

ROC Analysis of the Accuracy of Noncycloplegic Retinoscopy, Retinomax Autorefractor, and SureSight Vision Screener for Preschool Vision Screening

Gui-shuang Ying,¹ Maureen Maguire,¹ Graham Quinn,² Marjean Taylor Kulp,³ Lynn Cyert,⁴ and the Vision In Preschoolers (VIP) Study Group⁵

PURPOSE. To evaluate, by receiver operating characteristic (ROC) analysis, the accuracy of three instruments of refractive error in detecting eye conditions among 3- to 5-year-old Head Start preschoolers and to evaluate differences in accuracy between instruments and screeners and by age of the child.

METHODS. Children participating in the Vision In Preschoolers (VIP) Study ($n = 4040$), had screening tests administered by pediatric eye care providers (phase I) or by both nurse and lay screeners (phase II). Noncycloplegic retinoscopy (NCR), the Retinomax Autorefractor (Nikon, Tokyo, Japan), and the SureSight Vision Screener (SureSight, Alpharetta, GA) were used in phase I, and Retinomax and SureSight were used in phase II. Pediatric eye care providers performed a standardized eye examination to identify amblyopia, strabismus, significant refractive error, and reduced visual acuity. The accuracy of the screening tests was summarized by the area under the ROC curve (AUC) and compared between instruments and screeners and by age group.

RESULTS. The three screening tests had a high AUC for all categories of screening personnel. The AUC for detecting any VIP-targeted condition was 0.83 for NCR, 0.83 (phase I) to 0.88 (phase II) for Retinomax, and 0.86 (phase I) to 0.87 (phase II) for SureSight. The AUC was 0.93 to 0.95 for detecting group 1 (most severe) conditions and did not differ between instruments or screeners or by age of the child.

CONCLUSIONS. NCR, Retinomax, and SureSight had similar and high accuracy in detecting vision disorders in preschoolers across all types of screeners and age of child, consistent with previously reported results at specificity levels of 90% and 94%. (*Invest Ophthalmol Vis Sci.* 2011;52:9658-9664) DOI: 10.1167/iovs.11-8559

From the ¹Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania; ²The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the ³College of Optometry, Ohio State University, Columbus, Ohio; and the ⁴College of Optometry, Northeastern State University, Tahlequah, Oklahoma.

⁵The members of the Vision in Preschoolers Study Group are listed in Appendix A.

Supported by Grants U10EY12534, U10EY12545, U10EY12547, U10EY12550, U10EY12644, U10EY12647, U10EY12648, and R21EY018908 from the National Eye Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD.

Submitted for publication September 8, 2011; revised November 6, 2011; accepted November 11, 2011.

Disclosure: **G. Ying**, None; **M. Maguire**, None; **G. Quinn**, None; **M.T. Kulp**, None; **L. Cyert**, None

Corresponding author: Gui-shuang Ying, 3535 Market Street, Suite 700, Philadelphia, PA 19104; gsyng@mail.med.upenn.edu.

The Vision In Preschoolers (VIP) Study Group completed a two-phase study to evaluate preschool vision-screening tests for the detection of pediatric vision problems, including amblyopia, strabismus, significant refractive error, and reduced visual acuity in the absence of amblyogenic conditions. The multicenter, observational study was supported by the National Eye Institute. Screenings were conducted from 2002 to 2004 in five clinical centers. Phase I of the VIP Study, conducted in two consecutive years, was designed to compare the 11 screening tests administered by trained, certified, licensed eye care professionals (LEPs) in a mobile medical unit (controlled environment across all five clinical centers). Four tests including noncycloplegic retinoscopy (NCR), Retinomax Autorefractor (Nikon, Tokyo, Japan), SureSight Vision Screener (SureSight, Alpharetta, GA), and the Lea symbols VA test had the highest sensitivity for identifying preschool children with VIP-targeted vision conditions when overall specificity was set at either 90% or 94%.^{1,2} Phase II of the VIP Study was designed to compare the performance of the nurse screeners with that of the lay screeners in administering the best-performing preschool vision screening tests identified from phase I. Two screening tests for refraction (Retinomax and SureSight) were administered by trained and certified pediatric nurses and lay people in a Head Start school (typical screening environment). NCR was not included in phase II, because only eye care professionals have the skills and knowledge necessary to administer the test. When specificity was set at 90%, the Retinomax and the SureSight had similar screening performances whether administered by nurses or lay screeners, and performance was similar to the levels achieved when the LEPs screened children using the tests in the mobile medical unit.³ However, the sensitivity of each screening test was evaluated at only 90% or 94% specificity. Because there is a substantial debate concerning the particular specificity level that should be used in evaluating screening tests^{4,5} and because the best choice for specificity level may vary depending on the screening circumstances, analysis using receiver operating characteristic (ROC) curves is useful for evaluating the performance of screening tests over the full range of specificity levels.⁶

The main goal of this study was to comprehensively evaluate the ability of three screening tests of refractive error (NCR, Retinomax, and SureSight) to detect VIP-targeted pediatric vision conditions and to assess whether their performances differ between tests and type of screeners (nurse versus lay) and between younger (3-year-old) and older (4- and 5-year-old) preschoolers.

