Comparison of Preschool Vision Screening Tests as Administered by Licensed Eye Care Professionals in the Vision in Preschoolers Study

*The Vision in Preschoolers Study Group**

Purpose: To compare 11 preschool vision screening tests administered by licensed eye care professionals (LEPs; optometrists and pediatric ophthalmologists).

Design: Multicenter, cross-sectional study.

Participants: A sample (N = 2588) of 3- to 5-year-old children enrolled in Head Start was selected to over-represent children with vision problems.

Methods: Certified LEPs administered 11 commonly used or commercially available screening tests. Results from a standardized comprehensive eye examination were used to classify children with respect to 4 targeted conditions: amblyopia, strabismus, significant refractive error, and unexplained reduced visual acuity (VA).

Main Outcome Measures: Sensitivity for detecting children with ≥1 targeted conditions at selected levels of specificity was the primary outcome measure. Sensitivity also was calculated for detecting conditions grouped into 3 levels of importance.

Results: At 90% specificity, sensitivities of noncycloplegic retinoscopy (NCR) (64%), the Retinomax Autorefractor (63%), SureSight Vision Screener (63%), and Lea Symbols test (61%) were similar. Sensitivities of the Power Refractor II (54%) and HOTV VA test (54%) were similar to each other. Sensitivities of the Random Dot E stereoacuity (42%) and Stereo Smile II (44%) tests were similar to each other and lower ($P<0.0001$) than the sensitivities of NCR, the 2 autorefractors, and the Lea Symbols test. The cover–uncover test had very low sensitivity (16%) but very high specificity (98%). Sensitivity for conditions considered the most important to detect was 80% to 90% for the 2 autorefractors and NCR. Central interpretations for the MTI and iScreen photoscreeners each yielded 94% specificity and 37% sensitivity. At 94% specificity, the sensitivities were significantly better for NCR, the 2 autorefractors, and the Lea Symbols VA test than for the 2 photoscreeners for detecting ≥1 targeted conditions and for detecting the most important conditions.

Conclusions: Screening tests administered by LEPs vary widely in performance. With 90% specificity, the best tests detected only two thirds of children having ≥1 targeted conditions, but nearly 90% of children with the most important conditions. The 2 tests that use static photorefractive technology were less accurate than 3 tests that assess refractive error in other ways. These results have important implications for screening preschool-aged children. *Ophthalmology 2004;111:637–650 © 2004 by the American Academy of Ophthalmology.*

Amblyopia, strabismus, and significant refractive error are the most prevalent vision disorders of childhood.[1–5] Children with these conditions may benefit from early detection to allow treatment or to increase the frequency of follow-up eye care. Currently, there is considerable controversy concerning the best way to identify 3- to 5-year-old children with these conditions.[3,6,7] Guidelines requiring a comprehensive eye examination for each preschool-aged child have been adopted in some areas,[7,8] whereas others believe that screening may be a cost-effective method for identifying children who would benefit most from a comprehensive eye examination.[1,4–6,9–12] However, developing a rational approach to identifying preschool children with these conditions requires knowledge of the effectiveness of screening tests.[13]

To be effective, screening tests must be able to be administered successfully to a high proportion of preschool children (high testability), be able to identify a high proportion of children who have a vision disorder (high sensitiv-

Originally received: December 4, 2003.
Accepted: January 20, 2004. Manuscript no. 230828.

Supported by grants from the National Eye Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland: U10EY12534, U10EY12545, U10EY12547, U10EY12550, U10EY12644, U10EY12647, and U10EY12648.

The results and conclusions presented in this article represent the opinions of the authors and not necessarily those of the administrators, supervisory boards, governmental agencies, or any other entity operating a Head Start school that participated in the Vision in Preschoolers Study.

Reprint requests to The Vision in Preschoolers Study Center, The Ohio State University, College of Optometry, 320 West Tenth Avenue, PO Box 182342, Columbus, OH 43218-2342.

*The members of the Vision in Preschoolers Study Group are listed in the Appendix. Disclosures of financial interests of writing committee members are noted in the Appendix.

ISSN 0161-6420/04/$–see front matter
doi:10.1016/j.ophtha.2004.01.022

ity), and also be able to identify as normal a high proportion of children who do not have any disorder (high specificity). Although many studies of screening have been conducted in the preschool population,[2,3,13–15] few have included standardized, comprehensive eye examinations to provide a diagnosis for children passing screening as well as for children failing screening. In addition, the definition of the conditions targeted for detection has varied among studies. Thus, valid comparisons of alternative screening tests have not been possible in most cases.

The Vision in Preschoolers (VIP) Study is a multicenter, multidisciplinary clinical study to evaluate preschool screening tests. The screening tests evaluated in the VIP Study directly or indirectly assess visual acuity (VA), ocular alignment, and/or refractive error. Visual acuity testing, stereoacuity testing, cover testing, and noncycloplegic retinoscopy (NCR) are well-established, traditional screening methods that have been recommended by national and state agencies concerned with vision screening[1,3,6,9,16,17] and by the medical community[4,5,10–12,18] for screening of preschool-aged children. Autorefraction, static photorefraction, and video photorefraction are recently developed, technology-based tests that are currently being marketed for use with preschool-aged children.[19–24] The goal of the VIP Study is to provide data on the effectiveness of both traditional and technology-based tests for identification of preschool-aged children in need of a comprehensive eye examination for evaluation of amblyopia, strabismus, and/or significant refractive error.

To accomplish this goal, the VIP Study design incorporates a phased approach to the identification of accurate screening tests for preschool-aged children. Phase I activities evaluate the performance of the screening tests when they are administered to a selected population by licensed eye care professionals (LEPs; optometrists and pediatric ophthalmologists) in a controlled environment. Phase I provides a uniform comparison of screening tests as administered by personnel who understand the tests and are experienced in examining young children. Phase II activities will evaluate the performance of tests when they are administered to a selected population by pediatric nurses and laypeople in a more realistic screening environment. Phase III activities will evaluate the performance of the tests when they are administered to a general population of preschool-aged children in a realistic screening environment. The results of phase I are presented in this report.

Materials and Methods

Participants

Subjects were children who were enrolled in Head Start programs[25] in the vicinity of the 5 VIP clinical centers (Berkeley, California; Boston, Massachusetts; Columbus, Ohio; Philadelphia, Pennsylvania; Tahlequah, Oklahoma). Head Start is a national, comprehensive child development program that serves preschool children and their families. The goal of Head Start is to increase the school readiness of children from low income families. Vision in Preschoolers Study subjects were $\geq$3 and $<$5 years on September 1 of the academic year in which they were tested. To obtain a sample that was enriched with children who had vision problems, recruitment of children was based on the results of a regular vision screening conducted by local Head Start personnel. Screening procedures varied by site. In each year, the main goal was to recruit approximately 350 children who had $\geq$1 targeted conditions. To accomplish this goal, all children at participating Head Start centers who had failed the Head Start vision screening were asked to participate in the VIP Study, as were a randomly selected subset of children who had not failed the screening. The project was approved by the appropriate institutional board(s) associated with each center.

Screeners and Examiners

All testing was conducted by LEPs who had experience in eye care of young children and who had completed VIP Study–specific training and certification. In year 1 (October 2001 through June 2002), 25 LEPs conducted screening tests, and 35 LEPs conducted gold standard examinations (GSEs). In year 2 (October 2002 through June 2003), 29 LEPs served as screeners, and 45 LEPs served as gold standard examiners. Some examiners were certified to conduct both screening tests and GSEs; however, care was taken to ensure that examiners did not conduct both screening and a GSE on the same child.

Screening Procedures and Instruments

Procedures. Screenings were conducted in a customized VIP van. Across centers, vans were identically equipped, had standardized lighting, and were divided into 4 rooms for screening. During year 1, one room was used for Lea Symbols acuity screening (Precision Vision, Inc., La Salle, IL) followed by Retinomax autorefraction (Nikon, Inc., Melville, NY); a second room was used for Random Dot E (RDE; StereoOptical Co., Chicago, IL) stereoacuity testing followed by NCR; and a third room was used for cover testing followed by HOTV acuity screening (Precision Vision). During year 2, one room was used for Stereo Smile II stereoacuity testing (StereoOptical) followed by SureSight screening (Welch Allyn, Inc., Skaneateles Falls, NY); a second room was used for Power Refractor II (Plusoptix, Nuremberg, Germany) and Retinomax Autorefractor testing; and a third room was used for iScreen photoscreening (iScreen, Inc., Memphis, TN) and MTI photoscreening (Medical Technologies, Inc., Riviera Beach, FL). Screening tests were paired and ordered within each room to minimize the opportunity for the results of the first test to influence the results of the second test. When neither of the results of the 2 tests in a room was likely to influence the other, the testing order was randomly selected for each child. Children went to the testing rooms in a randomized order, with the exception that, in year 2, children went to the room with the photoscreeners last to prevent photographic flash afterimages from interfering with other screening tests. Children's spectacles were removed before screening. Vision in Preschoolers screeners did not have access to the Head Start screening results.

