could the conduct or interpretation of the index test have introduced bias?	-	-	-
Concerns regarding applicability: are there concerns that the index test, its conduct, or interpretation differ from the review question?	Tests used and testing procedure clearly reported and tests executed by personnel with sufficient training	Tests used are not validated or study personnel is insufficiently trained	Unclear tests (e.g. stereopsis-based tests but does not mention if a validated test is used) or unclear study personnel profile, background and training

REFERENCE STANDARD
Describe the reference standard and how it was conducted and interpreted

Table 2. QUADAS-2 assessment guidance *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Cover- uncover test performed by trained professionals, e.g. ophthalmologists, optometrists, orthoptists	Complete eye examination with cover- uncover test used as reference standard but not only the cover- uncover test used to judge on strabismus (e.g. visual acuity measure also used)	Complete eye examination used but unclear whether cover-uncover test used
Were the reference standard results interpreted without knowledge of the results of the index test?	Reference standard performed "blinded" or "independently and without knowledge of" index test results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding	Index test results available to those who conducted the reference standard; or the index test is part of the reference standard (e.g. visual acuity within a compete ophthalmic examination used as reference standard and visual acuity is also the index test analysed - this will be specific of each analysis)	Unclear whether results are interpreted independently.
Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias?			
Concerns regarding applicability: are there concerns that the target condition as defined by the reference standard does not match the review question?	Cover- uncover test used and testing procedure executed by personnel with sufficient training	Cover- uncover test used by personnel with inappropriate profile or insufficient training	Unclear study personnel profile, background and training.

FLOW AND TIMING
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2×2 table (refer to flow diagram): describe the time interval and any interventions between index test(s) and reference standard

Was there an appropriate interval between index test(s) and reference standard?	No more than three months between index and reference test execution, and no corrective intervention between assessments	More than three months between index and reference test execution	Unclear whether test results are executed within three months
Did all patients receive a reference standard?	The verification rate of index test-positive children is definitely higher than that of negative children (the opposite is unlikely)	All children receiving the index test are verified with the reference standard	Unclear whether all children receiving the index test are verified with the reference standard
Did all patients receive the same reference standard?	All children are verified with the cover- uncover test by trained professionals	Some children, i.e. positive children, are verified with the cover - uncover test by specialised personnel, while the others are	Unclear whether all children are verified with the cover- uncover test by trained professionals

Table 2. QUADAS-2 assessment guidance   *(Continued)*

		verified by personnel with lower level of training	
Were all patients included in the analysis?	The number of children included in the study does not match the number in analyses or children with undefined or borderline test results are excluded. However, children in whom one or more index tests are not performed because they are poorly cooperative can be excluded	The number of children included in the study does not match the number in analyses and children with undefined or borderline test results are excluded from the analyses	The number of children analysed, but not that included in the study, are reported; or unclear if there were inappropriate exclusions
Risk of bias: could the patient flow have introduced bias?	-	-	-
COMPARATIVE STUDIES (MULTIPLE INDEX TESTS)			
Were all tests performed on all patients, or randomly assigned?	All children received all index tests, or tests were randomly assigned	Not all children received all index tests and the assignment criterion was opportunistic or non-random (e.g. depending on test availability or type of professional)	Not all children received all index tests and the assignment criterion was unclear
Could the order in which the index tests were used affect the target condition or the interpretation of the alternative tests?	The order of presentation of the index test was random or alternate to avoid fatigue effects; or clear that no fatigue effect can arise	Several tests are delivered in a fixed order which can cause children to be less compliant with the second or later test	Unclear order of test presentation.

Table 3. Thresholds for analysis

Test type categories	Tests included	Output measure	Threshold to extract data
1) Tests which identify ocular misalignment	1.1) Corneal reflections tests: Hirschberg, Krimsky (prism reflection test). 1.2) Fundus reflections test: Brückner.	Prism dioptres (PD).	8 PD for horizontal deviations; 1 PD for vertical deviations (no published threshold identified)
2) Test of binocular function: stereopsis	Stereoacuity tests such as contour and random dot stereotests	Seconds of arc.	400 seconds of arc.
3) Tests designed to detect reduced ventral vision	3.1) Visual acuity tests, e.g. HOTV, LEA symbols, Keeler (previously Glasgow) crowded	LogMAR or logMAR equivalent.	0.2 logMAR.

Table 3. Thresholds for analysis   *(Continued)*

	logMAR, Sonksen crowded logMAR, crowded Kay picture test 3.2) Suppression tests. 3.3) Blur test.		
4) Automated refraction devices designed to report ocular misalignment	-	Millimetres of asymmetry or corneal reflections.	No published threshold identified.