METHODS

Subjects and Study Procedures

Details of the VIP Study design and screening tests have been published elsewhere and are briefly described here.^{1,3}

VIP participants were preschool children enrolled in Head Start in the five VIP clinical centers (Berkeley, California; Boston, MA; Columbus, OH; Philadelphia, PA; and Tahlequah, OK), with informed consent obtained from their parents or legal guardians, according to the Declaration of Helsinki. Head Start is a federal program for preschool children from low-income families. It promotes school readiness by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social, and other services to enrolled children and families.

Because Head Start admission criteria are defined nationally, screening Head Start children allowed the five VIP clinical centers across the country to test children from populations with similar socioeconomic characteristics. Furthermore, because the Head Start program requires all children to complete a vision screening performed by Head Start personnel or another organization within 45 days of the first day of school, the VIP Study was able to use the results of these initial classroom screenings to select an enriched sample by preferentially recruiting children who had already failed the Head Start local vision screening. All Head Start children who failed their local Head Start screening and a random sample of those who did not fail the local Head Start screening were targeted for informed consent by their parents for enrollment into the VIP Study. The children who completed the process of informed consent subsequently underwent the VIP vision screening tests and a comprehensive eye examination.

All children were 3, 4, or 5 years old when screened. In VIP Study phase I, trained, certified, licensed eye care professionals experienced in working with children tested with NCR, Retinomax, and SureSight in a mobile medical unit (controlled environment across all five clinical centers). In the VIP Study phase II, trained, certified pediatric nurse screeners and lay screeners administered the Retinomax and SureSight on-site at each child's Head Start school (real-world screening environment).

All the screened children had a comprehensive, standardized eye examination incorporating monocular threshold VA testing, cover test-

ing, stereopsis, and cycloplegic refraction. Results from the eye examination were used to classify children with respect to the four VIP-targeted conditions: amblyopia, strabismus, significant refractive error, and unexplained reduced visual acuity (Table 1). These conditions were categorized into three hierarchical groups according to the severity of the conditions. Group 1 conditions are those that it is important to detect and treat early. Group 2 conditions are important to detect early (but with less urgency than those in group 1). Group 3 conditions are less urgent, but nonetheless are clinically useful to detect. Children with more than one of the targeted conditions were included only in the group that corresponded to their most severe condition.

Statistical Analysis

The testability of each screening test was evaluated by defining "not testable" as no measurements obtainable in one or both eyes on a screening test after three attempts per eye. The proportion of screenings not testable was calculated for each screening test and for each age (3, 4, or 5 years). The comparisons of testability between screening tests and among age groups were performed using generalized estimating equations to account for the correlations from screening tests performed on the same children.

We assessed the accuracy of each screening test using ROC curve analysis. An ROC curve plots the sensitivity against the false-positive rate (i.e., $1 - \text{specificity}$), in which each point reflects values obtained at a different cutoff value from a continuous or ordinal measure. ROC analysis provides several advantages over sensitivity and specificity determination for a single cutoff. (1) Sensitivity and specificity are characteristics of a test at a particular cutoff value, whereas the ROC curve provides an overall picture of the characteristics of the test itself. (2) The tradeoff between sensitivity and specificity can be visualized on a ROC plot as the cutoff is shifted. (3) Calculating the area under the ROC curve (AUC) provides a summary of discriminative ability for a

TABLE 1. Hierarchy of VIP Targeted Disorders

Group 1: Very Important to Detect and Treat Early

Amblyopia

Presumed unilateral: ≥ 3 lines' interocular difference, a unilateral amblyogenic factor, and worse eye VA $\leq 20/64$

Suspected Bilateral: a bilateral amblyogenic factor, worse eye VA $< 20/50$ for 3-year-olds or $< 20/40$ for 4- and 5-year-olds, contralateral eye VA worse than 20/40 for 3-year-olds or 20/30 for 4- and 5-year-olds

Strabismus: constant in primary gaze

Refractive error

Severe anisometropia (interocular difference > 2 D hyperopia, > 3 D astigmatism, or > 6 D myopia)

Hyperopia ≥ 5.0 D

Astigmatism ≥ 2.5 D

Myopia ≥ 6.0 D

Group 2: Important to Detect Early

Amblyopia

Suspected Unilateral: 2-line interocular difference and a unilateral amblyogenic factor

Presumed Unilateral: ≥ 3 line interocular difference, a unilateral amblyogenic factor, and worse eye VA $> 20/64$

Strabismus: intermittent in primary gaze

Refractive error

Anisometropia (interocular difference: > 1 D hyperopia, > 1.5 D astigmatism, or > 3 D myopia)

Hyperopia > 3.25 and < 5.0 D and interocular difference (in SE) ≥ 0.5 D

Astigmatism > 1.5 and < 2.5 D

Myopia ≥ 4.0 and < 6.0 D

Group 3: Detection Is Clinically Useful

Unexplained Reduced VA

Bilateral: no bilateral amblyogenic factor, worse eye VA $< 20/50$ for 3-year-olds or $< 20/40$ for 4- and 5-year olds, contralateral eye VA worse than 20/40 for 3-year-olds or 20/30 for 4- and 5-year-olds

Unilateral: no unilateral amblyogenic factor, worse eye VA $< 20/50$ for 3-year-olds or $< 20/40$ for 4- and 5-year-olds or ≥ 2 line difference between eyes (except 20/16 and 20/25)

