Cost-Efficient Vision Screening for Astigmatism in Native American Preschool Children

Joseph M. Miller,[1,2,3] Velma Dobson,[1,2] Erin M. Harvey,[1,4] and Duane L. Sherrill[3]

PURPOSE. To design and test a cost-efficient, community-based vision screening program for a population of Native American preschool children in which there is a high prevalence of astigmatism.

METHODS. Based on analysis of vision screening and eye examination data from a preschool population with a 33% prevalence of astigmatism, comparative costs to conduct a 1000-child screening program with a target sensitivity of 90% were estimated for photoscreening, noncycloplegic autorefraction, autokeratometry, and Lea symbols distance visual acuity testing. Results of the cost analysis and examination of sensitivity and specificity data from the preschool population led to development of a hybrid screening program of autokeratometry and visual acuity screening with referral thresholds of 2.25 D of corneal astigmatism or inability to read a 20/63 Lea symbols line on two separate attempts. The screening program was prospectively implemented in a community-based screening of a similar cohort of 167 children, and its efficiency was evaluated by comparison to results of cycloplegic refraction.

RESULTS. The community-based screening showed 96.8% sensitivity and 79.2% specificity for detecting the presence of refractive astigmatism of 1.50 D or more.

CONCLUSIONS. Referring children who have at least 2.25 D of corneal astigmatism or acuity worse than 20/63 on two attempts, provides the high sensitivity and specificity associated with automated keratometry while maintaining an acuity component that can detect other causes of reduced acuity in the absence of astigmatism. (*Invest Ophthalmol Vis Sci.* 2003;44: 3756–3763) DOI:10.1167/iovs.02-0970

Some Native American tribes have a high prevalence of astigmatism, resulting in poor uncorrected visual acuity (Kershner RM, et al. *IOVS* 1984;25:ARVO Abstract 70).[1,2] Uncorrected astigmatism can also impair visual development, resulting in refractive amblyopia (below-normal best corrected visual acuity) and the subtype, meridional amblyopia, that arises because of the inability of the uncorrected astigmatic eye to focus orthogonally oriented lines simultaneously.[3,4] Although the relation between depth of amblyopia and risk factors, including the age at which an individual experienced

uncorrected astigmatism and the magnitude and duration of the uncorrected astigmatism, is poorly defined and a current topic of investigation,[5] studies have indicated that correction of astigmatism should be provided before the elementary school years if amblyopia is to be avoided.[4,6] Unlike hyperopia, in which, with sufficient accommodative effort, the child can bring the visual environment into focus, or myopia, in which the child can experience the world at near in sharp focus without any accommodative effort, uncorrected astigmatism prevents the child from seeing clearly at any time or at any distance.

There are two approaches that can be used to screen children for astigmatism. One is to use an instrument to detect physical characteristics of the eye that can cause a poorly focused image to form on the retina. This approach includes photoscreening[7,8] and autorefraction,[9,10] to screen for overall (refractive) astigmatism, and autokeratometry[11,12] to screen for corneal astigmatism, which is efficient because astigmatism among Native Americans has been shown to be primarily corneal in origin (Kershner RM, et al. *IOVS* 1984;25:ARVO Abstract 70).[11,13–16] The second, more conventional approach to vision screening is behavioral, in which the child is asked to "read" an eye chart, by naming or matching letters[17,18] or by naming or matching shapes.[19,20] With this method, if the child fails to identify targets with a small visual angle, he or she is presumed to have a problem with the visual system that warrants further study.[21–23] However, failure of a young child to identify small targets may also result from inattention or shyness, resulting in acuity estimates that can be falsely poor when in fact no organic cause of decreased visual acuity (e.g., refractive error) is present.

We have recently completed a study comparing the ability of four vision-screening methods to detect significant astigmatism in a population of preschool children participating in Head Start programs on the Tohono O'Odham reservation.[24] Photorefraction, autorefraction, autokeratometry, and measurement of visual acuity were compared by means of receiver operator characteristic (ROC) curves, and the screening utility of each method was rank ordered and characterized by maximum-likelihood estimates. The published ROC curves describe the expected sensitivity and specificity of each screening method when the method is used in a similar population of children. One important limitation of ROC curve analysis, however, is that, although the method allows comparison of screening methods and comparison of associated sensitivities and specificities, the optimum referral threshold is not directly specified. For specification of the optimum referral threshold, a more detailed analysis must be performed, and then the predictions of that analysis must be confirmed in a subsequent population.

The purpose of this article is to describe the methods used to design a cost-efficient, community-based vision screening protocol optimized for a preschool-aged population in which there is a high prevalence of astigmatism and then to report the results of the implementation of this protocol by Head Start program staff. First, the costs associated with implementing each of the four previously reported screening methods[24] were compared, and an economic analysis was used to determine

TABLE 1. Assumptions of the Economic Model

1. Prevalence		33%
2. Population size		1000 (200 per year for 5 years)
3. Target sensitivity		90%
4. Cost of eye examination		$50
5. Acquisition cost of screening instruments	MTIPS	$2000 (40*)
	NCARS	$15,000 (300*)
	KERS	$4000 (80*)
	LSVAS	$50 (1*)
6. Instrument specificity at 90% sensitivity	MTIPS	36%
	NCARS	96%
	KERS	80%
	LSVAS	50%

* Number of complete eye examinations equivalent to the cost of the instrument, at $50 per eye examination.

which screening method would be most appropriate for screening a specified number of subjects. Second, a screening protocol was developed for our target population, and a sensitivity analysis was performed on the dataset that led to the creation of the ROC curves,[24] to determine recommended referral levels and to predict the number of children who would be either correctly or incorrectly referred were that program to be implemented. Finally, we used the protocol to screen another cohort of 167 preschool children and compared the results with data from comprehensive examinations on these children to determine the actual efficiency of the program in a community-based screening effort.

