Comparison of Monocular Autorefraction to Comprehensive Eye Examinations in Preschool-aged and Younger Children

Alex R. Kemper, MD, MPH, MS; Lisa M. Keating, CO; Jorie L. Jackson, CO; Erika M. Levin, MD

Background: Monocular autorefraction is a newly available technology for vision screening that has been advocated to test young children. Such devices automatically determine the refractive state of each eye, but cannot directly detect amblyopia or strabismus.

Objective: To compare the results of a commercially available monocular autorefractor (SureSight; Welch Allyn Medical Products, Skaneateles Falls, NY) with findings from a comprehensive eye examination for significant refractive error, strabismus, and amblyopia.

Methods: Children 5 years and younger who were new patients attending a pediatric ophthalmology clinic were tested with the monocular autorefractor without dilation and underwent a comprehensive eye examination that included dilation.

Main Outcome Measures: The proportion of children who could be tested and the sensitivity and specificity of the screening.

Results: Of the 170 children enrolled (age, <3 years, n=80; age range, 3-5 years, n=90), 36% had abnormal eye examination findings. Most (84%) children 3 years or older could be tested compared with 49% of the children younger than 3 years (P<.001). Among those who were testable, for children younger than 3 years the sensitivity was 80% (95% confidence interval [CI], 44%-97%) and the specificity was 41% (95% CI, 24%-61%). For children aged 3 to 5 years, the sensitivity was 88% (95% CI, 68%-97%) and the specificity was 58% (95% CI, 43%-71%).

Conclusions: Our findings suggest that screening children aged 3 to 5 years with monocular autorefraction would identify most cases of visual impairment but would be associated with many false-positive results. For children younger than 3 years, testability was low and results were nonspecific.


Amblyopia, which affects 3% to 5% of all children, is decreased vision that remains even after any underlying ocular condition is corrected. Early detection of amblyopia is important because the success of treatment decreases with age. Common causes of amblyopia include strabismus (misalignment of the eyes), high hyperopia (far-sightedness), and anisometropia (refractive difference between each eye). Less common causes include ptosis and cataracts.

Preschool vision screening for the detection of amblyopia is widely endorsed; however, little is known about the optimal screening strategy. The rate of vision screening in the primary care practice setting is low, we believe in part because of unanswered questions about the feasibility of screening the vision of young children and because of concern about overreferral owing to the perceived low specificity of screening.

Automated and semiautomated vision screening devices may improve both the rate and quality of preschool vision screening in the primary care practice setting. Screening with these devices requires little participation on the part of the child and may be more accurate than traditional vision screening tests.

One example of this emerging technology is monocular autorefraction, in which a device determines the refractive state of each eye. Autorefraction only detects refractive error; not all cases of amblyopia, strabismus, cataracts, or ptosis have associated refractive error. A previous case series suggests that up to half of the cases of amblyopia caused by strabismus are not ac-
COMPRESSIVE EYE EXAMINATION

The comprehensive eye examination included evaluation of stereopsis and binocularity, ocular alignment, visual acuity and/or fixation preference, and cycloplegic refraction, as well as the findings of the dilated fundus examination and anterior segment examination. The nondilated portion of the examination was performed by 1 of 2 orthoptists (L.M.K. or J.L.J.) or a pediatric ophthalmologist (E.M.L.). The components of this portion of the examination were tailored to the age and ability of the child (eg, fixation preference by induced tropia test or visual acuity by Allen Pictures, HOTV Letters, or Snellen tests). The pediatric ophthalmologist was responsible for the dilated and medical portions of the comprehensive eye examination.

We used the criteria recommended by the American Association of Pediatric Ophthalmology and Strabismus for defining the specific factors that should be identified by screening and to define an abnormal examination result (anisometropia >1.5 D; any manifested strabismus; hyperopia >3.50 D; myopia >3.00 D; any media opacity >1 mm; astigmatism >1.5 D at 90° or 180° or >1.0 in oblique axis; piosis ≥1-mm margin reflex distance; or visual acuity per age-appropriate standards). Pearson χ² test statistic was used to assess for association between visual impairment and age.

TESTABILITY AND ACCURACY

A child was considered to be testable if both eyes had marginal or better reliability (≥5) after 2 screening attempts. Sensitivity and specificity were determined by comparing screening results to the results from diagnostic evaluation. Testability was recalculated after discarding the first 10% of screenings to determine the effect of screener experience. Means and 95% confidence intervals (CIs), based on a binomial probability distribution, were calculated. Pearson χ² test statistic was used to assess for an association between visual impairment and testability and to determine if discarding the first 10% of screenings altered testability. T Tests were used to evaluate the associations between age and the mean reliability score.