Refractive Error

Hyperopia > 3.25 D and < 5.0 D AND interocular difference in SE < 0.5 D

Myopia > 2.0 D and < 4.0 D

* Modified from Tables 2 and 3 of Reference 1.

screening test and allows quick comparison of discriminative ability among different screening tests. The AUC has a value from 0.5 to 1.0, where 1.0 represents perfect ability to discriminate between children without vision disorders and children with vision disorders, and 0.5 represents the discrimination resulting from pure chance.⁷ An AUC greater than 0.9 is considered excellent, greater than 0.8 to 0.9 very good, 0.7 to 0.8 good, 0.6 to 0.7 average, and <0.6 poor.⁸

NCR, Retinomax, and SureSight provide measures of refractive error (sphere, cylinder and axis for NCR and Retinomax, sphere and cylinder for SureSight) for each eye. However, our ROC analysis is child specific as in real screening settings—that is, if either eye fails the screening, the child is referred for a comprehensive eye examination. We used multivariate logistic regression techniques to perform the ROC analysis. The independent variables in the multivariate logistic regression model included four child-level values representing the most positive meridian, most negative meridian, cylinder, and difference in spherical equivalent between eyes. The dependent variable in the multivariate logistic regression model was the presence of VIP-targeted vision condition(s) within a child (yes/no). From the multivariate logistic regression model, we first estimated the probability of that a child would have the VIP-targeted vision condition(s), and then all the possible threshold values of the estimated probability were used as the pass/fail criteria for constructing the ROC curve.⁹ In the VIP, refractive error measurements could not be obtained in a small percentage of children (<2%) on a screening test. These were treated as screening failures in our ROC analyses, because the percentage of children with an ocular condition was at least two times higher for untestable children than for those who passed screening.¹⁰

In the VIP Study, the children who had failed the Head Start vision screening were preferentially recruited into the study and therefore were overrepresented. Statistical methods that account for the sampling weights were therefore used to generate accurate estimates of sensitivity and specificity.¹¹ To take the sampling weights into account in the ROC analysis, we used weighted logistic regression, with the weights calculated as the inverse sampling probability specific to each clinical center. The empirical AUC and its 95% confidence interval (CI) were then calculated using the ROC analysis functions available in commercial software (Procedure Logistic; SAS, ver. 9.2; SAS Institute Inc., Cary, NC). When the AUC was compared between screening tests administered to the same children or between two types of screeners (nurse versus lay) evaluating the same children, the statistical method described by DeLong et al.¹² was used to accommodate the correlation between the AUC estimates.

To compare the accuracy of a screening test between age groups, we performed ROC analyses separately for younger (3 years) and older (4 and 5 years) preschoolers and tested the difference in AUC by using the standard *z*-statistic for comparing two independent samples.

The above ROC analyses were performed for detecting any VIP-targeted condition (i.e., group 1, 2, and 3 conditions combined) and

the most severe group 1 conditions only. A two-sided $P < 0.05$ was considered to be statistically significant (SAS ver. 9.2; SAS Institute).

RESULTS

The VIP study enrolled 1142 preschoolers in phase 1 year 1, 1446 preschoolers in phase 1 year 2, and 1452 in phase II. In this enriched sample of 4040 Head Start preschoolers, 27% to 32% had any VIP-targeted condition, and 12% to 15% had group 1 VIP-targeted vision conditions. More than half (54%) of preschoolers were 4 years, 22% were 3 years, and 24% were 5 years of age at screening.

Less than 2% of the children were not testable, defined as no measurements obtained in one or both eyes on a screening test after three attempts per eye (Table 2). NCR was administered by pediatric eye care providers in 1142 children; 9 (0.79%) children were not testable. A total of 5476 Retinomax screenings were administered by pediatric eye care providers, nurse screeners, or lay screeners; no results were obtained in 19 (0.35%) tests. The percentage of SureSight screenings without test results (after three attempts allowed per eye) was 1.27%, which was significantly higher than that of Retinomax ($P < 0.0001$). When the percentage of screenings not testable on a screening test was analyzed for 3-, 4- and 5-year-olds separately, there was no statistically significant difference among age groups for each screening test. However, the percentage of children not testable was significantly higher with SureSight than with Retinomax in each age group ($P < 0.01$).

The ROC curves for detecting any VIP-targeted condition and group 1 conditions were shown in Figures 1 and 2, respectively. The ROC curves for all three tests were very similar when compared in the same VIP year, indicating their similarity in sensitivity and specificity across all possible cutoffs. Although their ROC curves crossed each other at some sensitivity/specificity points, the differences between screening tests in sensitivity at the same specificity all seemed small.

All three screening tests of refractive error had very high discriminatory power for detecting any VIP-targeted condition when administered by either licensed eye care professionals, pediatric nurse screeners, or lay screeners (Table 3). The AUC ranged from 0.83 to 0.88 for detecting any VIP-targeted condition and ranged from 0.93 to 0.95 for detecting most severe group 1 conditions. The differences in AUCs between NCR, the Retinomax, and the SureSight were all very small (with differences in AUC <0.02), particularly when they were administered in the same VIP year by the same type of screeners. None of the differences was statistically significant (all $P > 0.51$). The nurse and lay screeners had equal AUCs for detecting VIP-targeted conditions with the Retinomax and the SureSight.