APPENDICES

Appendix 1. Prevalence of strabismus

Study ID	Country/ Region	Definition of strabismus	Population	Strabismus prevalence
Friedman 2009	USA, Baltimore.	Any constant or intermittent heterotropia at near or distance fixation	Age 6 to 71 months, white.	3.3%
			Age 6 to 71 months, African-American.	2.1%
Preslan 1996	USA, Baltimore.	Manifest strabismic deviation with or without fixation preference, using alternate cover testing with the child fixating on an accommodative target at 33 cm	All children attending one school: preschool (125 children), kindergarten (213), first grade (165) and second grade (177); 75% African-American, 22% white, 3% other	3.1%
Traboulsi 2008	USA, Cleveland.	Manifest ocular deviation of any type or magnitude with or without fixation preference; any vertical phoria; any esophoria more than 8 prism dioptres (PD) or exophoria more than 15 PD at near or distance; any restricted eye movement or muscle overaction	Age 4 to 8 years. General population of Cleveland: 70.3% African American, 16.7% white, 10.4% Hispanic, 2.6% other	2.1%

Turacli 1995	Turkey, Ankara.	Not defined, but results report manifest and latent deviations, with manifest in 93% of all those classified as "strabismic"	Children attending randomly selected school in the Ankara urban area, age 5 to 12 years	2.5%
Wedner 2000	Tanzania, rural area	Not defined.	Tanzanian children and young people age 7 to 19 years.	0.5%
Matsuo 2007a	Japan, Okyama.	Not defined in abstract; full article not available.	Japanese children age 1.5 years.	0.01 to 0.12% 0.99%
			Japanese children age 3 years.	0.2 to 0.34%
Matsuo 2007b	Japan, Okyama	Not defined in abstract; full article not available	Japanese children age 6 to 12 years	0.99% in 2005, 1.28% in 2003
Graham 1974	UK	Manifest deviation, or exophoria of 9 PD or more, or esophoria of 7 PD or more, any hyperphoria	All children born in Cardiff between 1 January and 31 December in one year (not specified), age 5 to 6 years	manifest or large latent: 7%; manifest only: 5% large phoria: 1.3%

Appendix 2. Index tests

Corneal reflection test (Hirschberg) (Hirschberg 1881)

First described in 1881, the Hirschberg or corneal reflection test (CRT) uses the reflection of a light target by the surface of the cornea, also known as first Purkinje image, to evaluate whether the eye is fixing on a target. The normal position of the corneal light reflection (CR) is 0.5 mm nasal to the centre of the cornea, as the fovea is located temporal to the optical axis, resulting in a small angle between the visual axis and the optical axis of the eye (angle kappa).

In the literature there is some confusion around the terms visual axis, pupillary axis, optical axis, line of sight, angle kappa and angle lambda.

The visual axis is the line connecting the fovea and the nodal point of the eye and continuing anteriorly through the cornea.

The pupillary axis is the line perpendicular to the cornea that intersects the centre of the pupil. It is a clinical approximation of the visual axis.

The optical or anatomical axis of the eye connects the centre of the curvature of the cornea and the centre of the curvature of the posterior pole.

The line of sight is the line that connects the fixation point and the centre of the pupil; it is a clinical approximation of the optical axis.

The angle between visual and optical axis is called angle kappa. Landolt originally defined angle kappa as "the angle between the visual axis and the so-called central pupillary line (the pupillary axis)" (Emsley 1948). Lancaster then defined angle lambda as the angle between the pupillary axis and the line of sight. LeGrand finally re-defined angle kappa exactly the way Lancaster had defined angle lambda, stating that the nodal point of the eye is a theoretical concept, and that for all practical purposes the visual axis is identical to the line of sight (LeGrand 1980). In addition, angle lambda and angle kappa are nearly identical when the point of fixation is not very close to the eye.

Angle kappa is the angle between the visual and optical axis, or between the pupillary axis and the line of sight (LeGrand 1980). By convention it is normally positive.

Individuals with exotropia have higher angle kappa values than esotropic and orthotropic individuals (Basmak 2007). A large angle kappa may also give rise to pseudo-exotropia. When the fovea is situated nasal to the optical axis, such as in high myopia or ectopic

fovea, for example after retinopathy of prematurity, the angle kappa is negative, the CR is located temporal to the centre of the cornea, and pseudo-esotropia may be present.

In ocular misalignment, the CR is displaced - nasally in exotropia, temporally in esotropia (Hirschberg 1881). When the CR is located at the border of the pupil, the deviation is approximately 15 prism dioptres (PD). If it lies midway across the iris, the deviation measures around 30 PD, and when the reflection is near the limbus, around 45 PD. Hirschberg's original observations indicated a ratio of 12 to 14 PDs per millimetre displacement of the CR from the pupillary axis, the so-called Hirschberg ratio. Later evaluations of the Hirschberg test using photography to standardise measurements indicated a ratio of 19.5/1 (Wick 1980), 21/1 (Brodie 1987; DeRespinis 1989), 22/1 (Eskridge 1988) or 24/1 (Carter 1978) with little change from birth to adulthood (Hasebe 1998; Riddell 1994; Wick 1980). Photographs acquired whilst fixating with first the preferred eye, then the deviating eye, in primary position and in slightly eccentric fixation may allow a highly accurate measurement of the ocular misalignment (Romano 2006). Based mainly on reasons of photographic technique, some consider the limbus a more accurate landmark than the centre of the pupil (Barry 1997; Romano 2006); however displacement from the centre of the pupil remains the more commonly used value.

Use of the Hirschberg test as a screening test for ocular alignment has been recommended in young preverbal and also in pre-school children (daSilva 1991; Sansonetti 2004). To allow standardisation, videographic techniques have been proposed (Miller 1993). These can be applied when using video refractors or photoscreeners developed for the automated assessment of refractive errors (Griffin 1989; Hasebe 1995; Moghaddam 2012; Schaeffel 2002; Weinand 1998).

