Published in final edited form as:

Optom Vis Sci. 2010 February ; 87(2): 96–103. doi:10.1097/OPX.0b013e3181cc8fb9.

Impact of Confidence Number on Accuracy of the SureSight Vision Screener

Corresponding author: Gui-shuang Ying, Scheie Eye Institute, University of Pennsylvania, 3535 Market Street, Suite 700, Philadelphia, PA 19104, gsyng@mail.med.upenn.edu.

The Vision in Preschoolers Study Group

Executive Committee: Paulette Schmidt, OD, MS (Chair); Agnieszka Baumritter, MA; Elise Ciner, OD; Lynn Cyert, PhD, OD; Velma Dobson, PhD; Beth Haas; 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, RN, MA, CPNP; Gui-shuang Ying, PhD.
Writing Committee: Gui-shuang Ying, PhD (Chair), Velma Dobson, PhD; Graham Quinn, MD, MSCE; Maureen Maguire, PhD; Marjean Taylor Kulp, OD, MS; Lynn Cyert, PhD, OD.

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.

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

Boston, MA. *New England College of Optometry:* Bruce Moore, OD (PI/GSE); Joanne Bolden (PC); Sandra Umaña (PC/LS/PL); Amy Silbert (BPC); Nicole Quinn, OD (GSE); Heather Bordeaux, OD (GSE); Nancy Carlson, OD (GSE); Amy Croteau, OD (GSE); Micki Flynn, OD (GSE); Barry Kran, OD (GSE); Jean Ramsey, MD (GSE); Melissa Suckow, OD (GSE); Erik Weissberg, OD (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, RN (NS); Cynthia Peace-Pierre, RN (NS); Ann Saxbe, RN (NS); Vadra Tabb, RN (NS).

Columbus, OH. The Ohio State University College of Optometry: Paulette Schmidt OD, MS (PI); Marjean Taylor Kulp, OD, MS (Co-Investigator/GSE); Molly Biddle, MA (PC); Jason Hudson (BPC); Melanie Ackerman, OD (GSE); Sandra Anderson, OD (GSE); Michael Earley, OD, PhD (GSE); Kristyne Edwards, OD, MS (GSE); Nancy Evans, OD (GSE); Heather Gebhart, OD (GSE); Jay Henry, OD, MS (GSE); Richard Hertle, MD (GSE); Jeffrey Hutchinson, DO (GSE); LeVelle Jenkins, OD (GSE); Andrew Toole, OD, 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, RN (NS); Joy Cline, RN (NS); Sue Rund, RN (NS).

Philadelphia, PA. Pennsylvania College of Optometry: Elise Ciner, OD (PI/GSE); Angela Duson (PC/LS); Lydia Parke (BPC); Mark Boas, OD (GSE); Shannon Burgess, OD (GSE); Penelope Copenhagen, OD (GSE); Ellie Francis, PhD, OD (GSE); Michael Gallaway, OD (GSE); Sheryl Menacker, MD (GSE); Graham Quinn, MD, MSCE (GSE); Janet Schwartz, OD (GSE); Brandy Scombordi-Raghu, OD (GSE); Janet Swiatocha, OD (GSE); Edward Zikoski, OD (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, RN (NS); Margaret Lapsley, RN (NS); Cecilia Quenzer, RN (NS); Lynn Rosenbach, RN (NHC/NS).

Tahlequah, OK. Northeastern State University College of Optometry: Lynn Cyert, PhD, OD (PI/GSE); Linda Cheatham (PC/VD); Anna Chambless (BPC/PL); Colby Beats, OD (GSE); Jerry Carter, OD (GSE); Debbie Coy, OD (GSE); Jeffrey Long, OD (GSE); Shelly Rice, OD (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, RN (NS); Jody Larrison, RN (NS).

Study Center: Columbus, OH. The Ohio State University College of Optometry

Paulette Schmidt, OD, MS (PI); Beth Haas (Study Coordinator).

Coordinating Center: Philadelphia, PA. University of Pennsylvania, Department of Ophthalmology

Maureen Maguire, PhD (PI); Agnieszka Baumritter, MA (Project Director); Mary Brightwell-Arnold (Systems Analyst); Christine Holmes (AA); Andrew James (PR); Aleksandr Khvatov (PR); Lori O'Brien (AA); Ellen Peskin, MA (Project Director); Claressa Whearry (AA); Gui-shuang Ying, PhD (Biostatistician).

National Eye Institute: Bethesda, Maryland

Maryann Redford, DDS, MPH

Presented, in part, at the Association for Research in Vision and Ophthalmology Annual Meeting, Fort Lauderdale, Florida, April 2006.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Vision In Preschoolers Study Group

Abstract

Purpose—To assess the relation between the confidence number provided by the Welch Allyn SureSight Vision Screener and screening accuracy, and to determine whether repeated testing to achieve a higher confidence number improves screening accuracy in preschool children.

