VISION SCREENING FOR STRABISMUS AND AMBLYOPIA:

A CRITICAL APPRAISAL OF THE EVIDENCE

BCOHTA 95:1D APRIL 1995

B.C. Office of Health Technology Assessment Discussion Paper Series
Vision Screening for Strabismus and Amblyopia:  
A Critical Appraisal of the Evidence
ACKNOWLEDGMENTS

The contributions of the following individuals are gratefully acknowledged:

Internal review committee:

Samuel B. Sheps, MD, MSc, FRCPC

Head, Department of Health Care and Epidemiology
Faculty of Medicine
The University of British Columbia

Arminee Kazanjian, Dr. Soc.

Chair, BCOHTA Steering Committee
Associate Director, Centre for Health Services and Policy Research
Associate Professor, Department of Health Care and Epidemiology
University of British Columbia

Kathryn D. Friesen, MSc

Program Manager, BCOHTA
University of British Columbia

External Reviewers:

Dr. Ross A. Kennedy, MD, CM, FRCS(C)

Clinical Assistant Professor
Dept. of Ophthalmology
Faculty of Medicine
University of British Columbia

Dr. I.B. Pless, MD, FRCP(C)

National Health Scientist
Professor of Pediatrics & Epidemiology
McGill University
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................................................. ii  
FOREWORD .................................................................................................................... iii  
TABLE OF CONTENTS ...................................................................................................... iv  
LIST OF TABLES .............................................................................................................. vi  
EXECUTIVE SUMMARY .................................................................................................. 1  
INTRODUCTION .............................................................................................................. 4  

1. BACKGROUND AND SIGNIFICANCE ........................................................................... 5  
   1.1 Amblyopia ............................................................................................................... 6  
      1.1.1 Description and terminology ......................................................................... 6  
      1.1.2 Natural history .................................................................................................. 6  
      1.1.3 Epidemiology ................................................................................................... 6  
   1.2 Strabismus .............................................................................................................. 8  
      1.2.1 Description and terminology ......................................................................... 8  
      1.2.2 Natural history .................................................................................................. 8  
      1.2.3 Epidemiology ................................................................................................... 8  
   1.3 Treatment of strabismus and amblyopia ................................................................. 9  
      1.3.1 Amblyopia treatment ....................................................................................... 9  
      1.3.2 Strabismus treatment ....................................................................................... 10  
   1.4 Screening recommendations from various organizations .................................. 10  
   1.5 Current practice in British Columbia .................................................................... 12  

2. SCREENING TESTS FOR AMBLYOPIA AND STRABISMUS ....................................... 12  
   2.1 Tests for amblyopia .............................................................................................. 12  
   2.2 Tests for strabismus ............................................................................................. 13  
   2.3 Depth perception (stereopsis) tests ....................................................................... 13  
   2.4 Screening for precursors ..................................................................................... 14  
   2.5 The gold standard ophthalmological examination ............................................. 14
LIST OF TABLES

Table 1  Recommended vision screening by pediatricians and family physicians ................................................................. 11

Table 2a  Summary of research evaluating vision screening (as a clinical intervention): Israel ......................................................... 17

Table 2b  Summary of research evaluating vision screening (as a clinical intervention): U.K. ......................................................... 18

Table 3a  Summary of research evaluating vision screening as a diagnostic procedure: program assessment ........................................... 20

Table 3b  Summary of research evaluating vision screening as a diagnostic procedure: tester assessment ........................................... 20

Table 4  Results of critical appraisal of studies evaluating vision screening ....................................................................................... 21

Table 5  Vision screening for the early detection of strabismus and amblyopia in school-aged children evaluated against ten screening parameters ......................................................... 30
EXECUTIVE SUMMARY

Vision screening programs, conducted both prior to and at the time of school entry remain as
time-honoured efforts for the early detection of potentially vision-damaging ocular disorders,
particularly strabismus and amblyopia. At issue in this report, is whether the available scientific
evidence supports these vision screening efforts.

Definitions:

1) **Amblyopia** is a clinical condition characterized by decreased vision in one or both
    eyes without detectable anatomical damage either in the eye or the visual pathway.

2) **Strabismus** is a manifest deviation of the eyes from a single point of fixation.
    Strabismus is important because it is the most common cause of amblyopia.

The report documents that vision screening programs, although relatively harmless to conduct and
potentially contributing to the discovery of very serious visual disorders, cannot be supported by
direct medium and long-term effectiveness evidence. That is, controlled trials have not
established that children detected through any vision screening program are ultimately better off
because of the screening effort. Instead, evidence of testing validity and treatment efficacy for
vision screening can only be supported by indirect evidence. Indirect evidence means, in this
instance, that causality must be inferred through linking evidence for various program elements,
rather than established through a completed screening program evaluation. The purpose of this
report is to outline and critically examine this indirect evidence.

The report concludes that, while the weight of scientific evidence favours vision screening in
general and at a much earlier age than school entry (preferably under age 3) the evidence does not
establish whether a particular population will be better off with either a school entry or a
preschool program. Population benefit depends on the trade off between treatment effectiveness
and population coverage for any target age. Thus, instituting new or altering old programs
requires empirical study to establish optimum population benefit.

Screening programs for more common visual disorders, primarily refractive errors without
amblyopia, are not considered in this report because earlier intervention in these conditions is
unlikely to influence long-term visual outcome.1,2

The remainder of this summary outlines the literature reviewed and its relevance to the central
question:

**Central question:** Is screening for vision-threatening ocular conditions, primarily amblyopia
and strabismus, supported by direct or indirect scientific evidence?

