Interobserver Reliability of the Prism and Alternate Cover Test in Children With Esotropia

*Pediatric Eye Disease Investigator Group**

Objective: To determine 95% limits of agreement on a measurement and on a difference between 2 measurements for the prism and alternate cover test (PACT) at distance and at near fixation.

Methods: In a test-retest reliability study, 143 children aged 60 months or younger with esotropia were examined by 2 masked examiners on 1 or 2 occasions yielding 199 test-retest pairs for PACT at distance fixation and 239 test-retest pairs for PACT at near fixation.

Results: For angles greater than 20 prism diopters (PD), the 95% limits of agreement on a measurement and on a difference between 2 measurements were ±7.3 PD and ±10.4 PD, respectively, for PACT at distance and ±8.3 PD and ±11.7 PD, respectively, for PACT at near. For angles of 10 to 20 PD, the 95% limits of agreement on a measurement and on a difference between 2 measurements were ±4.1 PD and ±5.8 PD, respectively, for PACT at distance and ±3.3 PD and ±4.7 PD, respectively, for PACT at near.

Conclusion: In childhood esotropia, differences of 12 PD or more for angles greater than 20 PD and differences of 6 PD or more for angles between 10 PD and 20 PD are likely to indicate real change. Smaller differences could be real change but could also be due to measurement error.

Arch Ophthalmol. 2009;127(1):59-65

STRABISMUS SPECIALISTS MEASURE ocular misalignment to determine whether to perform surgery and how much surgery to perform. Nevertheless, little consideration has been given to the reliability of those measurements or to the magnitude of measurement error. It is unclear whether an isolated measurement is representative of the patient's true misalignment and whether a difference in measured angles over time is a real change or is within measurement error.

Few studies have addressed the issue of test-retest reliability in strabismus measurements. Holmes et al[1] studied adult patients with sixth nerve palsy and reported that the 95% limits of agreement for a difference between 2 measurements using the prism and alternate cover test (PACT) were 10.2 prism diopters (PD) at distance and 9.2 PD at near. Other test-retest studies have been limited to measuring phorias in normal adults.[2,3] There is a paucity of data on the reliability of strabismus measurements in young children.

We report interobserver test-retest reliability of the PACT and modified Krimsky test at near in children with esotropia aged 60 months or younger.

METHODS

The Pediatric Eye Disease Investigator Group (PEDIG) conducted this study at 23 community- and university-based clinical sites, and it was supported through a cooperative agreement with the National Eye Institute of the National Institutes of Health. This research adhered to the tenets of the Declaration of Helsinki and the respective institutional review boards approved the protocol and Health Insurance Portability and Accountability Act–compliant informed consent forms. The parent or guardian of each study participant gave written informed consent. The complete protocol is available at http://public.pedig.jaeb.org, and the major aspects are summarized below.

The present study was conducted as an ancillary study nested within an observational study evaluating the course of esotropia prior to surgical intervention in subjects with recent-onset esotropia. The test-retest study was conducted at follow-up visits 6 and 12 weeks after entry into the observational study. Subjects could participate in the test-retest study if 2 independent certified examiners were available to measure alignment at 1 or both visits.

The major eligibility criteria for the observational study included age between 2 and 60 months and onset of esotropia within the previous 6 months. Subjects who had esotropia onset before 6 months of age were classified as having infantile esotropia (IET) and, for eligi-

***Authors/Writing Committee:** The following members of the Pediatric Eye Disease Investigator Group take authorship responsibility for the study results: Jonathan M. Holmes, BM, BCh; Danielle L. Chandler, MSPH; Stephen P. Christiansen, MD; Eileen E. Birch, PhD; Erick Bothun, MD; Daniel Laby, MD; B. Michele Melia, ScM; Michael X. Repka, MD; David I. Silbert, MD; V. Lori Zeto.
Group Information: A complete list of the members of the Pediatric Eye Disease Investigator Group (PEDIG) participating in the study appears on page 64.

bility, were required to have a constant esotropia measuring more than 10 PD by PACT in the primary position at distance fixation or more than 10 PD by modified Krimsky test at near fixation if the PACT was not possible. Subjects who had esotropia onset at 6 months or older were classified as having acquired esotropia and were required to have a constant esotropia measuring more than 10 PD by simultaneous prism and cover test in the primary position at distance fixation while wearing appropriate refractive correction (if required). Full refractive correction was required, based on a cycloplegic refraction, if hyperopia was +3.00 diopters (D) or more in children with IET or +2.00 D or more in children with acquired esotropia, astigmatism was 3.50 D or more in IET or 2.50 D or more in acquired esotropia, or anisometropia was 2.00 D or greater. Lesser amounts of refractive error were corrected at investigator discretion. Of the subjects with acquired esotropia, those whose angle decreased less than 10 PD by PACT while wearing appropriate refractive correction and those who did not require spectacle correction were classified as having acquired nonaccommodative esotropia (ANAET). Subjects whose angle decreased 10 PD or more by PACT while wearing appropriate refractive correction were classified as having acquired partially-accommodative esotropia (APAET). Children were excluded if they had developmental delay in the judgment of the investigator, known central nervous system abnormalities, neuromuscular disease, or paralytic strabismus.