Methods—Lay and nurse screeners screened 1,452 children enrolled in the Vision in Preschoolers (VIP) Phase II Study. All children also underwent a comprehensive eye examination. Using statistical comparison of proportions, we examined sensitivity and specificity for detecting any ocular condition targeted for detection in the VIP study and conditions grouped by severity and by type (amblyopia, strabismus, significant refractive error, and unexplained decreased visual acuity) among children who had confidence numbers ≤ 4 (retest necessary), 5 (retest if possible), ≥ 6 (acceptable). Among the 687 (47.3%) children who had repeated testing by either lay or nurse screeners because of a low confidence number (<6) for one or both eyes in the initial testing, the same analyses were also conducted to compare results between the initial reading and repeated test reading with the highest confidence number in the same child. These analyses were based on the failure criteria associated with 90% specificity for detecting any VIP condition in VIP Phase II.

Results—A lower confidence number category was associated with higher sensitivity (0.71, 0.65 and 0.59 for ≤ 4 , 5, and ≥ 6 respectively, $p=0.04$) but no statistical difference in specificity (0.85, 0.85, and 0.91, $p=0.07$) of detecting any VIP-targeted condition. Children with any VIP-targeted condition were as likely to be detected using the initial confidence number reading as using the higher confidence number reading from repeated testing.

Conclusions—A higher confidence number obtained during screening with the SureSight Vision Screener is not associated with better screening accuracy. Repeated testing to reach the manufacturer's recommended minimum value is not helpful in preschool vision screening.

Keywords

confidence number; SureSight Vision Screener; screening accuracy; preschool vision screening

The Welch Allyn SureSight Vision Screener has been identified as a useful tool for screening refractive error in preschool children.¹⁻¹¹ This handheld autorefractor is based on a Shack-Hartmann wavefront analyzer, has a working distance of 35 cm, and can be used by individuals with minimal ophthalmic experience for quick and efficient screening of refractive error.² During measurement of an eye, the SureSight takes five to eight measurements, and then displays a reading for sphere and cylinder, along with a confidence rating from a low of 1 to a high of 9, indicating the number of good readings obtained and their consistency. When the confidence number is below the manufacturer's recommended minimum value of 6, the manufacturer recommends repeat testing if possible for confidence number of 5, and indicates that repetition of the test is necessary for confidence number of 4 or below.¹² When the SureSight is used with *preschool* children, the recommended confidence number can usually be obtained with retesting, but the confidence number of the initial reading is often below the manufacturer's recommended minimum value of 6. For example, in the Vision in Preschoolers (VIP) Phase II Study, 22.1% (23.0% by lay screeners, and 18.6% by nurse screeners) of eyes had a confidence number below 6 in the first attempt using the SureSight, and required repeated testing and the recommended confidence number was attained with retesting for 92.9% of eyes.² In the screening setting, especially when screening preschool children, repeated testing to achieve a higher confidence number can be time-consuming. Therefore, it is important to know whether it is worthwhile to perform repeated testing to achieve a higher confidence number when the first reading is below the manufacturer's recommended minimum value.

Harvey et al. recently reported that accuracy of measurement of astigmatism in three- through seven-year-old children was not significantly different for SureSight values with confidence numbers below 6 versus at or above 6.¹¹ However, to date, there are no reports concerning the impact of confidence number on the accuracy of the SureSight for more general screening of preschool children, nor concerning whether repeated testing to achieve a higher confidence number is worthwhile. The purpose of the present paper is to address the impact of confidence number of a SureSight reading on the screening accuracy (sensitivity and specificity) of the SureSight in preschool children. Specifically, we examine whether repeated testing achieves a higher confidence number and determine whether a higher confidence number achieved during retesting improves screening accuracy. Additionally, we examine whether certain ocular conditions are associated with obtaining a low confidence number.

METHODS

Details of the VIP Phase II Study design have been published previously,² and are thus described only briefly here.

Participants

Three- and four-year-old children (as of September 1, 2003) who were participants in Head Start were invited to enroll into the VIP Phase II Study through five VIP clinical centers (Berkeley, CA; Boston, MA; Columbus, OH; Philadelphia, PA; Tahlequah, OK). Created in 1965, Head Start is a federal government national school readiness program that provides comprehensive education, health, nutrition, and parent involvement services to low-income three- and four-year-old children and their families.

