Accuracy of the Welch Allyn SureSight for measurement of magnitude of astigmatism in 3- to 7-year-old children

Erin M. Harvey, PhD, Velma Dobson, PhD, Joseph M. Miller, MD, MPH, Candice E. Clifford-Donaldson, MPH, Tina K. Green, MS, Dawn H. Messer, OD, MPH, and Katherine A. Garvey, OD

PURPOSE
To evaluate the accuracy of the Welch Allyn SureSight in noncycloplegic measurements of astigmatism as compared to cycloplegic Retinomax K+ autorefractor measurements of astigmatism in children from a Native American population with a high prevalence of high astigmatism.

METHODS
Data are reported for 825 3- to 7-year-old children with no ocular abnormalities. Each child had a Retinomax K+ cycloplegic measurement of right eye astigmatism with a confidence rating ≥8 and 3 attempts to obtain a SureSight measurement on the right eye.

RESULTS
SureSight measurement success rates did not differ significantly across age or measurement confidence rating (<6 vs ≥6). Ninety-six percent of children had at least 1 measurement (any confidence), and 89% had at least 1 measurement with confidence at the manufacturer’s recommended value (≥6). Overall, the SureSight tended to overestimate astigmatism. If the SureSight measurement had any dioptric value (0.00 D to 3.00 D), astigmatism of 2.00 D or less was likely to be present. If the SureSight showed astigmatism beyond the instrument’s dioptric range (>3.00 D), Retinomax K+ measurements indicated that >2.00 D of astigmatism was present in 136 of 157 (86.6%). In cooperative children for whom the SureSight would not give a reading, 32 of 34 (94%) had >3.00 D of astigmatism.

CONCLUSIONS
The SureSight does not provide an accurate, quantitative measure of amount of astigmatism. However, it does allow accurate categorization of amount of astigmatism as ≤2.00 D, >2.00 D, or >3.00 D, and it has high measurement success rate in young children. (J AAPOS 2009;13:466-471)

Handheld autorefractors have the potential to be useful tools for measurement of refractive error in infants, toddlers, and preschoolers in both screening and research settings. However, when used without cycloplegia, autorefractors tend to have high variability across subjects in accuracy of sphere measurements in children,1-5 thus eliminating the feasibility of simply applying a single factor to correct for the overestimation of myopia/underestimation of hyperopia that occurs in the absence of cycloplegia.

In contrast, while some studies report significant differences between cycloplegic and non-cycloplegic measures of astigmatism, these differences tend to be of minimal clinical significance.1-8 Therefore, non-cycloplegic autorefraction may be useful in measuring astigmatism in infants and toddlers, in both screening5,8,9 and research5 settings.

The Welch Allyn SureSight has been used to measure refractive error in infants and young children with some success.4,10-19 However, one limitation of the SureSight is that the upper limit for measurement of astigmatism is only 3.00 D, although the instrument does provide an “out of range” indication if the astigmatism is beyond its range of measurement.

The purpose of the present study was to evaluate the accuracy of the SureSight in non-cycloplegic measurements of magnitude of astigmatism in 3- to 7-year-old children from a population with a high prevalence of astigmatism,20,21 in comparison to measurements of astigmatism obtained from the same children when tested under cycloplegic conditions using the Retinomax K+ autorefractor,
which has previously been shown to provide accurate measurement of astigmatism in this population.¹

**Methods**

**Subjects**

Subjects were 937 children who were recruited from the community, from the Head Start program, and from kindergarten and first grade classrooms on the Tohono O’odham reservation between September 2005 and December 2008.

This research followed the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the University of Arizona. Parents provided written informed consent prior to testing. This study conformed to the requirements of the United States Health Insurance Portability and Accountability Act (HIPAA).