Automated assessment of CR on digital photographs and videographs is currently in development (Almeida 2012; Model 2012; Yang, 2012).

Accuracy of the Hirschberg test may be in the range of ± 9 to 10 PD, which would make it unsuitable to detect or exclude microtropia. In orthoptic practice, Hirschberg and Krimsky tests are reserved for very young, preverbal patients or those with profound visual impairment which prevents fixation with the affected eye(s). The Hirschberg test is useful to demonstrate pseudo-strabismus in young children with a broad nasal bridge and epicanthal folds or in individuals with wide interpupillary distance.

Coaxial fundus reflex test (Brückner)

The Brückner test (Brückner 1965; Tongue 1981) is based on reflection of light by the fundus/retina at the back of the eye. A bright coaxial light source and observation system is used, usually a direct ophthalmoscope. Both eyes of the patient are simultaneously illuminated from a distance of around one metre. The observer notes any difference in brightness of the fundus reflex seen in the pupil through the ophthalmoscope. In the presence of strabismus the reflex is darker in the fixing eye than in the deviated eye. An additional, dynamic examination of pupil size, pupil reaction to light, and fixation movement of the eyes on alternating illumination can be useful to detect amblyopia (Tongue 1981). Whilst the manual Brückner test appears reliable at detecting strabismus, it may give rise to false positives, and its usefulness in screening for ocular misalignment is unclear (Griffin 1986; Griffin 1989). A recent evaluation of a modified Brückner test, performed using a streak retinoscope, indicated a sensitivity of 0.5, specificity of 0.98, negative predictive value of 0.97 and positive likelihood ratio of 20 to detect strabismus in a cohort of 343 children with a 5% prevalence of strabismus (Amitava 2012). Photographic and videographic versions of the Brückner test may also have high sensitivity and specificity to detect strabismus (Carrera 1993; Cibis 1994; Kaakinen 1979; Miller 1995; VanEenwyk 2008). Some photorefractors use the Brückner principle, as refractive errors cause a white crescent at the pupil border (Arnold 2000; Kothari 2007; Tongue 1987; Weinand 1998). These devices are commonly used in community screening programmes, operated by lay observers (Arnold 2000). Sensitivity and specificity appear higher on analysis of the coaxial fundus reflex on photographs than during manual performance of the test (Paysse 2001). The size of misalignment may affect the accuracy of the Brückner test, with lower sensitivity in small-angle esotropia (Graf 2012).

Stereovision tests

Stereovision or stereopsis is the perception of depth when viewing a scene with both eyes. As the visual axis of the right and left eye are at a slight angle to each other, the image seen by the right eye slightly differs from that seen by the left eye. This binocular disparity allows the brain to 'calculate' depth in the visual scene. True stereopsis requires perfect alignment and fusion of the foveal images from both eyes, known as central fusion. When assessing stereopsis, it is important to eliminate any monocular cues and to present stimuli whose three-dimensional qualities can only be perceived when foveal information from both eyes is integrated.

Tests of stereovision may indicate ocular misalignment. As other causes - such as uncorrected refractive errors and reduced visual acuity - also impair stereopsis, specificity for any particular cause may be poor.

Stereopsis is most commonly tested at near. Near stereotests fall into two categories: contour (Titmus fly, Wirt ring test); and random dot tests. Contour tests achieve horizontal image disparity by vectographic techniques and require polarised glasses to view a three-dimensional (3D) picture embedded in polarised filter sheets made from plastic. By stacking two of these sheets at a perpendicular angle, a separate image is shown to each eye. When viewed without the glasses, the picture can still be seen, but its 3D qualities can only be perceived through the polarising glasses. As contour tests gives some monocular cues to the position of the 3D shapes many clinicians prefer random dot tests for testing stereovision.

Random dot images do not contain any contour lines. Shapes can only be seen and depth can only be perceived when true binocular stereopsis with central (foveal) fusion is present. Random dot tests include the Frisby, Lang, TNO, Randot and Randot-E stereotests (Broadbent 1990; Lang 1983; Rosner 1984; Simons 1981b).

The **Frisby** test consists of three Perspex plates of different thickness. On each plate there are four square areas which contain triangular shapes apparently distributed in a random pattern. On one of the squares, some shapes arranged in a geometric pattern, such as a circle, are printed onto the back surface of the Perspex plate, whilst the remaining square is filled with triangles printed onto the front surface. The physical thickness of the Perspex plate and the distance between the shapes printed onto the front and the back of the plate induce horizontal image disparity. This test is simple to perform, does not require 3D viewers and is popular with children. Preverbal children may point onto the 3D shape or may direct their eyes towards it, similar to their response in preferential looking tests.

The **Lang** stereotest combines two methods of three-dimensional image perception: random dots and cylindrical gratings. The cylindrical gratings use a prismatic effect to achieve the slight horizontal image displacement required for a 3D effect. The advantage of this method is that it does not require special glasses for viewing. Essentially, the separation of the two images is achieved by a system of fine parallel cylindrical shapes. Beneath each cylinder are two fine strips of picture, one seen by the right, the other seen by the left eye. In the Lang test, random dot images hide simple shapes such as a star, a cat, a car. As it does not require 3D viewers, it is easy to use with children and is commonly used in vision screening programmes. Like the Frisby test, it can be used in preverbal children by observing their behavioural response.