METHODS

Economic Analysis of Four Screening Methods

Assumptions of the Economic Model. To determine which of four previously evaluated screening methods should be recommended for subsequent implementation in a community-based screening program, we performed an economic analysis based on assumptions of (1) the expected disease prevalence, (2) the number of children to be screened, (3) the target level of sensitivity, (4) the cost of an eye examination, (5) the acquisition cost of each screening instrument, and (6) the performance of each screening instrument. We assumed that the composition of the target population (Head-Start–aged children who were members of a Native American tribe that has a high prevalence of astigmatism) would remain similar to that observed in our previous work.[24] We also assumed that there would be rare instances of ocular disease in the target population, because among the children tested for our prior report, there were 3 who had ocular disease (data from these three were excluded from analyses) and 21 who had special needs as identified by the Head Start staff, of whom a significant number had ocular disease (data also excluded from analyses).

For assumption 1, the expected prevalence of astigmatism, we used 33%, which was the proportion of preschool children examined over a 2-year period (fall 1997 through fall 1998) who met the criteria for prescription of glasses for astigmatism ($\geq$2.00 D cylinder in a 3-year-old; $\geq$1.50 D cylinder in a 4- or 5-year-old).[24] During 4 years of screening children in the tribal Head Start program (fall 1997 through fall 2000), we tested 541 children, and so a reasonable assumption in an expanded screening program that accessed all 3- to 5-year-old children of the tribe over a period of approximately 5 years would be 1000 children screened, the number applied in assumption 2. After reviewing the results of our study[24] with the study's National Eye Institute–appointed Data Monitoring and Oversight Committee that included vision scientists, clinicians, and community representation, we selected a target sensitivity of 90% for each of the selected screening methods (albeit at very different specificities), and this percentage was used for assumption 3. For assumption 4, we set the cost of an eye

examination at $50, and did not adjust for inflation over time. The $50 estimate of the cost of an eye examination is based on an informal review of the cost in a retail setting and was also used in a preliminary report of data from this study.[9] For assumption 5, the acquisition cost of the Photoscreener (Medical Technology, Inc. [MTI], Riviera Beach, FL) was assumed to be $2000, the autorefractor (Retinomax K-Plus; Nikon, Inc., Melville, NY) to be $15,000, the autokeratometer (Nidek KM500; Marco Ophthalmic Instruments, Inc., Jacksonville, FL) to be $4000, and the Lea symbols distance visual acuity chart (Precision Vision, LaSalle, IL) to be $50. Although there are other photorefractors, autorefractors, autokeratometers, and preschool visual acuity charts that could be used in a screening environment and their acquisition costs may differ considerably from those of the instruments used in our analysis, we do not have performance data for these instruments and therefore cannot generate ROC curves for the target population. For assumption 6, estimates of instrument performance (specificity at the selected level of sensitivity) were based on analysis of ROC curves derived in our previous work.[24] These assumptions are summarized in Table 1, and are the basis of the economic model.

The ROC curves used for assumption 6 are illustrated in Figure 1. These curves were estimated by a maximum-likelihood estimate computer program.[25] Figure 1 illustrates the specificity achieved at 90% sensitivity, shown in Table 1.

The ROC curves reflect the ability of MTI photoscreening (MTIPS), noncycloplegic autorefraction screening (NCARS), autokeratometry screening (KERS), and Lea symbols visual acuity screening (LSVAS) to identify correctly whether eyeglasses should be prescribed for one of our Head Start program participants. In the initial design of our program, the decision to prescribe eyeglasses was based on the age of the
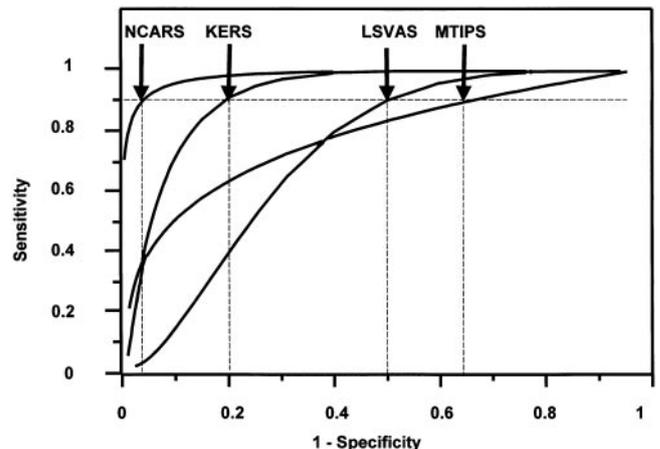

FIGURE 1. Receiver operator characteristic curves for the four evaluated screening techniques.[24] *Arrows:* each curve, as well as the point at which they achieve 90% sensitivity.