TABLE 2. Children Not Testable on NCR, Retinomax, and SureSight, Overall and by Age

| Screening Test | Not Testable*/Total Screenings, <i>n</i> (%) | | | | <i>P</i> (Comparison between Age Groups) |
|---|--|----------------|----------------|---------------|--|
| | All Ages Combined | 3-Year-Olds | 4-Year-Olds | 5-Year-Olds | |
| NCR | 9/1142 (0.79) | 1/215 (0.47) | 6/608 (0.99) | 2/319 (0.63) | 0.82 |
| Retinomax† | 19/5476 (0.35) | 7/1258 (0.56) | 8/2950 (0.27) | 4/1268 (0.32) | 0.48 |
| SureSight† | 55/4341 (1.27) | 18/1043 (1.73) | 23/2344 (0.98) | 14/954 (1.46) | 0.24 |
| <i>P</i> for comparison between two screening tests | | | | | |
| NCR vs. Retinomax | 0.11 | 0.86 | 0.09 | 0.51 | |
| NCR vs. SureSight | 0.14 | 0.046 | 0.99 | 0.18 | |
| SureSight vs. Retinomax | <0.0001 | 0.003 | 0.003 | 0.01 | |

* Not testable was defined as no measurements were obtained in one or both eyes after 3 attempts per eye.

† Nurse and lay screeners administered test to the same children in the VIP phase I; thus, screened children were counted twice in calculating the percentage of those not testable.

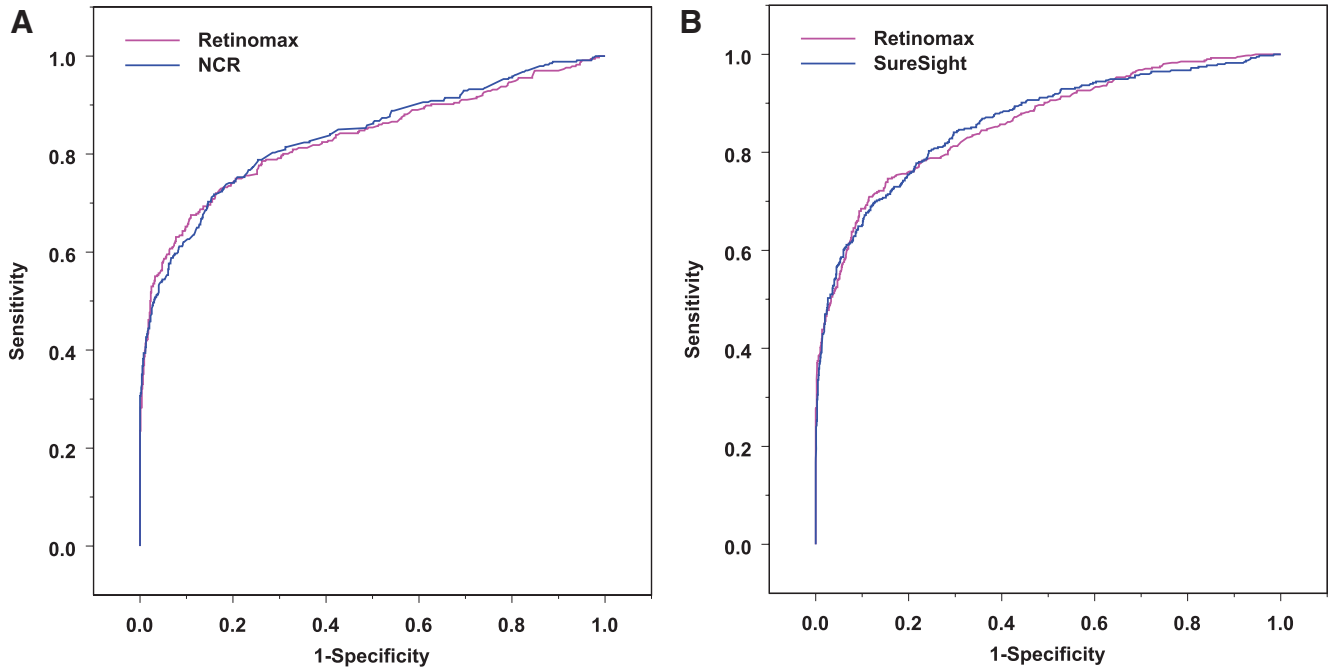


FIGURE 1. ROC curves for detecting any VIP-targeted condition in phase I, year 1 (A) and in phase I, year 2 (B).

Within each age group (3-year-olds vs. 4- and 5-year-olds), similar AUCs were obtained for each screening test for detecting any VIP-targeted condition (Table 4) and the group 1 conditions (data not shown). In both younger and older age groups, the differences in AUC between screening tests or between types of screeners were all small (< 0.08 in AUC), and none of the differences was statistically significant (all $P > 0.06$).

When the AUCs from the same screening tests were compared between the younger and older age groups (3-year-olds vs. 4- and 5-year-olds), all differences in AUC were small, were

within 0.07 for any VIP-targeted condition (Table 4), and were within 0.06 for group 1 conditions (data not shown). For detecting any VIP-targeted condition, the differences in AUC for the Retinomax between younger and older preschoolers in VIP phase II were statistically significant for both the nurse screeners (AUC difference = 0.07; 95% CI, 0.01-0.11) and the lay screeners (AUC difference = 0.05; 95% CI, 0.004-0.10), with 4- and 5-year-olds having higher AUCs than 3-year-olds.