The **Randot** and **Random dot E** stereotests require polarising glasses for viewing. Images of animals and geometric shapes are horizontally displaced using vectographic techniques. These tests allow fine grading of stereopsis, but not all children will like wearing the polarising glasses.

The **TNO** test creates a 3D effect by using red‑green anaglyphs. Red‑green anaglyphs are based on two images showing the same scene from a slightly different angle. One image is processed through a red, and the other image through a green or blue or mixed (cyan) filter. The resulting images are superimposed, but slightly offset. When viewing these pictures through glasses with one red and one green lens, a stereoscopic effect results.

Near stereotests are not sufficiently accurate to be used as standalone vision screening tests (Donahue 2013; Huynh 2005; Ohlsson 2002; Schmidt 2003; VIP 2005; VIP 2007). Testability is affected by age (Pai 2012; Schmidt 2003).

Distance stereopsis can be measured with the distance Frisby stereotest, a cabinet which houses Perspex plates which present random images at slightly different distances from the observer (Adams 2005; Holmes 2005; Kaye 2005), or with a distance Randot test (Fu 2006). Despite reports of high sensitivity to detect vision defects (Rutstein 2000), distance stereopsis has not been evaluated in vision screening programmes. Distance stereoacuity can be reduced in convergence excess esotropia and intermittent distance exotropia (Hatt 2008).

Visual acuity tests

Visual acuity (VA) is a measure of the spatial resolution of the visual system. Manifest strabismus causes a loss of VA in one eye by central suppression of the information from the deviating eye. Uncorrected refractive errors and ocular anomalies also cause a reduction in VA.

In older children and adults, VA is assessed by reading a chart of characters, or optotypes, at a defined distance. In very young children, assessment of visual acuity relies on observation of behavioural responses to visual targets.

Preferential looking cards showing patterns of high-contrast black and white stripes are used in children under the age of 2 years to determine "grating acuity" (Dobson 1978). In strabismic amblyopia, grating acuity is reduced to a lesser degree than linear letter acuity and results may overestimate the level of vision.

From the age of 2 years, single symbols such as Kay pictures can be used (Kay 1983). From the age of 3 years, crowded linear optotypes such as HOTV, or crowded Kay or Lea pictures can be used. These tests are often used in childhood vision screening programmes (Schmidt 2004; Hered 1997; Anonymous 2004). Crowded optotypes (several characters next to each other) viewed one eye at a time (monocularly), such as on HOTV or logMAR charts, are considered the 'gold standard' for visual acuity testing (Schmidt 2004). The use of crowded optotypes instead of single optotypes is particularly important in amblyopia screening, as single optotype testing can overestimate visual acuity. All visual acuity tests can be performed either by the child calling out the name of the picture or letter, or by

the child matching the target optotype with a chart held by a parent/guardian. More detailed letter‑optotype tests include the Keeler logMar and the Sonksen Silver logMar tests.

In order to both increase portability of charts and reduce variation of illumination levels, computer-based testing applications are available and used in some screening settings (Thomson 1999).

Visual acuity tests can be administered by any suitably trained person. In the UK, screening programmes are delivered by qualified orthoptists, health care technicians or school nurses trained by orthoptists (Hall 2003; UK National Screening Committee) In the USA, paediatric vision screening is usually performed by suitably trained nurses or lay screeners.

Autorefractors/Photorefractors

Autorefractors are instruments that measure the refractive state of the eye. Objective devices contain an optical system which determines the vergence of light reflected from the patient's retina. To avoid inducing accommodation, modern autorefractors use infrared rather than visible light.

Photorefraction analyses the reflection of light emitted from a small flashlight placed close to the camera lens. Three types of photorefraction have been developed: orthogonal, isotropic and eccentric (also called photoretinoscopy). Refractive errors result in certain patterns of photographic appearances, which vary with the degree to which the eye is defocused with respect to the plane of the camera (Howland 1974; Howland 2009). Photorefractors are mainly used to obtain refractive values. Some devices combine a photographic Brückner test and eccentric photorefraction to detect amblyogenic risk factors (Cibis 1994; VanEenwyk 2008). Several current photorefractors also detect strabismus as asymmetry of corneal light reflections (Arnold 2013; Dahlmann-Noor 2009a; Dahlmann-Noor 2009b; Moghaddam 2012; Silbert 2013).

The following table summarises possible test outcomes, pass/fail thresholds and examples of published screening studies that have used these tests. The variation of tests used in different studies for each group of index tests means that many specific tests have only been used in one or a small number of studies.

For the purpose of comparison all visual acuity thresholds have been converted to logMar, though this may not be entirely accurate.