Among the 1,452 children who completed the comprehensive eye examination (Gold Standard Eye Examinations (GSE)), 1,429 completed the screening with SureSight by both lay and nurse screeners and 1,449 completed the screening by either lay or nurse screeners (13 children completed screening by lay screeners only and 7 children completed screening by nurse screeners only). The research followed the tenets of the Declaration of Helsinki and was approved by the institutional review board of each clinical center. Written informed consent was obtained from parents prior to testing of each child.

SureSight Screening

The Welch Allyn SureSight Vision Screener (software version 2.12, Welch Allyn Medical Products, Skaneateles Falls, NY) is a hand-held instrument that measures refractive error monocularly along two meridians from a distance of 35 cm. During measurements, the child is encouraged to look at the fixation stimulus, which is a circle of eight flashing green LEDs surrounding a small, central red light. The SureSight takes five to eight measurements of the right eye, and then displays a reading for sphere and astigmatism along with a confidence number. The procedure is repeated for the left eye, and then the values for both eyes are printed. The confidence number, which indicates the number of good readings obtained and their consistency, ranges from 1 to 9, with higher numbers indicating better reliability.¹² The manufacturer's recommended minimum confidence number is 6. In the VIP Study, up to three readings per eye were performed when an initial confidence number below 6 was obtained. Repeated readings were permitted only on the eye(s) with confidence number less than 6. As recommended by the manufacturer for children younger than age 6 years, the SureSight was used in the "Child" mode, which adds a constant to the sphere value obtained, to correct for the accommodative response of the non-cycloplegic child during testing.

Comprehensive (GSE) Eye Examination

The GSEs were conducted in the VIP vans¹³ by optometrists and ophthalmologists who were experienced in providing care to children and were masked to the results of screening. The GSE included monocular threshold distance visual acuity (VA) assessment with crowded, single H, O, T, V optotypes using the Electronic Visual Acuity system,¹⁴⁻¹⁶ cover testing at distance and near, and cycloplegic retinoscopy, in order to determine whether a child had any of the four VIP targeted conditions (amblyopia, strabismus, significant refractive error, and unexplained reduced VA).^{1,2} Unilateral amblyopia was defined as a 3-line (presumed amblyopia) or a 2-line (suspected amblyopia) interocular acuity difference accompanied by strabismus and/or anisometropia. Reduced VA was defined as VA worse than 20/50 in 3-year-old children and worse than 20/40 in 4-year-old children. Bilateral amblyopia was defined as reduced VA and an amblyogenic factor in each eye (astigmatism $>2.5\text{D}$, hyperopia $>5.0\text{D}$, or myopia $>8.0\text{D}$). Significant refractive error was defined as astigmatism $>1.50\text{D}$, hyperopia $>3.25\text{D}$, myopia $>2.00\text{D}$, or anisometropia ($>1.00\text{D}$ interocular difference in hyperopia; $>3.00\text{D}$ interocular difference in myopia; $>1.50\text{D}$ interocular difference in astigmatism; antimetropic difference $>1.00\text{D}$ and one eye $>1.00\text{D}$ of hyperopia; antimetropic difference $>3.00\text{D}$ and one eye $>2.00\text{D}$ of myopia). GSE results were also used to determine the severity of the conditions, categorized into three hierarchical groups (Group 1, 2 and 3).¹ Group 1 is considered to be most severe, very important to detect, and should be treated early. Included in Group 1 are bilateral amblyopia, presumed unilateral amblyopia with worse eye VA 20/64 or poorer, constant strabismus, hyperopia $\geq 5.0\text{D}$, astigmatism $\geq 2.5\text{D}$, myopia $\geq 6.0\text{D}$, or severe anisometropia (interocular difference $>2\text{D}$ in hyperopia, $>3\text{D}$ in astigmatism, or $>6\text{D}$ in myopia). Included in Group 2 are suspected unilateral amblyopia, presumed unilateral amblyopia with worse eye VA better than 20/64, intermittent strabismus, hyperopia $>3.25\text{D}$ but $<5.0\text{D}$ with interocular difference in spherical equivalent (SE) $\geq 0.5\text{D}$, astigmatism $>1.5\text{D}$ but $<2.5\text{D}$, and myopia $\geq 4.0\text{D}$ but $<6.0\text{D}$. Included in Group 3 are bilateral or unilateral reduced VA, hyperopia $>3.25\text{D}$ but $<5.0\text{D}$ with interocular difference in SE $<0.5\text{D}$, or myopia $>2.0\text{D}$ but $<4.0\text{D}$.

Data Analysis

Two approaches were used to examine the impact of confidence number of SureSight readings on accuracy.