Instruments. Details concerning the screening instruments are provided in Table 1. Additional details are provided below. The manufacturer of one additional instrument, the VisiScreen (Vision Research, Inc., Birmingham, AL), was invited to participate in the study in year 2 but declined due to ongoing equipment modifications.

Lea Symbols Distance Visual Acuity Test. Visual acuity is tested monocularly with a modification of the MassVAT (Optom Vis Sci 74:S174, 1997) form of the Lea Symbols,[26] consisting of cards with linear arrays of either 4 (10/100 size) or 5 (other sizes) picture optotypes (square, circle, house, and apple). Optotypes are spaced at 1 optotype width and have a crowding bar surrounding

Table 1. Key Characteristics of Screening Tests Used in the Vision in Preschoolers Study

Test/Instrument*	Testing Distance	Stopping Rules for Each Eye	Maximum Number of Measurements per Eye	Possible Results
Lea Symbols distance visual acuity test	3 m	2 wrong on 1 optotype size, or 3 correct on smallest optotype size	4 optotypes per optotype size	10/100 3 yrs: 10/32, 10/25, 10/20 4 yrs: 10/25, 10/20, 10/16
HOTV distance visual acuity test	3 m	2 wrong on 1 optotype size, or 3 correct on smallest optotype size	4 optotypes per optotype size	10/100 3 yrs: 10/32, 10/25, 10/20 4 yrs: 10/25, 10/20, 10/16
Random Dot E stereoacuity test	0.5 m, 1 m, 1.5 m	2 wrong at a distance, or 4 correct at 1.5 m	5 presentations per distance	Nonstereo card only: 504, 252, 168 arc sec
Cover–uncover test	3 m, 40 cm	≥3 cover–uncover strokes	1	Strabismus, no strabismus
Noncycloplegic retinoscopy	3 m	1 measurement	1	Refractive error values
Retinomax Autorefractor	~4 cm	Reliability score ≥8	3	Refractive error values
Stereo Smile II stereoacuity test	40 cm	2 wrong at a disparity, or 4 correct at 120 arc sec	5 presentations per disparity	Nonstereo card only: 480, 240, 120 arc sec
Power Refractor II	~1 m	Value shown in green (valid reading)	3	Refractive error values in red or green
iScreen Photoscreener	~68 cm	Pupils in focus and diameter of ≥4 mm, good fixation in ≥1 eye in each image	3	Unreadable, abnormal, normal
MTI Photoscreener	~1 m	4 pupils in focus, pupils ≥4 mm in diameter, good fixation in ≥1 eye in each image	3	Unreadable, fail, pass
SureSight Vision Screener	~36 cm	Reliability score ≥6, no asterisk printed or asterisks on 2 readings	3	Refractive error values

the line at a distance of 0.5 optotype width. During a binocular pretest, the child must identify verbally, or by matching, each optotype presented singly at 1 m. Upon successful completion, the screener patches the child's left eye, moves to 3 m, and presents the 10/100 and age-specific cards. If the child correctly identifies 3 of 3 or 3 of 4 optotypes, the screener presents the next card. The procedure is repeated for the left eye, beginning with the 10/100 card. Vision in Preschoolers Study staff asked Head Start teachers to review the Lea Symbols with the children before the screening day.

HOTV Distance Visual Acuity Test. The procedures for the HOTV test are identical to those used for the Lea Symbols test, except that the optotypes are the letters *H, O, T,* and *V*.

Random Dot E Stereoacuity Test. The RDE test consists of a demonstration plate (a non–stereo raised E), a blank plate (a random dot pattern), and a plate displaying a random dot stereo E. The child wears polarizing glasses. The screener familiarizes the child with the raised E plate and then conducts a pretest by presenting the raised E plate paired with the blank plate at a distance of 50 cm. If the child correctly identifies the raised E plate on 4 of 4 or 4 of 5 presentations, the screener successively presents the stereo E plate paired with the blank test plate at 50 cm, 1 m, and 1.5 m as long as the child correctly identifies the E plate on 4 of 4 or 4 of 5 presentations. The screener varies the left–right and/or up–down position of the plate in a nonsystematic manner.

Cover–Uncover Test. Eye alignment is assessed using a cover–uncover test at both distance (3 m) and near (40 cm). The screener asks the child to look at a detailed, standardized fixation target and places a cover paddle over the child's left eye. The paddle is kept in front of the eye for approximately 3 seconds. The screener observes the unoccluded right eye to determine if refixation occurs. The cover–uncover stroke is repeated at least 3 times. The procedure is repeated, covering the right eye.

Noncycloplegic Retinoscopy. Noncycloplegic retinoscopy is used by eye care professionals to measure refractive error. The screener uses a streak retinoscope and a retinoscopy lens rack or handheld trial lenses. The child wears retinoscopy spectacles corresponding to the screener's working distance to control accommodation. Measurements, corrected for the screener's working distance, are obtained along the 2 principal meridia of each eye. The screener instructs the child to fixate on an animated target on a video screen.

Retinomax Autorefractor. The Retinomax is a handheld autorefractor that measures refractive error monocularly along 2 meridia. The screener places the instrument's headrest on the child's forehead, encourages the child to fixate the internal target, and focuses the mire in the center of the right pupil while ≥8 readings are taken automatically. The screener repeats the process for the left eye. The screener prints the refractive error and reliability rating for each eye. If the reliability reading is <8 (the manufacturer's recommended minimum value), the measurement is repeated.

Stereo Smile II Stereoacuity Test. The Stereo Smile II test consists of a demonstration plate (a nonstereo smile face on a background of random dots), a blank plate (a random dot pattern), and 3 plates each displaying a random dot stereo smile face of successively finer levels of stereopsis. The child wears Polaroid glasses. The screener conducts a pretest, by presenting the dem-

onstration plate paired with the blank plate. If the child correctly identifies the demonstration plate on 4 of 4 or 4 of 5 presentations, the screener successively presents the blank plate paired with the plates of finer disparity as long as the child correctly identifies 4 of 4 or 4 of 5 presentations. The screener varies the left–right and/or up–down position of the plate in a nonsystematic manner.

Power Refractor II. The Power Refractor II (version 3.11.01.24.00) is a tabletop video/photorefractor that binocularly measures refractive error in 8 meridia and measures eye alignment. When the child fixates on the red and green lights on the camera, the screener begins the measurement and continues until the refractive error in each eye and gaze deviation appears in green on a display or until the instrument times out. The screener prints the display image. If the refractive error displayed for either eye is red, the measurement of the highlighted eye(s) is repeated. If the output for either eye is again red, measurement may be made monocularly.

iScreen Photoscreener. The iScreen consists of an off-axis photorefractor connected to a laptop computer that binocularly measures refractive error in one meridian and measures eye alignment. The child sits in front of the camera unit, with his or her head against a headrest, and fixates on a red light. The screener positions the child so that the eyes are centered horizontally along a crosshair that is visible on the display and takes a photograph. If the image is not adequate, the screener repeats the process. The last image collected is transmitted to the iScreen scoring center in Memphis, Tennessee.

MTI Photoscreener. The MTI consists of an off-axis photorefractor in which the eccentric flash rotates 90° to allow binocular measurement of refractive error along 2 orthogonal axes and measurement of eye alignment. The screener holds the camera approximately 1 m from the child, positions 2 pointed target lights on the child's forehead, tells the child to watch the red lights on the camera, and takes a flash photograph. The flash automatically rotates 90°, and the screener takes a second flash photograph. The images are developed in approximately 30 seconds on instant film. If any of the criteria in Table 1 are not met, the screener takes additional photographs. All photographs for each child are sent to the Vanderbilt Ophthalmic Imaging Center for interpretation.

SureSight Vision Screener. The SureSight Vision Screener (version 2.12) is a handheld autorefractor that measures refractive error monocularly in 2 meridia based on wavefront technology.[27] The screener looks through the viewfinder, instructs the child to fixate the red lights on the front of the machine, and centers a crosshair on the pupil of the right eye. Auditory cues signal the screener to adjust the testing distance and signal the completion of measurements for the right eye. The screener repeats the process for the left eye and prints a refractive error and reliability rating for each eye. If the reliability reading is <6 (the manufacturer's recommended minimum value) or an asterisk is printed to denote a refractive error that exceeds the SureSight normal bounds, the measurement is repeated. If 2 consecutive asterisks are obtained for the same eye, then the measurement is not repeated.

Training of Screeners

Each year, a team of VIP Study personnel conducted a daylong training program for screeners at each local clinical center. During year 2, the training program included instruction by representatives from the manufacturers of the Power Refractor II, the iScreen Photoscreener, and the SureSight Vision Screener, as well as trainers on the MTI Photoscreener who were based at the Vanderbilt Ophthalmic Imaging Center. Screeners were observed by the local principal investigator or coinvestigator while testing ≥ 3 children on each screening test. Screeners completed written knowledge assessments and, in year 2, had to submit ≥ 3 readable images to the respective reading center for the iScreen Photoscreener and the MTI Photoscreener. During year 2, the Coordinating Center provided alerts on individual screeners with relatively high percentages of images judged as unreadable. These screeners were asked to review the corresponding screening test protocol.