To provide the failure criteria of each screening test, we pooled data from all VIP phases because of the similarity in the ROC across all phases and determined the sensitivities at sev-

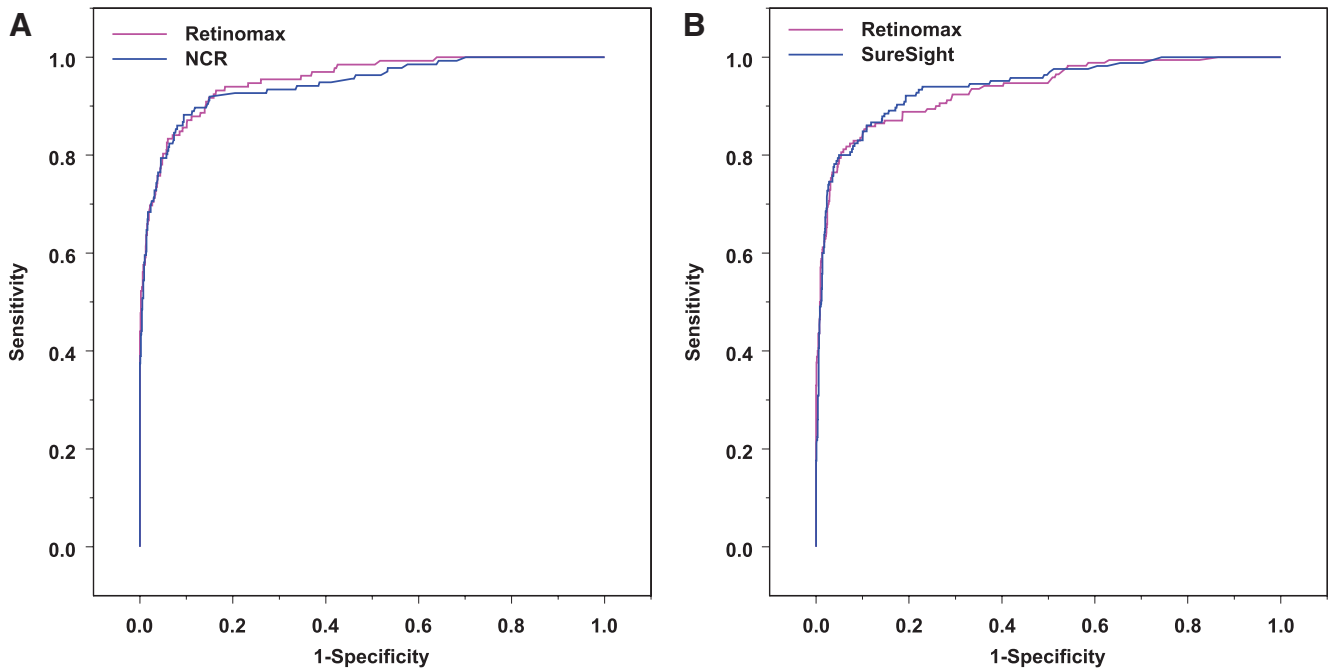


FIGURE 2. ROC curves for detecting VIP group 1 conditions in phase I, year 1 (A) and in phase I, year 2 (B).

TABLE 3. The AUC for Detecting Any VIP-Targeted Condition and Group 1 Conditions

| | Screeners | AUC (95% CI) | |
|---|-----------|----------------------------|--------------------|
| | | Any VIP-Targeted Condition | Group 1 Conditions |
| VIP Phase I, Year 1 (<i>n</i> = 1142) | | (<i>n</i> = 346) | (<i>n</i> = 139) |
| NCR | LEP | 0.84 (0.81-0.87) | 0.95 (0.92-0.97) |
| Retinomax | LEP | 0.83 (0.80-0.86) | 0.95 (0.94-0.97) |
| <i>P</i> for AUC comparison between instruments | | 0.75 | 0.51 |
| VIP Phase I, Year 2 (<i>n</i> = 1446) | | (<i>n</i> = 409) | (<i>n</i> = 172) |
| Retinomax | LEP | 0.86 (0.84-0.88) | 0.93 (0.91-0.96) |
| SureSight | LEP | 0.86 (0.84-0.88) | 0.94 (0.92-0.96) |
| <i>P</i> for AUC comparison between instruments | | 0.93 | 0.59 |
| VIP Phase II (<i>n</i> = 1452) | | (<i>n</i> = 462) | (<i>n</i> = 210) |
| Retinomax | Nurse | 0.88 (0.86-0.90) | 0.95 (0.93-0.97) |
| | Lay | 0.88 (0.86-0.90) | 0.95 (0.93-0.97) |
| <i>P</i> for AUC comparison between screeners | | 0.90 | 0.86 |
| SureSight | Nurse | 0.87 (0.85-0.90) | 0.95 (0.94-0.97) |
| | Lay | 0.87 (0.85-0.89) | 0.95 (0.93-0.97) |
| <i>P</i> for AUC comparison between screeners | | 0.60 | 0.88 |
| <i>P</i> for AUC comparison between instruments | | 0.88 | 0.94 |

LEP, licensed eye care professionals; Nurse, trained and certified pediatric nurse screeners; Lay, study trained lay screeners.

eral specificity levels ranging from 50% to 95%. The sensitivities for detecting any VIP-targeted eye conditions and group 1 conditions and their corresponding failure criteria are provided in Table 5. The cutoffs for failure criteria differed between screening tests, depending on the type of refractive error. For example, the referral criterion for amount of hyperopia was +2.75 D for the NCR, +1.75 D for the Retinomax, and +3.75 D for the SureSight, to achieve 90% overall specificity in detecting any VIP-targeted vision conditions. There were also differences in failure criteria between screening tests for myopia, astigmatism, and anisometropia.