The first (initial confidence number) approach was based on the initial confidence number for the right eye and the initial confidence number for the left eye. On the basis of the lower confidence number of these two readings, each child was classified into one of three groups: confidence number ≤ 4 (retest necessary), 5 (retest if possible), ≥ 6 (acceptable). The failure criteria associated with 90% specificity determined from VIP Phase II were used to classify each child as SureSight screening pass/failure.² For each of the three groups of children, the overall sensitivity and specificity for detecting any VIP targeted condition were calculated, as well as the sensitivity for detecting conditions grouped by severity (Group 1 (most severe), 2, 3 (least severe)) and by type (amblyopia, strabismus, significant refractive error, and unexplained decreased VA). Statistical comparisons for sensitivity and specificity among the three groups of children with increasing confidence number were made using a test for linear trend in proportions.

We also examined the relation between several ocular conditions as determined from the GSE and confidence number by comparing the proportions of children with each ocular condition among the three confidence number groups. The ocular conditions examined were: (i) Any VIP-targeted condition (amblyopia, strabismus, significant refractive error, or unexplained reduced VA), (ii) the three hierarchical groups of conditions based on severity (Groups 1, 2

and 3), (iii) each of the four individual VIP-targeted conditions, (iv) hyperopia, (v) myopia, and (vi) astigmatism.

The second approach (within-subjects comparison of initial versus highest confidence number) was conducted on data from 687 children who had at least two repeated readings by the same screener (either a lay or a nurse screener) because of a low confidence number for one or both eyes on the initial reading. This analysis addressed the question: “Does repeated testing by the same screener to achieve a higher confidence number improve sensitivity/specificity?”

To illustrate each analysis approach used, consider a child with an initial confidence number pair (OD, OS) of (4, 5), and with confidence number pairs from two subsequent testings of (5, 3) and (5, 8). The lower confidence number of the initial confidence number pair (4, 5) is 4, and the child was assigned to the group with confidence number of ≤ 4 . The initial readings associated with confidence number pair (4, 5) were used in defining screening pass/failure for this child in the first (initial confidence number) analysis approach. For the second analysis approach (initial versus highest), the comparison for this child was made between the readings associated with initial confidence number pair of (4, 5) vs. the highest confidence number pair of (5, 8) obtained during retesting.

Initially, analyses were conducted using results of screenings performed by lay screeners and screenings performed by nurse screeners. Because there was no substantial difference between results obtained for lay screeners and nurse screeners, data from lay and nurse screeners were combined, allowing increased statistical power and improved clarity of presentation. For the combined statistical analysis, correlation between results of screenings by lay and nurse screeners in the same child was adjusted by using the Generalized Estimating Equations (GEE).¹⁷ In the execution of GEE, data from each child were identified to the computing algorithm by specifying a unique identification number for each child, and an exchangeable working correlation structure was used to describe the correlation in data between lay screeners and nurse screeners.

Because multiple tests were performed to compare sensitivity of detecting hierarchically grouped conditions, and each of the individual VIP-targeted conditions, the Hochberg procedure (a less conservative, more powerful procedure than Bonferroni method) was used to adjust the p-values from multiple comparisons, and to control the overall type I error (0.05, two-sided).¹⁸ This procedure was conducted using PROC MULTTEST in SAS/STAT 9.1 (SAS Institute, Inc., Cary, NC).

RESULTS

Distribution of Confidence Number

Based on the confidence number from initial testing on each eye of a child, the lower confidence number ranged from 0 to 9, and initial testing provided a reliable measure of refraction (confidence number ≥ 6) in 71.2% of readings of both eyes of a child (Table 1, left column). However, through retesting of eyes with confidence number < 6 on initial testing, the percentage of children with a reliable measure (confidence number ≥ 6) of refraction in each eye was increased to 92.9% (Table 1, middle column). Among 832 readings with initial confidence number < 6 , 650 (78.1%) readings achieved confidence number ≥ 6 through retesting (Table 1, right column).

Comparisons of Screening Accuracy among 3 Groups of Confidence Number

Using the failure criteria that provided 90% specificity,² the comparisons of sensitivity and specificity for children with confidence numbers grouped by the lower confidence number between the two confidence numbers obtained for the left and right eyes of a child are shown

in Table 2. The sensitivity for the detection of any VIP-targeted condition was significantly higher with lower confidence number (0.71 for confidence number ≤ 4 , 0.65 for confidence number of 5, and 0.59 for confidence number ≥ 6 ; adjusted $p=0.04$, linear trend test). However, there was no significant difference in specificity among confidence number categories (adjusted $p=0.07$).

Sensitivity for detecting hierarchically grouped conditions did not differ significantly across confidence number categories after adjustment for multiple comparisons. However, sensitivity for detection of significant refractive error was highest in readings with confidence number ≤ 4 (adjusted $p=0.01$, linear trend test) (Table 2).