**Apparatus**

Two handheld autorefractors were used: the Welch Allyn SureSight (software version 2.16 and 2.20; Welch Allyn Medical Products, Skaneateles Falls, NY) and the Retinomax K+ (Nikon, Inc., Melville, NY; now manufactured by Right Manufacturing Co., Tokyo). The SureSight is based on a Shack-Hartmann wavefront analyzer and has a working distance of 35 cm. The fixation stimulus is a circle of eight flashing green LEDs surrounding a small, red central light. The SureSight takes 5-8 measurements of the eye, after which it displays a measurement of sphere, cylinder, and axis, along with a confidence rating from 1 to 9, indicating the reliability of the reading. According to the manufacturer, the SureSight measures astigmatism values ≤ 3.00 D, and displays a value of 9.99 for astigmatism values > 3.00 D. Two SureSight instruments were used for the study; both were serviced and calibrated by the manufacturer yearly.

The Retinomax K+ is a conventional autorefractor. It has a short working distance of about 5 cm from the eye, and the instrument is rested against the child’s forehead while measurements are made. The child is asked to look steadily at a fixation target (Christmas tree) within the instrument during measurements. The Retinomax K+ makes 8 measurements, from which it produces a representative refraction and a confidence rating from 1 to 10. The manufacturer indicates that the instrument can measure astigmatism values up to 8.00 D.

**Procedures**

Initially, each child had refractive error measured with the SureSight. Subjects sat facing the tester, who attempted to make 3 measurements (without regard to confidence ratings) of each of the child’s eyes. The measurements were used to calculate amount of astigmatism (based on the median of all measurements obtained for the eye).

Following SureSight measurements, each child underwent an eye examination, including measurement of refractive error with the Retinomax K+ at least 40 minutes after instillation of one drop of proparacaine, 0.5%, and two drops of cyclopentolate, 1%, in each eye. If the Retinomax K+ measurement for either eye had a confidence rating of <8, the tester attempted another measurement of the eye to obtain a confidence of at least 8 (3 attempts, maximum).

Analyses were conducted on data from each child’s right eye.

**Results**

**Subjects**

Of 937 children examined, the 825 who met the following criteria were included in the final analyses: 3 to 7 years of age on the date of the examination, right eye cycloplegic Retinomax K+ autorefraction with a measurement confidence ≥ 8, and no ocular abnormalities other than high refractive error upon examination. The mean age of the 825 children in the final sample was 4.9 years (SD 1.3). A summary of reasons for exclusion is provided in Table 1.

Results of cycloplegic Retinomax K+ measurements of astigmatism indicated that 459 children (55.6%) children had 0.00 D to 1.00 D of astigmatism, 171 (20.7%) had >1.00 to 2.00 D, 105 (12.7%) had >2.00 to 3.00 D, 53 (6.4%) had >3.00 to 4.00 D, 28 (3.4%) had >4.00 to 5.00 D, and 9 (1.1%) had >5.00 D of astigmatism. Results of cycloplegic Retinomax K+ measurements of spherical equivalent (SE) indicated that 1 (0.1%) had SE <−4.00 D, 6 (0.7%) had SE −4.00 to −2.00 D, 144 (17.5%) had SE −2.00 to <0, 31 (3.8%) had SE = 0, 539 (65.3%) had SE >0.00 to +2.00 D, 89 (10.8%) had SE > +2.00 D to +4.00 D, and 15 (1.8%) had SE > +4.00 D.

**Measurement Success Rates**

Table 2 shows (1) the percentages of eyes in which 3 SureSight measurements were obtained (including out-of-range 9.99 results); (2) the percentage of eyes in which at

---

¹ Due to experimenter error (1), instrument not available (1), poor cooperation (3), child left before measurement was made (3), child refused cycloplegia (7).

² Due to experimenter error (1), instrument not available (2).
least one SureSight measurement was obtained; (3) the percentage of eyes in which the SureSight would not take a measurement even though the child was cooperative, and (4) the percentage of eyes in which a SureSight measurement was not obtained because the child was uncooperative. Percentages are also shown for only those measurements that had the manufacturer’s recommended confidence of ≥6.