Gold Standard Examination Procedures

Monocular distance VA assessment, cover testing, and cycloplegic retinoscopy are used to determine if a child had amblyopia or reduced VA, strabismus, and/or significant refractive error. Evaluation of anterior segment and binocular indirect ophthalmoscopy is also performed to detect other possible causes of reduced VA. Other examination procedures such as color vision testing, NCR, stereoacuity testing, and ductions and versions are performed to provide a complete ocular assessment of the child. Gold standard examinations are conducted in the VIP van, which was reconfigured into 2 examination rooms. Examiners did not have access to the Head Start or VIP screening results.

Monocular Distance Visual Acuity Testing. Monocular distance threshold VA testing is conducted using the Electronic Visual Acuity tester (Jaeb Center for Health Research, Tampa, FL) at 3 m,[28] according to the protocol established by the Amblyopia Treatment Study.[29] The letters *H*, *O*, *T*, and *V* are presented as isolated optotypes with crowding bars surrounding the letter. Visual acuity between 20/16 and 20/800 can be measured. Children who wear spectacles are tested while wearing the spectacles. Children are retested on the same day after cycloplegic retinoscopy with their full cycloplegic refractive correction if their initial VA scores place them in a category other than normal (Table 2) and they have a refractive error in either eye of ≥ 0.5 diopters (D) of myopia, ≥ 2.0 D of hyperopia, or ≥ 1.0 D of astigmatism.

Cover Testing. Both a cover–uncover test and an alternating cover test are performed at distance (3 m) and near (40 cm). The procedure used for the cover–uncover test is identical to that used during screening. If a tropia is detected, the deviation is neutralized with loose prisms or a bar prism. In contrast to the screening protocol, if no deviation is detected at distance, the examiner then performs the alternating cover test at distance by occluding the left eye for 1 to 2 seconds, and then moving the occluder quickly to the right eye without allowing binocular fixation to occur. This alternating pattern is repeated ≥ 3 times. If no movement is detected, the examiner records that no tropia or phoria is present at distance. If movement is detected on the alternating cover test, the examiner measures the magnitude of the deviation using loose prisms or a prism bar. The procedure is repeated at near.

Cycloplegic Retinoscopy. Retinoscopy is performed 30 to 40 minutes after instillation of 1 drop of 0.5% proparacaine, followed by 1 drop each of 1% cyclopentolate and 0.5% tropicamide. A second set of the cycloplegic agents is instilled at the examiner's discretion. Retinoscopy is performed with the child wearing retinoscopy spectacles corresponding to the screener's working distance to control any residual accommodation. The child is instructed to fixate an animated video target presented at 3 m. The examiner uses a lens rack or handheld trial lenses to neutralize the refractive error in each eye. Measurements are obtained along the 2 principal meridia of each eye.

Training of Examiners

The examiners were trained by the VIP training team during a daylong program at the local clinical center. Examiners were observed performing GSE procedures on ≥ 2 preschool children and completed written knowledge assessments. During year 2, alerts were provided by the Coordinating Center to the clinical

Table 2. Definitions of Targeted Disorders in the Vision in Preschoolers Study

Targeted Disorder	Definitions*
Amblyopia	
Presumed unilateral	≥3-line interocular difference in VA *and* a unilateral amblyogenic factor[†]
Suspected unilateral	2-line interocular difference in VA *and* a unilateral amblyogenic factor[†]
Suspected bilateral	
3-year-olds	Worse than 20/50 in one eye, worse than 20/40 in the contralateral eye, *and* a bilateral amblyogenic factor[‡]
4- and 5-year-olds	Worse than 20/40 in one eye, worse than 20/30 in the contralateral eye, *and* a bilateral amblyogenic factor[‡]
Reduced VA[§]	
Bilateral	
3-year-olds	Worse than 20/50 in one eye, worse than 20/40 in the contralateral eye; no bilateral amblyogenic factor[‡]
4- and 5-year-olds	Worse than 20/40 in one eye, worse than 20/30 in the contralateral eye; no bilateral amblyogenic factor[‡]
Unilateral	
3-year-olds	Worse than 20/50 in only one eye *or* ≥2-line difference between the eyes (except 20/16 and 20/25); no unilateral amblyogenic factor[†]
4- and 5-year-olds	Worse than 20/40 in only one eye *or* ≥2-line difference between the eyes (except 20/16 and 20/25); no unilateral amblyogenic factor[†]
Strabismus	Any heterotropia in primary gaze
Significant refractive error	Cycloplegic refraction
Astigmatism	>1.50 D between principal meridians
Hyperopia	>3.25 D in any meridian
Myopia	>2.00 D in any meridian
Anisometropia	>1.00-D interocular difference in hyperopia; >3.00-D interocular difference in myopia; >1.50-D interocular difference in astigmatism; antimetropic[‖] difference >1.00 D and one eye >1.00 D of hyperopia; antimetropic[‖] difference >3.00 D and one eye >2.00 D of myopia

D = diopters; VA = visual acuity.
*Applied sequentially for amblyopia and reduced VA.
[†]Strabismus, anisometropia (as defined in Table 2), and a difference in spherical equivalent of ≥0.50 D when ≥1 eye had >3.50 D of hyperiopia were considered unilateral amblyogenic factors.
[‡]Astigmatism of >2.50 D, hyperopia of >5.00 D, or myopia of >8.00 D in each eye were considered bilateral amblyogenic factors.
[§]Reduced VA due to a cause other than amblyopia or refractive error; cause identified or not.
[‖]One eye hyperopic, 1 eye myopic.

centers concerning individual examiners with a relatively high percentage of children with incomplete results on the GSE. The local clinical center principal investigator was asked to review the protocol procedures with the examiner.

Classification of Children

Definitions for the targeted conditions of amblyopia, strabismus, or significant refractive error are provided in Table 2, along with definitions for a fourth category, unexplained reduced VA. This fourth category was added to account for children who have VAs below age-specific thresholds that are not attributable to an amblyogenic factor.

The targeted disorders were grouped by severity of the condition into a hierarchy consisting of 3 groups. The groups were defined by a consensus of the VIP Executive Committee, with input from the VIP Advisory Committee and without reference to study data. The conditions included in each group of the hierarchy are described in the lefthand column of Table 3. Group 1 conditions are considered very important to detect and treat early. Group 2 conditions are considered important to detect early (but with less urgency than group 1). Group 3 conditions are considered less urgent, but nonetheless are clinically useful to detect. Children with multiple conditions were classified on the basis of their worst condition.

Children who did not have any conditions on their GSE that would place them in groups 1, 2, or 3 were classified as normal or indeterminant. Results for a child were considered indeterminant if their GSE results included incomplete or missing information that prevented them from being assigned to a group or to the normal category.

Data Analysis

Data collection forms from the screening and GSE sessions and the interpretation of iScreen and MTI photoscreener images were data entered, edited, and analyzed at the Coordinating Center. Statistical computations were performed using SAS 8.2 (SAS Institute, Cary, NC).

Only data from children who were both screened and examined are included in this report. Children who did not have a complete VA test, cover test, or cycloplegic refraction on the GSE are excluded from the analyses of sensitivity and specificity. Sample sizes for individual screening tests in each year vary because children left the van before administration of all screening tests and because of data-recording errors. When multiple readings of refractive error were obtained, the first measurement with a reliability score or color code considered reliable by the manufacturer was used. If no reading met the criteria, the reading with the highest score was used.

Testability. For tests of VA or stereoacuity, children had to complete the pretest successfully to be considered testable. For the cover–uncover test, the screener had to make an assessment at each distance. For tests that immediately provided values for refractive error, screeners obtained measurements for the eyes of each child, regardless of the reliability score. For the MTI and iScreen photoscreeners, the screener had to obtain ≥1 image.

Specificity. Children without any targeted conditions were used to estimate specificity. Because children who had failed the Head Start vision screening were over-represented, a weighted estimate of specificity was used based on the proportion of children failing (1/6) or not failing (5/6) the Head Start screening.

Sensitivity. The overall sensitivity was calculated as the pro-

Table 3. Hierarchy of Vision in Preschoolers Targeted Disorders

Condition	n	%
Group 1: very important to detect and treat early	311	12.0
Amblyopia	92	3.6
Presumed unilateral *and* worse eye VA of ≤20/64	36	1.4
Suspected bilateral	56	2.2
Strabismus—constant	67	2.6
Refractive error	269	10.4
Severe anisometropia (interocular difference >2 D of hyperopia, >3 D of astigmatism, or >6 D of myopia)	43	1.7
Hyperopia ≥5.0 D	123	4.8
Astigmatism ≥2.5 D	146	5.6
Myopia ≥6.0 D	15	0.6
Group 2: important to detect early	229	8.8
Amblyopia	27	1.0
Suspected unilateral	20	0.8
Presumed unilateral *and* worse eye VA of >20/64	7	0.3
Strabismus—intermittent	29	1.1
Refractive error	202	7.8
Anisometropia, but not severe	64	2.5
Hyperopia of >3.25 D and <5.0 D *and* interocular difference in SE of ≥0.5 D	64	2.5
Astigmatism of >1.5 D and <2.5 D	113	4.4
Myopia of ≥4.0 D and <6.0 D	3	0.1
Group 3: detection clinically useful	215	8.3
Reduced VA	173	6.7
Bilateral	54	2.1
Unilateral	119	4.6
Refractive error	51	2.0
Hyperopia of >3.25 D and <5.0 D *and* interocular difference in SE of <0.5 D	45	1.7
Myopia of >2.0 D and <4.0 D	6	0.2
Normal	1833	70.9
Total no. of children	2588	

D = diopters; SE = spherical equivalent; VA = visual acuity.

portion of children who failed a screening test among children with ≥1 targeted conditions. Additional estimates of sensitivity were made for detection of each group of conditions in the VIP hierarchy (Table 3) and of the individual targeted conditions. Children who were not testable on a screening test were categorized as failing. Children who demonstrated testability but did not complete a test because of interruption, behavior problem, or examiner error were classified as failing the test if they had not yet completed the level required for passing.