Ocular Conditions Associated with Confidence Number

The prevalence of one or more of the VIP-targeted conditions was significantly higher among children with lower confidence number (≤ 4) as measured by either lay or nurse screeners (Table 3; $p<0.0001$). This was also true for prevalence of Group 1 (most severe) conditions ($p<0.0001$), amblyopia ($p=0.02$), strabismus ($p=0.02$) and significant refractive error (Table 3, $p=0.001$).

Further analysis of the association of confidence number with the prevalence of refractive error was performed. The confidence number of SureSight readings was positively associated with the presence of hyperopia, myopia (Figure 1 A and B) but was not associated with presence of astigmatism (Figure 1C). Children with confidence number ≤ 4 as measured either by lay or nurse screeners were more likely to be hyperopic ($p\text{-value}<0.0001$) or myopic ($p=0.047$) than were children with confidence numbers of 5 or ≥ 6 .

Impact of Confidence Number on Screening Accuracy: Results from Repeated Testing

Repeated testing was performed by lay screeners (461 cases) or nurse screeners (371 cases) in 687 children because the confidence number from the initial reading did not reach the manufacturer's recommended value in one or both eyes of the child. Of note, due to the violation of protocol, repeated testing was not performed in 30 cases (17 cases by lay screeners, and 13 cases by nurse screeners) of 29 children, although their initial reading had a confidence number < 6 . Among the 687 retested children, 145 (21.1%) children required retesting by both lay and nurse screeners, 316 (46.0%) required retesting by lay screeners only, and 226 (32.9%) required retesting by nurse screeners only. Testing could be repeated up to two additional times by each screener for a maximum total of 3 readings per eye per child. Among 728 eyes of 461 children who had repeated testing in one or both eyes by lay screeners, one additional reading was obtained in 490 (67.3%) eyes and two additional readings were obtained in 238 (32.7%) eyes; while among 629 eyes of 371 children who had repeated testing by nurse screeners, one additional reading was obtained in 407 (64.7%) eyes and two additional readings were obtained in 222 (35.3%) eyes. The confidence number of the repeated test was approximately 2 units better than initial testing, and 650 (78.1%) of retesting readings reached the manufacturer's recommended minimum value of 6 (Table 1). Children whose readings did not reach the minimum value of 6 by either lay or nurse screeners were significantly more likely than the remaining children in the retested group to have a VIP-targeted condition (45.8% vs. 32.8%, adjusted $p\text{-value}=0.02$), a Group 1 condition (28.6% vs. 14.6%, adjusted $p\text{-value}=0.002$), and significant refractive error (40.0% vs. 26.4%, adjusted $p\text{-value}=0.01$) (Table 4).

Using the 90% specificity failure criteria, sensitivities (based on repeated testing) for detecting any VIP-targeted condition, the hierarchically grouped conditions, and the four VIP-targeted conditions did not change substantially with retesting (range -0.02 to 0.06 , $p>0.05$); nor did specificity (0.90 vs. 0.90) (Table 5).

DISCUSSION

The SureSight Vision Screener is designed to determine refractive status in young children in a screening setting and has been used in several large scale studies.^{1-3, 9, 11} The manufacturer has included in the design of the instrument an estimate of the reliability of the measurements obtained and suggests that readings with a confidence number of 6 or greater (on a scale from 1 to 9) should be used to determine whether a child should be referred. The results of the present study indicate that even confidence numbers of less than 6 should be taken into consideration because such “unacceptable” confidence numbers may be an indicator that the child has an eye disorder that warrants a follow-up eye examination. Compared to a child with a confidence rating of 6 or more, a child with a confidence reading ≤ 4 was approximately 50% more likely to have one or more of the VIP targeted eye condition (amblyopia, strabismus, significant refractive error, or unexplained poor VA), about twice as likely to have a Group 1 (serious) eye condition, about 50% more likely to have amblyopia, about twice as likely to have strabismus, and about three times more likely to have significant refractive error, these increases are all statistically significant (Table 3). The results also indicate that, for this population of preschool children, there is no benefit in terms of accuracy of screening (sensitivity and specificity) to repeating the SureSight measurement to obtain a confidence number of 6 or greater (Table 5).

The findings of the current study are in agreement with the findings of Silverstein et al.,³ in which the SureSight was used to screen a population of more than 15,000 preschool children. Testability was very high (>99%) and, using a modification of the VIP referral criteria,⁹ the referral rate was 7.3%, with a positive predictive value of 48.2% for all children. When data were analyzed from only those children who had confidence numbers of less than 6 or estimated refractive errors exceeding the range of the instrument, the positive predictive value was greater than 70%.