Logistic regression analyses indicated no significant relation between measurement success rate, defined as success in obtaining at least one acceptable measurement, and subject age, when all measurements were considered acceptable (regardless of confidence rating) and also when only measurements with confidence rating ≥6 were considered. However, logistic regression analysis indicated a significant relation between measurement success rate and amount of astigmatism, defined by cycloplegic Retinomax K+ measurements (p < 0.001) when all measurements were considered and when only confidence measurements with confidence rating ≥6 were considered. Measurement success rates were lowest in children with high amounts of astigmatism (Table 2).

### Out-of-Range SureSight Values

A summary of SureSight values for each type of out-of-range measurement category is provided in Table 3. A SureSight astigmatism value of 9.99 (out-of-range per instrument, [ORi]) was obtained in all measurements of the right eye of 157 (19.0%) children, with 31 having all measurements with low confidence values (<6, ORiLC). In an additional 34 children, SureSight measurements could not be obtained, despite good instrument alignment (steady tone) and good cooperation by the child (out-of-range per tester [ORt]). Mean age did not differ significantly across the three out of range categories. However, ANOVA indicated that mean astigmatism, defined by cycloplegic Retinomax K+ measurements, differed significantly in the three groups ($F(2,190) = 84.18, p < 0.001$; ORiHC = 2.68 D [SD 0.60]; ORiLC = 3.39 D [SD 0.78]; ORt = 4.48 D [SD 1.06]). All pairwise post hoc comparisons were significant (p < 0.001).

Table 3 shows that if the SureSight measurement had any dioptric value (0.00 D to 3.00 D), astigmatism of 2.00 D or less was likely to be present. If the SureSight showed astigmatism beyond the instrument’s dioptric

### Table 2. Success with which right eye SureSight measurements were obtained by amount of right eye astigmatism (per cycloplegic Retinomax K+): percentages calculated including all SureSight measurements, regardless of confidence (top) and including only SureSight measurements with a confidence ≥6 (bottom)

<table>
<thead>
<tr>
<th>Amount of astigmatism</th>
<th>Number of acceptable SureSight measurements*</th>
<th>No acceptable SureSight measurement obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>At Least 1 N (%)</td>
<td>Cooperative N (%)</td>
</tr>
<tr>
<td>Any confidence*</td>
<td>746 (90.4)</td>
<td>789 (95.6)</td>
</tr>
<tr>
<td>0 to 1 D</td>
<td>442 (96.3)</td>
<td>458 (99.8)</td>
</tr>
<tr>
<td>&gt;1 to 2 D</td>
<td>167 (97.7)</td>
<td>169 (98.8)</td>
</tr>
<tr>
<td>&gt;2 to 3 D</td>
<td>100 (95.2)</td>
<td>105 (100)</td>
</tr>
<tr>
<td>&gt;3 to 4 D</td>
<td>33 (62.3)</td>
<td>45 (84.9)</td>
</tr>
<tr>
<td>&gt;4 to 5 D</td>
<td>4 (14.3)</td>
<td>11 (39.3)</td>
</tr>
<tr>
<td>≥5 D</td>
<td>0</td>
<td>1 (11.1)</td>
</tr>
<tr>
<td>Only confidence ≥6*</td>
<td>449 (54.4)</td>
<td>732 (88.7)</td>
</tr>
<tr>
<td>&lt;1 D</td>
<td>301 (65.6)</td>
<td>450 (98.0)</td>
</tr>
<tr>
<td>1 to &lt;2 D</td>
<td>100 (58.5)</td>
<td>166 (97.1)</td>
</tr>
<tr>
<td>2 to &lt;3 D</td>
<td>45 (42.9)</td>
<td>91 (86.7)</td>
</tr>
<tr>
<td>3 to &lt;4 D</td>
<td>3 (5.7)</td>
<td>23 (43.4)</td>
</tr>
<tr>
<td>4 to &lt;5 D</td>
<td>0</td>
<td>2 (7.1)</td>
</tr>
<tr>
<td>≥5 D</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Includes out-of-range measurements (9.99).