TABLE 2. Cost-Efficiency of Vision-Screening Methods

Number Screened	NCARS 96% Specificity	KERS 80% Specificity	LSVAS 50% Specificity	Refer All 0% Specificity
1	0	0	0	1
5	0	1	2	3
100	3	13	34	67
400	11	54	134	268
1000	27	134	335	670
2052	55	276	690	1380
2500	67	335	838	1675
5000	134	670	1675	3350

Data reflect the number of false-positive referrals at 90% sensitivity when the prevalence of need for glasses is 33%. Shaded areas indicate the sample size for which each screening modality is most efficient relative to the other modalities—that is, the sample size at which the added cost to purchase the next most expensive instrument is equal to the cost of the extra unnecessary examinations that would occur if the less-expensive screening modality were used.

child and the amount of refractive error present on cycloplegic examination.[5] In the implementation of the program, all children for whom glasses were prescribed met the criterion for prescription of glasses for astigmatism ($\geq$2.00 D for children younger than 48 months of age and $\geq$1.50 D for children 48 months of age or older[26]). Thus, the ROC curves essentially reflect the ability of each instrument to identify correctly the preschool children with astigmatism requiring spectacle correction. As approximately equal numbers of children were present in each age group (45% were younger than 48 months age), the ROC curves should be viewed as describing the ability to detect between 1.50 and 2.00 D of refractive astigmatism in a population of 3- to five-year-old children with astigmatism, 33% of whom needed spectacle correction.

Cost-Efficiency of Each Screening Instrument as Predicted by the Economic Model. In a screening program that is tightly coupled with a referral program, there is a direct correlation between false-positive referrals and associated costs, because each unnecessary referral either results in a charge (in a system in which patients seek and pay their own eye care provider) or loss of an appointment slot for another individual (in a capitated system, in which the costs are shared by all program participants). If level of sensitivity is specified (e.g., 90% in Table 1) and the prevalence of disease is known for the population (e.g., 33% in Table 1), it becomes possible to compare screening techniques on the basis of costs associated with these unwarranted examinations arising from referrals in which no disease is found. Until a sufficient number of children are screened to "pay for" a more expensive screening instrument that has a higher sensitivity (i.e., lower number of false referrals), a less expensive screening method is preferred.

Table 2 was constructed by rank ordering the assumed acquisition costs of each screening method and then comparing these acquisition costs to the specificity for each screening instrument that would be expected at the desired sensitivity. The number of false-positive referrals was calculated as the number of children screened $\times$ (1 − prevalence) $\times$ (1 − specificity). In all instances, the sensitivity was estimated to be 90%, and the prevalence was estimated to be 0.33.

Differences in operating costs (such as the relative amount of technician time each instrument requires, and the cost of consumable supplies, such as film for the photoscreener) were not included. Although there is no question that the least expensive instruments (visual acuity testing and photoscreening) are the most sensitive to the skill level of the tester, it appears that in the present study they would be appropriate only for very small screening volumes (Table 2). Furthermore, including the higher training costs would only make them more costly to administer. In addition, the logistics of screening require that the child be transported from the classroom to the screening station; that name and date of birth be verified; that paperwork be correctly filled out; and, after data collection, that the child be returned

to the classroom. This "overhead" is the same, regardless of screening instrument used, and results in a blunting of benefit of the very brief test times afforded by automated instruments.

Table 1 shows that the most expensive instrument (NCARS) also has the highest expected specificity (96%), and the second most expensive instrument (KERS) has the second highest expected specificity (80%). However, the third most expensive instrument (MTIPS) has a lower expected specificity (36%) than LSVAS (50%). Because MTIPS is more expensive to acquire than an eye chart, has associated recurring expenses in film (ignored in the present analysis), and has not been demonstrated to be less labor intensive than visual acuity screening, in the time required to photograph a child and interpret the image, it was immediately eliminated from further consideration for this specific population of Native American preschool children with astigmatism.

We then calculated the numbers of false-positive referrals that would occur when different numbers of children were screened with each of the three remaining screening instruments (NCARS, KERS, and LSVAS; Table 2). When the difference in acquisition cost between two instruments (in units of eye examinations, shown in Table 1) becomes less than the difference in the number of false-positive referrals between the two instruments, the more expensive screening modality becomes cost-efficient at that screening volume, compared with the next less-expensive instrument (represented by gray shading in Table 2). For example, screening of 2052 children would result in 55 false-positive referrals with NCARS, compared with 275 with KERS—a difference of 220 examinations at $50, or a total of $11,000. This is equal to the difference in cost ($11,000) between the NCARS and KERS instruments and is therefore the threshold number of screenings at which NCARS becomes more efficient than KERS. This threshold number is derived by dividing the incremental number of false-positive examination results to be eliminated (220 in this example) by the difference in specificity (0.96–0.80) $\times$ 1 − specificity (0.67, in this example). In the present example 2052 = 220/(0.16 $\times$ 0.67), after rounding. This screening volume is inversely proportional to the dollar amount assigned for an eye examination. For example, if the amount assigned to an eye examination is doubled to $100 from $50, then the number of subjects at which the autorefractor becomes cost-efficient is halved to 1026 from 2052.

It is interesting to note, as shown in Table 2, that for extremely small screening volumes (less than five children), it is most cost-efficient simply to send the children for an eye examination. For a very small screening program (fewer than 400 children to be screened), it is most cost-efficient to screen the children by LSVAS. For a screening program volume of at least 400 children, screening with an autokeratometer (KERS) produces fewer unnecessary eye examinations than the number associated with LSVAS, and this smaller number is sufficient to offset the higher capital acquisition cost of the autokeratom-

eter. For a large screening program of at least 2052 children, the autorefractor screening method (NCARS) is financially more advantageous than the autokeratometer (KERS) in the number of unneeded referrals generated compared with the higher acquisition cost.