In contrast to an analysis of the effect of confidence number on accuracy of screening with the Retinomax Autorefractor, which showed an increase in accuracy of screening (improvement in specificity) when measurements were repeated to obtain a confidence number at or above the manufacturer’s recommended value,¹⁹ the present analysis of SureSight Vision Screener data showed no improvement in either sensitivity or specificity resulting from repeated measurements that produced confidence numbers of 6 or greater (Table 5). This finding is in agreement with recent results from Harvey et al¹¹ indicating no difference in accuracy of SureSight measurements of astigmatism for confidence values <6 versus ≥ 6 .

The result from this study showed that repeated measures to achieve high confidence number did not improve the screening sensitivity in preschool children, implying that it may not be necessary to perform repeated measures if the confidence number is less than 6 in the initial measure. However, VIP data can not address the question of whether, among those who have confidence number ≥ 6 in the initial measure, repeated measures to obtain an even higher confidence number would improve the screening accuracy, because in the VIP Study, the repeated measures were performed only on eyes that had confidence number less than 6 on the initial measure. However, given that a single SureSight Vision Screener result is based on several assessments by the machine and the fact that repeated readings did not improve screening accuracy when the result was associated with a low confidence number, we do not think that repeat readings would yield substantially improved accuracy.

What are the implications of the results of this study? When using the SureSight Vision Screener in preschool children, the results suggest that it may be reasonable to take the first set of readings obtained from the child even if the confidence number is less than 6. This would simplify testing because repeat testing would not be needed. Furthermore, there does not appear

to be a disadvantage to this because sensitivity for detection of serious eye problems is higher for SureSight measurements with a confidence value of <6 than for SureSight readings with a confidence rating of ≥ 6 , and sensitivity for these children with confidence value of <6 did not improve with repeated testing.

Acknowledgments

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

REFERENCES

- Schmidt P, Maguire M, Dobson V, Quinn G, Ciner E, Cyert L, Kulp MT, Moore B, Orel-Bixler D, Redford M, Ying GS. Comparison of preschool vision screening tests as administered by licensed eye care professionals in the Vision In Preschoolers Study. *Ophthalmology* 2004;111:637–50. [PubMed: 15051194]
- The Vision in Preschoolers Study Group. Preschool vision screening tests administered by nurse screeners compared with lay screeners in the vision in preschoolers study. *Invest Ophthalmol Vis Sci* 2005;46:2639–48. [PubMed: 16043831]
- Silverstein E, Lorenz S, Emmons K, Donahue SP. Limits on improving the positive predictive value of the Welch Allyn SureSight for preschool vision screening. *J AAPOS* 2009;13:45–50. [PubMed: 18976944]
- Buchner TF, Schnorbus U, Grenzebach UH, Busse H. Examination of preschool children for ametropia: first experiences using a new hand-held autorefractor. *Strabismus* 2004;12:111–7. [PubMed: 15672934]
- Steele G, Ireland D, Block S. Cycloplegic autorefraction results in pre-school children using the Nikon Retinomax Plus and the Welch Allyn SureSight. *Optom Vis Sci* 2003;80:573–7. [PubMed: 12917576]
- Cordonnier M, De Maertelaer V. Comparison between two hand-held autorefractors: the Sure-Sight and the Retinomax. *Strabismus* 2004;12:261–74. [PubMed: 15545144]
- Iuorno JD, Grant WD, Noel LP. Clinical comparison of the Welch Allyn SureSight handheld autorefractor versus cycloplegic autorefraction and retinoscopic refraction. *J AAPOS* 2004;8:123–7. [PubMed: 15088044]
- Kemper AR, Keating LM, Jackson JL, Levin EM. Comparison of monocular autorefraction to comprehensive eye examinations in preschool-aged and younger children. *Arch Pediatr Adolesc Med* 2005;159:435–9. [PubMed: 15867116]
- Rowatt AJ, Donahue SP, Crosby C, Hudson AC, Simon S, Emmons K. Field evaluation of the Welch Allyn SureSight vision screener: incorporating the vision in preschoolers study recommendations. *J AAPOS* 2007;11:243–8. [PubMed: 17140828]
- Rogers DL, Neely DE, Chapman JB, Plager DA, Sprunger DT, Sondhi N, Roberts GJ, Ofner S. Comparison of the MTI Photoscreener and the Welch-Allyn SureSight autorefractor in a tertiary care center. *J AAPOS* 2008;12:77–82. [PubMed: 18314072]
- Harvey EM, Dobson V, Miller JM, Clifford-Donaldson CE, Green TK, Messer DH, Garvey KA. Accuracy of the Welch Allyn SureSight for measurement of magnitude of astigmatism in 3- to 7-year-old children. *J AAPOS* 2009;13:466–71. [PubMed: 19840726]
- Welch Allyn SureSight Vision Screener Operation Manual. [Accessed November 9, 2009]. Available at: <http://www.welchallyn.com/products/en-us/x-11-ac-100-000000001207.htm>
- The Vision in Preschoolers (VIP) Study Group. Development and implementation of a preschool vision screening program in a mobile setting. *NHSA Dialog* 2005;1:16–24.
- Holmes JM, Beck RW, Repka MX, Leske DA, Kraker RT, Blair RC, Moke PS, Birch EE, Saunders RA, Hertle RW, Quinn GE, Simons KA, Miller JM. The amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol* 2001;119:1345–53. [PubMed: 11545641]
- Moke PS, Turpin AH, Beck RW, Holmes JM, Repka MX, Birch EE, Hertle RW, Kraker RT, Miller JM, Johnson CA. Computerized method of visual acuity testing: adaptation of the amblyopia

