Published in final edited form as:
Optom Vis Sci. 2013 October ; 90(10): 1128–1137. doi:10.1097/OPX.0000000000000025.

Inter-tester Agreement in Refractive Error Measurements

Jiayan Huang, MS, **Maureen G. Maguire, PhD**, **Elise Ciner, OD**, **Marjean T. Kulp, OD, MS**, **Graham E. Quinn, MD, MSCE**, **Deborah Orel-Bixler, OD, PhD**, **Lynn A. Cyert, OD, PhD**, **Bruce Moore, OD**, **Gui-Shuang Ying, PhD**, and **for the Vision In Preschoolers (VIP) Study Group**[a]

Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania, Philadelphia, Pennsylvania (JH, MGM, G-SY), Salus University, Elkins Park, Pennsylvania (EC), 3College of Optometry, Ohio State University, Columbus, Ohio (MTK), 4The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (GEQ), School of Optometry, University of California, Berkeley, Berkeley, California (DO-B), 6College of Optometry, Northeastern State University, Tahlequah, Oklahoma (LAC), and 7New England College of Optometry, Boston, Massachusetts (BM)

Abstract

Purpose—To determine the inter-tester agreement of refractive error measurements between lay and nurse screeners using the Retinomax Autorefractor (Retinomax) and the SureSight Vision Screener (SureSight).

Methods—Trained lay and nurse screeners measured refractive error in 1452 preschoolers (3- to 5-years old) using the Retinomax and the SureSight in a random order for screeners and instruments. Inter-tester agreement between lay and nurse screeners was assessed for sphere, cylinder and spherical equivalent (SE) using the mean difference and the 95% limits of agreement. The mean inter-tester difference (lay minus nurse) was compared between groups defined based on child's age, cycloplegic refractive error, and the reading's confidence number using analysis of variance. The limits of agreement were compared between groups using the Brown-Forsythe test. Inter-eye correlation was accounted for in all analyses.

Results—The mean inter-tester differences (95% limits of agreement) were −0.04 (−1.63, 1.54) Diopter (D) sphere, 0.00 (−0.52, 0.51) D cylinder, and −0.04 (1.65, 1.56) D SE for the Retinomax; and 0.05 (−1.48, 1.58) D sphere, 0.01 (−0.58, 0.60) D cylinder, and 0.06 (−1.45, 1.57) D SE for the SureSight. For either instrument, the mean inter-tester differences in sphere and SE did not differ by the child's age, cycloplegic refractive error, or the reading's confidence number. However, for both instruments, the limits of agreement were wider when eyes had significant refractive error or the reading's confidence number was below the manufacturer's recommended value.

Conclusions—Among Head Start preschool children, trained lay and nurse screeners agree well in measuring refractive error using the Retinomax or the SureSight. Both instruments had similar inter-tester agreement in refractive error measurements independent of the child's age. Significant refractive error and a reading with low confidence number were associated with worse inter-tester agreement.

Corresponding author: Gui-shuang Ying, 3535 Market Street, Suite 700, Philadelphia, PA 19104, gsying@mail.med.upenn.edu.
[a]The members of the Vision in Preschoolers Study Group are listed in the Acknowledgments.

Presented in part at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, May 8, 2012.

Conflict of Interest: None.

Keywords

refractive error; Inter-tester agreement; preschool vision screening

The Vision In Preschoolers (VIP) Phase I Study established that when licensed eye care professionals (LEPs) administered 11 common screening tests in a controlled environment, the Retinomax Autorefractor (Retinomax) and the Welch Allyn SureSight Vision Screener (SureSight) were two of the four best performing tests, along with noncycloplegic retinoscopy and a Lea Symbols Distance Visual Acuity test (Precision Vision, Inc) for identifying targeted vision disorders in preschool children.[1] The VIP Phase II Study established that the Retinomax and the SureSight had similar screening performance in identifying vision disorders in preschool children when administered by either trained nurses or lay screeners, and their performance was similar to that of LEPs in Phase I.[2] Both the Retinomax and the SureSight are hand-held autorefractors, provide quick measurement of refractive error without need for cycloplegia, and are designed for use by both ophthalmic clinicians and individuals with minimal ophthalmic experience.[3]

The effectiveness of the Retinomax and the SureSight for identifying vision disorders has been well established,[4–8] and these instruments are already used in vision screening and clinical practice by clinicians, nurses and lay screeners. A few studies [9–14] have investigated the intra-tester agreement (i.e. agreement when administered by the same tester) and demonstrated that both the Retinomax and the SureSight had very good intra-tester agreement. However, the inter-tester agreement of these instruments for measuring refractive error has not yet been evaluated.

Assessing inter-tester agreement is important from both a clinical and research perspective, because children may undergo several vision screenings during their preschool years, possibly by different screening personnel of varied training and experience. It is essential that a screening test provide reproducible results, performed either by the same or different tester. Clinicians and researchers may also be interested in assessing change in refractive error over time and need guidelines on when real change may have occurred. In the VIP Phase II Study, trained lay and pediatric nurse screeners conducted vision screening using the Retinomax and the SureSight for each eye of 1452 preschoolers, thus provided us an excellent opportunity to evaluate the inter-tester agreement of refractive error for both instruments. The purpose of this paper is to evaluate inter-tester agreement between trained lay screeners and pediatric nurse screeners for measuring sphere, cylinder and spherical equivalent in a very large sample of 3- to 5-year-old preschool children (N=1,452).

MATERIALS AND METHODS

The VIP Study was a multi-center, cross-sectional, two-phased study sponsored by the National Eye Institute, to evaluate the effectiveness of vision screening tests in identifying preschool children who would benefit from a comprehensive eye examination. Phase II of the VIP Study was designed to compare the performance of lay screeners and nurse screeners in administering preschool vision screening tests. Details of the VIP Study have been published previously,[1,2] only the details of the screening tests (the Retinomax and the SureSight) related to this paper are described here.

Participants

Three- to five-year-old children attending Head Start were invited to enroll in the VIP Phase II Study through five clinical centers (Berkeley, CA; Boston, MA; Columbus, OH; Philadelphia, PA; Tahlequah, OK). All Head Start children who failed their local Head Start

screening and a random sample of those who did not fail the screening were invited to enroll in the VIP Study. Written informed consent was obtained from parents prior to screening each child. The research followed the tenets of the Declaration of Helsinki and was approved by the institutional review board of each clinical center.