Failure Criteria. Children failed a screening test if they met the failure criteria for one or both eyes. Four of the screening tests had predefined failure criteria (cover–uncover test) or failure criteria specified by the manufacturer or interpreter of the screening test (SureSight Vision Screener and MTI and iScreen photoscreeners). For other screening tests, specificity was first set at a minimum of 0.90, and failure criteria were selected to maximize the overall sensitivity for detecting any targeted condition. Failure criteria for the tests of VA and stereoacuity were age specific. If more than a set of failure criteria for a test involving refractive error provided the same level of sensitivity, the set with the highest sensitivity for detecting group 1 conditions was chosen. Because interpretations by central reading centers for the iScreen and MTI photoscreeners each provided a specificity of 0.94, failure criteria for the other tests were again developed with specificity set at 0.94 to facilitate comparison with the photoscreeners.

Statistical Comparisons. Pairwise comparisons of sensitivity between screening tests performed within the same year were made using the McNemar chi-square test for correlated data. When children completed only 1 of the 2 screening tests, a modification of the Mantel–Haenszel procedure was used.[30] Comparisons of sensitivity among screening tests performed in different years were made using the chi-square test of independence. Statistical comparisons involving the photoscreeners are based on the sensitivity value when specificity is set at 0.94. A large number of pairwise comparisons between screening tests may be of interest; if the conservative Bonferroni approach to multiple comparisons is applied, then comparisons associated with a P value of 0.05/66 = 0.0008 are statistically significant.

Results

Study Population

In years 1 and 2, 2211 children who had failed the Head Start vision screening and 1772 children who had not failed the screening were selected for enrollment. Consent was obtained and eligibility criteria fulfilled for 3121 of the 3983 children (78.3%). Screening was completed for 2780 (89.1%) of the children enrolled. Gold standard examinations were performed on 2666 (95.6%) of screened children, of whom 2588 (97.1%) completed VA testing, cover testing, and cycloplegic refraction. All children were 3, 4, or 5 years old when screened; however, few young 3-year-olds or old 5-year-olds were tested (Table 4). There was substantial diversity in ethnicity and race.

The distribution of children according to the hierarchy of VIP targeted disorders is displayed in Table 3. Significant refractive error, as defined in Table 2, was present in 284 (91.3%) of the 311 children with group 1 conditions and in 204 (89.1%) of the 229 children with a group 2 condition in the absence of a group 1 condition. Among the 2588 children, 755 (29.2%) had ≥1 targeted disorders: 163 (6.30%) had amblyopia, 110 (4.25%) had strabis-

Table 4. Characteristics of Children

Characteristic	n	%
Age at screening (mos)		
36–41	35	(1.4)
42–47	473	(18.3)
48–53	645	(24.9)
54–59	725	(28.0)
60–65	637	(24.6)
66–71	73	(2.8)
Gender		
Female	1291	(49.9)
Male	1297	(50.1)
Hispanic or Latino		
Yes	585	(22.6)
No	1876	(72.5)
Unknown	127	(4.9)
Racial category		
Black	1239	(47.9)
White	464	(17.9)
American Indian	199	(7.7)
Asian	94	(3.6)
Native Hawaiian/Pacific Islander	20	(0.8)
Other	112	(4.3)
Mixed race	101	(3.9)
Unknown	359	(13.9)
Total no. of children	2588	

mus, 539 (20.8%) had significant refractive error, and 246 (9.5%) had reduced VA as defined in Table 2. Significant refractive error was present in 58 (52.7%) of the 110 children with strabismus and in 30 (79.0%) of the 38 children with strabismic amblyopia.

Testability, Readability, and Number of Repeated Procedures

The proportion of children who were not testable was <1% for the majority of the screening tests (Table 5). The proportion untestable was highest (9.7%) for the RDE test, with 24 of 215 3-year-olds (11.2%), 63 of 608 4-year-olds (10.4 %), and 24 of 319 5-year-olds (7.5%) not testable.

Table 5. Children Not Testable on Screening Tests

Screening Test	n	Children Not Testable [n (%)]
Year 1		
Lea Symbols VA	1142	6 (0.5)
HOTV VA	1141	7 (0.6)
Random Dot E	1142	111 (9.7)
Cover–uncover	1141	24 (2.1)
Noncycloplegic retinoscopy	1142	9 (0.8)
Retinomax Autorefractor	1142	6 (0.5)
Year 2		
Stereo Smile II	1446	27 (1.9)
Power Refractor II*	1438	22 (1.5)
iScreen Photoscreener†	1439	2 (0.1)
MTI Photoscreener‡	1444	0 (0.0)
SureSight Vision Screener	1442	11 (0.8)
Retinomax Autorefractor	1435	2 (0.1)

VA = visual acuity.
*Thirty-two (2.2%) readings displayed red; invalid measurement.
†Thirty-seven (2.6%) images ungradable.
‡Eighty-four (5.8%) photographic sets ungradable.

Table 6. Failure Criteria for Visual Acuity (VA) and Stereoacuity Tests to Maximize Sensitivity When Specificity Was Set at 0.90

Test	Age (yrs)	Failure Criterion* (Inability to Pass)
Lea Symbols VA	3	10/32 line
	4	10/20 line
	5	10/20 line
HOTV VA	3	10/25 line
	4	10/25 line
	5	10/20 line
Random Dot E stereoacuity	3	Nonstereo card
	4	Stereo card at 50 cm (550 arc sec)
	5	Stereo card at 100 cm (252 arc sec)
Stereo Smile II stereoacuity	3	240–arc sec card
	4	240–arc sec card
	5	120–arc sec card

*Chosen to maximize overall sensitivity for detecting any targeted condition when specificity was set to 0.90.

Although Power Refractor II, iScreen, and MTI images were obtained for nearly all children, 32 of 1438 (2.2%) children had invalid Power Refractor II readings, 37 of 1439 (2.6%) children had unreadable iScreen Photoscreener images, and 84 of 1444 (5.8%) children had unreadable MTI Photoscreener images. One image set was obtained for 1301 of the 1438 (90.5%) children with the Power Refractor II, for 1011 of 1439 (70.3%) children with the iScreen Photoscreener, and for 919 of 1444 (63.6%) children with the MTI Photoscreener. Mean numbers of image sets per child were 1.14, 1.37, and 1.48, respectively, for the 3 instruments. In year 2, only 1 Retinomax Autorefractor reading was taken for 2523 of 2870 (87.9%) eyes (mean, 1.15 readings). Only 1 SureSight Vision Screener reading was taken for 1726 of 2884 (59.8%) eyes (mean, 1.64 readings).

Sensitivity of Screening Tests

Overview. Failure criteria when specificity was set at 0.90 are summarized in Table 6 for tests of VA and stereoacuity and in Table 7 for tests involving refractive error. Table 8 displays the sensitivity of the tests for detection of children who have any targeted condition, and for detection of children with group 1, group 2, or group 3 conditions. The *P* values for pairwise comparisons between screening tests when specificity is set at 0.90 are provided in Table 9 for detection of any targeted condition, and for detection of children with group 1 conditions. For example, the *P* value (0.70) for the comparison of the sensitivity for detecting ≥1 targeted conditions between the SureSight Vision Screener (0.63) and the NCR (0.64) is found in the SureSight row and the NCR column, whereas the *P* value (0.04) for the comparison of the sensitivity for detecting a group 1 condition between the 2 screening tests (0.81 vs. 0.90) is in the NCR row and the SureSight column. Table 10 provides sensitivity and specificity for each screening test for detection of amblyopia, reduced VA, strabismus, and significant refractive error. Table 11 displays the sensitivity of the iScreen and MTI photoscreeners and of the screening tests when failure criteria were determined with specificity set at 0.94. The sensitivity of the cover–uncover test was significantly ($P<0.0001$) lower than for all other tests for detecting children with any targeted condition or with a group 1 condition; these comparisons do not account for the higher specificity (0.98) of the cover–uncover test.