### Table 3. Magnitude of astigmatism present (per cycloplegic Retinomax K+) in children who had an out-of-range result for the right eye on the SureSight: shaded area is within the measurement range of SureSight, per the manufacturer

<table>
<thead>
<tr>
<th>Amount of astigmatism (Retinomax K+)</th>
<th>Out-of-range category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ORiHC* N (%)</td>
</tr>
<tr>
<td>0 to 1 D</td>
<td>0</td>
</tr>
<tr>
<td>&gt;1 to 2 D</td>
<td>20 (15.9)</td>
</tr>
<tr>
<td>&gt;2 to 3 D</td>
<td>80 (63.5)</td>
</tr>
<tr>
<td>&gt;3 to 4 D</td>
<td>24 (19.0)</td>
</tr>
<tr>
<td>&gt;4 to 5 D</td>
<td>2 (1.6)</td>
</tr>
<tr>
<td>≥5 D</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>126 (100.0)</td>
</tr>
</tbody>
</table>

*Out of range per instrument, with high confidence rating. Instrument provided a measurement of 9.99 for measurement of astigmatism, with a confidence rating ≥6.

*Out of range per instrument, with low confidence rating. Instrument provided a measurement of 9.99 for measurement of astigmatism, with a confidence rating <6.

*Out of range per tester. Child was cooperative, but instrument turned off immediately after auditory signal indicating a measurement was obtained or the instrument was aligned (steady alignment tone present) but “timed out” (turned off) before a measurement was provided.

---

*Journal of AAPOS*
Accuracy of SureSight Measurements of Astigmatism

This analysis included only data from the 621 children who had at least one SureSight measurement that was a dioptic value (ie, not 9.99). Figure 1 shows each child’s median SureSight astigmatism value compared with the child’s Retinomax K+ astigmatism value. Median SureSight astigmatism was calculated including all of each child’s dioptic value SureSight measurements, regardless of confidence rating.

SureSight measurements showed, on average, significantly higher amounts of astigmatism than did Retinomax K+ measurements (t(620) = 24.98; p < 0.001; mean SureSight measurement, 1.23 D [SD 0.86]; mean Retinomax K+ measurement, 0.73 D [SD 0.66]; mean difference, 0.50 D [SD 0.50]). The amount of astigmatism measured with the SureSight was significantly correlated with amount of astigmatism measured with the Retinomax K+ ($r^2 = 0.67, p < 0.001$). The following regression equation represents the relation between a child’s median SureSight measurement and that child’s Retinomax K+ measurement of astigmatism:

$$\text{Retinomax K+ Astigmatism} = -0.04 + 0.63 \times \text{median SureSight astigmatism}$$

ANOVA of the mean differences between SureSight and Retinomax K+ measurements of astigmatism indicated that there was a significant difference in accuracy by amount of astigmatism (F(2,617) = 52.66, p < 0.001). Post hoc analysis (with Bonferroni correction) indicated that accuracy was significantly poorer in the >1.00 D to 2.00 D group (n = 457; mean difference, 0.81 D [SD 0.45]) than in the 0 to 1.00 D (n = 148; mean difference, 0.42 D [SD 0.40]) and the >2.00 D to 3.00 D groups (n = 13; mean difference, 0.18 [SD 0.70]), but agreement did not differ between the 0 to 1.00 D and >2.00 D to 3.00 D groups. Table 4 summarizes agreement between SureSight and Retinomax K+ in terms of categorical data.

**Discussion**

The results of the present study provide data on the validity of the SureSight in measurement of magnitude of astigmatism in young children.