Selection and Training of Screeners

Lay screeners were individuals with at least a high school degree and had at least 2 years of experience working with young children. Nurse screeners either were either pediatric nurses or had at least 3 years of experience in a pediatric setting. Sixteen lay screeners and fifteen nurse screeners conducted the screening testing.

All screeners completed a day-long, local training program, conducted by a team of VIP Study personnel. The program included an overview of the VIP Study, instruction and practice with each screening instrument, and review of data collection procedures, research ethics and confidentiality. After several practice screening sessions, lay and nurse screeners were observed by the local principal investigator or co-investigator while testing at least two children 3 to 5 years of age. All screeners completed human subjects training, passed written knowledge assessments, and were certified as screeners for the VIP Study.

Screening Instruments

Two hand-held autorefractors, the Retinomax Autorefractor (Nikon Retinomax K, Nikon Inc, Tokyo, now manufactured by Righton Ophthalmic Instruments, Tokyo) and the SureSight Vision Screener (software version 2.12, Welch Allyn, Inc.) as described below, were used to measure refractive error of the right eye and then the left eye.

The Retinomax Autorefractor

The Retinomax Autorefractor measures refractive error in each eye along two meridians. Measurements can be made in auto measurement mode, continuous measurement mode, or quick mode. Auto measurement mode was used in the VIP Study. The screener placed the instrument's headrest on the child's forehead, encouraged the child to fixate the internal target, and focused the mire in the center of the right pupil while up to eight measured values were taken automatically by the autorefractor. The screener then repeated the process for the left eye. Based on up to eight measured values, the instrument calculated a single representative reading (sphere, cylinder and axis) for each eye along with a confidence number. The confidence number indicates the variability of measured values, ranging from 1 to 10, with larger confidence numbers indicating better reliability (i.e., lower variability). If there are fewer than three valid measured values, the confidence number cannot be calculated for a reading and "E" (Error) is automatically shown instead of a confidence number. The manufacturer's recommended minimum confidence number is 8. Per the protocol of the VIP Study, up to three repeated readings per eye were performed in order to achieve a confidence number of 8 or higher. The repeated testing was performed only on the eye(s) with an initial confidence number below 8. If the confidence number from all three readings was below 8, no further testing was performed, and the reading with the highest confidence number was used.

The possible range of the Retinomax measurements is −18 to +23 Diopters (D) for sphere, −12 to 12 D for cylinder and 0 to 180 degree for axis. Measurements are displayed in increments of 0.25 D for sphere and cylinder, and 1 degree for axis.

The SureSight Vision Screener

The SureSight Vision Screener measures refractive error in each eye along two meridians from a distance of 35 cm. Measurements can be made in 'child mode' or 'adult mode'. Child

mode was used (as recommended by the manufacturer for children younger than age 6) to add a constant to the sphere value obtained for correcting the accommodative response with non-cycloplegic testing. During the test, the screener encouraged the child to look at an internal fixation stimulus, a circle of eight flashing green LEDs surrounding a small central red light, while up to eight measured values were taken by the SureSight for the right eye. The screener then repeated the process for the left eye. Base on up to eight measured values, the instrument calculated a single representative reading (sphere and cylinder) for each eye along with a confidence number. The confidence number indicates the variability of measured values, ranging from 1 to 9, with larger confidence numbers indicating better reliability (i.e., lower variability). The manufacturer's recommended minimum confidence number is 6 for the SureSight. In the VIP Study, up to three repeated readings per eye were taken when an initial confidence number was below 6. The repeated testing was performed only on the eye(s) with an initial confidence number below 6. If the confidence number from all three readings was below 6, no further testing was performed, and the reading with the highest confidence number was used.

The possible range of the SureSight measurements is −5.0 to +6.0 D for sphere, −4 to 4 D for cylinder. A +9.99 or −9.99 is used to indicate a reading outside the unit's measurement range. Measurements are displayed in increments of 0.125 D for sphere and cylinder. The screening version of the SureSight, which does not provide the cylinder axis, was used in the VIP Study.

Screening Environment and Procedures

Screenings were performed inside local Head Start centers, in areas provided by each school, such as classrooms, hallways, cafeterias or nurses' offices. Each child was tested by a lay screener and a nurse screener with the Retinomax and the SureSight. Children were randomly assigned to either the lay or nurse screener for conducting screening first. Each screener conducted the screening of the Retinomax and the SureSight, with test order also randomly assigned. A coordinator attended each screening and removed any child's spectacles before testing so that screeners did not know if a child habitually wore spectacles. No cycloplegic dilation was performed for measuring refractive error using the Retinomax or the SureSight. Lay screeners and nurse screeners used separate data collection forms and were masked to results from the other screener.

The Gold Standard Examinations (GSEs)

The GSEs were conducted in the VIP vans[15] by optometrists and ophthalmologists who were experienced in providing care to children and were masked to the results of the screening. The GSEs included monocular threshold distance visual acuity assessment with crowded, single H, O, T, V optotypes using the Electronic Visual Acuity system,[16] cover testing at distance and near, and cycloplegic retinoscopy. Results from the GSEs were used to determine whether a child had amblyopia, strabismus, significant refractive error, and/or unexplained reduced VA.[1–2]

Statistical Analysis

For the assessment of inter-tester agreement of refractive error, the first reading with a confidence number considered acceptable by the manufacturer was used in the analysis. If no acceptable readings were obtained, the reading with the highest confidence number was used. The readings associated with confidence number of "E" (Error) were excluded from statistical analysis. Readings with out of range values (either +9.99 or −9.99 in sphere or cylinder) were also excluded from the analysis of inter-tester agreement, because their true values were unknown. The negative cylinder notation was used for both the Retinomax and

the SureSight. Spherical equivalent (SE) was calculated as sphere plus half the magnitude of the cylinder power.