Based on the economic analysis summarized in Table 2 and on the characteristics and number of children to be screened (Table 1), we selected the autokeratometer (KERS) as the primary instrument in our preschool screening protocol.

Design of a Hybrid Screening Program

The Need to Incorporate a Visual Acuity Test into an Autokeratometry Screening Program. In our target population, autokeratometry screening is very efficient in predicting the need for eyeglasses, because the predominant cause of refractive error is astigmatism, and most of the astigmatism is corneal in origin.[24] However, it is possible that other significant disease may exist that would be completely missed by keratometry screening, because a keratometer measures only the shape of the anterior surface of the cornea for the presence of corneal astigmatism. Because visual acuity testing is recommended as a preschool vision screening tool by pediatric ophthalmologists, optometrists, and pediatricians,[21–23] we decided to add visual acuity screening with a single line of letters to our keratometry screening protocol to detect more rare (in this population) defects of the visual system, such as cataract and strabismic amblyopia.

Hybrid Screening Program Procedure. The hybrid screening procedure is based on an initial screening using autokeratometry, followed by visual acuity screening of children who pass autokeratometry screening. Only children who pass autokeratometry screening participate in visual acuity screening. Children who fail autokeratometry screening are referred for a comprehensive eye examination, and if problems with acuity exist, it is assumed that they will be found in the course of the eye examination. If a child passes autokeratometry screening and the initial visual acuity screening, no more testing is needed and the child's screening result is "pass." However, a child who passes autokeratometry screening but fails the initial acuity screening has acuity rescreened on a separate day. Only those who fail both acuity screenings are referred for a comprehensive eye examination. Thus, initial screening with autokeratometry reduces the pool of children who must have acuity screening performed, whereas the requirement for retesting children who pass autokeratometry screening and fail the first acuity screening reduces the number of children who are referred as a result of an acuity deficit related to poor performance rather than to an ocular problem.

Specification of Referral Criteria for the Hybrid Screening Program. To determine referral thresholds for the hybrid vision screening program, an analysis was conducted on data from the 379 children, 36 to 63 months of age, who participated in the study comparing photorefraction, autorefraction, autokeratometry, and visual acuity screening.[24] None of the children had special needs as defined by the Head Start program and none had evidence of ocular disease, based on a comprehensive eye examination. As indicated in the previous publication,[24] the research adhered to the tenets of the Declaration of Helsinki, parents provided informed consent before testing, and the research was approved by the institutional review board of the University of Arizona.

Testing was conducted at Head Start sites on the Tohono O'Odham reservation. Each child was tested with each of the four screening procedures, followed by a comprehensive eye examination including cycloplegic refraction, during the fall of the 1997/1998 or 1998/1999 academic year. Among the 379 children, 1 had incomplete autorefraction data, 1 had incomplete autokeratometry data, 30 were unable to perform visual acuity testing, and 3 were wearing spectacles during visual acuity testing, which rendered their acuity noninformative for screening purposes.

Three hundred eight of the 379 children participated in follow-up testing in the spring of the same academic year. The follow-up session included measurement of visual acuity, followed by a comprehensive

eye examination, including cycloplegic refraction. Of the 308 children who participated in spring testing, 4 were unable to perform visual acuity testing, and 50 were wearing spectacles during visual acuity testing, which rendered their acuity noninformative for screening purposes.

The first step in the determination of referral criteria was to define disease-positive status (significant refractive astigmatism warranting spectacle correction). A single criterion of astigmatism 1.50 D or more in one or both eyes, based on the results of the cycloplegic refraction, was used. The definition of disease-positive status had to include only refractive error, because subjects having other disease were excluded from the dataset used for analysis. This definition of disease-positive status differs from that in our earlier work, in which age-dependent criteria for prescribing glasses were used.[5,24,26] The effect of this change is to provide a slightly more conservative definition for the younger children, to provide a single criterion regardless of age. Of the 379 subjects, 142 (37.5%) met this criterion for disease-positive status. No children had significant hyperopia (>4.50 D) or significant myopia (>2.50 D if <48 months of age; >1.50 D if >48 months) in the absence of the specified amount of astigmatism.

The next step in determining referral thresholds for the hybrid screening program was to write a computer program that could analyze the dataset under a variety of scenarios. The input to the computer program was the amount of refractive astigmatism in the eye with the higher amount of astigmatism (from cycloplegic refraction), the amount of corneal astigmatism in the eye with the higher amount of corneal astigmatism (from KERS), and the visual acuity that was obtained from the eye with the worse acuity, as measured by trained testers using a Lea symbols chart mounted in a retroilluminated cabinet at a 3-m test distance, on up to two occasions (LSVAS). Children who were present during both the fall and spring testing sessions contributed both initial and follow-up visual acuity results for each eye to the data analysis, whereas children who were absent in either the spring or the fall contributed only a single visual acuity result.

We assumed in this analysis that a community-based measurement of corneal astigmatism using KERS would not differ from that we obtained as part of our testing program, because the autokeratometer does not seem to require a great deal of practice or skill on the part of the operator to obtain reliable readings. For analysis of visual acuity results, we assumed that a child whose visual acuity was measured to be a certain value (e.g., 20/32) under the test conditions of our program would be able to read a single line of letters administered by a community tester if that line were of the same size or larger, after warm-up testing with a single line of larger (20/80) letters. We recognize that there are differences in performance when a child has more extensive practice with larger letters as in a full-threshold measurement, and that there may be additional differences that arise from using a single-line presentation. This difference, if present, would become apparent when the protocol was prospectively evaluated.