DISCUSSION

By using ROC analysis of the rich data from the VIP Study, we extended the evaluation of the accuracy of the three best-performing tests of refractive error (NCR, Retinomax, and SureSight) in detecting VIP-targeted vision conditions among 3- to 5-year-old Head Start preschoolers. We found that these tests had very good AUCs (0.83-0.88) for detecting any VIP-targeted condition and excellent AUCs (0.93-0.95) for detecting the

most severe group 1 conditions. These three tests of refractive error performed equally well in detecting VIP-targeted vision conditions, extending our report of similarity among the tests when overall specificity was set at 90% and 94%. Nurse screeners and lay screeners performed equally well in administering Retinomax and SureSight. We also found that these tests performed similarly well among younger (3-year-old) and older (4- and 5-year-old) preschoolers in most comparisons.

We previously reported the failure criteria for the three tests of refractive error at 90% and 94% specificity.¹⁻³ These failure criteria were determined by empirically searching for the combinations of four refractive error measures (most positive meridian, most negative meridian, cylinder, and the difference in spherical equivalent between eyes) that maximized the sensitivity at 90% or 94% specificity. We now also provide the failure criteria at specificity levels ranging from 50% to 95%. These failure criteria are the basis for choosing the screening tool and optimal failure criteria to maximize the objective of a particular screening program (e.g., low percentage of referral of normal children or a high rate of detection of children with vision problems). A single, universal set of screening failure criteria

TABLE 4. The AUC for Detecting any VIP-Targeted Eye Conditions by Age

| | Screeners | AUC (95% CI) | | <i>P</i> (AUC Comparison between Age Groups) |
|---|-----------|-------------------|--------------------|--|
| | | 3-Year-Olds | 4- and 5-Year-Olds | |
| VIP Phase I, Year 1 (<i>n</i> = 1142) | | (<i>n</i> = 215) | (<i>n</i> = 927) | |
| NCR | LEP | 0.82 (0.76-0.89) | 0.84 (0.81-0.87) | 0.67 |
| Retinomax | LEP | 0.88 (0.81-0.94) | 0.82 (0.79-0.86) | 0.16 |
| <i>P</i> for AUC comparison between instruments | | 0.24 | 0.33 | |
| VIP Phase I, Year 2 (<i>n</i> = 1446) | | (<i>n</i> = 293) | (<i>n</i> = 1153) | |
| Retinomax | LEP | 0.87 (0.81-0.92) | 0.86 (0.83-0.88) | 0.81 |
| SureSight | LEP | 0.84 (0.78-0.90) | 0.87 (0.84-0.89) | 0.48 |
| <i>P</i> for AUC comparison between instruments | | 0.41 | 0.67 | |
| VIP Phase II (<i>n</i> = 1452) | | (<i>n</i> = 377) | (<i>n</i> = 1075) | |
| Retinomax | Nurse | 0.83 (0.79-0.88) | 0.90 (0.87-0.92) | 0.02 |
| | LAY | 0.85 (0.80-0.89) | 0.90 (0.87-0.92) | 0.03 |
| <i>P</i> for AUC comparison between screeners | | 0.62 | 0.93 | |
| SureSight | Nurse | 0.86 (0.81-0.90) | 0.88 (0.86-0.91) | 0.29 |
| | LAY | 0.85 (0.80-0.89) | 0.88 (0.85-0.90) | 0.21 |
| <i>P</i> for AUC comparison between screeners | | 0.67 | 0.73 | |
| <i>P</i> for AUC comparison between instruments | | 0.45 | 0.29 | |

TABLE 5. The Failure Criteria and their Corresponding Sensitivities at the Various Specificity Levels