ASSOCIATION BETWEEN SCREENING AND DIAGNOSIS

Correlation coefficients were used to assess the strength of the association between the refractive error derived from the autorefactor’s directions. During screening, the device was held 35 cm from the eye to be screened. Once aligned with the eye, the device automatically takes 3 measures over approximately 3 seconds. The device evaluates for myopia and hyperopia, measured as dipters (D) of sphere, and astigmatism, measured as dipters of cylinder. There is a “child calibration” mode for use when screening children younger than 7 years, which adjusts the measure of spherical error and adds report of anisometropia, as measured by the difference in spherical equivalent [sphere D +1/2 (cylinder D)] between the 2 eyes. Children are identified for referral based on preset criteria (sphere of −1.0 D or less or +2.0 D or more; cylinder +1.0 D or more; or the difference in the spherical equivalent ≥1.0 D). The device also reports reliability of screening, based on agreement among the repeated measures. Up to 2 screening attempts were made. If the reliability score for both eyes was marginal or poor (≤5 on a scale from 1 [worst] to 9 [best]), a second screening attempt was made; or if the reliability score for only 1 eye was marginal or poor (≤5), the screener could rescreen either both eyes or only the eye with the low reliability score. If only 1 eye was rescreened, the difference in spherical equivalent was manually calculated.

MONOCULAR AUTOREFRACTION

Each child was screened by a study investigator—either an orthoptist (L.M.K. or J.L.J.) or a pediatric ophthalmologist (E.M.L.)—using SureSight. The screening followed the manufacturer’s directions. During screening, the device was held 35 cm from the eye to be screened. Once aligned with the eye, the device automatically takes 3 measures over approximately 3 seconds. The device evaluates for myopia and hyperopia, measured as dipters (D) of sphere, and astigmatism, measured as dipters of cylinder. There is a “child calibration” mode for use when screening children younger than 7 years, which adjusts the measure of spherical error and adds report of anisometropia, as measured by the difference in spherical equivalent [sphere D +1/2 (cylinder D)] between the 2 eyes. Children are identified for referral based on preset criteria (sphere of −1.0 D or less or +2.0 D or more; cylinder +1.0 D or more; or the difference in the spherical equivalent ≥1.0 D). The device also reports reliability of screening, based on agreement among the repeated measures. Up to 2 screening attempts were made. If the reliability score for both eyes was marginal or poor (≤5 on a scale from 1 [worst] to 9 [best]), a second screening attempt was made; or if the reliability score for only 1 eye was marginal or poor (≤5), the screener could rescreen either both eyes or only the eye with the low reliability score. If only 1 eye was rescreened, the difference in spherical equivalent was manually calculated.

COMPRESSIVE EYE EXAMINATION

The comprehensive eye examination included evaluation of stereopsis and binocularity, ocular alignment, visual acuity and/or fixation preference, and cycloplegic refraction, as well as the findings of the dilated fundus examination and anterior segment examination. The nondilated portion of the examination was performed by 1 of 2 orthoptists (L.M.K. or J.L.J.) or a pediatric ophthalmologist (E.M.L.). The components of this portion of the examination were tailored to the age and ability of the child (eg, fixation preference by induced tropia test or visual acuity by Allen Pictures, HOTV Letters, or Snellen tests). The pediatric ophthalmologist was responsible for the dilated and medical portions of the comprehensive eye examination.

We used the criteria recommended by the American Association of Pediatric Ophthalmology and Strabismus for defining the specific factors that should be identified by screening and to define an abnormal examination result (anisometropia >1.5 D; any manifested strabismus; hyperopia >3.50 D; myopia >3.00 D; any media opacity >1 mm; astigmatism >1.5 D at 90° or 180° or >1.0 in oblique axis; piosis ≥1-mm margin reflex distance; or visual acuity per age-appropriate standards). Pearson χ² test statistic was used to assess for association between visual impairment and age.

TESTABILITY AND ACCURACY

A child was considered to be testable if both eyes had marginal or better reliability (≥5) after 2 screening attempts. Sensitivity and specificity were determined by comparing screening results to the results from diagnostic evaluation. Testability was recalculated after discarding the first 10% of screenings to determine the effect of screener experience. Means and 95% confidence intervals (CIs), based on a binomial probability distribution, were calculated. Pearson χ² test statistic was used to assess for an association between visual impairment and testability and to determine if discarding the first 10% of screenings altered testability. T Tests were used to evaluate the associations between age and the mean reliability score.