**Answer for direct evidence:**

No. Controlled trials have not established the overall efficacy or effectiveness of vision screening
programs.
the better eye. Treatment effectiveness evidence from long-term observational studies from England and Israel cited above, for example, show that earlier vision screening, ideally under 3 years of age, results in shorter and more effective treatment for amblyopia. As well, the same studies describe the diminishing benefit from treatment for these screen detected children beginning at age 4 to 5 years.7-10

Indirect evidence also shows that the degree of sophistication of the tester and the ophthalmic equipment required for pre-school vision screening increases rapidly with decreasing age. The primary problem is that amblyopia detection without sophisticated optometric equipment and training involves unilateral acuity testing; testing dependent on children's cooperation and comprehension. Several relatively weak cohort studies have shown that, other than newer automated photographic techniques, the more knowledgeable and skilled the tester, the more effective the program of amblyopia detection.21,22

However, while the weight of scientific evidence, albeit indirect, supports vision screening both at an earlier age than school entry and using sophisticated tests and testers, this does not necessarily mean that a program with these characteristics provides the optimum visual outcome for a particular population. The optimum visual outcome benefit for any population will depend on population coverage, the availability of testers and tests, as well as the likelihood of treatment compliance. Central among these issues for British Columbia is the trade off between vision screening at school entry, when most children are efficiently gathered together, versus screening at an earlier age, when complete population coverage is less likely but treatment is known to be more effective. Thus, it is recommended that any move to eliminate current or establish new vision screening programs should depend on careful empirical evidence from visual outcome program evaluations.
knowledge claims in this area and insight about some of the problems with screening program evaluations.

A review of the current efficacy and effectiveness evidence on vision screening is also reasonable at this time because this evidence is not likely to be clarified in the foreseeable future. On the one hand, the quality of direct evidence of screening program benefits will not likely improve because of the ethical problems associated with conducting a controlled trial. That is, delaying potential vision-saving treatment in a cohort of children, a step necessary for a controlled trial, has rarely occurred in the past and seems increasingly unlikely in the future. As a result no definitive clinical trials designed to measure the effect of screening on long-term visual outcome are known to be underway or are planned. On the other hand, the most valid, albeit mostly indirect observational evidence comes from studies conducted more than a decade and often two decades ago, perhaps when families were more likely to remain in one district long enough for adequate long-term assessments. The problems associated with mobility of populations is, if anything, worsening. Nevertheless, while clinical trials will likely remain unethical and any study will likely prove difficult longitudinally, carefully-conducted cohort studies could be used to study vision screening programs. For example, this design could be used to compare existing with innovative programs in single or among matched populations in the same or among different districts over time.

In summary, the overall conclusion that vision screening programs have not been proven beneficial using scientifically-sound research remains solid and well-accepted in the literature found in various Western countries. Therefore, knowledge claims regarding and recommendations for vision screening programs from any professional group are based on opinion, not long-term outcome evidence.

Our recommendation, therefore, is that, if existing programs are to continue or new programs are planned, they should only be instituted as fully-evaluated pilot projects, using the best study design possible, in order to assess particular testers, tests and treatments.

With any change in the provision of vision screening in the province, a whole new series of public policy problems will likely arise; not the least of which is whether public health or primary care practitioners should be responsible for providing and/or administering vision screening programs.

1. BACKGROUND AND SIGNIFICANCE

Amblyopia and strabismus are closely related pathogenetically and clinically. For example, on the one hand, poor visual acuity due to amblyopia in one eye can result in manifest deviation of that eye from normal alignment and result in the clinical diagnosis of strabismus. On the other hand, prolonged deviation of one eye from normal alignment could result in cortical suppression of the visual input from that eye which could become amblyopic. In addition, the presence of either amblyopia or strabismus can result in a loss of binocular vision, also known as fine depth perception or stereopsis.

The close inter-relationship between three eye parameters -- acuity, alignment, and stereoaucity -- have made natural history and epidemiological studies particularly difficult. That is, not only are the conditions often coincidental but they may also have codetermined etiology. Thus
(ii) anisometropia - inability to focus eyes simultaneously causing suppression of the image in one eye; associated with hyperopia.

(iii) large ametropia - both eyes may become amblyopic because of a failure to form a focused image in either eye.

(iv) occlusion amblyopia - for example, complete blepharoptosis or capillary hemangioma.

The epidemiology of amblyopia remains particularly problematic because this term remains too vague for rigorous assessment of even the most basic disease parameters. Hillis\textsuperscript{17} provides a detailed analysis of the problems associated with epidemiologic amblyopia research.

Amblyopia is a common problem, with a generally-accepted prevalence estimate of 2\%.\textsuperscript{17} Oliver \textit{et al.}\textsuperscript{15} provide one of the few population-based estimates of amblyopia prevalence at different age groups:

<table>
<thead>
<tr>
<th>Age Group (years)</th>
<th>Prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5 - 2</td>
<td>0.4</td>
</tr>
<tr>
<td>2 - 3</td>
<td>0.6</td>
</tr>
<tr>
<td>3 - 4</td>
<td>1.0</td>
</tr>
<tr>
<td>4 - 5</td>
<td>2.1</td>
</tr>
<tr>
<td>5 - 6</td>
<td>1.6</td>
</tr>
</tbody>
</table>

These results suggest a tendency for prevalence to increase with age. However, this tendency likely reflects nothing more than the difficulty of diagnosing children under age 3 years.