PROCEDURE

Alignment measurements were made by PACT at distance and near fixation by 2 independent certified examiners. Certification consisted of reading a manual of procedures and completing a brief multiple-choice test regarding the details of the manual. For subjects who were unable to perform PACT testing (ie, some of the younger children), the misalignment was measured using the modified Krimsky method at near fixation. The examiners performed their measurements at least 15 minutes apart but within 1 hour of each other, and each examiner was masked to the other's measurements at the given visit and to all measurements from the previous visit (ie, neither examiner reviewed the patient's alignment history prior to obtaining measurements). In addition to the standard loose plastic prism set (1-PD increments from 1 to 10 PD, 2-PD increments from 10 to 20 PD, and 5-PD increments from 20 to 50 PD), additional prisms of magnitudes of 22.5 PD, 27.5 PD, 32.5 PD, 37.5 PD, 42.5 PD, and 47.5 PD were manufactured for the study (Gulden Ophthalmics, Elkins Park, Pennsylvania), allowing the measurements to be made to the nearest 2.5 PD for angles of esotropia measuring more than 20 PD.

For the PACT, the examiners placed a plastic prism in the frontal plane position before one eye and alternately occluded the eyes with a cover, observing the refixation movement of the just-uncovered eye on an age-appropriate accommodative target. The prism power was gradually increased until the direction of the refixation movement of the just-uncovered eye reversed. The prism power was then reduced until no further refixation movement of the fellow eye was seen (ie, neutralization). The magnitude of prism that either neutralized the deviation or was closest to neutralization was recorded. Examiners were specifically instructed not to split prisms for angles less than 50 PD and not to stack prisms. For deviations greater than 50 PD, the examiners split the prisms between eyes such that the prism power was approximately equal for each eye. If no deviation was present, a situation that could potentially occur given that this ancillary study was conducted at follow-up visits, a value of 0 PD was recorded.

For the modified Krimsky measurement at near, performed if a younger child was unable to cooperate with the PACT, the examiners used a light at one-third meter and placed increasing power prisms before the fixating eye until the corneal light reflex of the nonfixating eye was symmetrical to that of the corneal light reflex in the fixating eye.

Each examiner was asked to classify the deviation at each visit as either constant, variable (constant, but with an angle of deviation that appeared to change at the same fixation distance during the examination), or intermittent (the manifest deviation was not present during some portion of the examination). Pairs of measurements were excluded from analysis if 1 or both examiners classified the subjects as having a variable or intermittent deviation at the time of the examination (19 pairs for PACT at distance, 26 for PACT at near, and 4 for modified Krimsky at near).

STATISTICAL ANALYSIS

Subjects were excluded from our analysis when both measurements were less than 10 PD because the prism steps between 1 and 10 PD were in 1-PD increments and there were too few data to permit separate meaningful analysis (n=11 for PACT at distance; n=2 for PACT at near).

Different prism increments were used to measure smaller (10-20 PD) and larger (>20 PD) angles. Therefore, we analyzed pairs of measurements in 2 strata, one where the prism increment used for measurement was 2 PD (angles between 10 and 20 PD) and a second where the prism increment was 2.5 PD (angles >20 PD). If one measurement was between 10 and 20 PD and the other measurement was more than 20 PD, the test-retest pair was analyzed in the more than 20-PD stratum.

The difference between test and retest (second examiner's results minus the first examiner's results) was calculated for each pair of measurements. The 95% limits of agreement and 95% confidence intervals (CIs) were then calculated separately for (1) a measurement and (2) a difference between 2 measurements, based on the standard error of measurement.[4,5] The standard error of measurement was derived from a repeated-measures analysis-of-variance model using the measurements as the dependent variable. This model includes an adjustment for the correlation between measurements for 6- and 12-week visits from the same individual, but not for possible correlation in the measurement errors of the same individual. In other words, an assumption of the model is that measurement error is a property of the test and not of the individuals being tested.[6] To confirm this assumption, we also performed an analysis limited to the first test-retest pair for each subject and verified that the results were similar. All analyses were conducted separately according to the testing method (PACT at distance, PACT at near, and modified Krimsky at near).