| | Hyperopia | Myopia | Astigmatism | Anisometropia | Group 1 Conditions | Any VIP-Targeted Eye Conditions | Specificity for Any VIP-Targeted Eye Conditions |
|------------------|------------------|--------|-------------|---------------|--------------------|---------------------------------|---|
| <i>NCR</i> | 2.00 | 1.00 | 0.50 | 1.00 | 0.98 | 0.88 | 0.50 |
| | 1.75 | 0.25 | 0.75 | 0.75 | 0.96 | 0.84 | 0.60 |
| | 2.00 | 0.25 | 1.00 | 1.00 | 0.96 | 0.81 | 0.70 |
| | 2.50 | 2.00 | 1.00 | 1.00 | 0.96 | 0.76 | 0.80 |
| | 2.50 | 1.25 | 1.25 | 1.00 | 0.92 | 0.71 | 0.85 |
| | 2.75 | 2.75 | 1.25 | 1.50 | 0.90 | 0.64 | 0.90 |
| | 2.50 | 2.00 | 2.00 | 2.00 | 0.85 | 0.56 | 0.95 |
| | <i>Retinomax</i> | 1.50 | 2.00 | 0.75 | 0.75 | 0.96 | 0.90 |
| 0.75 | | 2.50 | 0.75 | 2.25 | 0.96 | 0.88 | 0.60 |
| 0.75 | | 2.50 | 1.00 | 2.25 | 0.95 | 0.83 | 0.70 |
| 1.25 | | 4.00 | 1.00 | 1.75 | 0.92 | 0.77 | 0.80 |
| 1.00 | | 3.75 | 1.25 | 2.25 | 0.91 | 0.73 | 0.85 |
| 1.75 | | 3.75 | 1.25 | 2.75 | 0.87 | 0.68 | 0.90 |
| 1.75 | | 4.00 | 1.75 | 2.75 | 0.83 | 0.58 | 0.95 |
| <i>SureSight</i> | | 2.25 | 0.50 | 0.75 | 1.00 | 0.98 | 0.91 |
| | 3.25 | 0.75 | 0.75 | 1.75 | 0.95 | 0.88 | 0.60 |
| | 2.50 | 0.75 | 1.25 | 1.25 | 0.95 | 0.83 | 0.70 |
| | 3.25 | 0.75 | 1.25 | 2.25 | 0.90 | 0.77 | 0.80 |
| | 4.25 | 0.75 | 1.25 | 3.00 | 0.87 | 0.72 | 0.85 |
| | 3.75 | 0.75 | 1.75 | 2.75 | 0.82 | 0.65 | 0.90 |
| | 5.00 | 1.00 | 2.00 | 4.00 | 0.77 | 0.55 | 0.95 |

does not meet the needs of all screening programs. Screenings conducted by Lions Clubs, pediatricians in their offices, Head Start examiners (serving populations with limited access to the health care system), and school-entrance examiners may best achieve their goals by aiming for different combinations of sensitivity and specificity. Some programs may have the resources to provide treatment for only amblyopia or the most severe vision conditions, whereas others may serve groups with full access to health care and strive also to identify less severe conditions. For example, the Alaska Blind Discovery Project requires very high specificity (few false positives) because transporting a child to a confirmatory examination is very expensive (\$1000) for Alaskans who live in remote villages.⁴ On the other hand, the Head Start program seeks to identify as many children as possible (i.e., higher sensitivity and lower specificity) while children are enrolled in their program, because Head Start can assist with obtaining the needed diagnosis and treatment.

We found that NCR, the Retinomax, and the SureSight performed similarly well in both younger and older preschoolers. These tests require relatively little cooperation from the child. The testability of these three tests was very high (>98%) and similar in younger and older preschoolers, although the SureSight had a higher percentage of children unable to perform the test than did the Retinomax (1.27% vs. 0.35%). The lower maturity of the 3-year-olds did not appear to have any impact on the accuracy of test results. This finding is consistent with our previous finding that 3-year-old children perform as well as 4- to 5-year-old children on the Lea Symbols or HOTV screening tests of visual acuity.¹³

Given the increasingly widespread use of NCR, the Retinomax, and the SureSight in vision screening settings and clinical practice,¹⁴⁻¹⁶ our results on the similarity of their overall screening accuracy imply that there are no appreciable differences in their suitability for screening, whether used with failure criteria calibrated to 90%, 94% or to any other level of specificity. However, users should use the failure criteria particular to each screening test. Selection of the most appropriate test for a particular screening program can be made on the

basis of the cost of the testing and the requirements for the skill and training of the screener.

Acknowledgments

The authors remember the late Velma Dobson, PhD, for her many contributions to the VIP Study.

References

1. Vision in Preschoolers (VIP) Study Group. Comparison of preschool vision screening tests as administered by licensed eye care professionals in the Vision in Preschoolers Study. *Ophthalmology*. 2004;111:637-650.
2. Vision in Preschoolers (VIP) Study Group. Sensitivity of screening tests for detecting vision in preschoolers-targeted vision disorders when specificity is 94%. *Optom Vis Sci*. 2005;82:432-438.
3. Vision in Preschoolers (VIP) Study Group. Preschool vision screening tests administered by nurse screeners compared to lay screeners in the Vision in Preschoolers Study. *Invest Ophthalmol Vis Sci*. 2005;46:2639-2648.
4. Arnold RW. Vision in Preschoolers Study (letter). *Ophthalmology*. 2004;111:2313.
5. Vision in Preschoolers (VIP) Study Group. Reply to: Vision in Preschoolers Study. *Ophthalmology*. 2004;111:2313.
6. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8:283-298.
7. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
8. Choi B. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *Am J Epidemiol*. 1998;148:1127-1132.
9. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *J Am Stat Assoc*. 1993;88:1350-1355.
10. Vision in Preschoolers (VIP) Study Group. Children unable to perform screening tests in Vision in Preschoolers Study: proportion with ocular conditions and impact on measures of test accuracy. *Invest Ophthalmol Vis Sci*. 2007;48:83-87.
11. Roberts G, Rao JNK, Kumar S. Logistic regression analysis of sample survey data. *Biometrika*. 1987;74:1-12.