Prevalence estimates in an untreated population of 4 to 6 year old children range from .8 to 5\%, with the most widely cited estimate of 2\% (for discussion see Hillis\textsuperscript{17} and von Noorden\textsuperscript{34}). These authors suggest that the wide discrepancy is likely the result of variation in screening program sensitivity rather than actual variation in population prevalence.

Prevalence calculations for amblyopia are extremely problematic and must be viewed with caution especially for children younger than 3-4 years of age. This is in part due to variations in disease definitions and testing methods. The major problem, however, is that for children in this age group, acuity testing is often very difficult (best described by Ingram;\textsuperscript{35} also Campbell\textsuperscript{36}). For example, at the time of detection of visual impairment, it is difficult to distinguish sensory cortical suppression from difficulties with testing younger children, either with or without corrective lenses. Moreover, even after treatment of suspected mild to moderate amblyopia, improved vision in the most affected eye often remains suspect.

Amblyopia is most commonly associated with strabismus. For example, in one unselected group of 38,000 children screened between 1-1/2 and 3 years of age, 41\% of the 1.3\% of children with strabismus had amblyopia.\textsuperscript{13}

At all ages infants with unilateral strabismus had a higher prevalence of amblyopia than those with alternating strabismus.\textsuperscript{37}
Acquired esotropia, the most common form of strabismus, usually develops between 1 1/2 and 3 years of age, but may develop in younger and older children.\(^1\)

Incidence falls off rapidly after age 5. For example, only two new cases of strabismus were found in a study of 1530 seven-year-old children who had previously been examined at age four.\(^42\)

### 1.3 Treatment of Strabismus and Amblyopia

Treatment of strabismus and amblyopia are considered separately in this section, although in practice these conditions often are treated concurrently.

#### 1.3.1 Amblyopia Treatment

The goals of amblyopia treatment are to acquire normal vision in both eyes, normal eye alignment, and ultimately normal depth perception. The main principle guiding treatment is to improve use of the affected eye by impeding use of the better eye. This is usually accomplished by occlusion, but other treatments are employed including penalization.\(^{43,44}\) Occlusion and penalization are described as follows:

Occlusion is usually achieved using an adhesive patch. It may be full or part-time. If full-time occlusion is used, care is taken that treatment amblyopia does not occur in the better eye. Occlusion is usually maintained until vision becomes normal, or until visual acuity is stable for at least 3 months. Further occlusion may be required due to the fact that, in 50% of cases, deterioration in acuity occurs.\(^{41}\)

Penalization, a relatively uncommon technique, involves blurring the vision in the better eye using atropine or blurring lenses alone or in combination. Penalization is usually used if occlusion fails or for maintenance therapy.\(^1\)

The greatest clinical difficulty with amblyopia therapy is deciding when to stop. This is particularly true in younger children where testing visual acuity is problematic. Thus, clinicians have difficulty determining if visual acuity has improved to a maximum, or alternately, if treatment is stopped, whether visual acuity is degenerating.

Acuity is usually treated before strabismus in part because it can be accomplished immediately in the office with corrective lenses. Strabismus, in contrast, is usually delayed due to complicating factors such as, in the Canadian context, hospital operating room schedules. There are also instances where improvements in acuity may lead to better alignment.\(^{32}\) Subsequent treatment may require correction of residual strabismus.

Many ophthalmology practice-based descriptive studies show dramatic amblyopia treatment benefit. Epelbaum et al, in a typical retrospective chart review of 407 patients with strabismic amblyopia,\(^{19}\) noted that recovery of acuity of the amblyopic eye was maximal when occlusion was initiated before age 3 and decreased as a function of age until it was null by 12 years:
2) The 1988 U.S. Preventative Services Task Force recommends that:
"vision screening is recommended for all children once before entering school, preferably at age 3 or 4. Routine vision testing is not recommended as a component of the periodic health examination of asymptomatic schoolchildren."48

3) The Vision Advisory Committee to the BC. Ministry of Health, a committee including representatives from both ophthalmology and optometry recommends that:
"that all children have an eye examination by an eye doctor (ophthalmologist or optometrist) by age 3 years." (personal communication)

Table 1
Recommended Vision Screening by Pediatricians and Family Physicians

<table>
<thead>
<tr>
<th>Age</th>
<th>Examination</th>
<th>Referral Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newborn</td>
<td>Penlight examination of cornea</td>
<td>Any ocular pathology</td>
</tr>
<tr>
<td></td>
<td>Nystagmus</td>
<td>Rule out nystagmus</td>
</tr>
<tr>
<td></td>
<td>Red reflexes</td>
<td>Abnormal red reflexes or white reflex</td>
</tr>
<tr>
<td>By 6 months</td>
<td>Objection to occlusion</td>
<td>Fixation to light or small toys</td>
</tr>
<tr>
<td></td>
<td>Penlight examination</td>
<td>Nystagmus; any ocular pathology</td>
</tr>
<tr>
<td></td>
<td>Strabismus</td>
<td>Corneal light reflex test, cover text</td>
</tr>
<tr>
<td></td>
<td>Red reflexes</td>
<td>Abnormal red reflexes or white reflex</td>
</tr>
<tr>
<td>3.5 years</td>
<td>Visual acuity</td>
<td>Acuity of 20/40 or less in one or both eyes</td>
</tr>
<tr>
<td>5 years or older</td>
<td>Visual acuity</td>
<td>Acuity of 20/30 or less in one or both eyes</td>
</tr>
<tr>
<td></td>
<td>Strabismus</td>
<td>Corneal light reflex test, cover text</td>
</tr>
<tr>
<td></td>
<td>Fundus examination</td>
<td>Any ocular pathology</td>
</tr>
<tr>
<td></td>
<td>Fundus examination</td>
<td>Any ocular pathology</td>
</tr>
</tbody>
</table>