**Measurement Success Rates**

Measurement success rates did not differ significantly across the 3- to 7-year-old age range. Compared to some previous studies, the measurement success rate of 89% for high confidence measurements in this age range is low.$^4,10,13,18$ However, it is likely that this is due to the relatively large number of children in this sample (11%) who had astigmatism beyond the instrument’s upper measurement range. The majority of the measurement failures (70/93, 75%) occurred in children who had >3.00 D of astigmatism. Previous studies have reported a wide range (33% to 100%) of measurement success rates in infants.
The finding in the present study that D of astigmatism was present in most children (96.8%), necessarily more than 3.00 D. The presence of more than 2.00 D of astigmatism, but not of-range measurements are likely to be associated with (and 2.75 D. Remaining 11 children having astigmatism between 1.50 D was present in only 8 of 19 (42%) children, with the remaining 11 children having astigmatism between 1.50 and 2.75 D. Silverstein and colleagues reported that 92 of 120 (76.7%) of 1- to 5-year-old children on whom they obtained an out-of-range measurement failed their gold standard examination (failure was defined as presence of AAPOS-determined amblyogenic factors, which included with or against the rule astigmatism >1.50 D or oblique astigmatism >1.00 D). It is interesting to note that Silverstein and colleagues found that children with low confidence out-of-range measurements were more likely to fail the gold standard examination (27/28, 96.4%) than were children with high confidence out of range measurements (65/92, 70.7%), similar to the findings in the present study. Additional evidence that the SureSight overestimates magnitude of astigmatism is found in a study of preschool screening, in which 70.9% of children with a SureSight astigmatism reading of 1.50 D or more, 62.5% of children with a SureSight reading of 1.75 D or more, and 50% of children with a SureSight reading of 2.20 D or more did not have astigmatism >1.50 D. In the present study, there were 34 children (4.1%) who were cooperative and with whom the instrument was aligned (per the instrument’s auditory feedback signal) but no measurements were obtained. Astigmatism of 3.00 D or greater was present in nearly all of these children (32/34, 94.1%), and 10 of 34 (29.4%) had extremely high levels of astigmatism (≥5.00 D).

Significance of Out-of-Range Measurements
In the present study, we examined the relation between out-of-range measurements provided by the instrument (ie, a reading of 9.99), and amount of astigmatism present. If the out of range readings had a high confidence rating, astigmatism >2.00 D was present in most children (84%) (Table 3). Astigmatism >3.00 D was present in only 21%, even though the manufacturer states that the SureSight measures astigmatism up to 3.00 D. If the out-of-range readings had a low confidence rating (<6), >2.00 D of astigmatism was present in most children (96.8%), and astigmatism >3.00 was present in 61.3%. Thus out-of-range measurements are likely to be associated with the presence of more than 2.00 D of astigmatism, but not necessarily more than 3.00 D.

A previous study that examined the relation between “out of range (9.99)” astigmatism results and amount of astigmatism present found that astigmatism ≥3.00 D was present in only 8 of 19 (42%) children, with the remaining 11 children having astigmatism between 1.50 and 2.75 D. Silverstein and colleagues reported that 92 of 120 (76.7%) of 1- to 5-year-old children on whom they obtained an out-of-range measurement failed their gold standard examination (failure was defined as presence of AAPOS-determined amblyogenic factors, which included with or against the rule astigmatism >1.50 D or oblique astigmatism >1.00 D). It is interesting to note that Silverstein and colleagues found that children with low confidence out-of-range measurements were more likely to fail the gold standard examination (27/28, 96.4%) than were children with high confidence out of range measurements (65/92, 70.7%), similar to the findings in the present study. Additional evidence that the SureSight overestimates magnitude of astigmatism is found in a study of preschool screening, in which 70.9% of children with a SureSight astigmatism reading of 1.50 D or more, 62.5% of children with a SureSight reading of 1.75 D or more, and 50% of children with a SureSight reading of 2.20 D or more did not have astigmatism >1.50 D.

Accuracy of Astigmatism Measurements
The results of the present study indicate that when it provides a numeric value for astigmatism, the SureSight tends to overestimate astigmatism by approximately one third. This is somewhat greater than reported previously. Results of the present study also indicated that measurement accuracy varied by amount of astigmatism present, indicating that applying a simple “correction” to eliminate measurement bias is not likely to significantly increase overall accuracy.