The computer program varied referral thresholds for both KERS and LSVAS. If the child passed KERS at the target referral level, the threshold visual acuity for the worse eye from the first visual acuity test session in which the child produced uncorrected acuity results was evaluated to determine whether it was equal to or better than the target LSVAS referral level. If the first visual acuity measured in a child who had passed autokeratometry screening was worse than the target screening acuity threshold or the child was unable to perform acuity testing, the computer program would then check for the presence of a second acuity score, which would have been collected at the spring follow-up session. If the child's second acuity score was equal to or better than the target screening acuity threshold, the computer program scored the child's acuity results as a "pass." If the child had a spring acuity score that was worse than the target screening acuity threshold, or the child was unable to perform acuity testing, the computer scored the child's acuity results as a "fail." If the child was absent for the spring follow-up session or was tested in spectacles at the spring follow-up session, then the child was assigned the status of having provided incomplete data for the analysis. Also assigned to

TABLE 3. Subject Classification by Different Referral Thresholds When the Hybrid Screening Algorithm Is Applied to the Data of the Cohort of Preschool Children Tested in the Study

Referral Criteria		Children with Significant Refractive Error (Cyl ≥ 1.50 D) $n = 142$			Children without Significant Refractive Error (Cyl < 1.50 D) $n = 237$			Overall Percentage of Children Who Would Be Referred*	
Keratometry (D)	Acuity† (LogMAR)	TP	FN	IP	TN	FP	IN	Lower	Upper
1.75	0.3	140	0	2	71	154	12	78%	81%
	0.4	139	2	1	96	133	8	72%	74%
	0.5	139	2	1	102	129	6	71%	73%
	0.6	139	2	1	106	127	4	70%	72%
	0.7	139	2	1	111	124	2	69%	70%
2.00	0.3	136	0	6	96	124	17	69%	75%
	0.4	133	4	5	130	95	12	60%	65%
	0.5	132	6	4	140	87	10	58%	61%
	0.6	132	8	2	147	83	7	57%	59%
	0.7	132	8	2	152	80	5	56%	58%
2.25	0.3	134	0	8	111	104	22	63%	71%
	0.4	130	6	6	152	70	15	53%	58%
	0.5	129	9	4	165	59	13	50%	54%
	0.6	129	11	2	174	54	9	48%	51%
	0.7	129	11	2	180	50	7	47%	50%
2.50	0.3	127	1	14	115	99	23	60%	69%
	0.4	120	11	11	160	61	16	48%	55%
	0.5	119	17	6	174	49	14	44%	50%
	0.6	119	20	3	184	43	10	43%	46%
	0.7	119	21	2	190	39	8	42%	44%
2.75	0.3	121	1	20	119	93	25	56%	68%
	0.4	114	14	14	167	52	18	44%	52%
	0.5	110	23	9	183	38	16	39%	46%
	0.6	105	32	5	194	32	11	36%	40%
	0.7	104	35	3	200	28	9	35%	38%

The shaded row represents the screening referral thresholds that were applied in the prospective assessment of the hybrid screening program.
TP, true positive; FN, false negative; IP, incomplete positive; TN, true negative; FP, false positive; IN, incomplete negative.
* Lower = (TP + FP)/(total number of subjects). Upper = (TP + IP + FP + IN)/(total number of subjects).
† LogMAR and Snellen fraction acuities are related by 0.3 logMAR = 20/40 Snellen, 0.4 = 20/50, 0.5 = 20/63, 0.6 = 20/80, and 0.7 = 20/100.

"incomplete data" status was the one child in whom no keratometry measurement was obtained. All three children who wore spectacles for acuity testing in the fall had sufficient corneal astigmatism to fail autokeratometry screening at all threshold levels tested and therefore they were not assigned to incomplete data status, based on visual acuity results.

Table 3 lists the observed attributes of the children by retrospective application of the hybrid screening program. Children who had significant refractive error, herein defined as 1.50 D or more refractive astigmatism, were classified as either having a true-positive (failed keratometry, or passed keratometry and failed two acuity tests), false-negative (passed keratometry and at least one acuity test), or incomplete-positive (passed keratometry and failed the fall acuity test, but did not have results from the spring acuity test) screening classification. Children without significant refractive error were classified as true-negative (passed keratometry and at least one acuity test), false-positive (failed keratometry, or passed keratometry and failed two acuity tests), or incomplete-negative (passed keratometry and failed the fall acuity test, but did not have results from the spring acuity test) screening classification.

Table 3 also lists the percentage of children who would be referred for follow-up examination as a result of application of the hybrid screening program. The high percentage of referrals is consistent with the high disease positive rate in this sample, in which 37.5% (142/379) of the children needed referral.

Table 4 shows the range of sensitivities and specificities that result from analysis of the data set at various referral levels at or above the target sensitivity of 90%. Lower- and upper-bound estimates are based on exclusion and inclusion, respectively, of the incomplete subjects in the numerator for the calculation (see Table 4 for formulas).