The agreement for the confidence number and out of range occurrence between lay and nurse screeners were evaluated by percent of agreement and the Kappa statistic. The Bland-Altman plots[17] were used for evaluating inter-tester agreement between lay and nurse screeners for sphere, cylinder and spherical equivalent (SE). Inter-tester agreement was quantified using the mean inter-tester difference and its 95% limits of agreement, which were estimated by the mean difference ± 1.96 standard deviations of the inter-tester difference. The difference was calculated as the measured value from lay screener minus the measured value from nurse screener, with a positive difference indicating the value from the lay screener is larger (i.e., more plus) than that from the nurse screener. The mean inter-tester difference provides an indication of measurement bias, while the 95% limits of agreement provide an indication of variability of inter-tester difference. The subgroup analyses of inter-tester agreement were performed by child's age (3-year-old vs. 4-year-olds vs. 5-year-olds), presence of significant refractive error (defined as hyperopia >3.25 D, myopia >2.0 D, astigmatism >1.5 D, or anisometropia >1.0 D) based on the GSE,[1–2] and the confidence number associated with refractive error measurements (meeting the manufacturer's recommended level by both lay and nurse screener versus not meeting the manufacturer's recommended level by either lay or nurse screener). The comparisons of inter-tester agreement between subgroups were performed for mean inter-tester difference using analysis of variance, and for variability of inter-tester difference using the Brown-Forsythe test.[18] Because refractive errors from both eyes of a child were included in the analyses, generalized estimating equations (GEE) [19] were used to adjust for inter-eye correlation in the comparison of both mean and variability of inter-tester difference. Two-sided $P<0.05$ was considered to be statistically significant. All the statistical analyses were performed in SAS v9.3 (SAS Institute Inc, Cary, NC).

RESULTS

Study Population

Among 1452 children in the VIP Phase II study, 377 (26.0%) were 3-year-olds, 793 (54.6%) 4-year-olds, and 282 (19.4%) 5-year-olds. Based on the GSEs performed by optometrists and ophthalmologists, 31% of children had at least one VIP-targeted vision disorder, 7% had amblyopia, 3% had strabismus, 27% had significant refractive error, and approximately 4% children wore spectacles.

Screening by the Retinomax and the SureSight

The flowchart for screening with the Retinomax administered by lay and nurse screeners is shown in Figure 1. Among 1452 children targeted for screening, 1433 (98.7%) children (2849 eyes) completed screening with the Retinomax by both lay and nurse screeners, thus providing measurements of refractive error for assessing inter-tester agreement of the Retinomax. The flowchart for screening with the SureSight administered by lay and nurse screeners is shown in Figure 2. Among 1452 children screened, 1404 (96.7%) children (2729 eyes) completed screening with the SureSight by both lay and nurse screeners, thus providing valid measurements of refractive error for assessing inter-tester agreement of the SureSight. Among 1452 children (2904 eyes) screened by the SureSight, out of range (less than −5.0 or greater than +6.0 D) in sphere occurred in 21 (0.72%) eyes measured by lay screeners and in 17 (0.59%) eyes measured by nurse screeners (5 eyes by both lay and nurse screeners); out of range (less than −4.0 D or greater than +4.0 D) in cylinder occurred in 46 (1.58%) eyes measured by lay screeners and in 58 (2.00%) eyes measured by nurse screeners (29 eyes by both lay and nurse screeners). Overall, out of range in either sphere or

cylinder occurred in 66 (2.27%) eyes measured by lay screeners, and in 75 (2.58%) eyes measured by nurse screeners (p=0.28). There were no occurrences of out-of-range values for the Retinomax.

Agreement in the Confidence Number of the Reading

Among the 2849 refractive error readings of the Retinomax from lay screeners and nurse screeners, respectively, the lay and nurse screeners had 95.9% agreement (Kappa = 0.09, 95% CI: 0.01 to 0.16) in confidence number when confidence numbers were grouped below (lay: 2.3%; nurse: 2.3%) versus equal to or above the manufacturer's recommended value of 8 (lay: 97.7%; nurse: 97.7%). The mean difference (95% limits of agreement) for the reading's confidence number between lay and nurse screeners was 0.03 (−1.3, 1.4).

Among 2729 refractive error readings of the SureSight from lay screeners and nurse screeners respectively, the lay and nurse screeners had 94.0% agreement (Kappa = 0.10, 95% CI: 0.03 to 0.16) in confidence number when confidence numbers were grouped below (lay: 3.8%; nurse: 3.4%) versus equal to or above the manufacturer's recommended value of 6 (lay: 96.2%; nurse: 96.6%). The mean difference of confidence number between lay and nurse screeners (95% limits of agreement) was 0.09 (−2.3, 2.5).

Overall Inter-tester Agreement of Refractive Error Measures

When refractive error was measured with the Retinomax, the mean inter-tester differences (lay minus nurse) were −0.04 D for sphere, 0.00 D for cylinder and −0.04 D for SE (Table 1). Their 95% limits of agreement were within 1.65 D for both sphere and SE, and within 0.52 D for cylinder. When refractive error was measured with the SureSight, the mean inter-tester differences (lay minus nurse) were 0.05 D for sphere, 0.01 D for cylinder and 0.06 D for SE (Table 1), and their 95% limits of agreement within 1.58 D for sphere and SE, and within 0.60 D for cylinder.

The inter-tester differences of sphere, cylinder and SE, plotted against the average reading from lay and nurse screeners, are shown in Bland-Altman plots (Figure 3). The mean inter-tester difference (central horizontal line) and lower and upper limit of 95% agreement (dashed horizontal lines) are shown on each plot. The plots did not show any specific pattern in the inter-tester difference, suggesting the random variation in the inter-tester difference.

Agreement of Refractive Error Measures by Age

The effect of the child's age (3-year-old vs. 4-year-olds vs. 5-year-olds) on mean inter-tester differences and 95% limits of agreement were evaluated (Tables 2, 3). For the Retinomax, the mean inter-tester differences did not differ by age for sphere and SE (inter-tester differences in each age group all within 0.10 D for sphere and SE, p>0.50 for all). The mean inter-tester difference for the cylinder was small for 3-year-olds, 4 year-olds and 5 year-olds (0.01 D, 0.01 D, −0.04 D respectively), yet statistically significant (p=0.006, Table 2). The mean inter-tester difference for the SureSight also did not differ by the child's age, with the mean differences all within 0.10 D for sphere and SE, and within 0.05 D for cylinder (Table 3; p>0.70 for all). The width of the 95% agreement limits (which are determined by the variance of the distribution of inter-tester differences) of sphere, cylinder and SE also did not differ by age of preschoolers for both the Retinomax and the SureSight (p>0.10 for all).

Agreement of Refractive Error Measures by Cycloplegic Refractive Error from Gold Standard Examination

The effect of magnitude of refractive error (based on cycloplegic refraction from the gold standard examination) on agreement was analyzed using presence/absence of significant refractive error and by categorizing refractive error into 4 groups. Presence/absence of

significant refractive error was defined as hyperopia >3.25 D, myopia >2.0 D, astigmatism >1.5 D, or anisometropia >1.0 D. The mean inter-tester differences of sphere and SE were similar between eyes with and without significant refractive error for the Retinomax. The mean of inter-tester differences was within 0.05 D for both sphere and SE ($p>0.05$, Table 2). The mean of inter-tester difference of cylinder was slightly larger, yet statistically significant in eyes with significant refractive error than eyes without significant refractive error (0.04 D vs. −0.02 D, $p=0.001$, Table 2). For the SureSight, the mean inter-tester differences of sphere, cylinder and SE were similar between eyes with and without significant refractive error (within 0.05 D for sphere, 0.01 D for cylinder and 0.06 for SE, $p>0.05$, Table 3).