Table 7. Failure Criteria for Retinoscopy, Photorefraction, and Autorefractor Screening Tests to Maximize Sensitivity

Instrument	Hyperopia	Myopia	Astigmatism	Anisometropia*
Noncycloplegic retinoscopy	≥2.75 D	≥2.75 D	≥1.25 D	≥1.50 D
Power Refractor II	≥3.50 D	≥3.00 D	≥2.00 D	≥1.50 D
SureSight Vision Screener†				
Manufacturer	>2.00 D	>1.00 D	>1.00 D	>1.00 D SE
VIP Study	≥4.00 D	≥1.00 D	≥1.50 D	≥3.00 D
Retinomax Autorefractor				
Year 1	≥1.50 D	≥2.75 D	≥1.50 D	≥2.00 D
Year 2	≥1.50 D	≥2.75 D	≥1.50 D	≥1.75 D

D = diopters; SE = spherical equivalent; VIP = Vision in Preschoolers.
With the exception of criteria set by the manufacturer, failure criteria were chosen to maximize overall sensitivity for detecting any targeted condition when specificity was set to 0.90 or above.
*The maximum of intereye differences in the power of the most positive meridian, the most negative meridian, and the magnitude of cylinder was used to determine presence of anisometropia for all tests, except in the manufacturer's criteria for SureSight.
†Used in child mode, which adds a correction for accommodation.

Lea Symbols Visual Acuity. Overall sensitivity was similar to the sensitivity of NCR and the Retinomax in year 1, but lower for detection of children with group 1 conditions, amblyopia, and refractive error. The Lea Symbols VA test had sensitivity for detecting children with any targeted condition statistically significantly higher than sensitivities of the RDE and Stereo Smile tests and the photoscreeners, and sensitivity for detecting children with group 1 conditions higher than that of the RDE.

HOTV Visual Acuity. Specificity for HOTV VA is 0.89, not 0.90, because children with incomplete test results were not included in the determination of failure criteria but were classified as passing the test if they had read beyond the failure criterion. Overall sensitivity and group 1 sensitivity were somewhat lower for the HOTV VA test than for the Lea Symbols VA test (Table 8; 0.54 vs. 0.61 for any condition and 0.72 vs. 0.77 for group 1), but the associated *P* values were not significant (Table 9; 0.03 and 0.27, respectively).

Random Dot E Stereotest. Overall sensitivity and group 1 sensitivity were lower than sensitivity values obtained with the Retinomax, SureSight, Power Refractor II, Lea Symbols Visual Acuity test, and HOTV test (overall sensitivity only). Because of the relatively high proportion of untestable children, the data were reanalyzed with these children considered as passing. To achieve at least 0.90 specificity, the failure criteria were changed to inability to pass the stereo card at 100 cm for ages 3 and 4 years, and inability to pass the stereo card at 150 cm for age 5 years. Slightly higher sensitivities were achieved (overall sensitivity, 0.44; groups 1, 2, and 3 sensitivity, 0.62, 0.37, and 0.27, respectively, and specificity, 0.91).

Cover–Uncover Test. Except for detection of children with strabismus (Table 10), the cover–uncover test showed very low sensitivity for detection of ocular disorders. In year 1, 19 (73%) of the 26 children with a constant strabismus were detected by the screening cover–uncover test, and 10 (45%) of the children with intermittent strabismus were detected. The cover–uncover test had statistically significantly lower sensitivity than all other screening tests for detection of any targeted condition or group 1 condition.

Additional analyses showed that pairing the cover–uncover test with NCR yielded small increases in sensitivity beyond the level of NCR alone. An additional 16 of 346 (4.6%) children with a

Table 8. Sensitivity by Vision in Preschoolers Hierarchy of Conditions with Specificity Set to 0.90 for Tests without Established Failure Criteria

Screening Test	Sensitivity				Specificity (n = 796)
	Any condition (n = 346)	*Group 1 (n = 139)*	*Group 2 (n = 108)*	*Group 3 (n = 99)*	
Year 1					
Lea Symbols VA	0.61	0.77	0.57	0.41	0.90
HOTV VA	0.54	0.72	0.41	0.44	0.89
Random Dot E	0.42	0.59	0.33	0.27	0.90
Cover–uncover	0.16	0.24	0.13	0.06	0.98
Noncycloplegic retinoscopy	0.64	0.90	0.63	0.29	0.90
Retinomax Autorefractor	0.63	0.87	0.63	0.30	0.90
Year 2	*(n = 409)*	*(n = 172)*	*(n = 121)*	*(n = 116)*	*(n = 1037)*
Stereo Smile II	0.44	0.72	0.30	0.20	0.91
Power Refractor II	0.54	0.72	0.43	0.39	0.90
iScreen Photoscreener	0.37	0.57	0.24	0.20	0.94
MTI Photoscreener	0.37	0.55	0.27	0.19	0.94
SureSight Vision Screener*	0.85	0.96	0.90	0.63	0.62
SureSight Vision Screener	0.63	0.81	0.68	0.29	0.90
Retinomax Autorefractor	0.64	0.88	0.55	0.37	0.90

VA = visual acuity.
*Screening failure defined as recommended by the manufacturer.

Table 9. *P* Values Associated with Pairwise Comparisons of Sensitivity between Screening Tests with Specificity Set at 0.90

Sensitivity	NCR	Retinomax, Year 2	Retinomax, Year 1	SureSight	Lea VA	HOTV VA	Power Refractor II	Stereo Smile	Random Dot E
Any condition	0.64	0.64	0.63	0.63	0.61	0.54	0.54	0.44	0.42
Group 1	0.90	0.88	0.87	0.81	0.77	0.72	0.72	0.72	0.59
NCR		0.72	0.37	0.04	0.0015	<0.0001	<0.0001	<0.0001	<0.0001
Retinomax, year 2	1.00		0.86	0.01	0.0093	0.0004	<0.0001	<0.0001	<0.0001
Retinomax, year 1	0.73	0.88		0.21	0.02	0.0004	0.0019	0.0009	<0.0001
SureSight	0.70	0.50	0.88		0.40	0.06	0.03	0.02	<0.0001
Lea VA	0.23	0.36	0.37	0.65		0.27	0.36	0.30	0.0007
HOTV VA	0.0039	0.0090	0.0027	0.03	0.03		1.00	1.00	0.02
Power Refractor II	0.0060	0.0007	0.01	0.0048	0.08	0.94		0.90	0.02
Stereo Smile	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0068	0.0015		0.02
Random Dot E	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0002	0.0010	0.51	

NCR = noncycloplegic retinoscopy; VA = visual acuity.
P values in lower left triangle are for comparisons of sensitivity for detection of all conditions; those in upper right triangle are for comparison of sensitivity for detection of group 1 conditions only. If the conservative Bonferroni approach to multiple comparisons is applied and all pairwise comparisons are considered of interest, then comparisons associated with a *P* value of 0.05/66 = 0.0008 would be considered statistically significant.

targeted condition (8 with strabismus) would be detected by addition of the cover–uncover test. Specificity for combining the 2 screening tests in this way decreased to 0.88. If the failure criteria for NCR were reset so that the specificity was 0.90 for the combination of tests, then overall sensitivity was 0.66, sensitivity for group 1 conditions was 0.93, and the sensitivities for amblyopia and strabismus were 0.87 and 0.71, respectively.

Noncycloplegic Retinoscopy. As shown in Tables 8 and 10, no test had better overall sensitivity or better sensitivity for detecting children with group 1 conditions or significant refractive error than NCR. Noncycloplegic retinoscopy had statistically significantly higher sensitivity for detection of any targeted condition than the tests of stereoacuity and the photoscreeners, and higher sensitivity for detecting children with group 1 conditions than the same set of screening tests plus the Power Refractor II and the HOTV VA test.

Retinomax Autorefractor. In year 1, performance of the Retinomax Autorefractor was very similar to the performance of NCR for each group of conditions in the VIP hierarchy (Table 8) and for each of the targeted conditions (Table 10). In year 2, with a less stringent criterion for anisometropia, there was no other screening test with higher sensitivity in detecting group 1 conditions or strabismus, refractive error, or reduced VA when specificity was at least 0.90. When the failure criteria developed in year 1 were applied to the data from year 2, very similar results were achieved (overall sensitivity, 0.64; group 1 sensitivity, 0.88; sensitivity for amblyopia, 0.85; and specificity, 0.90). The Retinomax had sensitivity for any condition and for group 1 conditions statistically significantly higher than those of all other tests in year 2 except the SureSight.

Stereo Smile II Test. The sensitivities of the Stereo Smile II test for detecting children with any targeted condition, group 1 conditions, amblyopia, and strabismus were higher than the corresponding sensitivities of the RDE test, but not to a statistically significant degree. No other tests had statistically significantly higher sensitivity for detection of children with strabismus.