ASSOCIATION BETWEEN SCREENING AND DIAGNOSIS

Correlation coefficients were used to assess the strength of the association between the refractive error derived from the autore-
factor and the cycloplegic refraction. Logistic regression analysis was used to determine the association between a positive finding on each component of the screening test and the odds of having either a true-positive referral or a false-positive referral. These logistic regression models were adjusted for the child’s age.

EVALUATION OF RELIABILITY SCORE AND REFERRAL CRITERIA

We evaluated the effect of increasing the minimum acceptable reliability score on testability and accuracy. We also evaluated the effect of modifying the manufacturer’s referral criteria by developing a receiver-operator characteristic (ROC) curve to understand the trade-off between sensitivity and specificity. To generate the ROC curve, we modified those criteria most strongly associated with false-positive referrals based on the logistic regression analysis. The area under the ROC curve was calculated as another measure of test accuracy, as recommended for the evaluation of vision screening tests.10

STATISTICAL ANALYSIS

All statistical analyses were performed with Stata 8.1 (StataCorp LP; College Station, Tex).

RESULTS

PARTICIPANT SELECTION AND CHARACTERISTICS

There were 176 children eligible for participation in this study, of which 170 (97%) enrolled. Among the participants, 80 (47%) were in the younger group (<3 years old) and 90 (53%) were in the older group (3-5 years old).

Overall, 36% of the children had an abnormal comprehensive eye examination result (Table 1). Amblyopia, based on the comprehensive eye examination result, was more common among children aged between 3 and 5 years than among the younger children. The overall prevalence of refractive error and strabismus did not vary by age.

Among the 30 children with strabismus, 13 (43%) did not have a refractive error. These 13 children were evenly distributed by age (<3 years, n=6; 3-5 years, n=7). Six (46%) of these 13 children also had amblyopia.

TESTABILITY AND ACCURACY

Overall, 68% (n = 115) of the children were successfully tested according to the minimum reliability score recommended by the manufacturer (Table 2). Older children were much more likely to be successfully tested than younger children (84% vs 49%, P <.001). Most children who were successfully screened (75%) required only 1 screening attempt. This did not vary with age (P = .60). Discarding the first 10% of screening attempts did not alter testability (P = .51).

The overall mean (SD) reliability score was 7.0 (0.9). The mean reliability score was higher for older children compared with younger children (7.2 vs 6.6; P <.001). The mean reliability was lower among those with an abnormal examination result for both younger children than 3 years (6 vs 6.7, P = .02) and those aged between 3 and 5 years (6.9 vs 7.3, P = .03).

Association Between Nondilated Autorefraction and Dilated Refraction

The correlations between the components of monocular autorefraction and the corresponding components of the dilated manual refraction were similar: right sphere, 0.51 D, P <.001; left sphere, 0.58 D, P <.001; left cylinder, 0.62, P <.001; right cylinder, 0.61, P <.001; the difference in spherical equivalent between each eye was 0.44, P <.001.

Levels of sphere or cylinder or differences in spherical equivalent between each eye at or exceeding the manufacturer’s threshold were associated with greater odds of an abnormal result on a comprehensive eye examination (Table 3). However, exceeding the manufacturer’s threshold for cylinder was associated with more than a 9-fold increase in the odds of having a false-positive result (Table 3).
**Table 3. Odds of a True-positive or False-positive Monocular Autorefraction Screening Result Based on Screening Component, Adjusted for Age***

<table>
<thead>
<tr>
<th>Factor</th>
<th>True-positive Result</th>
<th>False-positive Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere, D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>−1.0 to +2.0</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>−1.0 or less or +2.0</td>
<td>7.57 (2.42-23.63)</td>
<td>0.88 (0.27-2.81)</td>
</tr>
<tr>
<td>Cylinder, D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to +1.0</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>+1.0 or more</td>
<td>4.46 (1.49-13.31)</td>
<td>9.64 (3.84-24.20)</td>
</tr>
<tr>
<td>Difference in spherical equivalent, D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to +1.0</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>+1.0 or more</td>
<td>3.38 (1.02-11.22)</td>
<td>1.57 (0.46-5.28)</td>
</tr>
</tbody>
</table>

Abbreviation: D, diopter.

*Data are given as the odds ratio (95% confidence interval).