After discussion with our Data Monitoring and Oversight Committee, the referral criteria shaded in gray marked in Tables 3 and 4 were selected for implementation and testing in the prospective, community-based evaluation of the hybrid screening program. The selected referral thresholds appeared to be reasonable for several reasons. First, the keratometry screening threshold of 2.25 D agreed with our previous findings that corneal astigmatism exceeds refractive astigmatism in this population by an average of 0.75 D,[11] so it would seem reasonable that if we were screening for the presence of 1.50 D of refractive astigmatism, we would use a referral threshold of 2.25 D of corneal astigmatism. Second, among the five candidate visual acuity thresholds (20/40, 20/50, 20/63, 20/80, and 20/100) that produced sensitivity of 90% or greater when paired with the 2.25 D keratometry referral criterion, the 20/80 and 20/100 referral thresholds offered the highest predicted ranges of specificity, and the specificities for the 20/100 referral threshold were only one to three percentage points better than those for the 20/80 threshold. The 20/80 referral threshold was selected, because it offered the more conservative threshold.

In screening for a 20/80 acuity referral threshold, a child would be presented with a 20/63 line. If the child could see that line, it would imply that he/she could see the 20/80 line as well. If the child could not see the 20/63 line, then the acuity would be presumed to be 20/80 or worse, warranting referral.

Finally, although the design target sensitivity was 90%, the set of referral criteria selected were anticipated to produce 91% to 92% sensitivity. This was to allow for any performance changes that might exist when screening tests were administered by a community-based tester rather than by a member of the research team and when both acuities were performed within a few days or weeks, as opposed to being separated by months. Thus, data from our original subject pop-

TABLE 4. Sensitivity and Specificity of Different Referral Thresholds When the Hybrid Screening Algorithm Is Applied to the Data of the Cohort of Preschool Children Tested in the Study

Referral Criteria		Sensitivity*		Specificity†	
Keratometry (D)	Acuity (logMAR)	Lower Bound	Upper Bound	Lower Bound	Upper Bound
1.75	0.3	99%	100%	30%	35%
	0.4	98%	99%	41%	44%
	0.5	98%	99%	43%	46%
	0.6	98%	99%	45%	46%
	0.7	98%	99%	47%	48%
2.00	0.3	96%	100%	41%	48%
	0.4	94%	97%	55%	60%
	0.5	93%	96%	59%	63%
	0.6	93%	94%	62%	65%
	0.7	93%	94%	64%	66%
2.25	0.3	94%	100%	47%	56%
	0.4	92%	96%	64%	70%
	0.5	91%	94%	70%	75%
	0.6	91%	92%	73%	77%
	0.7	91%	92%	76%	79%
2.50	0.3	89%	99%	49%	58%
	0.4	85%	92%	68%	74%
	0.5	84%	88%	73%	79%
	0.6	84%	86%	78%	82%
	0.7	84%	85%	80%	84%
2.75	0.3	85%	99%	50%	61%
	0.4	80%	90%	70%	78%
	0.5	77%	84%	77%	84%
	0.6	74%	77%	82%	86%
	0.7	73%	75%	84%	88%

The shaded row represents the screening referral thresholds that were applied in the prospective assessment of the hybrid screening program. Abbreviations are defined in Table 3.
* Lower bound = TP/(TP + FN + IP), Upper Bound = (TP + IP)/(TP + FN + IP).
† Lower bound = TN/(TN + FP + IN), Upper Bound = (TN + IN)/(TN + FP + IN).

ulation suggest that the hybrid screening protocol would be expected to perform in this population with a sensitivity between 91% and 92% and specificity between 73% and 78%.

Prospective Evaluation of the Hybrid Screening Protocol

To evaluate the efficiency of the hybrid screening protocol and the selected referral thresholds, a manual of procedures was developed, and the Head Start Health Services Coordinator was trained in the use of the Nidek KM-500 autokeratometer (Marco Ophthalmic, Inc.), the Massachusetts Visual Acuity Test (MASSVAT) Lea symbols (Precision Vision, Inc., La Salle, IL), which is a flip chart with one line of symbols per page,[27] and the hybrid screening protocol. The Health Services Coordinator then independently implemented the hybrid screening program in a cohort of 167 Head Start program participants during the fall of 2000. On a separate day, each child also underwent a complete eye examination including cycloplegic refraction by the research team.[5,24] The research adhered to the tenets of the Declaration of Helsinki, informed consent was obtained from parents or guardians of

all subjects, and the research was approved by the institutional review board of the University of Arizona.

RESULTS

Table 5 lists the results that were obtained when the hybrid screening program was implemented using the listed referral criteria. Sensitivity of 96.8% and specificity of 79.2% were obtained in detecting the presence of significant astigmatism ($\geq$1.50 D) in this population.

There were three false negative misclassifications. Each of these three astigmatic children had 1.50 D of refractive astigmatism in one eye and either 1.00 D ($n = 2$) or 0.50 D ($n = 1$) of refractive astigmatism in the other eye, but had less than 2.25 D keratometric astigmatism in either eye, when measured by autokeratometry. Eighteen false-positive misclassifications occurred in children who failed autokeratometry screening, with 2.25 D or more corneal astigmatism in one eye, while having less than 1.50 D of refractive astigmatism in the more

TABLE 5. Hybrid Screening Program Referral Criteria and Results of Implementation of the Program by the Head Start Health Services Coordinator in 167 Preschool Children

Referral criterion for Nidek KM-500 autokeratometer results	Refer at 2.25 D of corneal astigmatism in either eye
Referral criterion for Lea symbols Massachusetts Visual Acuity Test	Refer if child cannot read at least 3 of 5 letters of 20/63 size with each eye on up to two attempts
Definition of disease positive status	Refractive astigmatism $\geq$ 1.50 D in either eye.
Number with disease positive status/number screened	71/167 (42.5%)
Sensitivity achieved	68/71 (96.8%)
Specificity achieved	76/96 (79.2%)

astigmatic eye. The amount of refractive astigmatism present in the more astigmatic eye in the 18 false-positive misclassifications was 0.25 D ($n = 1$), 0.50 D ($n = 4$), 0.75 D ($n = 4$), 1.00 D ($n = 5$), or 1.25 D ($n = 4$).