Table 10. Sensitivity by Condition Type* with Specificity Set to 0.90 for Tests without Established Failure Criteria

	Sensitivity				
Screening Test	*Amblyopia* (*n* = 75)	*Reduced VA* (*n* = 132)	*Strabismus* (*n* = 48)	*Refractive Error* (*n* = 240)	**Specificity (n = 796)**
Year 1					
Lea Symbols VA	0.76	0.58	0.56	0.70	0.90
HOTV VA	0.73	0.48	0.65	0.59	0.89
Random Dot E	0.63	0.38	0.60	0.47	0.90
Cover–uncover	0.27	0.06	0.60	0.16	0.98
Noncycloplegic retinoscopy	0.85	0.47	0.56	0.81	0.90
Retinomax Autorefractor	0.85	0.50	0.65	0.78	0.90
Year 2	(*n* = 88)	(*n* = 114)	(*n* = 62)	(*n* = 299)	(*n* = 1037)
Stereo Smile II	0.77	0.30	0.68	0.51	0.91
Power Refractor II	0.80	0.43	0.55	0.61	0.90
iScreen Photoscreener	0.62	0.27	0.50	0.43	0.94
MTI Photoscreener	0.63	0.24	0.65	0.42	0.94
SureSight Vision Screener†	0.98	0.70	0.92	0.92	0.62
SureSight Vision Screener	0.89	0.43	0.59	0.75	0.90
Retinomax Autorefractor	0.85	0.45	0.69	0.76	0.90

VA = visual acuity.
*Children may have more than 1 condition.
†Screening failure defined as recommended by the manufacturer.

Table 11. Comparison of Sensitivity for the iScreen and MTI Photoscreeners with Failure Criteria for Other Tests Set to Provide Specificity of 0.94

	Any Condition			Group 1 Conditions		
		P Values			*P Values*	
	Sensitivity	*iScreen*	*MTI*	*Sensitivity*	*iScreen*	*MTI*
Year 1						
Lea Symbols VA	0.49	0.0004	0.0004	0.65	0.10	0.06
HOTV VA	0.36	0.94	0.94	0.48	0.14	0.25
Random Dot E	0.22	<0.0001	<0.0001	0.30	<0.0001	<0.0001
Noncycloplegic retinoscopy	0.57	<0.0001	<0.0001	0.87	<0.0001	<0.0001
Retinomax Autorefractor	0.52	<0.0001	<0.0001	0.81	<0.0001	<0.0001
Year 2						
Stereo Smile II	0.33	0.17	0.17	0.57	0.91	0.73
Power Refractor II	0.36	0.93	0.90	0.56	0.87	0.90
iScreen Photoscreener	0.37	—	0.93	0.57	—	0.71
MTI Photoscreener	0.37	0.93	—	0.55	0.71	—
SureSight Vision Screener	0.51	<0.0001	<0.0001	0.75	0.0003	<0.0001
Retinomax Autorefractor	0.52	<0.0001	<0.0001	0.81	<0.0001	<0.0001

VA = visual acuity.

Power Refractor II. Failure criteria for the refractive error results are in Table 7, and the criterion for gaze deviation was ≥10.25°. Sensitivity to detect any targeted condition was statistically significantly lower than that of the Retinomax. Sensitivity to detect group 1 conditions was lower than sensitivities of NCR and the Retinomax (Table 8).

iScreen Photoscreener. Central interpretation of the images was associated with a specificity of 0.94 and an overall sensitivity of 0.37. Sensitivity was 0.57 for group 1 conditions and 0.62 for amblyopia. These values are nearly identical to those obtained with the MTI Photoscreener. When failure criteria for the other tests were set to obtain 0.94 specificity (Table 11), the iScreen sensitivity for any targeted condition was statistically significantly below the values for NCR, the Retinomax Autorefractor, SureSight Vision Screener, and Lea Symbols VA test. The sensitivity for detecting group 1 conditions was statistically significantly below the values for NCR, the Retinomax Autorefractor, and SureSight Vision Screener.

MTI Photoscreener. Central interpretation of the photographs was associated with a specificity of 0.94 and an overall sensitivity of 0.37. Sensitivities were 0.55 for group 1 conditions, 0.63 for amblyopia, and 0.65 for strabismus. The set of screening tests with statistically significantly higher sensitivity was the same for the MTI Photoscreener as for the iScreen Photoscreener.

SureSight Vision Screener. The failure criteria specified by the manufacturer of the instrument were associated with high sensitivity (0.85), but low specificity (0.62) for all conditions (Table 8). When specificity was set at 0.90 to allow a more uniform comparison with the other tests, the sensitivities for detecting any condition, amblyopia, and refractive error were similar to the sensitivities of the Retinomax in year 2. The SureSight had statistically significantly higher sensitivity than the tests of stereoacuity and the photoscreeners for detection of any targeted condition, and higher sensitivity than the RDE test and the photoscreeners for group 1 conditions.

Discussion

Phase I of the VIP Study is the first comprehensive investigation of an array of currently relevant preschool vision screening tests. The research design allows valid comparisons of the accuracy of the tests in detecting a defined set of significant and prevalent vision disorders. The vision screening tests were administered by LEPs (optometrists and pediatric ophthalmologists) under controlled, standardized conditions. The tests evaluated included recently developed photorefractive devices and autorefractors, as well as traditional tests of VA, binocularity, and refractive error. More than 2500 3- to 5-year-old children enrolled in Head Start centers in 5 geographically and culturally diverse areas of the United States participated. All children underwent a GSE to allow classification of their vision status using standardized definitions of targeted conditions.

Screening tests are often assessed using receiver operator characteristic curves that display all possible combinations of sensitivity and specificity achievable through varying the failure criteria. Some of the screening tests had established failure criteria that provided one combination of sensitivity and specificity (1 point on a receiver operator characteristic curve). The refractive error screening tests have multidimensional failure criteria (degree of hyperopia, myopia, astigmatism, and anisometropia) that yield multiple values of sensitivity for values of specificity. Thus, the screening tests could not all be evaluated by comparing receiver operator characteristic curves. Instead, we chose to compare tests with specificity set at 2 particular levels of interest. A specificity of 0.90 was chosen as a level in the range of interest for mass screening. A specificity of 0.94 was chosen because 2 of the tests with established failure criteria achieved specificity of 0.94. The fact that the ranking of the screening tests is the same at each specificity level greatly reduces the likelihood that our conclusions would change for alternative levels of specificity.

Detection of Children with Targeted Conditions

The results showed that NCR, autorefraction (Retinomax Autorefractor and SureSight Vision Screener), and monoc-

ular Lea Symbols VA testing, when administered by LEPs, were the most accurate in distinguishing children who had ≥1 targeted vision disorders from children who did not have the conditions (Table 8).

When we examined sensitivity (the ability to detect children who have the targeted conditions) and specificity (the ability to identify children who do not have any of the targeted conditions) of each screening test for detection of each of the targeted conditions separately, the most accurate tests for detection of children with amblyopia were NCR, the Retinomax Autorefractor, and the SureSight Vision Screener (Table 10). The most accurate tests for detection of strabismus were monocular HOTV VA testing, the Retinomax Autorefractor, and the Stereo Smile II test. For detection of significant refractive error, the most accurate tests were the same as the best tests for detection of children with amblyopia: NCR, the Retinomax Autorefractor, and the SureSight Vision Screener.

Detection of Children by Severity of the Condition

Although amblyopia, strabismus, and significant refractive error all merit detection and further evaluation by an eye care professional, each condition can range in severity. To take this into account, the VIP Executive Committee developed a hierarchy that ranked the importance of detection of different levels of severity of each of the targeted conditions (Table 3). When each screening test was evaluated with respect to the hierarchy of severity of the targeted conditions, the tests that were best in detecting group 1 conditions (very important to detect and treat early) were NCR, the Retinomax Autorefractor, SureSight Vision Screener, and monocular Lea Symbols test (Table 9). For all screening tests, the sensitivity was lower for detection of group 2 (important to detect early) and group 3 (detection clinically useful) conditions than for detection of group 1 conditions (Table 9). However, the best tests for detecting group 2 conditions were NCR, the Retinomax Autorefractor, SureSight Vision Screener, and monocular Lea Symbols VA test. The best tests for detecting group 3 conditions were the 2 VA tests, Power Refractor II, and Retinomax Autorefractor.

In some screening settings, very high sensitivity may be preferred over high specificity, particularly for group 1 conditions. Differences among tests may be enhanced under this requirement. For example, if detecting 95% of children with group 1 conditions is the criterion, the specificities for NCR, the Retinomax Autorefractor, SureSight Vision Screener, and Lea Symbols VA test are 0.81, 0.70, 0.65, and 0.38, respectively. This ranking of the tests is the same ranking as determined when specificity was set at 0.90 and 0.94.

Refractive Error and the Targeted Conditions

A striking finding of this initial phase of the VIP Study is the accuracy of NCR, the Retinomax Autorefractor, and SureSight Vision Screener in detecting children who have ≥1 targeted conditions, as well as the most severe of these conditions, defined as group 1 in the hierarchy of targeted disorders. It is not surprising that these tests are accurate at picking up high refractive error; however, it is not obvious that these would be appropriate choices for detection of strabismus and strabismic amblyopia. In agreement with previous reports based on other populations of children,[31,32] examination of the VIP study population revealed that both strabismus and strabismic amblyopia were frequently associated with the presence of significant refractive error.

Screening Tests Assessing Visual Acuity and Stereopsis

Traditionally, VA testing has been an integral part of vision screening programs. In this phase of the VIP Study, monocular VA testing with Lea Symbols performed well overall and reasonably well in the detection of amblyopia, strabismus, and significant refractive error, as well as in the detection of group 1 conditions. However, testing VA was no better than NCR, the Retinomax Autorefractor, and SureSight Vision Screener in detecting any of the targeted conditions or in detecting group 1 conditions.