The influence of factors on the magnitude of absolute test-retest differences (a measure of the magnitude of variability) was examined in linear regression models, using generalized estimating equations[7] to account for correlation between the 6- and 12-week measurements for the same subject. The effect of measurement precision on test-retest differences was assessed using a 2-level categorical variable indicating either 10- to 20-PD angles measured in 2-PD increments or more than 20-PD angles measured in 2.5-PD increments. Within each level of measurement precision, the effect of the magnitude of angle size on test-retest differences was assessed using continuous angle size based on the average of test and retest measurements. For angles of more than 20 PD (where we had sufficient numbers of subjects to evaluate other factors), we also examined the effects of sex, race, age (as a continuous variable), esotropia type (IET, ANAET, or APAET), whether spectacles were worn, and whether amblyopia was present. Presence of amblyopia was defined by optotype visual acuity when possible (2 or more logarithms of the minimum angle

of resolution lines' difference in visual acuity) or by fixation preference (inability to maintain fixation with the nonpreferred eye for ≥3 seconds, through a blink, or through a smooth pursuit).

All reported *P* values are 2-tailed. Analyses were conducted using SAS version 9.1 (SAS Institute, Cary, North Carolina).

RESULTS

SUBJECTS AND EXAMINATIONS

Between August 2004 and August 2007, 143 subjects completed PACT measurements as part of this test-retest study; 37% had IET and 63% had acquired esotropia. Subjects ranged in age from 2.1 to 60.2 months at the time of the first test-retest measurement, with a mean (SD) age of 22.2 (15.0) months. A total of 66 subjects (46%) were female and 122 (85%) were white. Refractive correction was worn by 85 subjects (59%) and 57 (40%) had amblyopia. Test-retest data were collected for 115 subjects (80%) at both the 6- and 12-week examinations, for 13 subjects (9%) at the 6-week visit only, and for 15 subjects (10%) at the 12-week visit only, yielding 199 test-retest pairs for PACT at distance and 239 test-retest pairs for PACT at near. The distribution of average angle size for the test-retest pairs is shown in **Table 1**.

Table 1. Average Angle of Deviation

Average Angle of Deviation, PD[a]	No. (%) PACT at Distance (n=199)	No. (%) PACT at Near (n=239)
Mean (SD) angle	32.5 (14.3)	37.2 (14.8)
Range	10 to 82.5	10 to 85
10-20	47 (24)	35 (15)
21-30	63 (32)	56 (23)
30-39	40 (20)	60 (25)
40-49	26 (13)	45 (19)
50-59	15 (8)	25 (10)
60-69	6 (3)	13 (5)
70-69	1 (1)	3 (1)
≥80	1 (1)	2 (1)

Abbreviations: PACT, prism and alternate cover test; PD, prism diopters.
[a] Refers to the average of the initial and retest measurements for the given testing method.

PRISM AND ALTERNATE COVER TEST

Compared with the initial test, the retest measurement (performed within an hour of the initial test) ranged from a decrease of 15 PD to an increase of 15 PD (mean, 0.3 PD) for PACT at distance and from a decrease of 17.5 PD to an increase of 17.5 PD (mean, −0.1 PD) for PACT at near. The mean difference between the initial and retest measurements was not statistically different from zero (mean, 0.32 PD; 95% CI, −0.35 to 0.99 PD for PACT at distance; mean, −0.09 PD; 95% CI, −0.80 to 0.62 PD for PACT at near). These findings suggest there was no appreciable fatigue effect or dissociation effect from test to retest.

For PACT at distance, 46 test-retest pairs (23%) were for angles of 10 to 20 PD and 153 (77%) were for angles larger than 20 PD. The mean (SD) absolute value of the test-retest differences was 2.0 (2.1) PD (median, 2 PD; interquartile range, 0-4 PD) for angles of 10 to 20 PD and 3.4 (4.1) PD (median, 2.5 PD; interquartile range, 0-5.0 PD) for angles larger than 20 PD. The initial test and the retest yielded the same value in 17 cases (37%) for angles of 10 to 20 PD and in 75 cases (49%) for angles larger than 20 PD (**Table 2**). The number of cases with an absolute value test-retest difference within 5 PD and 10 PD was 42 (91%) and 46 (100%), respectively, for angles of 10 to 20 PD and 126 (82%) and 143 (93%), respectively, for angles larger than 20 PD.

For PACT at near, 34 test-retest pairs (14%) were for angles of 10 to 20 PD and 205 (86%) were for angles larger than 20 PD. The mean (SD) absolute value of the test-retest differences was 1.5 (1.9) PD (median, 0 PD; interquartile range, 0-2 PD) for angles of 10 to 20 PD and 4.2 (4.3) PD (median, 5 PD; interquartile range, 0-5 PD) for angles larger than 20 PD. The initial test and the retest yielded the same value in 18 cases (53%) for angles of 10 to 20 PD and in 74 cases (36%) for angles larger than 20 PD (Table 2). The number of cases with an absolute value test-retest difference within 5 PD and 10 PD was 33 (97%) and 34 (100%), respectively, for angles of 10 to 20 PD and 159 (78%) and 191 (93%), respectively, for angles larger than 20 PD.

Limits of Agreement

For angles larger than 20 PD, the 95% limits of agreement on a measurement were ±7.3 PD for PACT at distance and ±8.3 PD for PACT at near (**Table 3**). For a difference between 2 measurements, the 95% limits of agreement were ±10.4 PD for PACT at distance, and ±11.7 PD for PACT at near (Table 3).