	Possible measurements	Threshold for diagnosis of manifest strabismus	Use in studies
Type 1 tests: Direct identification of ocular alignment			
Corneal reflection tests			
Hirschberg test	Categorical: central, 15 PD (pupil border), 30 PD (mid-iris), 45 PD (limbus)	Any displacement from centre of pupil.	
Cover test (Gold standard)			
Cover- uncover test	Dichotomous: refixation movement present/absent.	Any refixation movement.	Traboulsi 2008: manifest ocular deviation of any type or magnitude with or without fixation preference
			Friedman 2009: constant or intermittent tropia of any magnitude at distance (6 m) or near (40 cm) fixation; if only testable at one distance and no strabismus on that test: non-strabismic
			Schmidt 2004: at 3 m and 40 cm: strabismus = any heterotropia in primary gaze

(Continued)

Simultaneous prism and cover test	Categorical: prisms used to neutralise deviation.	Any use of prism.	
Alternate cover test	Dichotomous: refixation movement present/absent.	Demonstrates manifest plus latent strabismus.	Traboulsi 2008: any vertical phoria; any esophoria more than 8 PD or exophoria more than 15 PD at near or distance
Prism and alternate cover test	Categorical: Prisms used to neutralise deviation.	Quantifies manifest plus latent strabismus.	Traboulsi 2008: Any vertical phoria; any esophoria more than 8 PD or exophoria more than 15 PD at near or distance
Type 2 test: Tests of binocular function: control and stereoacuity			
Contour tests: Titmus	Fly 3,600 sec of arc, animals 400, 200, 100, Wirt circles 800, 400, 200, 140, 100, 80, 60, 50, 40 seconds of arc at 40 cm	Fly 3,600 sec of arc; animals 400, 200, 100; Wirt rings 800 to 40 seconds of arc	Traboulsi 2008: less than 400 seconds of arc.
Random dot stereotests			
TNO	480, 240, 120, 60, 30, 15 seconds of arc.	At age 5 years: greater than 60 seconds of arc.	Manufacturer recommends 240 seconds of arc as "fail" threshold, as 95% of amblyopes are unable to see this figure
Lang II	600, 400, 200 seconds of arc at 40 cm.		
Frisby	At 30 cm viewing distance: 600, 300 150 seconds of arc. At 40 cm viewing distance: 340, 170 and 85 seconds of arc.		
Randot	500 to 20 seconds of arc.	At age 5 years: greater than 60 seconds of arc.	
Random dot E at 0.5, 1 and 1.5 m	Schmidt 2004: non-stereocard only; 504; 252; 168 arc seconds		Schmidt 2004: non-stereocard only; 504; 252; 168 seconds of arc
Stereo Smile II at 40 cm	Schmidt 2004: non-stereocard only; 480; 240; 120 arc seconds.		Schmidt 2004: non-stereo card only; 480; 240; 120 seconds of

(Continued)

			arc.
Type 3 tests: **Visual acuity tests**			
Single-surrounded HOTV			Friedman 2009: • uniocular amblyopia = 2-line interocular difference in best-corrected visual acuity (BCVA), 0.20 logMar or worse in the worse eye, plus at least one amblyogenic risk factor (strabismus, previous strabismus surgery, anisometropia consistent with the eye with worse VA), evidence of past or present visual axis obstraction. • bilateral amblyopia = bilateral BCVA worse than 0.40 logMar if age 30 to 47 months, or worse than 0.30 logMar if age ≥ than 48 months, with either bilateral visual axis obstruction or bilateral ametropia.
HOTV at 3 m		Schmidt 2004: 3 years: 10/100; 10/32; 10/25; 10.20. 4 years: 10/100; 10/25; 10/20; 10/16.	Schmidt 2004: • suspected unilateral amblyopia = 2 line interocular difference and a unilateral amblyogenic factor (strabismus, anisometropia). • suspected bilateral amblyopia: 3 year olds: worse than 0.40 logMar in one eye, worse than 0.30 logMar in the contralateral eye, and a bilateral amblyogenic factor (refractive error); 4 year olds: worse than 0.3 logMar in one eye, worse than 0.18 logMar in the contralateral eye, and a bilateral amblyogenic risk factor (refractive error).
LEA symbols at 3 m		Schmidt 2004: 3 years: 10/100; 10/32; 10/25; 10.20.	

		4 years: 10/100; 10/25; 10/20; 10/16.	
Snellen letters			Traboulsi 2008: screen fail if visual acuity difference of 2 Snellen lines between eyes or visual acuity less than 0.30 logMar in either eye
Allen figures			Traboulsi 2008: screen fail if visual acuity difference of 2 Snellen lines between eyes or visual acuity less than 0.30 logMar in either eye
Keeler (previously Glasgow) crowded logMar			
Sonksen			
Crowded Kay Picture Test			
Type 4 tests: refraction devices that report ocular misalignment			
Plusoptix Vision Screener	Assessment of corneal reflections (asymmetry in millimetres or ratio)		

Other tests used in the diagnosis of strabismus, but not in primary care or community screening settings delivered by lay screeners or primary care professionals

Krimsky test

The prism reflection test is a modification of the Hirschberg test. The patient fixes a spotlight at a near position (33 cm). Prisms are placed in front of the fixing eye, with the apex pointing in the direction of the deviation. This shifts the CR towards the centre of the pupil. The prism which positions the CR in the centre of both pupils indicates the angle of deviation (Krimsky 1951). This test is typically used in children or adults who are not cooperative enough to undertake a Prism Cover Test or in cases when the visual acuity is poor and the patient is unable to move the eye to take up fixation. The 95% limit of agreement of inter-observer variability for the Krimsky test has been reported as 6.1 PD (Yang, 2012). Neutralisation of the deviation with prisms requires orthoptic expertise, and this test is not used by lay screeners or non-ophthalmic professionals in community screening programmes.