Another integral part of many vision screening programs is assessment of binocularity through cover testing and/or assessment of stereopsis. The cover–uncover test, which is used specifically to detect manifest strabismus, did not perform better in a screening setting than any other test in identifying children with strabismus. The longer period of observation of the child by the GSE examiner, the monocular assessment of VA before cover testing, and the addition of the alternating cover test to the examination sequence may have enhanced the detection of tropias, especially intermittent tropias, during the GSE.

The tests of stereoacuity were not as accurate as NCR, the Retinomax Autorefractor, and SureSight Vision Screener in detecting children with ≥1 targeted conditions or children with group 1 conditions. However, no other screening test was more accurate than the Stereo Smile II test in detection of strabismus. Comparison of the results of the 2 tests of stereoacuity showed that the Stereo Smile II test had higher testability than the RDE test, especially for younger children. However, when children who were unable to perform the test were classified as failing, the 2 tests performed similarly in detecting any targeted condition. The sensitivity of the Stereo Smile II test for detecting children with group 1 conditions was higher than the sensitivity of the RDE test (72% vs. 59%), but not to a statistically significant degree ($P = 0.02$).

Summary

Overall, the results suggest that personnel skilled in NCR can detect approximately two thirds of children with ≥1 targeted disorders, and 90% of those with group 1 conditions, while referring 10% of normal children for an eye examination (90% specificity). In addition, the performance of the SureSight Vision Screener and Retinomax Autorefractor, when used by highly skilled personnel, is similar to that of NCR. However, additional research is needed to provide an evaluation of the autorefractors in the hands of

individuals less experienced in evaluating ocular conditions and in more realistic screening environments.

Results of this study indicate that, in the hands of eye care professionals, the Lea Symbols test was nearly as accurate as NCR and the 2 autorefractors in detecting children who have ≥1 targeted conditions. The accuracy of the HOTV VA test was not as high as that of the Lea Symbols test, but the difference was not statistically significant. An important consideration with respect to screening with VA tests is the length of time needed to obtain monocular VA results in preschool children, which is considerably greater than that needed to obtain results from NCR or autorefraction.

Since the introduction of the first photorefractors in the late 1970s, there has been considerable interest in the use of this technology for large-scale screening of children.[14,33–36] The results of the present study show that the iScreen and MTI photorefractors are not as accurate as NCR, the SureSight and Retinomax autorefractors, or monocular Lea Symbols VA testing in detecting whether a child has ≥1 targeted conditions.

Although the results of this study provide insight into which vision screening tests are most accurate in detecting the most prevalent vision disorders in preschool-aged children, caution should be used in generalizing these results. In this initial phase of the VIP Study, all testing was conducted in the controlled environment of a mobile medical van, and tests were administered by highly skilled personnel (optometrists and pediatric ophthalmologists) using a standardized protocol. In addition, all personnel underwent a certification procedure that included extensive didactic and hands-on instruction by an experienced training team that traveled to all sites and that, in year 2, included representatives of the manufacturers of the Power Refractor II, iScreen and MTI photoscreeners, and SureSight Vision Screener. These features enhance the comparison of the performance of the screening tests performed within years 1 and 2 but may not provide accurate estimates of the sensitivity and specificity of the screening tests when administered in more realistic screening environments by less trained personnel.

The positive predictive value (proportion of children having a targeted condition among those referred for a comprehensive examination) is also an important feature in evaluating screening tests. However, in addition to the above caveats on generalizability of results, this initial phase of the VIP Study tested a sample of children that, by design, had an over-representation of children with vision problems. Nevertheless, if estimates of the prevalence of groups 1, 2, and 3 conditions are extrapolated from VIP Study data, the positive predictive value is approximately 50% to 60% for each of the tests evaluated.

The VIP Study results presented here are relevant to the recommendations on preschool screening by professional, governmental, and private organizations concerned with vision in children.[1,4–6,9–12] Phase I of the VIP Study has provided evaluative data from which policy makers can begin to determine for their constituents the most reasonable approach to identifying preschool children with amblyopia, strabismus, and significant refractive error. More data will be provided as phases II (pediatric nurse screeners and lay screeners in realistic screening environments) and III (primary screening tests in a more general population) are completed. Furthermore, information on prevalence of the targeted conditions and on costs and benefits of early detection needs to be considered in determining the role of preschool vision screening in eliminating preventable loss of vision in our children.

References

1. U.S. Public Health Service. Vision screening in children. Am Fam Physician 1994;50:587–90.
2. Moore B. Eye Care for Infants and Young Children. Boston: Butterworth-Heinemann; 1997:175–89.
3. Ciner EB, Schmidt PP, Orel-Bixler D, et al. Vision screening of preschool children: evaluating the past, looking toward the future. Optom Vis Sci 1998;75:571–84.
4. American Optometric Association Consensus Panel on Pediatric Eye and Vision Examination. Pediatric Eye and Vision Examinations. 2nd ed. St. Louis: American Optometric Association; 2002.
5. American Academy of Ophthalmology Preferred Practice Patterns Pediatric Ophthalmology Panel. Pediatric Eye Evaluations. Preferred Practice Pattern. San Francisco: American Academy of Ophthalmology; 2002.
6. Hartmann EE, Dobson V, Hainline L, et al, Maternal and Child Health Bureau and National Eye Institute Task Force on Vision Screening in the Preschool Child. Preschool vision screening: summary of a task force report. Pediatrics 2000; 106:1105–16.
7. Zaba JN, Johnson RA, Reynolds WT. Vision examinations for all children entering public school—the new Kentucky law. Optometry 2003;74:149–58.
8. Ky 2000 Regular Sess, House Bill No. 706 (as enacted), March 28, 2000.
9. Gerali PS, Flom MC, Raab EL. Report of Children's Vision Screening Task Force. Schaumburg, IL: National Society to Prevent Blindness; 1990.
10. The American Association for Pediatric Ophthalmology and Strabismus. Eye care for the children of America. J Pediatr Ophthalmol Strabismus 1991;28:64–7.
11. National Association of School Nurses. Vision Screening Guidelines for School Nurses. Rev ed. Scarborough, ME: National Association of School Nurses, Inc.; 1995.
12. Committee on Practice and Ambulatory Medicine, Section on Ophthalmology, American Association of Certified Orthoptists, et al. Eye examination in infants, children and young adults by pediatricians. Pediatrics 2003;111:902–7.
13. Kemper AR, Margolis PA, Downs SM, Bordley WC. A systematic review of vision screening tests for the detection of amblyopia. Pediatrics 1999;104:1220–2.
14. Simons K. Preschool vision screening: rationale, methodology and outcome. Surv Ophthalmol 1996;41:3–30.
15. Snowdon SK, Stewart-Brown SL. Preschool vision screening. Health Technol Assess 1997;1(8):i–iv, 1–83.
16. Ehrlich MI, Reinecke RD, Simons K. Preschool vision screening for amblyopia and strabismus. Programs, methods, guidelines, 1983. Surv Ophthalmol 1983;28:145–63.
17. Ciner EB, Dobson V, Schmidt PP, et al. A survey of vision screening policy of preschool children in the United States. Surv Ophthalmol 1999;43:445–57.
18. Green M, Palfrey JS, eds. Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents. 2nd