For angles of 10 to 20 PD, the 95% limits of agreement on a measurement were ±4.1 PD for PACT at distance, and ±3.3 PD for PACT at near (Table 3). For a difference between 2 measurements, the 95% limits of agreement were ±5.8 PD for PACT at distance, and ±4.7 PD for PACT at near (Table 3).

Factors Influencing the Magnitude of Test-Retest Variability

Bland Altman plots[4] (**Figure 1** A and B) show the relationship between the angle size and magnitude of test-retest difference by plotting the test-retest difference against the average of the initial and retest measurements.

Test-retest differences were greater in angles larger than 20 PD (angles measured in 2.5-PD increments) than in angles of 10 to 20 PD (angles measured in 2-PD increments) (mean, 3.4 vs 2.0 PD; $P = .001$ for PACT at distance; mean, 4.2 vs 1.5 PD; $P < .001$ for PACT at near). Nevertheless, within each of these stratum, the magnitude of test-retest differences was not associated with the size of the measured deviation for either PACT at distance ($P = .41$ for angles of 10 to 20 PD; $P = .23$ for angles >20 PD) or PACT at near ($P = .20$ for angles of 10 to 20 PD; $P = .42$ for angles >20 PD).

Of the angles larger than 20 PD (angles measured in 2.5-PD increments), the magnitude of the test-retest dif-

Table 2. Absolute Value of Test-Retest Differences According to Prism Increment Used for Measurement[a]

	PACT at Distance				PACT at Near			
	Angles 10 to 20 PD (n=46)		Angles >20 PD (n=153)		Angles 10 to 20 PD (n=34)		Angles >20 PD (n=205)	
Absolute Value of Test-Retest Difference, PD	No. (%)	Cumulative No. (%)	No. (%)	Cumulative No. (%)	No. (%)	Cumulative No. (%)	No. (%)	Cumulative No. (%)
0	17 (37)	17 (37)	75 (49)	75 (49)	18 (53)	18 (53)	74 (36)	74 (36)
1-5	25 (54)	42 (91)	51 (33)	126 (82)	15 (44)	33 (97)	85 (41)	159 (78)
6-10	4 (9)	46 (100)	17 (11)	143 (93)	1 (3)	34 (100)	32 (16)	191 (93)
11-15	0 (0)	46 (100)	10 (7)	153 (100)	0 (0)	34 (100)	12 (6)	203 (99)
16-20	0 (0)	46 (100)	0 (0)	153 (100)	0 (0)	34 (100)	2 (1)	205 (100)

Abbreviations: PACT, prism and alternate cover test; PD, prism diopters.
[a]Within each testing method, strata are based on the prism increment used for measurement (2-PD increments for angles between 10 and 20 PD and 2.5-PD increments for angles >20 PD). If one measurement was between 10 and 20 PD and one measurement was >20 PD, the test-retest pair was analyzed in the 2.5-PD increment stratum.

Table 3. Limits of Agreement on a Measurement and on a Difference Between Measurements

	Limits of Agreement (95% CI)[a]	
Measurement Method	Measurement	Difference Between 2 Measurements
PACT at distance		
Angles 10-20 PD	±4.1 (±2.8 to ±5.3)	±5.8 (±4.0 to ±7.5)
Angles >20 PD	±7.3 (±6.0 to ±8.6)	±10.4 (±8.5 to ±12.2)
PACT at near		
Angles 10-20 PD	±3.3 (±2.1 to ±4.6)	±4.7 (±2.9 to ±6.5)
Angles >20 PD	±8.3 (±7.0 to ±9.6)	±11.7 (±9.9 to ±13.6)
Modified Krimsky at near[b]		
Angles >20 PD	±9.4 (±4.4 to ±14.5)	±13.3 (±6.2 to ±20.8)

Abbreviations: CI, confidence interval; PACT, prism and alternate cover test; PD, prism diopters.
[a]95% CIs are the 95% CIs on the limits of agreement.
[b]Note that no angles measured by modified Krimsky at near were 10 to 20 PD.

ference was not associated with patient age for either PACT at distance or for PACT at near (eTable; http://www.archophthalmol.com) ($P=.78$ and $P=.91$, respectively) (**Figure 2**). Esotropia type (IET, ANAET, APAET), sex, race, and presence of amblyopia also did not influence the magnitude of test-retest difference. There was an association with spectacle wear, as subjects wearing spectacles had larger test-retest differences for PACT at near than those not wearing spectacles (mean, 4.9 PD vs 3.2 PD; $P=.004$), though not for PACT at distance (mean, 3.8 PD vs 2.9 PD; $P=.19$).