For the Retinomax, the width of the 95% limits of agreement was greater in sphere, cylinder and SE among eyes with significant refractive error than eyes without significant refractive error (all $p<0.01$, Table 2). For the SureSight, the width of the 95% limits of agreement was greater for cylinder ($p<0.0001$) but not for sphere and SE ($p>0.05$) among eyes with significant refractive error than eyes without significant refractive error (Table 3).

When cycloplegic refractive error measured during the gold standard examination was analyzed as spherical equivalent and grouped into four groups (myopia ≤ −0.5 D, emmetropia −0.5 D to 1 D, mild hyperopia 1 to 2 D, moderate to severe hyperopia > 2 D), the mean inter-tester difference of refractive error from the Retinomax did not differ by levels of refractive error, but the 95% limits were significantly larger when the eye was either myopic or hyperopic ($p<0.0001$, Table 2). Similarly for the SureSight, the mean inter-tester difference of refractive error did not differ by levels of refractive error, but the 95% the limits were significantly larger when the eye was either myopic or hyperopic ($p<0.0001$, Table 3).

Agreement of Refractive Error Measures by the Confidence Number of the Reading

The confidence number of the reading by either nurse or lay screeners was below the manufacturer's recommended value in 123 (4.3%) eyes measured with the Retinomax and in 185 (6.8%) eyes measured with the SureSight.

For both instruments, the mean inter-tester differences were small whether or not readings from lay or nurse screeners had a confidence number below the manufacturer's recommended value ($p>0.05$, Tables 2, 3). However, the 95% limits of agreement in sphere, cylinder and SE were greater when the confidence numbers from either nurse or lay screeners were below the manufacturer's recommended value ($p<0.05$, Tables 2, 3). This was particularly true for the Retinomax, the width of 95% agreement limits of sphere and SE were more than doubled when the confidence number of the reading was below the manufacturer's recommended value ($p<0.01$, Table 2).

DISCUSSION

This study evaluated the inter-tester agreement between trained lay and nurse screeners for measuring refractive error using the Retinomax and the SureSight. Based on data from large numbers of Head Start preschoolers across five clinical centers, this study found that lay and nurse screeners agreed well in measuring refractive error in a screening setting. Child's age, refractive error, and confidence number of autorefractor reading had little impact on the mean inter-tester difference. However, the variation of inter-tester differences is larger (i.e., wider limits of agreement) when children had significant refractive error, or when the confidence number of the reading was below the manufacturer's recommended value.

Hand-held autorefractors have gained wide use for vision screening in preschoolers because they provide quick readings of refractive error without involvement of highly trained clinical

personnel and without need for cycloplegic dilation. The Retinomax and the SureSight are hand-held autorefractors, and have been widely used in vision screening and clinical practice for measuring refractive error by personnel with various levels of training. Previous VIP papers have reported that both the Retinomax and the SureSight correlated well with gold standard eye examinations for detecting vision disorders whether tests were administrated by licensed eye care professionals (sensitivity of 63% to 64% at specificity of 90%, and 51% to 52% at 94% specificity),[1,20] by trained nurses or lay screeners (sensitivity of 62% to 68% for the Retinomax, and 61% to 64% for the SureSight, at 90% specificity).[2] Other researches also have indicated that the Retinomax and the SureSight provided valid measures of refractive error in young children. [4–8] However, data on their intra-tester or inter-tester agreement are scarce. A few, small-sample studies evaluated the intra-tester (i.e., test-retest) agreement of the two instruments in a variety of age groups,[9–14] and results suggested good intra-tester agreement for both instruments in the preschool age group (Table 4). The means of intra-tester differences from these studies were all within 0.15 D for the Retinomax, and within 0.20 D for the SureSight, except one study in very young children (age 2 to 12 months old) that showed large mean differences of 1.8 D for sphere and 1.3 D for cylinder.[12] Our study evaluated inter-tester agreement between two types of testers (trained lay screeners vs. trained pediatric nurse screeners) for measuring sphere, cylinder and SE on a very large sample of 3- to 5-year-old preschoolers (N=1,452). The mean inter-tester differences from our study are comparable to the mean intra-tester differences reported in the literature. However, the 95% limits of inter-tester agreement tend to be larger than that for intra-tester difference. This may be due to the additional variation introduced by the second tester in the inter-tester agreement. These findings on the inter-tester agreement provide valuable information on the expected difference of measuring refractive error when testing is performed by two different screeners using the Retinomax or the SureSight. Considering all the findings of intra-tester agreement from previous studies and inter-tester agreement of the current study for the Retinomax and the SureSight, both instruments seem to provide very consistent measures of refractive error in preschoolers, whether the test is administered by the same screener or by a different screener, supporting their use for detecting the change of refractive error over time.

The range of refractive error measures from the Retinomax and the SureSight are very different. The possible range of sphere is −18 to +23 D for the Retinomax versus −5.0 to +6.0 D for the SureSight. The possible range of cylinder is −12 to 12 D for the Retinomax versus −4 to 4 D for the SureSight. Because of the smaller range of sphere and cylinder from the SureSight, an out of range reading from the SureSight occurred in approximately 1% of eyes for sphere and 2% of eyes for cylinder, while no measurements were out of range using the Retinomax. We excluded the out of range readings from the assessment of inter-tester agreement, because their true value of sphere or cylinder is unknown in these cases. The exclusion of these out of range values may lead to underestimation of the limits of inter-tester agreement for the SureSight. Out of range reading did not occur in the Retinomax because it allows a wider range of sphere and cylinder. Thus, any direct comparison of inter-tester agreement between the Retinomax and the SureSight needs to consider these differences.