To evaluate the potential negative effect on specificity that occurred by including the visual acuity screening component in the hybrid screening program, we compared the performance of the hybrid screening program with the performance that would have been obtained if autokeratometry screening alone had been used. Sensitivity remained at 96.8%, whereas specificity increased to 80.2% because there was one child without astigmatism who passed autokeratometry screening but failed visual acuity screening twice. Thus, it appears that the hybrid screening program, which combines autokeratometry and visual acuity screening does so with no significant loss of specificity in this population (compared with autokeratometry screening alone).

DISCUSSION

We have described a method of economic analysis that can be used to compare alternative methods for screening a population for vision problems, to determine which method is most advantageous for a given number of patients to be screened. This economic analysis led to the selection of autokeratometry screening as the preferred method to be used in a community-based program designed to screen a preschool-aged population with a known high prevalence of astigmatism. We then described the need for a hybrid screening program that incorporates a visual acuity screening component, and performed a retrospective analysis of data collected from preschool-aged children in the study population to estimate referral thresholds for this hybrid screening program to achieve a desired level of sensitivity and specificity. Finally, we prospectively evaluated this hybrid screening program in a cohort of Head Start children, when the testing was performed independently by a community-based screener. The results of the community-based hybrid screening program agree closely with model predictions, and the method appears to be innovative in that it allows for a component of visual acuity screening at a level of specificity that is usually not associated with visual acuity testing of preschool-aged children.

The hybrid screening program that was implemented was able to incorporate visual acuity screening without significant degradation of specificity, because the purpose of the acuity screen is not to detect the need for spectacles, but rather the presence of significant eye problems that may arise from reasons other than astigmatism. How useful is it to detect acuity worse than 20/63, the level at which referral is instituted in the hybrid screening program? In the Amblyopia Treatment Study, which included children between 3 and 7 years of age with moderate strabismic or anisometropic amblyopia (acuity of 20/100 or better in the amblyopic eye),[28] the amblyopic eye had acuity of 20/80 or worse at the time of enrollment in 65% of children less than 4 years of age, in 49% of children from 4 to 5 years of age, and in 45% of children from 5 to 6 years of age. Thus, a referral acuity threshold of vision worse than 20/63 might be expected to have about a 50% chance of detecting moderate amblyopia arising from either strabismus or anisometropia.

The decision to screen preschool-aged children for a visual acuity value other than the commonly recommended value of 20/40 was made with some trepidation. The ROC curve shows that visual acuity screening alone, when performed with 90% sensitivity, has only 50% specificity. We have shown in this population[9] that when the conventionally recommended referral criterion of worse than 20/40 acuity in either eye or a

two-line difference between eyes is applied, sensitivity of 91% and specificity of 44% was found. We modified the conventional recommendation in two ways. First, we changed the criterion to visual acuity worse than 20/63 in either eye and eliminated the intereye difference as a referral criterion. Second, we gave each child a second chance to recognize three of five 20/63 Lea symbols in a line. Preschool children frequently ($>30\%$ of time) improve a line or more on the second test of acuity,[18] and so a second attempt to reach the minimum visual acuity threshold improves the specificity, but not to the degree that use of an automated instrument does. The hybrid screening method presented allows visual acuity testing to be performed for the detection of significant visual acuity decrease that occurs because of a mechanism that cannot be detected by the automated keratometry screener. We believe that as new modalities develop for detecting risk factors for amblyopia, there remains a role for visual acuity testing as part of the screening process. Only visual acuity measurement tests the global visual performance of the child, but the variability in test performance both within a child and between children makes acuity screening at 20/40 inherently unreliable, compared with other automated methods of screening. We have demonstrated a hybrid screening protocol that effectively incorporates both behavioral and structural testing.

In this unique population of children with astigmatism, the Photoscreener (MTI, Riviera Beach, FL) was not cost-efficient at any volume for the target sensitivity of 90%. It appears that this is a consequence of the relatively low specificity achieved for a comparably high sensitivity, and the tendency for trained raters to demonstrate significantly higher specificities (in this population, 88%–89%) than sensitivities (in this population, 57%–58%).[8,24] Our ROC analysis suggested that at 90% sensitivity, only 35% specificity would be expected. Examination of the ROC curve (Fig. 1) demonstrates how the Photoscreener becomes more efficient than LSVAS for sensitivities lower than approximately 80%. It is not apparent how to recalibrate the photograph interpretation process, for there is no ready method to adjust referral criteria in photoscreening, given the nonlinear interaction between pupil size, magnitude of refractive error, and crescent size.[29]

In summary, we have designed and validated a hybrid screening program to detect the need for spectacle wear in a population of Native American children who have a high prevalence of astigmatism (approximately 33%). The hybrid screening program detects the need for spectacle correction through the use of automated instrumentation, while retaining the ability to detect other visual pathway conditions that result in reduced acuity that are not detected by keratometry. Although the screening protocol introduced here is appropriate only for a preschool population with a high prevalence of astigmatism, the methodology used to design the protocol can be applied in other populations with problems that warrant screening.