ed rev. Arlington, VA: National Center for Education in Maternal and Child Health; 2002:307.
19. Morgan KS, Johnson WD. Clinical evaluation of a commercial photorefractor. Arch Ophthalmol 1987;105:1528–31.
20. Ottar WL, Scott WE, Holgado SI. Photoscreening for amblyogenic factors. J Pediatr Ophthalmol Strabismus 1995;32:289–95.
21. el-Defrawy S, Clarke WN, Belec F, Pham B. Evaluation of a hand-held autorefractor in children younger than 6. J Pediatr Ophthalmol Strabismus 1998;35:107–9.
22. Choi M, Weiss S, Schaeffel F, et al. Laboratory, clinical, and kindergarten test of a new eccentric infrared photorefractor (PowerRefractor). Optom Vis Sci 2000;77:537–48.
23. Kennedy RA, Thomas DE. Evaluation of the iScreen digital screening system for amblyogenic factors. Can J Ophthalmol 2000;35:258–62.
24. Schimitzek T, Wesemann W. Clinical evaluation of refraction using a handheld wavefront autorefractor in young and adult patients. J Cataract Refract Surg 2002;28:1655–66.
25. U.S. Department of Health and Human Services, Administration for Children and Families. About Head Start Available at: http://www2.acf.dhhs.gov/programs/hsb/about/index.htm. Accessed November 29, 2003.
26. Hyvärinen L, Näsänen R, Laurinen P. New visual acuity test for preschool children. Acta Ophthalmol (Copenh) 1980;58: 507–11.
27. Liang J, Grimm B, Goelz S, Bille JF. Objective measurement of wave aberrations of the human eye with the use of a Hartmann-Shack wave-front sensor. J Opt Soc Am A Opt Image Sci Vis 1994;11:1949–57.
28. Moke PS, Turpin AH, Beck RW, et al. Computerized method of visual acuity testing: adaptation of the Amblyopia Treatment Study visual acuity testing protocol. Am J Ophthalmol 2001;132:903–9.
29. Holmes JM, Beck RW, Repka MX, et al. The Amblyopia Treatment Study visual acuity testing protocol. Arch Ophthalmol 2001;119:1345–53.
30. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. Am J Epidemiol 1989;130:371–8.
31. Atkinson J, Braddick O, Bobier B, et al. Two infant vision screening programmes: prediction and prevention of strabismus and amblyopia from photo- and videorefractive screening. Eye 1996;10:189–98.
32. Repka MX, Beck RW, Holmes JM, et al, Pediatric Eye Disease Investigator Group. A randomized trial of patching regimens for treatment of moderate amblyopia in children. Arch Ophthalmol 2003;121:603–11.
33. Arnold RW, Gionet EG, Jastrzebski AI, et al. The Alaska Blind Child Discovery project: rationale, methods and results of 4000 screenings. Alaska Med 2000;42:58–72.
34. Donahue SP, Johnson TM, Leonard-Martin TC. Screening for amblyogenic factors using a volunteer lay network and the MTI Photoscreener. Initial results from 15,000 preschool children in a statewide effort. Ophthalmology 2000;107:1637–44, discussion 1645–6.
35. Tong PY, Bassin RE, Enke-Miyazaki E, et al. Screening for amblyopia in preverbal children with photoscreening photographs: II. Sensitivity and specificity of the MTI Photoscreener. Ophthalmology 2000;107:1623–9.
36. Committee on Practice and Ambulatory Medicine, Section on Ophthalmology, American Academy of Pediatrics. Use of photoscreening for children's vision screening. Pediatrics 2002;109:524–5.

Appendix: The Vision in Preschoolers Study Group

Executive Committee

Paulette Schmidt, OD, MS (Chair)*; Elise Ciner, OD†; Lynn Cyert, PhD, OD; Velma Dobson, PhD; Marjean Taylor Kulp, OD, MS; Maureen Maguire, PhD; Bruce Moore, OD; Deborah Orel-Bixler, PhD, OD; Ellen Peskin, MA; Graham Quinn, MD, MSCE; Maryann Redford, DDS, MPH; Janet Schultz, MA, CPNP.

Advisory Committee (Voting Members)

Roy Beck, MD, PhD (Chair); Susan Cotter, OD; Jonathan Holmes, MD; Melvin Shipp, OD, DrPH; Sheila West, PhD.

Writing Committee

Paulette Schmidt, OD, MS (Chair)*; Maureen Maguire, PhD; Velma Dobson, PhD; Graham Quinn, MD, MSCE; Elise Ciner, OD†; Lynn Cyert, PhD, OD; Marjean Taylor Kulp, OD, MS; Bruce Moore, OD; Deborah Orel-Bixler, PhD, OD; Maryann Redford, DDS, MPH; Gui-shuang Ying, MS.

Participating Centers

AA = Administrative Assistant; BPC = Back-up Project Coordinator; GSE = Gold Standard Examiner; LPS = LEP Screener; LS = Lay Screener; NHE = Nurse/Health Coordinator; PC = Project Coordinator; PI = Principal Investigator; PL = Parent Liaison; PR = Programmer; VD = Van Driver.

Berkeley, California. University of California Berkeley School of Optometry. Deborah Orel-Bixler, PhD, OD (PI/GSE/LPS); Pamela Qualley, MA (PC); Dru Howard (BPC/PL); Sarah Fisher, PhD, OD (GSE/LPS); Darlene Fong, OD (GSE); Cindy Hsiao,OD (GSE); Selim Koseoglu, MD (GSE); A. Mika Moy, OD (GSE); Sharyn Shapiro, OD (GSE); Lisa Verdon, OD (GSE); Tonya Watson, OD (GSE); Sara Frane, OD (LPS); Nina Friedman, OD, MS (LPS); Jennifer Seino, OD (LPS); Sean McDonnell (LS/VD); Erika Paez (LS); Coriemae Perea (LS); Darlene Sloan (LS); Evelyn Smith (LS); Leticia Soto (LS); Angela Stelly-Leonard (LS/NHE).

Boston, Massachusetts. New England College of Optometry. Bruce Moore, OD (PI/GSE/LPS); Joanne Bolden (PC); Sandra Umaña (PC); Justin Smith (BPC); Nicole Quinn, OD (GSE/LPS); Nancy Carlson, OD (GSE/LPS); Melissa Suckow, OD (GSE/LPS); Amy Croteau, OD (GSE); Barry Kran, OD (GSE); Jean Ramsey, MD (GSE); Erik Weissberg, OD (GSE); Daniel Kurtz, PhD, OD (LPS); Daniel Laby MD (LPS); Stacy Lyons, OD (LPS); Mar-

*Dr Schmidt was principal investigator for a grant (<$20 000) issued to The Ohio State University Research Foundation from Welch Allyn, Inc. for testing a prototype of the SureSight Vision Screener.

†Pennsylvania College of Optometry receives royalties of less than $10 000 annually on sales of the Stereo Smile test.

thedala Chery (LS/PL); Leticia Gonzalez (LS/PL); Edward Braverman (LS/VD); Susan Crowley (NHE); Paul Dennehy (VD); Benny Jaramillo (VD).

Columbus, Ohio. The Ohio State University College of Optometry. Paulette Schmidt, OD, MS (PI); Marjean Taylor Kulp, OD, MS (Coinvestigator/GSE/LPS); Molly Biddle (PC); Jo Haynes (PC); Jason Hudson (BPC); Kristyne Edwards, OD, MS (GSE/LPS); Heather Gebhart, OD (GSE/LPS); Ann Hickson, OD, MS (GSE/LPS); LeVelle Jenkins,OD (GSE/LPS); Sandra Anderson, OD (GSE); Nancy Evans, OD (GSE); Jay Henry, OD, MS (GSE); Richard Hertle, MD (GSE); Jeffrey Hutchinson, DO (GSE); Andrew Toole, OD, MS (GSE); Michael Earley, OD, PhD (LPS); Sherry Crawford, OD, MS (LPS); Kathy Reuter, OD (LPS); Keith Johnson (LS/VD); Beth Haas (LS); Tonya James (LS); Denise Martin (LS); Sandra Dorton (NHE); Youlanda Grace (NHE); Trina Hisle (NHE); Cheryl Jones (NHE); Betty Smith (NHE); Robert Bower (VD).

Philadelphia, Pennsylvania. Pennsylvania College of Optometry. Elise Ciner, OD (PI/GSE/LPS); Angela Duson (PC/LS); Lydia Parke (BPC); Mark Boas, OD (GSE/LPS); Shannon Burgess, OD (GSE/LPS); Penelope Copenhaven, OD (GSE/LPS); Ellie Francis, PhD, OD (GSE/LPS); Michael Gallaway, OD (GSE/LPS); Graham Quinn, MD (GSE/LPS); Janet Schwartz, OD (GSE/LPS); Brandy Scombordi-Raghu, OD (GSE/LPS); Janet Swiatocha, OD (GSE/LPS); Edward Zikoski, OD (GSE/LPS); Jennifer Lin, MD (GSE); Sheryl Menacker, MD (GSE); Rose Little (LS/PL); Geneva Moss (LS/PL); Jose Figueroa (LS/VD); Barbara Hall (LS); Eric Nesmith (LS); Gwen Gold (BPC/NHE/PL); Deborah Ciner (PL); Elizabeth Jordan (PL); David Harvey (VD).

Tahlequah, Oklahoma. Oklahoma Northeastern State University College of Optometry. Lynn Cyert, PhD, OD (PI/GSE/LPS); Linda Cheatham (PC/VD); Anna Chambless (BPC; PL); Colby Beats, OD (GSE); Debbie Coy, OD (GSE); Jeffrey Long, OD (GSE); Jerry Carter, OD (GSE); Shelly Rice, OD (GSE); James Dunn, OD (LPS); Elisabeth Harrington, OD (LPS); Leslie Trimble, OD (LPS); Shelly Dreadfulwater (LS/PL); Cindy McCully (LS/PL); Rod Wyers (LS/VD); Edith Bingham (LS); Vicky Taylor (LS); Glenda Byfield (PL); Pat Gower (NHE); Kathryn Roastingear (NHE); Elizabeth Ross (NHE).

Study Center: Columbus, Ohio. The Ohio State University College of Optometry. Paulette Schmidt, OD, MS (PI); Beth Haas (Study Coordinator).

Coordinating Center: Philadelphia, Pennsylvania. University of Pennsylvania, Department of Ophthalmology. Maureen Maguire, PhD (PI); Ellen Peskin, MA (Project Director); Mary Brightwell-Arnold (Systems Analyst); Christine Holmes (AA); Andrew James (PR); Aleksandr Khvatov (PR); Lori O'Brien (AA); Claressa Whearry (AA); Gui-shuang Ying, MS (Biostatistician).

National Eye Institute: Bethesda, Maryland. Maryann Redford, DDS, MPH.