In the subgroup analyses, we examined the impact of children's age, cycloplegic refractive error, and confidence number of readings on the inter-tester agreement of refractive error measurements, by comparing their mean inter-tester difference and their 95% limits of agreement. Among 3 to 5-years-old preschoolers, this study found that the child's age has no substantial impact on the mean inter-tester difference from both instruments. However, we found that cycloplegic refractive error (either myopic or hyperopic) was associated with wider limits of inter-tester agreement for both the Retinomax and the SureSight, and confidence numbers below the manufacturer's recommended value were significantly

associated with wider limits of inter-tester agreement for both the Retinomax and the SureSight. Our previous work also demonstrated that lower confidence numbers were associated with worse sensitivity and specificity for detecting vision disorders using the Retinomax.[21] These data suggest that when the confidence number is below the manufacturer's recommended value, repeated testing is recommended to obtain higher confidence numbers and a more reliable measure of refractive error.

In the VIP study, the lay screeners and nurse screeners received the same training for performing vision screening using the Retinomax and the SureSight provided by a team of VIP Study personnel. How the screeners are trained may impact their ability to obtain reliable measures of refractive error using screening instruments. It is reasonable to assume that the inter-tester differences may increase if the testers are not trained, or are not trained in the same way. It is also important to note that the Retinomax has three test modes (normal mode, quick mode, and auto mode), and the SureSight has two modes (child mode and adult mode). The VIP study used the auto mode for the Retinomax and child mode for the SureSight. The findings from using these modes may not be generalizable to the other modes.

The strengths of this study include the large sample size, inclusion of preschool children with various vision disorders (amblyopia, strabismus, astigmatism, significant refractive error), and the standard training of screeners and application of same screening protocol to both lay and nurse screeners. Also, because the Retinomax and the SureSight were measured without cyloplegic refraction, the results more accurately represent the expected results in when conducting preschool vision screening using the Retinomax and the SureSight.

In conclusion, the evaluation of inter-tester agreement from a large sample of VIP participants demonstrated that trained lay and nurse screeners agree well in measuring refractive error when using either the Retinomax or the SureSight on preschool children in a screening setting. These results are also consistent with the main findings from the VIP Study that the Retinomax and the SureSight are similarly effective when used by nurse and lay screeners. While the preschooler's age, refractive error, and confidence number of the autorefractor reading have little impact on the mean inter-tester differences of refractive error measurements, the agreement limits for inter-tester difference are greater in eyes with significant refractive error or when the confidence number is below the manufacturer's recommended number.

Acknowledgments

This study is 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.

The Vision in Preschoolers Study Group

Executive Committee: Paulette Schmidt, OD, MS (Chair); Agnieshka 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.

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 Bordeau, 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 Copenhaven, 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); Agnieshka 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

REFERENCES

1. 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. Vision In Preschoolers (VIP) Study Group. Ophthalmology. 2004; 111:637–650. [PubMed: 15051194]
2. 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. [PubMed: 16043831]
3. Cordonnier M, De Maertelaer V. Comparison between two hand-held autorefractors: the SureSight and the Retinomax. Strabismus. 2004; 12:261–274. [PubMed: 15545144]
4. Ying GS, Maguire M, Quinn G, Kulp MT, Cyert L. ROC analysis of the accuracy of Noncycloplegic retinoscopy, Retinomax Autorefractor, and SureSight Vision Screener for preschool vision screening. Invest Ophthalmol Vis Sci. 2011; 52:9658–9664. [PubMed: 22125281]
5. Paff T, Oudesluys-Murphy AM, Wolterbeek R, Swart-van den Berg M, de Nie JM, Tijssen E, Schalij-Delfos NE. Screening for refractive errors in children: the plusoptiX S08 and the Retinomax K-plus2 performed by a lay screener compared to cycloplegic retinoscopy. J AAPOS. 2010; 14:478–483. [PubMed: 21168070]
6. 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; 11:243–248. [PubMed: 17140828]
7. Lang D, Leman R, Arnold AW, Arnold RW. Validated portable pediatric vision screening in the Alaska Bush. A VIPS-like study in the Koyukon. Alaska Med. 2007; 49:2–15. [PubMed: 17479731]
8. 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; 13:466–471. [PubMed: 19840726]
9. Harvey EM, Miller JM, Wagner LK, Dobson V. Reproducibility and accuracy of measurements with a hand held autorefractor in children. Br J Ophthalmol. 1997; 81:941–948. [PubMed: 9505815]
10. Harvey EM, Miller JM, Dobson V, Tyszko R, Davis AL. Measurement of refractive error in Native American preschoolers: validity and reproducibility of autorefraction. Optom Vis Sci. 2000; 77:140–149. [PubMed: 10772231]
11. Cordonnier M, Dramaix M. Screening for refractive errors in children: accuracy of the hand held refractor Retinomax to screen for astigmatism. Br J Ophthalmol. 1999; 83:157–161. [PubMed: 10396190]
12. Adams RJ, Dalton SM, Murphy AM, Hall HL, Courage ML. Testing young infants with the Welch Allyn suresight non-cycloplegic autorefractor. Ophthalmic Physiol Opt. 2002; 22:546–551. [PubMed: 12477019]
13. Rosenfield M, Chiu NN. Repeatability of subjective and objective refraction. Optom Vis Sci. 1995; 72:577–579. [PubMed: 8539026]

14. Shoemaker JA. Repeatibility of Welch Allyn SureSight autorefraction in adults. Invest Ophthalmol Vis Sci. 2000; 41(suppl):S301.

15. 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.

16. 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; 119:1345–1353. [PubMed: 11545641]

17. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999; 8:135–160. [PubMed: 10501650]

18. Brown MB, Forsythe AB. Robust tests for the equality of variances. J AM Stat Assoc. 1974; 69:364–367.

19. Liang KY, Zeger SL. Longitudinal data-analysis using generalized linear-models. Biometrika. 1986; 73:13–22.

20. Ying GS, Kulp MT, Maguire M, Ciner E, Cyert L, Schmidt P. Sensitivity of screening tests for detecting vision in preschoolers-targeted vision disorders when specificity is 94%. The Vision In Preschoolers Study Group. Optom Vis Sci. 2005; 82:432–438. [PubMed: 15894920]

21. Vision In Preschoolers (VIP) Study Group. Impact of confidence number on the screening accuracy of the retinomax autorefractor. Optom Vis Sci. 2007; 84:181–188. [PubMed: 17435531]