12. DeLong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.
13. Vision in Preschoolers (VIP) Study Group. Effect of age using Lea Symbols or HOTV for preschool vision screening. *Optom Vis Sci*. 2010;87:87-95.
14. Rowatt AJ, Donahue SP, Crosby C, Hudson AC, Simon S, Emmons K. Field evaluation of the Welch Allyn SureSight Vision Screener: incorporating the VIP Study recommendations. *J AAPOS*. 2007;11:243-248.
15. Steele G, Ireland D, Block S. Cycloplegic autorefraction results in preschool children using Nikon Retinomax Plus and the Welch Allyn SureSight. *Optom Vis Sci*. 2003;80:573-577.
16. Zhao J, Mao J, Luo R, Li F, Pokharel GP, Ellwein LB. Accuracy of noncycloplegic autorefraction in school-age children in China. *Optom Vis Sci*. 2004;81:49-55.

APPENDIX A

The Vision in Preschoolers Study Group

Executive Committee. Paulette Schmidt (Chair), Agnieszka Baumritter, Elise Ciner, Lynn Cyert, Velma Dobson (deceased), Beth Haas, Marjean Taylor Kulp, Maureen Maguire, Bruce Moore, Deborah Orel-Bixler, Ellen Peskin, Graham Quinn, Maryann Redford, Janet Schultz, and Gui-shuang Ying.

Participating Centers

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

School of Optometry, University of California Berkeley, Berkeley, CA. Deborah Orel-Bixler (PI/GSE), Pamela Qualley (PC), Dru Howard (BPC/PL), Lempi Miller Suzuki (BPC), Sarah Fisher (GSE), Darlene Fong (GSE), Sara Frane (GSE), Cindy Hsiao-Threlkeld (GSE), Selim Koseoglu (GSE), A. Mika Moy (GSE), Sharyn Shapiro (GSE), Lisa Verdon (GSE), Tonya Watson (GSE), Sean McDonnell (LS/VD), Erika Paez (LS), Darlene Sloan (LS), Evelyn Smith (LS), Leticia Soto (LS), Robert Prinz (LS), Joan Edelstein (NS), and Beatrice Moe (NS).

New England College of Optometry, Boston, MA. Bruce Moore (PI/GSE), Joanne Bolden (PC), Sandra Umaña (PC/LS/PL), Amy Silbert (BPC), Nicole Quinn (GSE), Heather Bordeau (GSE), Nancy Carlson (GSE), Amy Croteau (GSE), Micki Flynn (GSE), Barry Kran (GSE), Jean Ramsey (GSE), Melissa Suckow (GSE), Erik Weissberg (GSE), Marthedala Chery (LS/PL), Maria Diaz (LS), Leticia

Gonzalez (LS/PL), Edward Braverman (LS/VD), Rosalyn Johnson (LS/PL), Charlene Henderson (LS/PL), Maria Bonila (PL), Cathy Doherty (NS), Cynthia Peace-Pierre (NS), Ann Saxbe (NS), and Vadra Tabb (NS).

College of Optometry, The Ohio State University, Columbus, OH. Paulette Schmidt (PI), Marjean Taylor Kulp (Co-Investigator/GSE), Molly Biddle (PC), Jason Hudson (BPC), Melanie Ackerman (GSE), Sandra Anderson (GSE), Michael Earley (GSE), Kristyne Edwards (GSE), Nancy Evans (GSE), Heather Gebhart (GSE), Jay Henry, MS (GSE), Richard Hertle (GSE), Jeffrey Hutchinson (GSE), LeVelle Jenkins (GSE), Andrew Toole, MS (GSE), Keith Johnson (LS/VD), Richard Shoemaker (VD), Rita Atkinson (LS), Fran Hochstedler (LS), Tonya James (LS), Tasha Jones (LS), June Kellum (LS), Denise Martin (LS), Christina Dunagan (NS), Joy Cline, RN (NS), and Sue Rund (NS).

Pennsylvania College of Optometry, Philadelphia, PA. Elise Ciner (PI/GSE), Angela Duson (PC/LS), Lydia Parke (BPC), Mark Boas (GSE), Shannon Burgess (GSE), Penelope Copenhagen (GSE), Ellie Francis, PhD (GSE), Michael Gallaway (GSE), Sheryl Menacker (GSE), Graham Quinn (GSE), Janet Schwartz (GSE), Brandy Scombordi-Raghu (GSE), Janet Swiatocha (GSE), Edward Zikoski (GSE), Leslie Kennedy (LS/PL), Rosemary Little (LS/PL), Geneva Moss (LS/PL), Latricia Rorie (LS), Shirley Stokes (LS/PL), Jose Figueroa (LS/VD), Eric Nesmith (LS), Gwen Gold (BPC/NHC/PL), Ashanti Carter (PL), David Harvey (LS/VD), Sandra Hall, RN (NS), Lisa Hildebrand (NS), Margaret Lapsley (NS), Cecilia Quenzer (NS), and Lynn Rosenbach (NHC/NS).

College of Optometry, Northeastern State University, Tahlequah, OK. Lynn Cyert (PI/GSE), Linda Cheatham (PC/VD), Anna Chambless (BPC/PL), Colby Beats (GSE), Jerry Carter (GSE), Debbie Coy (GSE), Jeffrey Long (GSE), Shelly Rice (GSE), Shelly Dreadfulwater, (LS/PL), Cindy McCully (LS/PL), Rod Wyers (LS/VD), Ramona Blake (LS/PL), Jamey Boswell (LS/PL), Anna Brown (LS/PL), Jeff Fisher (NS), and Jody Larrison (NS).

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

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

National Eye Institute, Bethesda, MD. Maryann Redford.