MODIFIED KRIMSKY TEST AT NEAR

Twenty-three pairs of test-retest data for modified Krimsky at near were collected from 13 subjects with IET who ranged in age from 3.8 to 9.7 months (mean, 6.6 months). These subjects had the modified Krimsky test at near performed throughout the study because, at enrollment, PACT testing could not be performed. All 23 test-retest pairs were for angles larger than 20 PD. Compared with the initial test, the retest measurements ranged from a decrease of 10 PD to an increase of 15 PD (mean, 1.5 PD) and the test-retest difference was 0 PD in 7 cases (30%), 1 to 5 PD in 9 cases (39%), 6 to 10 PD in 4 cases (17%), and 11 to 15 PD in 3 cases (13%). The 95% limits of agreement were ±9.4 PD for a measurement and ±13.3 PD for a difference between 2 measurements (Table 3).

COMMENT

In this interobserver test-retest variability study of PACT measurements in childhood esotropia, we found 95% limits of agreement for a single measurement to range from 3.3 PD to 8.3 PD, depending on angle magnitude and fixation distance (Table 3). The 95% limits of agreement for a difference between 2 measurements ranged from 4.7 PD to 11.7 PD, again depending on angle magnitude and fixation distance.

There are few published data with which to compare our results. In adult patients with sixth nerve palsy, Holmes et al[1] reported 95% limits of agreement on a difference between 2 measurements of 10.2 PD for PACT at distance and 9.2 PD for PACT at near. Johns et al[2] reported 95% limits of agreement on a difference between 2 measurements of 4 PD for PACT at near in small or moderate phorias in adults. Our study differs from these studies in patient age and type of deviation.

Our findings have important practical implications. Any measurement of ocular alignment is the sum of the true deviation and the measurement error. The 95% limits of agreement for a measurement pertain to uncertainty around a single measurement. Similarly, for any observed difference in measurement over time, we cannot know what component of the difference is measurement error and what component is real change. We can only have some degree of certainty that there is a component of real change if the observed difference exceeds the limits of agreement for a difference between 2 measurements. The limits of agreement for a difference between 2 measurements are greater than the limits of agreement for a single measurement, because a difference incorporates measurement error for both measurements. Taking our results for PACT testing at distance and at near together, 6 PD might be a practical threshold for real change in angles between 10 and 20 PD and

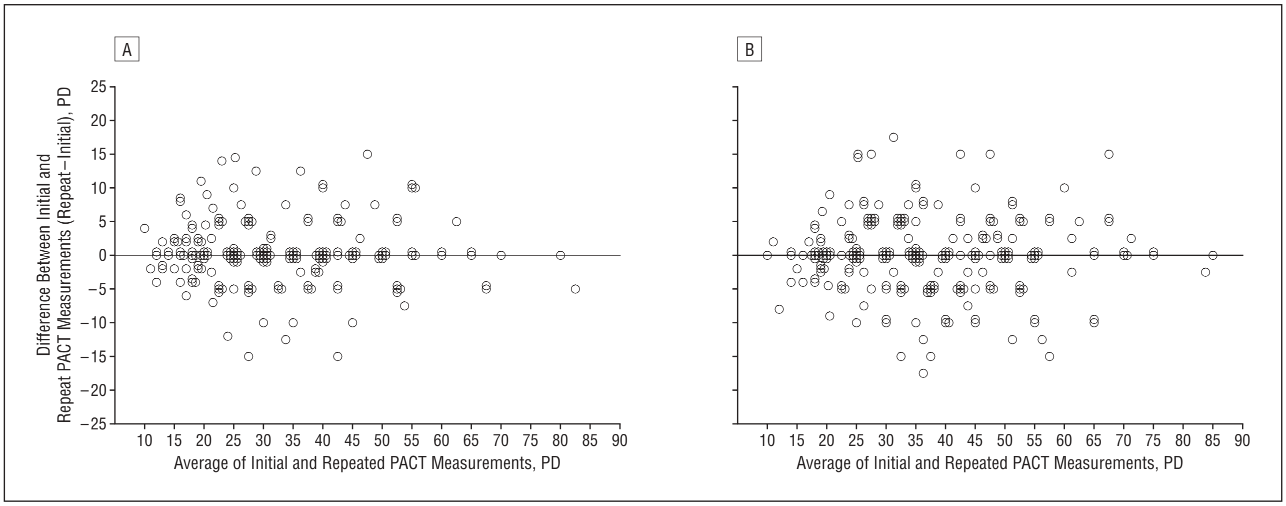

Figure 1. Bland Altman plot of test-retest difference vs average of test and retest measurements for prism and alternate cover test (PACT) at distance (A; n=199) and near fixation (B; n=239). The solid horizontal line indicates no difference between the test and retest; PD, prism diopters.