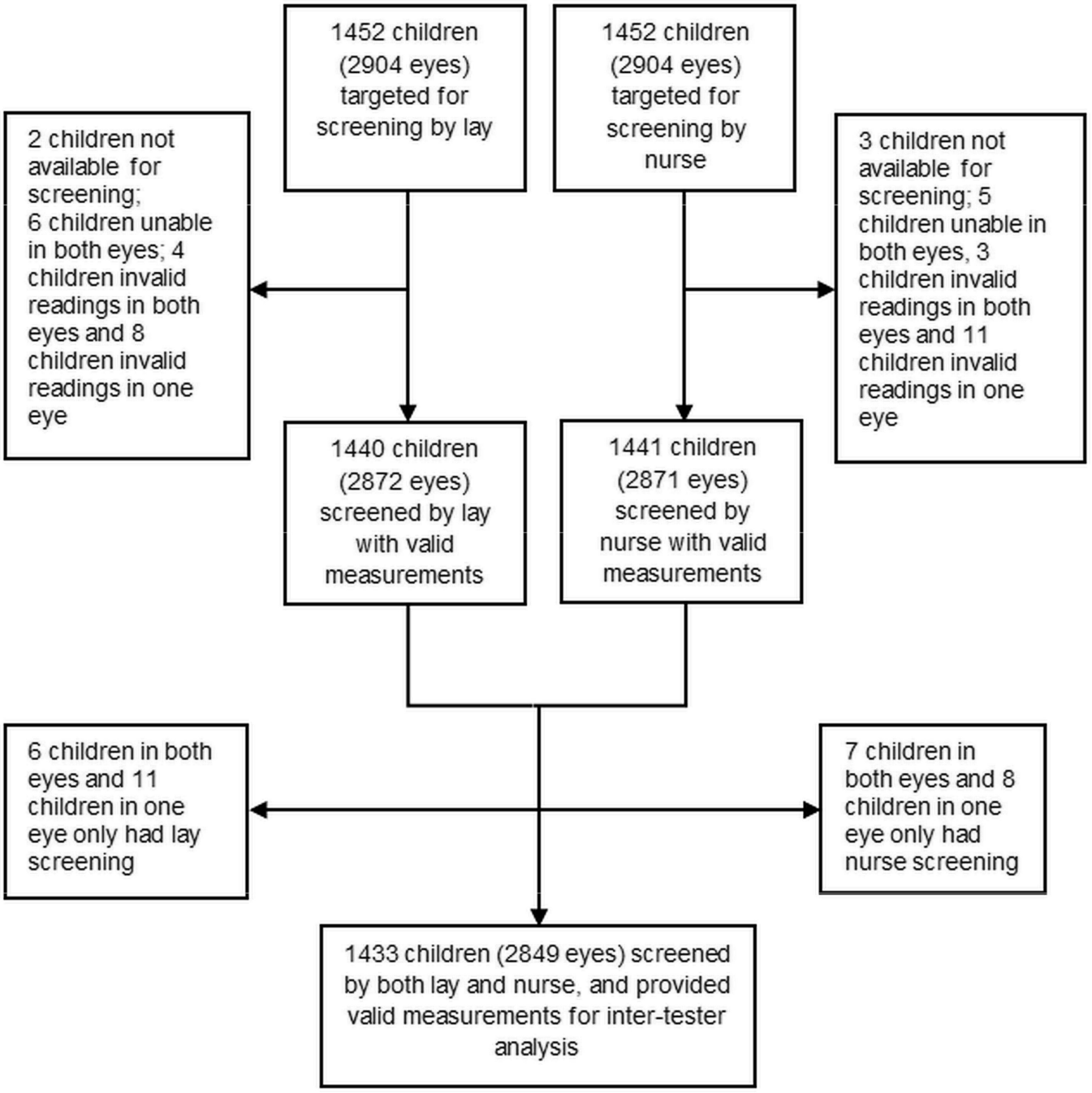

Figure 1.
The flowchart for the analyzable refractive error measurements from the Retinomax.