(After Crouch and Kennedy)49:196

4) The BC Association of Optometrists, in conjunction with the Canadian Association of Optometrists (COA):
"encourages the implementation of effective vision screening programs for young children. The earlier in life such programs can be applied the greater the potential benefit to the child. Irrespective of the performance of vision screening programs,
usually illiterate, children some form of "E" game or letter matching game (HOTV) is used. The child is asked to point to appropriate letters on a near card, or in the direction that the E is pointing. The "E"s and letters are charted in diminishing sizes. These tests are generally accepted as having high specificity, above 95%, however their sensitivity is low at 30%. Testing for amblyopia, therefore, often involves additional tests for stereoscopic vision, which are discussed below.

2.2 Tests for Strabismus

The most common screening tests for strabismus are the Hirschberg light reflex and the Cover/Uncover tests. The Hirschberg test simply involves shining a light towards the eyes of a child and observing the symmetry of light reflected from the cornea. While simple to perform and requiring little child cooperation, it is considered difficult to interpret and is seldom used alone.

For the Cover/Uncover test, the child fixates on an object while the examiner alternately covers each eye. A movement of the uncovered eye indicates strabismus. MacFarlane, for example, reports typical Cover/Uncover efficacy of: sensitivity 43.5%, specificity 100%, and positive predictive value 100%. Others have confirmed the low test sensitivity of the Cover/Uncover test. The problem is that only about 50% of strabismus is grossly visible using the crude screening tests such as Cover/Uncover tests. Detection of what are sometimes termed "micro-strabismic" children requires more sophisticated ophthalmic equipment and training.

2.3. Depth perception (stereopsis) tests

Stereoscopic vision is one of the desired functional end points of normal bilateral visual acuity and alignment. The Titmus Stereofly and the Random Dot E tests are two of the more commonly-used stereoscopic vision tests. In both instances, a child wearing polarized glasses is asked to identify objects or letters on cards visible only with symmetrical optical images.

Testing for this "higher level" vision function is considered more efficacious than simple acuity tests for the identification of strabismus and amblyopia. The advantages cited include relative simplicity, reliability, and the fact that no verbal response is required. Since the introduction of these tests in the 1970s, they have often been proposed as the method of choice for mass screening. However, evaluation has resulted in concerns regarding efficacy primarily because the "pass" criteria can be varied, particularly with the Titmus test, resulting in variable sensitivities and some difficulty in comparing rates from different studies. Testing children under three years old has generally been unsuccessful due to the child's inability to cooperate.

The Random Dot E test and the Titmus Stereofly are more effective than visual acuity tests in detecting strabismus, amblyopia or amblyogenic conditions in preschool children. Early efficacy studies showed the Random Dot E test more sensitive than the Titmus Stereofly although the false negatives are still considered unacceptably high.

Controversy remains, however. Hope, for example, compared two populations of children, one at school (n=100, ages 5 to 15 years) and one in preschool (n=168, ages 3 to 4 years). In the preschool group, the test was found to have an unacceptably low positive predictive value (17%).

B.C. OFFICE OF HEALTH TECHNOLOGY ASSESSMENT
Vision Screening
3. APPRAISAL OF RESEARCH REPORTING PRIMARY DATA

3.1 METHODOLOGY

3.1.1 Search Strategy

Electronic searches were conducted using the National Library of Medicine MEDLINE (1985 to present), HEALTH (1975 to present), HSTAR, CIJE, RIE, Nursing (1983 to present) and PSYCH INFO. Key search terms used were "school health services, vision screening, vision tests, program evaluation, amblyopia, strabismus, therapy and treatment." The search was limited to English language journals and studies with humans. Current content database searches were continued until the time of publication.

A fugitive or non-electronic search was also conducted. Material that was examined included reference lists in published articles, reports from other technology assessment offices, and other agencies involved in systematic review of clinical evidence. Finally, clinical practice input was solicited from a practicing orthoptist providing vision screening, optometrists, a pediatric ophthalmologist, and representatives from the BC Optometry Association. Optometrists were particularly helpful in providing reference material regarding epidemiology, treatment effectiveness, and screening program research.

3.1.2 Inclusion Criteria and Exclusion Criteria

Studies included in the critical appraisal process were those reporting primary data evaluating screening for early detection of vision threatening ocular disorders in a geographically defined pediatric population. The screening function was defined in the broadest sense to mean testing to separate children into disease or risk categories. Evaluation was loosely defined as an effort to compare a screened to an unscreened population or to follow an entire screen positive and/or screen negative population over time.