treatment study visual acuity testing protocol. *Am J Ophthalmol* 2001;132:903–9. [PubMed: 11730656]

16. Kulp MT, Dobson V, Peskin E, Quinn G, Schmidt P, The Vision in Preschoolers (VIP) Study Group. The electronic visual acuity tester: testability in preschool children. *Optom Vis Sci* 2004;81:238–44. [PubMed: 15097765]
17. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
18. Hochberg Y. A sharper Bonferroni procedure for multiple significance testing. *Biometrika* 1988;75:800–3.
19. The Vision in Preschoolers (VIP) Study Group. Impact of confidence number on the screening accuracy of the retinomax autorefractor. *Optom Vis Sci* 2007;84:181–8. [PubMed: 17435531]

Figure 1.

Confidence number vs. child's refraction. Confidence number groups were defined based on the lower number of two readings from each child: (1) the reading from the child's right eye and (2) the reading from the child's left eye. Confidence number ≤ 4 group includes children whose confidence number was ≤ 4 for testing conducted by either lay or nurse screeners, Confidence number equal to 5 group includes children whose confidence number was 5 from one screener and ≥ 5 from another screener, while ≥ 6 group includes children whose confidence number was ≥ 6 both for testing conducted by lay screeners and testing conducted by nurse screeners. Refraction was determined based on the cycloplegic retinoscopy, conducted as part of a comprehensive eye examination. Confidence number is significantly associated with

hyperopia ($p < 0.0001$, Figure A) and myopia ($p = 0.047$, Figure B), but not significantly associated with astigmatism ($p = 0.51$, Figure C).

Table 1

Distribution of confidence numbers of SureSight readings by lay and nurse screeners on 1,449 children (Number of readings=2,878 including 1,442 readings by lay screeners and 1,436 readings by nurse screeners) and on 687 children who underwent retesting (Number of readings=832 including 461 readings by lay screeners and 371 readings by nurse screeners)

Confidence Number	Lower Confidence Number between Eyes of a Child on Initial Testing		Lower Confidence Number between Eyes, Based on Highest Confidence Number Obtained for Each Eye for All Readings from a Child		Lower Confidence Number between Eyes, Based on Highest Confidence Number of Repeated Readings among retested Children [‡]	
	n	%	n	%	n	%
0*	4	0.14				
1	3	0.10				
2	47	1.63	11	0.38	9	1.08
3	82	2.85	16	0.56	20	2.40
4	263	9.14	52	1.81	59	7.09
5	463	16.1	126	4.38	94	11.3
6	665	23.1	995	34.6	327	39.3
7	775	26.9	1,002	34.8	225	27.0
8	453	15.7	543	18.9	88	10.6
9	123	4.27	133	4.62	10	1.20
Total	2,878		2,878		832	

* Although the SureSight manufacturer user's manual states that the confidence number ranges from 1 to 9, confidence number of 0 was seen in 4 readings from 4 different children by different screeners. The manufacturer did not answer our inquiry concerning a confidence number of 0.

[‡] Retesting was not performed in 30 cases, in which the initial confidence number was less than 6.

Table 2

Comparison of sensitivity and specificity for the SureSight among three confidence number groups based on the lower confidence number between eyes for each child from initial testing

Confidence number of reading [*]						
	Children (N)	≤ 4 (n=399)	5 (n=463)	≥ 6 (n=2,016)	P-value [†]	Adjusted p-value [‡]
Sensitivity						
Any Condition	459	0.71	0.65	0.59	0.005	0.04
Group 1	207	0.84	0.85	0.80	0.26	0.52
Group 2	144	0.63	0.59	0.54	0.59	0.59
Group 3	108	0.46	0.28	0.31	0.13	0.51
Amblyopia	98	0.89	0.82	0.75	0.03	0.18
Strabismus	47	0.61	0.53	0.43	0.17	0.51
Refractive Error	377	0.80	0.73	0.66	0.001	0.01
Reduced VA	117	0.62	0.58	0.49	0.15	0.51
Specificity [§]	990	0.85	0.85	0.91	0.01	0.07

^{*} Lower confidence number of two readings from each child: (1) the reading from the child's right eye and (2) the reading from the child's left eye.