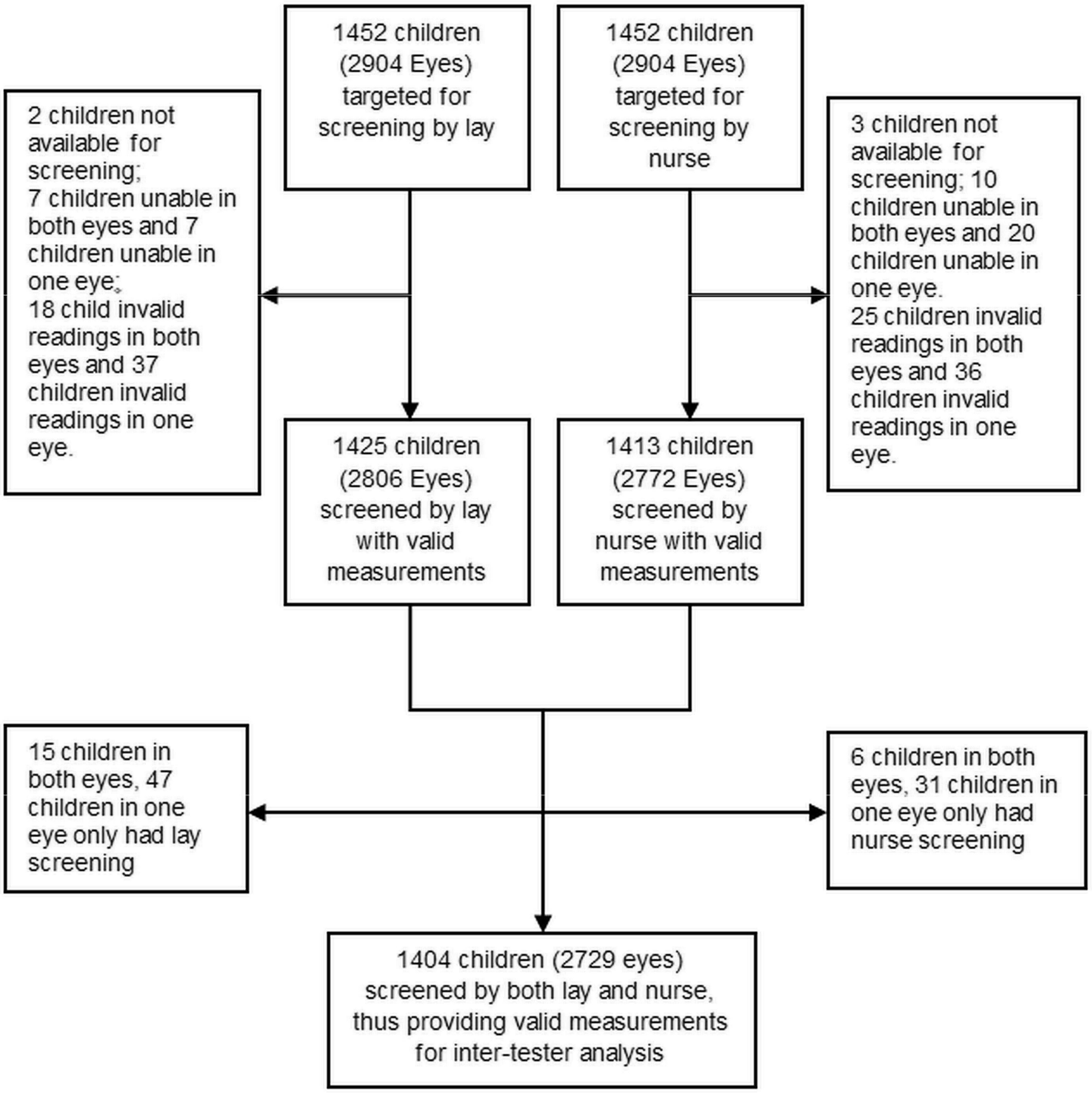

Figure 2.
The flowchart for the analyzable refractive error measurements from the SureSight.

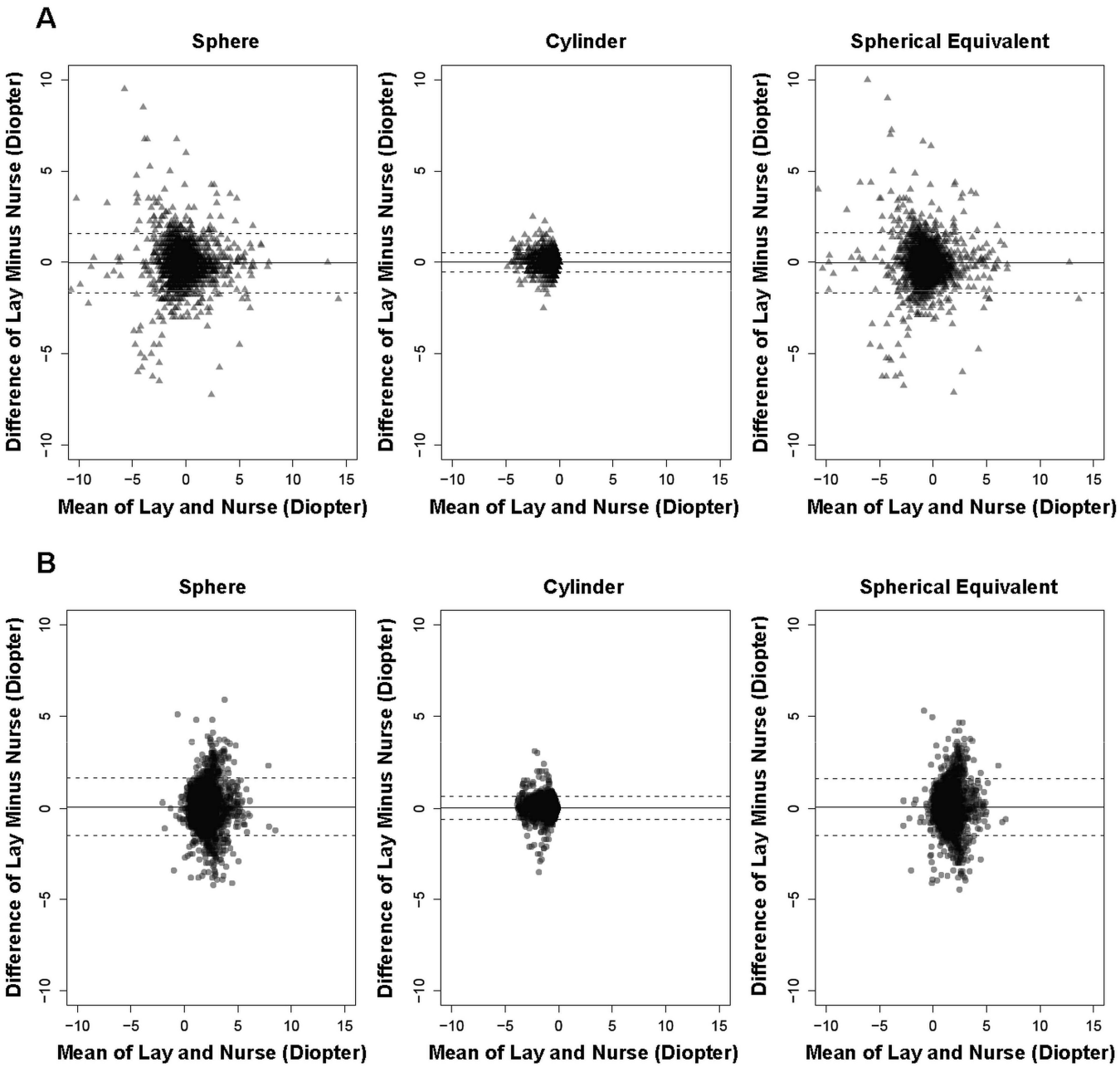

Figure 3.

Bland-Altman plots for the inter-tester agreement of refractive error measurements (sphere, cylinder and spherical equivalent) between lay and nurse screeners from: **(A)** the Retinomax (N=2849 eyes); **(B)** the SureSight (N=2729 eyes).

Table 1

Inter-tester agreement of refractive error measurements between lay and nurse screeners for the Retinomax and the SureSight.

	Retinomax (N=2849 eyes)		SureSight (N=2729 eyes)	
	Mean (SD) of difference (Lay – Nurse)	**95% limits of agreement**	**Mean (SD) of difference (Lay – Nurse)**	**95% limits of agreement**
Refractive error measurements (Diopter)				
Sphere	−0.04 (0.81)	(−1.63, 1.54)	0.05 (0.78)	(−1.48, 1.58)
Cylinder	0.00 (0.26)	(−0.52, 0.51)	0.01 (0.30)	(−0.58, 0.60)
Spherical equivalent	−0.04 (0.82)	(−1.65, 1.56)	0.06 (0.77)	(−1.45, 1.57)

SD: Standard Deviation.

Table 2

Inter-tester agreement between lay and nurse screeners by age group, presence of significant refractive error and confidence number for the Retinomax (N=2849 eyes).