The requirement that a study evaluates screening resulted in elimination of descriptive studies, also known as case series or cross-sectional surveys. This meant excluding, from critical appraisal, most of the published literature including the recent report on school vision screening in British Columbia. Similar program descriptions have been noted for other areas of Canada including Ontario, Manitoba, and New Brunswick as well as other countries such as the United Kingdom and the United States. The bibliography lists approximately 50 observational studies that were systematically reviewed but excluded from the critical appraisal process.

The requirement that the study evaluated screening of a geographic population reflects the orientation of this report towards assessing primary or mass screening. Excluded are studies of children already within the health care system such as ophthalmologic or optometric clinic-based populations.

Background material on pathology, epidemiology and treatment was also reviewed, as were a number of reports providing expert clinical opinions.
Table 2a. Summary of research evaluating vision screening as a clinical intervention: Israel

<table>
<thead>
<tr>
<th>Study/Setting</th>
<th>Population/Criteria</th>
<th>Age range</th>
<th>Study design</th>
<th># enrolled</th>
<th>Screen test/testers</th>
<th>Prelim. findings</th>
<th>Gold standard Test/testers</th>
<th>Diagnosis</th>
<th>Treatment / duration</th>
<th>Outcome</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neumann 1971 Pilot Project</td>
<td>8000 offered</td>
<td>1 to 2-1/yr. old</td>
<td>Prospective Descriptive Focus on Strabismus and Amblyopia</td>
<td>6400</td>
<td>Orthoptists: Observation Light Reflex Cover test</td>
<td>342 failed</td>
<td>Opth.: &quot;Full&quot; exam No details</td>
<td>166 treated 133 (2.1%) S 104 new 29 known 33 other PPV (all) - 48% PPV (S) - 39%</td>
<td>28 - CUE, assumed A</td>
<td>21 of 22 became alternators 6 lost to follow-up</td>
<td>See Neumann 1987 for long-term outcome</td>
</tr>
<tr>
<td>Neumann 1987 Follow-up</td>
<td>Consecutive cases of CUE</td>
<td>7 mo. to 3 yrs. (initial screen)</td>
<td>Prospective Analytic</td>
<td>19,000</td>
<td>See above</td>
<td>97 failed</td>
<td>Opth: acuity, stereo acuity, alternation</td>
<td>97 CUS 78 esotropia and followed to age 6 No initial VA</td>
<td>Glasses Occclusion Minimum 3-6 yrs. of the diagnosed over age 2 yrs. - 50% had VA 6/12 or better</td>
<td>Convenience sample 19/97 children lost to follow-up</td>
<td></td>
</tr>
<tr>
<td>Friedman 1980 Child Welfare Clinics</td>
<td>98% pop.</td>
<td>1 to 2-1/yr. old</td>
<td>Prospective Analytic</td>
<td>38,000</td>
<td>Orthoptists (all inspection light reflex cover text)</td>
<td>2009 failed</td>
<td>Opth: full exam no details Includes refraction S - (498) 1.3% CUS - (208) 0.5% (Amblyopia assumed) Other (78) .2%</td>
<td>Not described</td>
<td>See Friedman 1985</td>
<td>Program description</td>
<td></td>
</tr>
<tr>
<td>Friedman 1980 Results of Screen Detection Non-Strabismic Amblyopia Marked amblyopia followed to 7-10 yrs.</td>
<td>1 to 2-1/yr. old</td>
<td>1974-77</td>
<td>Prospective Analytic</td>
<td>39/15,084</td>
<td>See Friedman 1980 Rapid retinoscopy screening</td>
<td>See Friedman 1980</td>
<td>Opth.</td>
<td>A - 6/12 or less, or difference of 2 lines on eye chart</td>
<td>Until age 7 Glasses Penalization</td>
<td>20 (61%) residual A 6 (15%) bilateral A 11 (35%) A 6/12-6/15 7 (17%) A 6/20-6/24 2 (5%) A 6/90 8 full stereoptes</td>
<td>Biased by required follow-up length</td>
</tr>
</tbody>
</table>

**LEGEND:**
A = Amblyopia
S = Strabismus
CUE = Constant Unilateral Esotropia
CUS = Constant Unilateral Strabismus
GP = General Practitioner
VA = Visual Acuity
6/6 = normal acuity; 6/9 = 2/3 normal; 6/12 = 1/2 normal, etc.
PPV = Positive Predictive Value
### Table 3a. Summary of research evaluating vision screening as a diagnostic procedure: Program assessment