References

1. Garber JM. High corneal astigmatism in Navajo school children and its effect on classroom performance. *J Am Optom Assoc.* 1981;52: 583–586.

2. Dobson V, Tyszko RM, Miller JM, Harvey EM. Astigmatism, amblyopia, and visual disability among a Native American population. In: *Vision Science and Its Applications. Technical Digest.* Washington, DC: Optical Society of America; 1996;139–142.

3. Freeman RD, Mitchell DE, Millodot M. A neural effect of partial visual deprivation in humans. *Science.* 1972;175:1384–1386.

4. Mitchell DE, Freeman RD, Millodot M, Haegerstrom G. Meridional amblyopia: evidence for modification of the human visual system by early visual experience. *Vision Res.* 1973;13:535–558.

5. Miller JM, Dobson V, Harvey EM, Sherrill DL. Astigmatism and amblyopia among Native American children (AANAC): design and methods. *Ophthalmic Epidemiol.* 2000;7:187–207.

6. Cobb SR, MacDonald CF. Resolution acuity in astigmats: evidence for a critical period in the human visual system. *Br J Physiol Opt.* 1978;32:38–49.

7. Ottar WL, Scott WE, Holgado SI. Photoscreening for amblyogenic factors. *J Pediatr Ophthalmol Strabismus.* 1995;32:289–295.

8. Donahue SP. Johnson TM. Leonard-Martin TC. Screening for amblyogenic factors using a volunteer lay network and the MTI photoscreener: initial results from 15,000 preschool children in a statewide effort. *Ophthalmology.* 2000;107:1637–1644.

9. Miller JM, Harvey EM, Dobson V. Visual acuity screening versus noncycloplegic autorefraction screening for astigmatism in Native American preschool children. *J Am Assoc Pediatr Ophthalmol Strabismus.* 1999;3:160–165.

10. Pizarello L, Tilp M, Tiezzi L, Vaughn R, McCarthy J. A new school-based program to provide eyeglasses: ChildSight. *J Am Assoc Pediatr Ophthalmol Strabismus.* 1998;2:372–374.

11. Dobson V, Miller JM, Harvey EM. Corneal and refractive astigmatism in a sample of 3-to 5-year-olds with a high prevalence of astigmatism. *Optom Vis Sci.* 1999;76:855–860.

12. Harvey EM, Miller JM, Dobson V. Reproducibility of corneal astigmatism measurements with a hand held keratometer in preschool children. *Br J Ophthalmol.* 1995;79:983–990.

13. Maples WC, Atchley JW, Hughes J. Refractive profile of Navajo children. *J Behav Optom.* 1996;7:59–64.

14. Goss DA. Astigmatism in American Indians: prevalence, descriptive analysis, and management issues. In: Goss DA, Edmondson LL, editors. *Eye and Vision Conditions in the American Indian.* Yukon, OK: Pueblo Publishing Press; 1990:61–76.

15. Heard T, Reber N, Levi D, Allen D. The refractive status of Zuni Indian children. *Am J Optom Physiol Opt.* 1976;53:120–123.

16. Lyle WM, Grosvenor T, Dean KC. Corneal astigmatism in Amerind children. *Am J Optom Arch Am Acad Optom.* 1972;49:517–524.

17. Friendly DS. Preschool visual acuity screening tests. *Trans Am Ophthalmol Soc.* 1978;76:383–480.

18. Holmes JM, Beck RW, Repka MX et al. The amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol.* 2001;119:1345–1353.

19. Allen H. A new picture series for preschool vision testing. *Am J Ophthalmol.* 1957;44:38–41.

20. Hyvärinen L, Näsänen R, Laurinen P. New visual acuity test for pre-school children. *Acta Ophthalmol.* 1980;58:507–511.

21. American Association for Pediatric Ophthalmology and Strabismus. Eye care for the children of America. *J Pediatr Ophthalmol Strabismus.* 1991;28:64–67.

22. Scheiman MM, Amos CS, Ciner EB, Marsh-Tootle W, Moore BD, Rouse MW. *Pediatric Eye and Vision Examination: Optometric Clinical Practice Guideline.* St. Louis: American Optometric Association; 1994:1–45.

23. Committee on Practice and Ambulatory Medicine, Section on Ophthalmology. Eye examination and vision screening in infants, children, and young adults. *Pediatrics.* 1996;98:153–157.

24. Miller JM, Dobson V, Harvey EM, Sherrill DL. Comparison of preschool vision screening methods in a population with a high prevalence of astigmatism. *Invest Ophthalmol Vis Sci.* 2001;42:917–924.

25. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially paired datasets. *Med Decis Making.* 1998;18:110–121.

26. Miller JM, Harvey EM. Spectacle prescribing recommendations of AAPOS members. *J Pediatr Ophthalmol Strabismus.* 1998;35:51–52.

27. Moore B, Weissberg EE, Mayer DL. Testability of the Massachusetts visual acuity test: a screening version of the LEA visual acuity test (Abstract). *Optom Vis Sci.* 1997;74(suppl):174.

28. The Pediatric Eye Disease Investigator Group. The clinical profile of moderate amblyopia in children younger than 7 years. *Arch Ophthalmol.* 2002;120:281–287.

29. Gekeler F, Schaeffel F, Howland HC, Wattam-Bell J. Measurement of astigmatism by automated infrared photoretinoscopy. *Optom Vis Sci.* 1997;74:472–482.