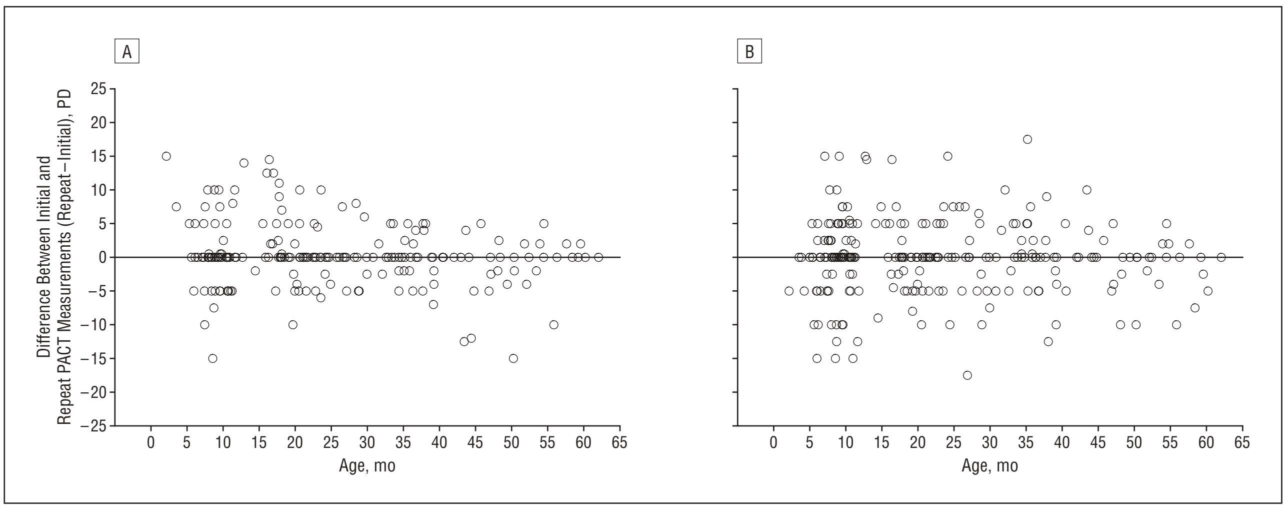

Figure 2. Plot of test-retest difference vs age in months for prism and alternate cover test (PACT) measurements at distance (A; n=199) and near fixation (B; n=239). The solid horizontal line indicates no difference between test and retest; PD, prism diopters.

12 PD might be a practical threshold for real change in angles greater than 20 PD.

Regarding deviation magnitude and measurement error, we did find that smaller deviations (10 to 20 PD), measured using smaller increments (2 PD), had less measurement error than larger deviations (>20 PD), measured in 2.5-PD increments. Nevertheless, within a stratum (10 to 20 PD and >20 PD), we could not confirm a relationship between angle size and measurement error. It was not possible to completely separate the potential effects of measurement precision and angle size, however, because smaller prism steps were used for smaller angles and larger steps for larger angles.

For angles greater than 20 PD, we had sufficient data to analyze the effect of a number of factors on variability (sex, race, age, esotropia type [IET, ANAET, and APAET], whether spectacles were worn, and whether amblyopia was present). The lack of an age effect was surprising. In children who could cooperate for PACT testing, we would have predicted that measurement error would be greater in the younger children owing to issues of inattention, fatigue, and increased difficulty of examination. The comparison of measurement error between types of esotropia is confounded by age, but because children with IET are younger than children with acquired esotropia (ANAET and APAET) by definition, we had too little overlap in age to perform a meaningful age-adjusted analysis of esotropia type.

The wearing of spectacles might have been expected to affect measurement error either adversely or favorably. If the neutralization endpoint was more difficult to define through spectacles, then we might have expected increased measurement error. In contrast, if the use of hyperopic spectacles reduced the accommodative demands or reduced the variability of the accommodative state, then we might have expected decreased variability. In fact, we found spectacle wear (compared with no spectacle wear) was associated with larger test-retest differences for PACT at near, though not for PACT at distance. We cannot explain this discrepancy; the finding at near may be due to chance, given that we evaluated several factors.

The Pediatric Eye Disease Investigator Group (PEDIG)

Clinical Sites

Clinical Sites that participated in this protocol are listed in order by the number of patients enrolled into the study. An asterisk indicates the center received support for this project from an unrestricted grant from Research to Prevent Blindness Inc, New York, New York.