<table>
<thead>
<tr>
<th>Study/Setting</th>
<th>Criteria/Population</th>
<th>Age range</th>
<th>Study design</th>
<th># % enrolled</th>
<th>Screen test/testers</th>
<th>Prelim. findings</th>
<th>Gold standard Test/testers</th>
<th>Diagnostic outcome</th>
<th>PPV/NPV</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feldman 1980 Canada Preschools &amp; schools</td>
<td>Representative samples of geographic areas</td>
<td>5-1/2 yrs</td>
<td>Prospective cohort</td>
<td>78 (10%) failed</td>
<td>Ophthalmic exam</td>
<td>No results</td>
<td>N/C</td>
<td>Best prospective cohort study</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>matched:</td>
<td>6-12 months earlier</td>
<td>19 (2.5%) glasses 24 (3%) VA 20/50</td>
<td>Ophthalmic exam</td>
<td>N/C</td>
<td>Best prospective cohort study</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>socio-economic status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kohler and Stigmar G. 1973/1978 Sweden Community-based School Setting</td>
<td>All children starting school in one city over 3 year period</td>
<td>7 yrs.</td>
<td>Prospective cohort</td>
<td>112 (15%) failed</td>
<td>Ophthalmic exam</td>
<td>No results</td>
<td>N/C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>experimental group</td>
<td>12 (1.6%) glasses 43 (6%) VA 20/50</td>
<td>Ophthalmic exam</td>
<td>N/C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>school nurses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Linear E chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Failed: acuity &lt; .9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>control group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unscreened</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Becher 1992 Canadian Provincial (Nova Scotia)</td>
<td>Random representative sample from entire province</td>
<td>4-1/2 to 5-1/2 yrs</td>
<td>Prospective Analytic</td>
<td>802 (6.8%) total 200 screen positive, tive randomly-selected</td>
<td>Orthoph. and Ophthalmic test 157 (76.5%) seen</td>
<td>11 (7%) fail 146 (93%) pass 43 not seen</td>
<td>NPV 97.6% Vision-threatening conditions 1 ambyopia missed</td>
<td>Result may be high 25% sample not tested</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P.H. Nurse HOTV Stereo-acuity test</td>
<td>36 (80%) seen</td>
<td>18 Fail 18 Pass</td>
<td>PPV 50% Result suspect 20% not tested</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11,814</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LEGEND:**
- **VA** = Visual Acuity
- **NC** = Not Calculated
### Table 4. Results of critical appraisal of studies evaluating vision screening

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Representative and generalizable to screening population</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>2. Community setting</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>3. Specified and appropriate inclusion/exclusion criteria</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>4. Adequate gold standard control</td>
<td>G</td>
<td>P</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>5. Adequate description of tests</td>
<td>G</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>P/G</td>
<td>G</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>7. All subjects eligible and enrolled, accounted for</td>
<td>P</td>
<td>G</td>
<td>G</td>
<td>P</td>
<td>G</td>
<td>F</td>
<td>P</td>
<td>F</td>
<td>P/F</td>
<td>F</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>8. Appropriate calculations of sensitivity and specificity</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>9. Appropriate calculations of prevalence, positive and negative predictive values</td>
<td>G</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>G</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>F</td>
</tr>
<tr>
<td>10. Appropriate use of statistical tests</td>
<td>G</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>G</td>
<td>NC</td>
<td>NC</td>
<td>NC</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

G = good (criterion was adequately met), F = fair (criterion was inadequately met), P = poor (criterion was not met), NC = not calculated
applied to the correct individuals is a necessary component of an assessment of screening initiated treatment efficacy and effectiveness.

Researchers used both intervention and diagnostic study designs to assess vision screening. Studies using an "intervention" design are listed in Table 2a and b; 2a for Israel, and 2b for the United Kingdom. Studies using a "diagnostic" design are listed in Table 3a and b; 3a for program assessments and 3b for tester assessments.

The actual study quality using these designs, while variable, was quite poor. Moreover, the problems with study quality were reciprocal. That is, on the one hand, while the intervention studies followed children over time, these children were not assessed relative to an explicit or a historical control group. On the other hand, the diagnostic studies, three of which included some form of control group, did not follow children long enough to adequately compare treatment outcome for the two groups.

3.2.2 Evidence of Diagnostic Efficacy

Three research teams studying diagnostic efficacy removed most major sources of bias by including a control group in their study design. Incidentally, as mentioned above, all three of these studies examined children at the time of school entry.

In two studies, a previously-preschool-screened population was re-examined at school entry and compared to a cohort of previously unscreened children.

Feldman et al provided stronger evidence because the two cohorts were matched for age, sex, and socio-economic status. In this study, a pediatric nurse tested children at school entry (mean age of 65 months) using the "illiterate E" visual acuity test. They showed a 50% lower overall prevalence of untreated eye disorders and 79% fewer moderate to severe vision problems in the previously screened group.

Kohler et al, using an unmatched and poorly-described comparison group, also showed that "the risk of finding a significant eye disorder in a school entrant was more than 6 times greater for a child who was not examined in his preschool years, and the risk of finding an amblyotic child was more than ten times greater."

A third study, de Becker et al, although using a different study design, also provided evidence that their screening program of 4-1/2 to 5-1/2 year old children accurately detected visual disorders. These researchers randomly selected 200 screen negative children and subjected 78.5% of them to gold standard ophthalmologic examination. The study, which was designed to calculate the negative predictive value of their screening program, showed incidentally that their program had greatly reduced the prevalence of vision threatening ocular disorders in the screen negative population. This was shown by the significant decrease in prevalence of visual disorders among the 200 randomly selected screen negative children.

The explicit end-point of most diagnostic studies evaluating vision screening (Tables 3a and 3b) was the number of cases of visual disorders detected, both minor refractive errors, and the more
children and no one achieved normal vision. The relatively poor outcomes shown for these usually severe forms of amblyopia is consistent with other studies of non-screen detected children with these forms of amblyopia. Thus, some severe forms of amblyopia, even when discovered through screening at one year of age, are largely resistant to treatment.

Second, long-term visual outcome of screen detected constant unilateral strabismus cases, presumably with amblyopia, was explicitly studied.12,13 These studies are best considered together as they are based on the same screen detected population. The authors report the visual outcome for 78 of 93 screen detected cases of unilateral esotropia detected when the children were seven months to 3 years of age and followed until they were at least 6 years. While only final, not initial, acuity was reported, 50% achieved near normal and 67% acceptable (at least 50% of normal) vision in the strabismic eye.