Prism reflection test

The prism reflection test is similar to the Krimsky test, but the prism is placed in front of the deviating eye. Some authors found the prism reflection test to have low accuracy (Choi 1998).

Controlled binocular acuity (CBA) test of strabismus

Some types of strabismus can be controlled by increased accommodation, i.e. 'over-focusing'. In intermittent distance exotropia (IDEX), individuals may use accommodative convergence to control the exodeviation; hence over-accommodating and losing clarity of vision in the distance (Walsh 2000). To measure controlled binocular acuity, the patient reads an optotype chart whilst the examiner observes the patient's ocular alignment by corneal reflections and noting at which optotype size one eye deviates.

Suppression tests

Visual information from the deviating eye is suppressed at the level of the visual cortex of the brain (Smith 1985). Binocular suppression tests may have high specificity and may be useful for strabismus and amblyopia screening (Pott 1998; Simons 1996b). Tests are based on dissociation of the images seen by each eye, for example by using polarised lenses (Nuzzi 1986; Pott 1998; Pott 2003; Prakash 1996). Worth's four-light test, Bagolini striated lenses or the synoptophore are used to detect the presence of suppression. The 4-prism dioptre (PD) prism test is used to detect central suppression. The area of suppression within the field of vision can be approximated by the synoptophore. Density of suppression can be quantified by using neutral density filters or the Sbisa bar (Bagolini filter bar) (McCormick 2002). Whilst used in orthoptic practice to determine the risk of developing double vision in adult patients undergoing amblyopia treatment, suppression measurement is not used in paediatric screening.

Fusion tests

Orthoptists may test children's ability to overcome a 4, 15 or 20 PD prism to maintain binocular single vision (prism reflex test); failure to overcome the prism can indicate weakness of fusional control. The prism fusion range uses the same principle, but uses a prism bar, allowing a more detailed measurement of prisms that can be overcome. Measurements are typically in steps as dictated by prism bar or loose prism used (2 to 20 PD usually in 2 PD increments, then 20 to 45 PD in 5 PD increments).

Blur test

The blur test aims to detect low hypermetropia, which may be associated with potentially decompensating strabismus.

Appendix 3. The Cochrane Library search strategy

#1 MeSH descriptor: [Vision Tests] explode all trees

#2 MeSH descriptor: [Mass Screening] explode all trees

#3 (test* or screen* or diagnos* or assess*) near/4 (vision or visual*)

#4 (test* or screen* or diagnos* or assess*) near/4 program*

#5 (test* or screen* or diagnos* or assess*) near/4 (communit* or population)

#6 (test* or screen* or diagnos* or assess*) near/4 hospital*

#7 (test* or screen* or diagnos* or assess*) near/4 (nursery or preschool* or school*)

#8 (test* or screen* or diagnos* or assess*) near/4 (referred or referal or monit*)

#9 (test* or screen* or diagnos* or assess*) near/4 (orthoptist* or ophthalmologist* or optometrist*)

#10 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9

#11 MeSH descriptor: [Strabismus] explode all trees

#12 strabism* or squint*

#13 esotrop* or exotrop*

#14 hypertrop* or hypotrop*

#15 microtrop*

#16 MeSH descriptor: [Amblyopia] this term only

#17 amblyop*

#18 lazy near/3 eye*

#19 #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18

#20 MeSH descriptor: [Infant] this term only

#21 MeSH descriptor: [Child, Preschool] this term only

#22 MeSH descriptor: [Child] this term only

#23 infan* or child* or toddler* or boy* or girl* or paediatric* or pediatric* or nursery or nurseries or kindergarten* or pre school* or preschool* or school age or schoolage*

#24 #20 or #21 or #22 or #23

#25 MeSH descriptor: [Photogrammetry] this term only

#26 MeSH descriptor: [Photography] this term only

#27 MeSH descriptor: [Blinking] this term only

#28 corneal light reflection*

#29 corneal reflection*

#30 CLRT

#31 (cover or uncover) near/3 test*

#32 fundus reflection*

#33 (Hirschberg* or Krimsky* or Bruckner*) near/5 test*

#34 MeSH descriptor: [Retinoscopy] this term only

#35 retinoscop*
#36 stereoscopy or stereotest*
#37 Randot
#38 random dot
#39 two pencil test
#40 Frisby* or Titmus*
#41 TNO or FD2
#42 HOTV or LEA
#43 suppression near/3 test*
#44 autorefractor* or photorefractor* or Plusoptix or Retinomax
#45 #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44
#46 #10 and #19 and #24 and #45