**Figure.** Receiver-operator characteristic (ROC) curve for monocular autorefraction demonstrating the trade-off between sensitivity and specificity with changes in the criteria for referral. Point 1 represents the manufacturer’s criteria (sphere of −1.0 diopter [D] or less or +2.0 D or more; cylinder of +1.0 D or more; or the difference in the spherical equivalent of 1.0 D or more). Point 2 represents increasing the manufacturer’s criteria for cylinder (+1.4 D). Point 3 represents omitting the cylinder from the manufacturer’s criteria. Point 4 is omitting both sphere and cylinder and leaving only the criteria for difference in spherical equivalent.

**EVALUATION OF RELIABILITY SCORE AND REFERRAL CRITERIA**

Increasing the minimum acceptable reliability score further decreased the proportion of the younger children who could be tested (Table 2); therefore, we did not have sufficient statistical power to accurately determine the effect of increasing the acceptable reliability score on test accuracy among younger children (Table 2). Increasing the minimum reliability score to 6 for the older children had no statistically significant effect on testability, sensitivity, or specificity. Increasing the score to 7 decreased the proportion of older children who were testable by 47% (Table 2).

Findings from the logistic regression models suggest that increasing the threshold for referral based on the cylinder component would increase test specificity. We generated an ROC curve (Figure) based on the manufacturer’s criteria for referral (point 1), increasing the cylinder threshold to +1.4 D or higher (point 2), omitting the cylinder criteria altogether (point 3), and basing referral only on the difference in spherical equivalent (point 4). The area under the ROC curve was 0.79 (95% CI, 0.70-0.88). Increasing the threshold for cylinder referral (point 2) increased the point estimate for specificity to 68% from 52% and decreased the point estimate for sensitivity to 79% from 85%. These differences were not statistically significant, however.

In this study, most of the children 3 to 5 years of age were testable with the monocular autorefractor, and most, but not all, cases of visual impairment were detected. The low specificity suggests that monocular autorefraction in the primary care practice setting would lead to a large number of false-positive referrals. We identified revisions to the manufacturer’s referral criteria that may decrease the number of false-positive referrals; however, implementing these criteria may cause some cases to be missed.

The sensitivity and specificity that we observed when testing children aged 3 to 5 years were similar to those previously reported. However, the proportion of children who could be successfully tested was lower (84% vs >99%). Although this difference may be because of differences in training, greater experience with the device in our study was not associated with increased testability.

Within our sample, monocular autorefration identified for referral some of the children who had strabismus with no refractive error. Because monocular autorefraction cannot identify strabismus, these important referrals were likely a result of the low specificity of the device. A separate test of stereopsis (eg, the Random-Dot stereo E test) to identify isolated strabismus may be useful in conjunction with monocular autorefraction. However, increasing the number of vision screening tests would increase the difficulty of screening and the number of false-positive referrals.

Current screening recommendations call for preschool-aged children to be assessed with a test of visual acuity and a test of stereopsis. There are no high-quality data regarding the accuracy of this combined testing strategy to compare with our findings regarding monocular autorefraction.

Our study findings suggest that monocular autorefractor has limited usefulness for screening children younger than 3 years. Because of the low testability rate, more cases of visual impairment would be missed than identified, and there would be an even larger number of false-positive referrals.

The major limitation of our study is that the device was evaluated in ophthalmology clinics, not in the primary care practice setting. We chose to evaluate the device in this setting to increase the feasibility of the study because all children were already slated to receive a comprehensive eye examination and because screening in a high-prevalence setting decreases the required sample size. There are 3 ways that evaluation in an ophthalmology clinic may bias the findings from this study. First, this study may overestimate the proportion of
children who are testable because those doing the screening were familiar with vision screening and may have better strategies to optimize the vision screening process than those in a busy primary care practice setting. Second, the higher prevalence of visual impairment in ophthalmology clinics may bias measures of accuracy. Third, those performing the comprehensive eye examination were not masked to the outcome of screening. Because of the low specificity of screening, we do not feel that this knowledge significantly biased the examination.

Based on the available data, it is challenging to make a vision screening strategy for use in pediatric clinics. Few data are available regarding the accuracy of vision screening tests, and none of these data were collected in the primary care setting. To ensure best visual outcomes while preserving limited health care resources, future studies are needed to evaluate the outcomes of preschool-aged vision screening as implemented in the primary care setting, including assessment of the relative cost-effectiveness of the various available strategies.

Accepted for Publication: June 9, 2004.
Correspondence: Alex R. Kemper, MD, MPH, MS, Department of Pediatrics, 6E08-300 North Ingalls Bldg, Ann Arbor, MI 48109-0436 (kempera@med.umich.edu).
Acknowledgment: Welch Allyn Medical Products lent us the SureSight monocular autorefractors used in this study.

REFERENCES