In addition to the two series mentioned above, long-term visual outcome of screen detected disorders including both strabismic and straight eyed amblyopia was also studied by Ingram et al19 and Oliver and Nawratzki.15,16 Ingram et al followed the children initially screened at age 3 1/2 until at least age 5 and some children until age 11. They found that, while acknowledging the difficulties of diagnosis and treating amblyopia, "most children achieved normal corrected visual acuity."

Oliver and Nawratzki, in a methodologically much weaker study,15,16 found that 67% of the 1 1/2 to 6 year old children with screen detected amblyopia achieved "good visual acuity."

In summary, without cohort studies, the issue that remains is to what extent these screen detected children have better ultimate visual outcome due to earlier detection and treatment of amblyopia.

3.2.4 Evidence relating screening age to long-term visual outcome

Several studies explicitly examined the influence of screening age on diagnostic efficacy and long-term visual outcome. Two study types are identified: (1) studies that show lack of benefit from older, school entry, screening programs; and (2) studies that show increasing benefit from screening younger children.

Ingram et al7,8 found vision screening at the time of school entry provided no improvement in "the incidence or severity of amblyopia 3 or more years later."8:236 Although they were only descriptive studies, they are included here because a screen detected population was followed for a significant length of time. These studies found only 15 of the 120 screen detected children had improvement. As the authors note, without a control group the amount of deterioration prevented cannot be estimated. Ingram et al,10 in a methodologically-similar study, reports that visual acuity improved two lines on a standard eye chart in only 16 children out of the 108 children identified through school entry vision screening.

There is weak to moderate evidence to support the notion that significantly greater benefit is achieved if treatment is undertaken before three years of age, and perhaps as young as 18 months old. Neumann et al12 provide the best evidence of treatment benefit of children detected under
4. DISCUSSION OF FINDINGS

Vision screening programs, while sharing the common goal of facilitating normal binocular visual development, vary in almost every other sense, including tests, testers, and testing frequency, as well as referral patterns and treatments applied. The almost complete lack of standardization, particularly the lack of treatment outcome definitions, makes programs very difficult to evaluate and compare. Nevertheless, some general observations can be made and are included in the sections that follow.

4.1 The assumption of benefit from ophthalmologic interventions

A basic assumption underlying this research is that virtually all ophthalmologic interventions designed to diagnose and treat amblyopia and strabismus provide at least some benefit and relatively little harm. This assumption was based on various factors:

First, diagnostic investigations following screening tests require no invasive tests and can be repeated as often as necessary. In particular, false positive tests do not lead to vision threatening or patient damaging treatments. However, false positive tests do have significant financial and psychological cost to children and their families.

Second, it is assumed that some treatment benefit occurs even in the circumstance where visual acuity is only marginally improved, or degeneration does not occur, in the most amblyopic eye and strabismus correction is only for cosmetic reasons. This assumption can be made for amblyopia because treatments involve relatively harmless, if carefully monitored, occlusion therapy of the better eye. Surgical correction of alignment, however, is not without anesthetic and technical risks.

A high or even marginal benefit to harm treatment ratio, however, itself provides no justification for the screening effort. A positive benefit/harm ratio is merely a necessary, not a sufficient condition to support these programs. Screening effort justification rests on whether earlier treatment in the natural history of these disorders improves the benefit to harm ratio.

4.2 Data synthesis to reach overall conclusions

In the absence of direct evidence of overall vision screening program benefit using a controlled trial study design, it has been necessary to critically appraise indirect evidence of vision screening efficacy. This has meant examining studies of the various elements of a vision screening program: tests, testers, diagnostic accuracy, and treatment efficacy. One other possible method to evaluate screening program efficacy is to combine studies examining the same population over time.

Two series of studies, one from the United Kingdom7-10 and one from Israel,11-14 provide this opportunity. In both instances, combining the studies allows a cohort of children, screened from an acceptably-representative geographic population, to be followed reasonably long-term.

The strengths and limitations of the individual studies were detailed in the preceding section. Overall, the studies were poor and the evidence for vision screening was weak. The evidence was
5. Summary and Conclusion

Table 5 summarizes this vision screening assessment in relation to the 10 screening parameters outlined by Wilson and Jungner. Vision screening, which shifts treatment to the latent phase of childhood visual disorders, remains largely unproved by direct experimental evidence. Direct evidence requires an ethically unacceptable controlled trial in which children are randomly assigned to and maintained in screened and non-screened groups. However, other research designs did provide some indirect and partial evidence that vision screening results in at least moderate patient benefit and very little harm. The following points are also worth noting:

a) An adequate positive predictive value can be achieved for amblyopia and strabismus across a wide range of tests, testers and age groups.

b) All studies reviewed here, while methodologically weak, consistently show, except in the most severe forms of amblyopia, improvement in visual acuity from the time of diagnosis.

c) No studies showed evidence of physical harm to visual functioning.

d) School entry screening showed the weakest, while age 3 to 4 screening showed the best evidence of treatment efficacy and effectiveness.

Therefore, it seems reasonable to assume that moving the diagnosis back in the natural history through screening efforts is justified particularly in the preschool age groups.