		Mean (95% limits) of Inter-tester Difference (Lay – Nurse)		
		Sphere	**Cylinder**	**Spherical equivalent**
Age	3 years old (n=722)	−0.08 (−1.78, 1.61)	0.01 (−0.47, 0.50)	−0.08 (−1.80, 1.64)
	4 years old (n=1569)	−0.03 (−1.58, 1.53)	0.01 (−0.53, 0.54)	−0.02 (−1.60, 1.55)
	5 years old (n=558)	−0.03 (−1.54, 1.48)	−0.04 (−0.52, 0.43)	−0.05 (−1.58, 1.48)
	P-value for comparing mean	0.57	0.006	0.60
	P-value for comparing 95% limits	0.70	0.82	0.82
Significant refractive error	Yes (n=737)	−0.04 (−1.93, 1.84)	0.04 (−0.57, 0.66)	−0.02 (−1.93, 1.89)
	No (n=2112)	−0.04 (−1.51, 1.42)	−0.02 (−0.48, 0.45)	−0.05 (−1.54, 1.44)
	P-value for comparing mean	1.00	0.001	0.61
	P-value for comparing 95% limits	0.002	<0.0001	0.002
Spherical equivalent from gold standard exams	≤ −0.5 D (n=125)	0.11 (−1.90, 2.11)	0.06 (−0.77, 0.90)	0.14 (−1.90, 2.17)
	> −0.5, ≤ 1 D (n=1104)	−0.03 (−1.46, 1.40)	−0.01 (−0.56, 0.53)	−0.04 (−1.49, 1.41)
	>1, ≤ 2 D (n=1057)	−0.04 (−1.65, 1.57)	0.00 (−0.51, 0.51)	−0.04 (−1.67, 1.59)
	> 2 D (n=563)	−0.10 (−2.25, 2.05)	0.01 (−0.54, 0.55)	−0.10 (−2.28, 2.09)
	P-value for comparing mean	0.43	0.35	0.38
	P-value for comparing 95% limits	<0.0001	<0.0001	<0.0001
Confidence number	<8 (n=123)	0.04 (−3.28, 3.37)	−0.02 (−0.93, 0.90)	0.03 (−3.40, 3.47)
	≥8 (n=2726)	−0.05 (−1.57, 1.48)	0.00 (−0.51, 0.50)	−0.05 (−1.59, 1.50)
	P-value for comparing mean	0.55	0.65	0.61
	P-value for comparing 95% limits	0.002	<0.0001	0.003

Table 3

Inter-tester agreement between lay and nurse screeners by age group, presence of significant refractive error and confidence number for the SureSight (N=2729 eyes)

		Mean (95% limits) of Inter-tester Difference (Lay – Nurse)		
		Sphere	**Cylinder**	**Spherical equivalent**
Age	3 years old (n=697)	0.07 (−1.57, 1.71)	0.03 (−0.55, 0.61)	0.08 (−1.54, 1.70)
	4 years old (n=1503)	0.05 (−1.47, 1.57)	0.005 (−0.62, 0.63)	0.05 (−1.45, 1.55)
	5 years old (n=529)	0.04 (−1.36, 1.44)	0.004 (−0.52, 0.52)	0.04 (−1.34, 1.42)
	P-value for comparing mean	0.88	0.51	0.77
	P-value for comparing 95% limits	0.16	0.12	0.14
Significant refractive error	Yes (n=641)	0.05 (−1.51, 1.61)	0.01 (−0.67, 0.70)	0.05 (−1.51, 1.61)
	No (n=2088)	0.05 (−1.46, 1.57)	0.01 (−0.55, 0.57)	0.06 (−1.43, 1.55)
	P-value for comparing mean	0.88	0.98	0.88
	P-value for comparing 95% limits	0.24	<0.0001	0.18
Spherical equivalent from gold standard exams	≤ −0.5 D (n=108)	−0.16 (−1.62, 1.29)	0.02 (−0.91, 0.95)	−0.15 (−1.79, 1.49)
	> −0.5, ≤1 D (n=1073)	0.06 (−1.37, 1.49)	0.02 (−0.60, 0.63)	0.07 (−1.32, 1.45)
	>1, ≤ 2 D (n=1036)	0.05 (−1.67, 1.76)	0.01 (−0.61, 0.63)	0.05 (−1.64, 1.74)
	> 2 D (n=512)	0.10 (−1.73, 1.94)	−0.01 (−0.63, 0.61)	0.10 (−1.72, 1.92)
	P-value for comparing mean	0.08	0.74	0.18
	P-value for comparing 95% limits	<0.0001	<0.0001	<0.0001
Confidence number	<6 (n=185)	0.05 (−2.17, 2.26)	0.04 (−1.02, 1.10)	0.06 (−2.07, 2.20)
	≥6 (n=2544)	0.05 (−1.48, 1.59)	0.01 (−0.58, 0.59)	0.06 (−1.46, 1.58)
	P-value for comparing mean	0.95	0.60	0.84
	P-value for comparing 95% limits	0.02	0.004	0.02

Table 4

Comparison of studies evaluating the reproducibility of refractive error in children and adults.

Instrument	Study/reference	Age Range	Cycloplegia used	Mode of instrument	# of subjects	Repeated measures from	Mean difference (SD)		
							Sphere	Cylinder	SE
Retinomax	Harvey et al., 1997	11 to 93 months	Yes	NA	N=47	Test-retest	0.15 (0.19)	0.18 (0.17)	--
	Harvey et al., 2000	3 to 6 years	No	Normal	N=36	Test-retest	0.00 (0.51)	−0.10 (0.25)	−0.05 (0.50)
			Yes				−0.04 (0.32)	0.03 (0.31)	−0.03 (0.22)
	Cordonnier et al., 1999	9 to 36 months	No	Quick	N=299	Test-retest (OD)	0.09 (1.35)	0.04 (0.46)	--
						Test-retest (OS)	0.13 (1.23)	0.02 (0.42)	--
	Current Study	*3 to 5 years*	*No*	*Auto*	*N=1452*	*Lay and nurse screeners*	*−0.04 (0.81)*	*0.00 (0.26)*	−0.04 (0.82)
SureSight	Adams et al., 2002	2 to 12 months	No	--	N=74	Test-retest	1.78 (0.79)	1.33 (0.74)	--
	Rosenfield et al., 1995	23 to 60 years	No	--	N=12	Test-retest	0.16 (0.02)	0.19 (0.03)	0.14 (0.02)
	Shoemaker et al.,	Average 46 years	No	--	N=21	Test-retest	0.18 (0.18)	0.09 (0.11)	--
	Current study	*3 to 5 years*	*No*	*Child*	*N=1452*	*Lay and nurse screeners*	*0.05 (0.78)*	*0.01 (0.30)*	0.06 (0.77)