Appendix 4. MEDLINE (Ovid) search strategy

1. exp vision tests/
2. exp mass screening/
3. ((test$ or screen$ or diagnos$ or assess$) adj4 (vision or visual$)).tw.
4. ((test$ or screen$ or diagnos$ or assess$) adj4 program$).tw.
5. ((test$ or screen$ or diagnos$ or assess$) adj4 (communit$ or population)).tw.
6. ((test$ or screen$ or diagnos$ or assess$) adj4 hospital$).tw.
7. ((test$ or screen$ or diagnos$ or assess$) adj4 (nursery or preschool$ or school$)).tw.
8. ((test$ or screen$ or diagnos$ or assess$) adj4 (referred or referral or monit$)).tw.
9. ((test$ or screen$ or diagnos$ or assess$) adj4 (orthoptist$ or ophthalmologist$ or optometrist$)).tw.
10. or/1-9
11. exp strabismus/
12. (strabism$ or squint$).tw.
13. (esotrop$ or exotrop$).tw.
14. (hypertrop$ or hypotrop$).tw.
15. microtrop$.tw.
16. amblyopia/
17. amblyop$.tw.
18. (lazy adj3 eye$).tw.
19. or/11-18
20. Infant/
21. child, preschool/
22. child/
23. (infan$ or child$ or toddler$ or boy$ or girl$ or paediatric$ or paediatric$ or nursery or nurseries or kindergarten$ or pre school$ or preschool$ or schoolage$ or school age$).tw.
24. or/20-23
25. Photogrammetry/
26. Photography/
27. Blinking/
28. corneal light reflection$.tw.
29. corneal reflection$.tw.
30. CLRT.tw.
31. ((cover or uncover) adj3 test$).tw.
32. fundus reflection$.tw.
33. ((Hirschberg$ or Krimsky$ or Bruckner$) adj5 test$).tw.
34. Retinoscopy/
35. retinoscop$.tw.
36. (stereoscopy or stereotest$).tw.

37. Randot.tw.
38. random dot.tw.
39. two pencil test.tw.
40. (Frisby$ or Titmus$).tw.
41. (TNO or FD2).tw.
42. (HOTV or LEA).tw.
43. (suppression adj3 test$).tw.
44. (autorefractor$ or photorefractor$ or Plusoptix or Retinomax).tw.
45. or/25-44
46. 10 and 19 and 24 and 45

Appendix 5. Embase (Ovid) search strategy

1. exp vision test/
2. exp mass screening/
3. ((test$ or screen$ or diagnos$ or assess$) adj4 (vision or visual$)).tw.
4. ((test$ or screen$ or diagnos$ or assess$) adj4 program$).tw.
5. ((test$ or screen$ or diagnos$ or assess$) adj4 (communit$ or population)).tw.
6. ((test$ or screen$ or diagnos$ or assess$) adj4 hospital$).tw.
7. ((test$ or screen$ or diagnos$ or assess$) adj4 (nursery or preschool$ or school$)).tw.
8. ((test$ or screen$ or diagnos$ or assess$) adj4 (referred or referral or monit$)).tw.
9. ((test$ or screen$ or diagnos$ or assess$) adj4 (orthoptist$ or ophthalmologist$ or optometrist$)).tw.
10. or/1-9
11. exp strabismus/
12. (strabism$ or squint$).tw.
13. (esotrop$ or exotrop$).tw.
14. (hypertrop$ or hypotrop$).tw.
15. microtrop$.tw.
16. amblyopia/
17. amblyop$.tw.
18. (lazy adj3 eye$).tw.
19. or/11-18
20. Infant/
21. preschool child/
22. child/
23. (infan$ or child$ or toddler$ or boy$ or girl$ or paediatric$ or pediatric$ or nursery or nurseries or kindergarten$ or pre school$ or preschool$ or schoolage$ or school age$).tw.
24. or/20-23
25. Eye photography/
26. Medical photography/
27. Photography/
28. Blinking/
29. Hirschberg cornea light reflection test/
30. corneal light reflection$.tw.
31. corneal reflection$.tw.
32. CLRT.tw.
33. ((cover or uncover) adj3 test$).tw.
34. fundus reflection$.tw.
35. ((Hirschberg$ or Krimsky$ or Bruckner$) adj5 test$).tw.
36. Retinoscopy/
37. retinoscop$.tw.
38. (stereoscopy or stereotest$).tw.
39. Randot.tw.

40. random dot.tw.
41. two pencil test.tw.
42. Frisby test/
43. Revised Frisby Davis Distance test/
44. (Frisby$ or Titmus$).tw.
45. TNO stereotest/
46. (TNO or FD2).tw.
47. (HOTV or LEA).tw.
48. (suppression adj3 test$).tw.
49. (autorefractor$ or photorefractor$ or Plusoptix or Retinomax).tw.
50. or/25-49
51. 10 and 19 and 24 and 50

Appendix 6. CINAHL (EBSCO) search strategy

S34 S28 and S33
S33 S29 or S30 or S31 or S32
S32 infan* or child* or toddler* or boy* or girl* or paediatric* or pediatric* or nursery or nurseries or kindergarten* or pre school* or preschool* or schoolage* or school age*
S31 (MH "Child+")
S30 (MM "Child, Preschool")
S29 (MH "Infant+")
S28 S18 and S27
S27 S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26
S26 lazy N3 eye
S25 amblyop*
S24 (MM "Amblyopia")
S23 microtrop*
S22 hypertrop* or hypotrop*
S21 esotrop* or exotrop*
S20 strabism* or squint*
S19 (MM "Strabismus")
S18 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13 or S14 or S15 or S16 or S17
S17 optometrist* N4 (test* or screen* or diagnos* or assess*)
S16 ophthalmologist* N4 (test* or screen* or diagnos* or assess*)
S15 orthoptist* N4 (test* or screen* or diagnos* or assess*)
S14 monit* N4 (test* or screen* or diagnos* or assess*)
S13 referral N4 (test* or screen* or diagnos* or assess*)
S12 referred N4 (test* or screen* or diagnos* or assess*)
S11 school* N4 (test* or screen* or diagnos* or assess*)
S10 preschool* N4 (test* or screen* or diagnos* or assess*)
S9 nursery N4 (test* or screen* or diagnos* or assess*)
S8 hospital* N4 (test* or screen* or diagnos* or assess*)
S7 population N4 (test* or screen* or diagnos* or assess*)
S6 communit* N4 (test* or screen* or diagnos* or assess*)
S5 program* N4 (test* or screen* or diagnos* or assess*)
S4 visual* N4 (test* or screen* or diagnos* or assess*)
S3 vision N4 (test* or screen* or diagnos* or assess*)
S2 (MM "Vision Screening")
S1 (MH "Vision Tests+")