[§] From linear trend test, with correlation of measures from same child adjusted by GEE.

[‡] Adjusted by the Hochberg procedure.¹⁸

[†] Using the failure criteria at 90% specificity determined from Phase II of VIP.²

Table 3

Comparison of prevalence of ocular conditions among three confidence number groups of children based on the lower confidence number for a child from initial testing

Ocular conditions	Children (N)	Confidence number of reading*			P-value [§]	Adjusted p-value [‡]
		≤ 4 in either lay or nurse screener (n=355) n (%)	5 from one lay or nurse screener, and ≥ 5 from another screener (n=350) n (%)	≥ 6 by both lay and nurse screener (n=744) n (%)		
Any Condition	459	144 (40.6)	111 (31.7)	204 (27.4)	<0.0001	<0.0001
Group 1	207	80 (22.5)	50 (14.3)	77 (10.4)	<0.0001	<0.0001
Group 2	144	30 (8.45)	39 (11.1)	75 (10.1)	0.48	0.48
Group 3	108	34 (9.58)	22 (6.29)	52 (6.99)	0.21	0.42
Amblyopia	98	121 (34.1)	90 (25.7)	166 (22.3)	0.003	0.02
Strabismus	47	38 (10.7)	23 (6.57)	37 (4.97)	0.004	0.02
Refractive Error	377	21 (5.97)	11 (3.14)	15 (2.02)	0.0002	0.001
Reduced VA	117	37 (10.4)	26 (7.43)	54 (7.26)	0.19	0.42

* Lower confidence number of two readings from each child: (1) the reading from the child's right eye and (2) the reading from the child's left eye.

[§] Fisher's exact test.

[‡] Adjusted by the Hochberg procedure.¹⁸

Table 4

Comparison of prevalence of ocular conditions between two groups of children based on the lower confidence number for a child in retesting

	Children (N)	Confidence number of reading*			Adjusted p-value [‡]
		<6 in retesting by either lay or nurse screener (n=168) n (%)	≥ 6 in retesting by both lay and nurse screeners (n=519) n (%)	P-value [§]	
Any Condition	247	77 (45.8)	170 (32.8)	0.003	0.02
Group 1	124	48 (28.6)	76 (14.6)	0.0002	0.002
Group 2	69	19 (11.3)	50 (9.63)	0.56	0.56
Group 3	54	10 (5.95)	44 (8.48)	0.33	0.56
Amblyopia	58	22 (13.1)	36 (6.94)	0.02	0.10
Strabismus	30	9 (5.42)	21 (4.05)	0.51	0.56
Refractive Error	204	67 (40.0)	137 (26.4)	0.001	0.01
Reduced VA	62	18 (10.7)	44 (8.48)	0.44	0.56

* Lower confidence number of two readings from each child: (1) the reading with the highest confidence number from the child's right eye and (2) the reading with the highest confidence number from the child's left eye.

§ Fisher's Exact test.

‡ Adjusted by the Hochberg procedure.¹⁸

Table 5

Comparison of sensitivity and specificity between readings with confidence number <6 in initial testing vs. confidence number based on the lower of (1) the highest confidence number reading from the right eye and (2) the highest confidence number from the left eye, based on repeated testing in 687 children in whom >1 SureSight readings were taken in one or both eyes by lay or nurse screeners

	Children (N)	Reading with confidence number <6 in initial testing (N=832 readings)	Reading with highest confidence from repeated testing* (N=832 readings)	P-value [§]
Sensitivity				
Any Condition	247	0.61	0.63	0.29
Group 1	124	0.82	0.83	0.65
Group 2	69	0.45	0.51	0.17
Group 3	54	0.30	0.31	0.83
Amblyopia	58	0.82	0.88	0.16
Strabismus	30	0.51	0.49	0.78
Refractive Error	204	0.69	0.71	0.34
Reduced VA	62	0.49	0.50	0.81
Specificity [†]				
	440	0.90	0.90	0.67

* Reading is the lower confidence number of two readings from each child: (1) the reading with the highest confidence number from the child's right eye and (2) the reading with the highest confidence number from the child's left eye, based on the repeated testing.

[§]From the chi-square test with correlation between measures from same children adjusted by GEE.

[†]Using the failure criteria at 90% specificity determined from Phase II of VIP.²