Ophthalmic Associates, Anchorage, Alaska (50): Mary Diane Armitage, Robert W. Arnold; *Pediatric Ophthalmology of Erie, Erie, Pennsylvania (23):* Rhonda M. Hodde, Nicholas A. Sala, V. Lori Zeto; *Vanderbilt Eye Center, Nashville, Tennessee (18)*:* Ronald J. Biernacki, Sean Donahue, Joey Martin, David G. Morrison, Neva J. Palmer; *Family Eye Group, Lancaster, Pennsylvania (13):* Noelle S. Matta, David I. Silbert, Eric L. Singman; *University of Minnesota, Minneapolis (13)*:* Erick D. Bothun, Stephen P. Christiansen, Sara J. Downes, Kathy M. Hogue, Kim S. Merrill, C. Gail Summers; *Wolfe Clinic, West Des Moines, Iowa (13):* Ashley D. Andreassen, Lisa M. Fergus, Donny W. Suh, Rhonda J. Swisher; *Texas Children's Hospital, Houston (7):* Jane Covington Edmond, Evelyn A. Paysse, Gina Rosby-Obeng, Kimberly G. Yen; *Wake Forest University, Winston-Salem, North Carolina (6):* Angela Z. Moya, Richard Grey Weaver Jr; *North Carolina Eye, Ear & Throat, Durham (5):* Lynelle Perez, Kim T. Pojman, Joan Therese Roberts, Marguerite J. Sullivan; *Casey Eye Institute, Portland, Oregon (4)*:* Kimberley A. Beaudet, Daniel J. Karr, Ann U. Stout, David T. Wheeler; *Wilmer Institute, Baltimore, Maryland (3)*:* Carole R. Goodman, Xiaonong Liu, Michael X. Repka; *Mayo Clinic, Rochester, Minnesota (3)*:* Jonathan M. Holmes, Brian G. Mohney; *Cardinal Glennon Children's Hospital, St Louis, Missouri (3):* Oscar A. Cruz, Bradley V. Davitt, Emily A. Miyazaki; *Department of Ophthalmology, Children's Hospital Boston, Boston, Massachusetts (2):* Linda R. Dagi, Sarah E. MacKinnon, Mariette Tyedmers; *The Emory Eye Center, Atlanta, Georgia (1):* Scott R. Lambert, Rachel A. Robb; *Katherine Ann Lee, MD, PA, Boise, Idaho (1):* Katherine A. Lee, Bonita R. Schweinler; *University of Michigan, Brighton (1):* Jorie L. Jackson, Erika M. Levin; *Pediatric Ophthalmology, PC, Grand Rapids, Michigan (1):* Patrick J. Droste, Robert J. Peters; *Midwest Eye Institute, Indianapolis, Indiana (1):* Jay G. Galli, Gavin J. Roberts, Naval Sondhi, Derek T. Sprunger; *Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (1):* Brian J. Forbes, Alexandra Huebner, Graham E. Quinn, Jaclyn Sharman; *Family Eye Care of the Carolinas, Pinehurst, North Carolina (1):* Michael John Bartiss, Keith P. Poindexter; *University of California, Department of Ophthalmology & Vision Science, Sacramento (1):* Maedi M. Bartolacci, Mary O'Hara; *Eye Care Group, PC, Waterbury, Connecticuit (1):* Andrew J. Levada, Cheryl Schleif, Tabitha L. Walker.

Esotropia Treatment Study Steering Committee:

Eileen E. Birch, Erik D. Bothun, Danielle L. Chandler, Stephen P. Christiansen, Sean Donahue, Donald F. Everett, Richard W. Hertle, Jonathan M. Holmes, Noelle S. Matta, Brian G. Mohney, Sandy Owens, Graham E. Quinn, Michael X. Repka, David K. Wallace, David R. Weakley, Richard Grey Weaver Jr.

PEDIG Coordinating Center:

Roy W. Beck, Brian D. Becker, Gladys N. Bernett, Nicole M. Boyle, Christina M. Cagnina-Morales, Debora A. Cagnina, Esmeralda L. Cardosa, Danielle L. Chandler, Laura E. Clark, Sharon R. Constantine, Katrina L. Dawson, Quayleen Donahue, Michelle D. Drew, Mitchell Dupre, Kyle L. Baccus Horsley, Heidi A. Gillespie, Raymond T. Kraker, Stephanie V. Lee, Lee Anne Lester, Shelly T. Mares, Holly McCombs, B. Michele Melia, Pamela S. Moke, Stephanie Morgan-Bagley, Jim L. Pyner, Diana E. Rojas, Nicole T. Reese, Heidi J. Strayer, Dana Williams.

National Eye Institute - Bethesda, Maryland:

Donald F. Everett.

PEDIG Executive Committee:

Roy W. Beck, Eileen E. Birch, Stephen P. Christiansen, Susan A. Cotter, Sean P. Donahue, Donald F. Everett, Jonathan M. Holmes, Darren L. Hoover, Pamela A. Huston, Raymond T. Kraker, Michael X. Repka, Nicholas A. Sala, Mitchell M. Scheiman, David K. Wallace.

PEDIG Data and Safety Monitoring Committee:

William Barlow, Edward G. Buckley, Barry Davis, Marie D. Diener-West, Velma Dobson, John L. Keltner, Hana Osman, Earl A. Palmer, Dale L. Phelps, Stephen W. Poff, Richard A. Saunders, Lawrence Tychsen.