Appendix 7. Web of Science Conference Proceedings Citation Index - Science (CPCI-S) search strategy

#20 #15 AND #18 AND #19

#19 TS= (infan* or child* or toddler* or boy* or girl* or paediatric* or pediatric* or nursery or nurseries or kindergarten* or pre school* or preschool* or schoolage* or school age*)

#18 #16 OR #17

#17 TS= (esotrop* or exotrop* or hypertrop* or hypotrop* or microtrop*)

#16 TS= (strabimus or strabismic or squint*)

#15 #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 OR #14

#14 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 optometrist*)

#13 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 ophthalmologist*)

#12 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 orthoptist*)

#11 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 monitor*)

#10 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 referral*)

#9 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 referred*)

#8 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 hospital*)

#7 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 population*)

#6 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 communit*)

#5 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 program*)

#4 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 visual*)

#3 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 vision)

#2 TS=mass screening

#1 TS= vision test

Appendix 8. BIOSIS Previews search strategy

#20 #15 AND #18 AND #19

#19 TS= (infan* or child* or toddler* or boy* or girl* or paediatric* or pediatric* or nursery or nurseries or kindergarten* or pre school* or preschool* or schoolage* or school age*)

#18 #16 OR #17

#17 TS= (esotrop* or exotrop* or hypertrop* or hypotrop* or microtrop*)

#16 TS= (strabimus or strabismic or squint*)

#15 #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 OR #14

#14 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 optometrist*)

#13 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 ophthalmologist*)

#12 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 orthoptist*)

#11 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 monitor*)

#10 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 referral*)

#9 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 referred*)

#8 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 hospital*)

#7 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 population*)

#6 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 communit*)

#5 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 program*)

#4 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 visual*)

#3 TS= ((test* or screen* or diagnos* or assess*) NEAR/4 vision)

#2 TS=mass screening

#1 TS= vision test

Appendix 9. MEDION search strategy

Database was searched on ICPC code field. Using code "f" for ophthalmology.

Appendix 10. ARIF search strategy

strabismus OR amblyopia

Appendix 11. ISRCTN search strategy

(strabismus OR amblyopia) AND (test OR screen OR diagnosis OR assess)

Appendix 12. ClinicalTrials.gov search strategy

(strabismus OR amblyopia) AND (test OR screen OR diagnosis OR assess)

Appendix 13. ICTRP search strategy

strabismus OR amblyopia = Condition AND test OR screen OR diagnosis OR assess = Intervention

Appendix 14. Glossary

Accommodation: mechanism by which an eye focuses on a near object; accommodation involves contraction of the ciliary muscle, which relaxes the fibres holding the lens inside the eye; the lens then assumes a more rounded shape.

Binocular: seeing with both eyes.

Convergent: appearance of one or both eyes deviated inwards/towards the nose.

Cornea: the clear window at the front of the eye.

Cycloplegic: using pharmacological agents (eyedrops) to paralyse the ciliary muscle in the eye to prevent accommodation when carrying out a test for glasses.

Dioptre (D): the unit of measurement describing the optical power of a lens.

Divergent: appearance of one or both eyes deviated outwards/towards the temple.

Esotropia: appearance of one or both eyes deviated inwards/towards the nose; same as convergent.

Exotropia: appearance of one or both eyes deviated outwards/towards the temple; same as divergent.

Fundus: the structures at the back of the eye which change the visual information from light to electrical signals and transmit them to the brain.

Horizontal: ocular misalignment in the horizontal plane, i.e. the eyes are at the same level, but one or both eyes are deviated towards one side.

Hypermetropia: far-sightedness.

Latent: strabismus not present when both eyes are open and fixing on a target, but can be demonstrated by interrupting binocular viewing (seeing with both eyes).

Manifest: strabismus present when both eyes are open.

Myopia: short-sightedness.

Prism dioptre (PD); unit of measurement describing the power of a prism that aligns the eyes.

Refraction: test for glasses.

Strabismus: ocular misalignment.

Vertical: ocular misalignment in the vertical plane, i.e. one eye appears higher or lower than the other.

CONTRIBUTIONS OF AUTHORS

All authors helped to draft the protocol. Sarah Hull and Vijay Tailor reviewed the abstracts provided by the search as well as full articles which potentially met our inclusion criteria. They extracted relevant data, wrote the Results section and updated this article from protocol to full review. Annegret Dahlmann-Noor provided a third opinion on study inclusion where required. Gianni Virgili reviewed the analysis and provided critical review of findings. Jugnoo Rahi and Christine Schmucker critically reviewed the final version of this article.

DECLARATIONS OF INTEREST

The authors have no interests to declare.

SOURCES OF SUPPORT