However, despite the SureSight’s limits in accuracy of measuring specific magnitude of astigmatism, the present study indicates that the SureSight can be useful in categorizing amount of astigmatism present. The data from Tables 3 and 4, taken together, indicate that when the SureSight provides any dioptric value measurement (0.00 – 3.00 D), it is very likely that astigmatism of 2.00 D or less is present, whereas when the SureSight provides any form of out-of-range indication (ORi or ORt), it is likely that astigmatism >2.00 D is present. For the sample, ≤2.00 D of astigmatism was present, per the Retinomax K+, 97.4% of the time when a dioptric value SureSight measurement was obtained. When an out of range result was obtained, > 2.00 D was present 87.9% of time, with the percentage increasing to 95.4% when the out-of-range

Table 4. Accuracy of SureSight by amount of astigmatism present (per cycloplegic Retinomax K+)

<table>
<thead>
<tr>
<th>Median SureSight astigmatism measurement</th>
<th>0 to 1 D N (%)</th>
<th>&gt;1 to 2 D N (%)</th>
<th>&gt;2 to 3 D N (%)</th>
<th>&gt;3 to 4 D N (%)</th>
<th>&gt;4 to 5 D N (%)</th>
<th>Total N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 1 D</td>
<td>328 (98.8)</td>
<td>0 (0.9)</td>
<td>0</td>
<td>1 (0.3)</td>
<td>332 (100.0)</td>
<td></td>
</tr>
<tr>
<td>&gt;1 to 2 D</td>
<td>120 (72.7)</td>
<td>44 (26.7)</td>
<td>1 (0.6)</td>
<td>0</td>
<td>165 (100.0)</td>
<td></td>
</tr>
<tr>
<td>&gt;2 to 3 D</td>
<td>9 (7.3)</td>
<td>101 (81.5)</td>
<td>12 (9.7)</td>
<td>1 (0.8)</td>
<td>124 (100.0)</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>457 (73.6)</td>
<td>148 (23.8)</td>
<td>13 (2.1)</td>
<td>2 (0.3)</td>
<td>621 (100.0)</td>
<td></td>
</tr>
</tbody>
</table>

and toddlers. The finding in the present study that confidence rating was not related to accuracy of measurements suggests that it will be possible to improve success rates by using measurements with low, as well as high, confidence ratings.
measurement was either low confidence or was indicated by the instrument turning itself off.

Conclusions
For studies of refractive error prevalence in child populations, a single handheld instrument that can provide accurate measurements and uniformly high measurement success rates across age would be ideal. Use of the same instrument reduces measurement variability due to differences in methods, and consistently high success rates reduce the chance that prevalence differences across age may be due to age-related sampling differences in the children on whom measurements can be obtained. The results of the present study suggest that the SureSight achieves consistently high measurement success rates across age, that it can be used reliably to screen children for specific amounts of astigmatism (eg, >2.00 D), and that it may be useful for describing prevalence of high astigmatism in populations of young children.

Acknowledgments
Supported by grant U10 EY13153 (EMH) from the National Eye Institute of the National Institutes of Health, Department of Health and Human Services, and by Research to Prevent Blindness (unrestricted funds to the Department of Ophthalmology and Vision Science (JMM), a Walt and Lilly Disney Award for Amblyopia Research (JMM), and a Career Development Award (EMH)). The authors thank the Tohono O'odham Nation, the Tohono O'odham Early Childhood Head Start Program, the Indian Oasis/Babouquvarri School District, the Bureau of Indian Affairs Office of Indian Education Programs (BIA OIEP, Papago/Pima Agency), the San Xavier Mission School, the parents and children who participated in the study, and our NIH/NEI Data Monitoring and Oversight Committee (Robert Hardy, PhD (chair), Morgan Ashley, Donald Everett, MA, Jonathan Holmes, MD, Andrew Lorentine, and Karla Zadnik, OD, PhD).

References