We also found that amblyopia was not associated with increased measurement error despite the expectation that poorer visual acuity might increase variability of measurements. In the younger children who were unable to perform optotype visual acuity testing, the diagnosis of amblyopia by fixation preference may be prone to classification error.[8,9]

Part of our finding of measurement error might have been due to a truly variable nature of the deviation in a proportion of our subjects. Although we specifically excluded subjects who had an obviously variable deviation, it is possible that intrinsic variability of the condition might still contribute to test-retest differences. Finally, some of the test-retest differences might have been due to variation in prism measurement technique, despite a specific protocol for measurement of the deviation, including proper positioning of plastic prisms.[10]

For the modified Krimsky test at near, our point estimate for 95% limits of agreement for a difference between 2 measurements (±13 PD) was only slightly greater than that for the PACT test at near for angles greater than 20 PD (12 PD). Nevertheless, our finding for the modified Krimsky method at near should be interpreted with caution because these data were collected for only 13 subjects; therefore the estimate has a very wide CI (±5.8 to ±20.8 PD).

Data from the current PACT test-retest study can now be used to set criteria for stability and instability to be used in the overall prospective observational study of angle instability in IET and acquired esotropia. We plan to use Monte Carlo simulation methods to create separate hypothetical data sets for subjects with stable deviations and unstable deviations, where the differences are the sum of actual change and measurement error. From these simulations we will be able to explore criteria for defining stable and unstable misalignment.

Our test-retest study has a number of potential weaknesses. We used 2 independent masked examiners; therefore our observed measurement error has components of intraobserver error and interobserver variability. If it were possible to obtain an unbiased estimate of intraobserver measurement error, it would likely be less than the interobserver error we are currently reporting. In addition, inclusion of patients in the analysis of PACT at near who could not complete PACT at distance might have biased the results toward a higher estimate of measurement error in the analysis for near. Nevertheless, results from a separate analysis (not shown) limited to subjects who could complete both PACT at distance and near showed that the limits of agreement did not differ from the primary analysis. Patient fatigue, disinterest at the retest, and disruption of fusion might all be expected to increase test-retest differences, but we found no evidence of larger angles of misalignment on the second test. Another limitation is that because the measurement scale is stepped rather than continuous, we would expect that in some cases our limits of agreement would encompass slightly more or less than 95% of the measurement error distribution.

In childhood esotropia, when the angle is greater than 20 PD, a difference of 12 PD or more from one PACT measurement to the next is likely to represent real change. When the angle is between 10 to 20 PD, a difference of 6 PD or more is likely to represent real change. Caution should be taken when making clinical judgments based on differences in PACT of less than these magnitudes because, although such small differences do not exclude the possibility of real change, they could be due to measurement error.

Submitted for Publication: March 25, 2008; final revision received August 13, 2008; accepted August 14, 2008.
Correspondence: Jonathan M. Holmes, BM, BCh, c/o Jaeb Center for Health Research, 15310 Amberly Dr, Ste 350, Tampa, FL 33647 (pedig@jaeb.org).
Financial Disclosure: None reported.
Funding/Support: This study was supported by grants EY011751 and EY015799 from the National Eye Institute of National Institutes of Health, Department of Health and Human Services.
Additional Information: The eTable is available at http://www.archophthalmol.com.

REFERENCES

1. Holmes JM, Leske DA, Hohberger GG. Defining real change in prism-cover test measurements. *Am J Ophthalmol*. 2008;145(2):381-385.
2. Johns HA, Manny RE, Fern K, Hu YS. The intraexaminer and interexaminer repeatability of the alternate cover test using different prism neutralization endpoints. *Optom Vis Sci*. 2004;81(12):939-946.
3. Rainey BB, Schroeder TL, Goss DA, Grosvenor TP. Inter-examiner repeatability of heterophoria tests. *Optom Vis Sci*. 1998;75(10):719-726.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.
5. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci*. 1993;34(1):120-129.
6. Dunn G. *Design and Analysis of Reliability Studies: the Statistical Evaluation of Measurement Errors.* New York, NY: Halsted Press; 1989:114-129.
7. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986;42(1):121-130.
8. Cotter SA, Varma R, Tarczy-Hornoch K, et al. Relationship of fixation preference to amblyopia in preschool children: the Multi-Ethnic Pediatric Eye Disease Study [ARVO abstract 2380]. *Invest Ophthalmol Vis Sci*. 2007;48(5):E-abstract 2380.
9. Sener EC, Mocan MC, Gedik S, Ergin A, Sanac AS. The reliability of grading the fixation preference test for the assessment of interocular visual acuity differences in patients with strabismus. *J AAPOS*. 2002;6(3):191-194.
10. Thompson JT, Guyton DL. Ophthalmic prisms: measurement errors and how to minimize them. *Ophthalmology*. 1983;90(3):204-210.

From the Archives of the ARCHIVES

In order to make the iris less sensitive during operation, Rudin injects a small quantity of cocaine in bichloride solution beneath conjunctiva at the upper margin of the cornea. The vesicle there produced is pushed round the cornea with a pledget of cotton. In eleven operations the result was satisfactory in eight, and partially so in two. In one prolonged operation the patient complained of pain. (Koller's method –Ed.)

Reference: *Arch Ophthalmol*. 1897;26:466.