Available evidence points to younger ages as the optimal time to screen for potential vision-threatening ocular conditions. School entry vision screening as a program directed at older children is more difficult to justify in terms of the scientific evidence. However, it may be that, in a particular population and setting, comprehensive school entry screening may in fact provide better overall population benefit than more limited screening at an earlier age. Similar to almost any other aspect of this area of medicine, the cost/benefit of eliminating and/or substituting vision screening programs requires empirical study.
APPENDIX A
Data Synthesis of Israeli and United Kingdom Studies

In the following analysis, studies are grouped according to country: six from Israel and four from the United Kingdom.

Israel

Four of the six Israeli studies reported prospective observational results of a vision screening program for children age 1 to 2 1/2 years, tested primarily by orthoptists and ophthalmologists visiting Child "Welfare" Clinics in Haifa, Israel over a 15 year period, beginning in 1968.

The initial report, and more completely a subsequent report, describe the screening program including the population screened, the examiners involved, and the techniques employed. The initial program reported in these studies was almost exclusively suited to detection of strabismus and used the following techniques:

1) history taking from parents
2) external inspection of the eyes
3) corneal reflex test
4) the cover test

Neumann et al reported on, among other things, the strabismus and strabismic amblyopia treatment results from screening 6400 children during the first 18 months of the program. Of the 342 screen positive children, 133 (2.1%) were found to have strabismus, 29 of which were already under care. Short term treatment efficacy is described for the 28 children that had previously undiagnosed constant unilateral strabismus. Constant unilateral strabismus was assumed associated with amblyopia since amblyopia testing is not possible in this age group. Twenty-one of the 22 children, followed for four to nine months became "alternators," meaning their eyes did not shift during the cover test.

The subsequent report presents more complete results of the first eight years of this screening program involving 38,000 children which was 98% of the children born in this area during the study period. In the larger study, 498 children (1.3%) had constant unilateral strabismus and presumably amblyopia. The long-term visual outcome and stability of children found to have constant unilateral strabismus in the earliest screen detected cohort (1969-72) were reported in Nuemann et al. A total of 78 of the 97 screen detected cases remained in attendance until at least the age of six years.

The researchers also assessed long term stability of treatment by examining 27 of the children 4 to 10 years following the end of treatment in their eye clinic. Treatment included a combination of glasses, occlusion of the fixating eye, the use of atropine drops, as well as surgery. They found that 82% of the children achieved normal or near normal visual acuity (6/12 or better). Furthermore, they found 59% retained their vision when examined 4 to 10 years later. No children developed full stereoscopic vision. While the study conclusions regarding the optimal age for screening are invalid because of the significant number of children lost to follow up, the
Ingram further illustrated the limitation of school entry vision screening programs. This study was a retrospective review of the 108 cases referred to ophthalmologists from school vision screening in Kettering from 1976 to 1978. Referral criteria were visual acuity less than 6/18 with or without strabismus. These years were chosen because 1976 was the first year records were kept and after 1978 many of the children in this district were prescreened near age one with refraction. The cases were drawn from seven general medical practices around a hospital referral eye clinic. The purpose was to compare initial with final acuity of the worst eye from the time of referral to last known acuity. Treatments were not specified in detail, rather they were referred to as the "usual" correction of refractive errors with glasses, as well as occlusion and surgery for amblyopia and strabismus when "advisable and practical". Duration of follow-up was not specified. Improvement of more than two lines on a Snell or Sheridan-Gardiner chart was found in 16 children. Thirteen of 18 children with severe amblyopia did not improve. Ingram estimates 900 children were born each year in the practices from which the sample was derived. He notes that only 16/900 improved (0.6%). The study does not indicate how many of these 900 children were vision tested in school.

Ingram et al provide a prospective analysis of a screening program for 1507, 3 and 1/2 year old children, followed until 5 to 11 years old. This represented between 70 to 75% of the population in this district. Screening at this age detected 41% of visual defects. Twenty-six percent had been detected earlier and 33% presented after 3 and 1/2 years old. Treatment included corrective lenses and occlusion therapy. Decision regarding treatment was very problematic for ophthalmologists because of difficulties distinguishing blurred vision from either uncorrected errors of refraction or amblyopia requiring occlusion. With ambiguous acuity in 50% of cases, improvement could not be assessed. In addition, failure of occlusion to result in improvement could not be concluded because the child may not have needed the occlusion in the first place. Corrective lenses were very successful but only 3 of 16 children identified as having decreased acuity at the initial age achieved normal vision. It was noted that 9 of 10 children with severe amblyopia presented prior to screening and never benefited from any therapy. The authors conclude, based on this and other work, that therapy seems to depend on the presence or absence of hypermetropia at age 1 rather than timing of diagnosis and therapy.

Discussion

Despite the absence of a control group, these reports collectively provide some evidence of population based screening program benefit because they show improved visual outcome in a substantial portion of the screen detected children. As mentioned in section 1.1.2, despite lack of natural history evidence, expert opinion considers improvement extremely unlikely in the absence of treatment. Furthermore, because of the subtlety of the amblyopia and strabismus found, detection of these conditions is considered very unlikely outside of these vision screening programs. Indeed, as the authors note, their detection would likely otherwise occur much later in the natural histories of the conditions.


20 Ingram RM, Walker C. Refraction as a means of predicting squint or amblyopia in preschool siblings of children known to have these effects. Br J Ophthalmol 1979;63:238-42.


Ingram RM, Walker C. Refraction as a means of predicting squint or amblyopia in preschool siblings of children known to have these effects. Br J Ophthalmol 1979;63:238-42.


