Published in final edited form as:

J AAPOS. 2012 April ; 16(2): 120–124. doi:10.1016/j.jaapos.2011.11.008.

Variability of angle of deviation measurements in children with intermittent exotropia

Sarah R. Hatt, DBO, David A. Leske, MS, Laura Liebermann, CO, Brian G. Mohney, MD, and Jonathan M. Holmes, BM, BCh

Department of Ophthalmology, Mayo Clinic, Rochester, MN

Abstract

Background—A change in the angle of deviation is often used to monitor change in severity of intermittent exotropia over time; nevertheless, thresholds for a clinically significant change in angle have not been determined. We analyzed variability due to test–retest differences and short-term variability in the condition to provide thresholds as a guide for assessing clinically significant, long-term change in angle of intermittent exotropia.

Methods—Twenty-six children with intermittent exotropia (median age, 7; range, 1–13 years) underwent repeat prism and alternate cover test measures during 3 or 4 examinations (2 hours apart) over the course of a day; 95% repeatability coefficients were derived to determine test–retest differences at distance and near fixation.

Results—Derived 95% repeatability coefficients at distance were 3.4^{Δ} (95% CI, 0.7^{Δ} – 6.2^{Δ}) for angles 20^{Δ} and 7.2^{Δ} (95% CI, 4.4^{Δ} – 9.9^{Δ}) for angles $>20^{\Delta}$; at near, 6.6^{Δ} (95% CI, 3.7^{Δ} – 9.6^{Δ}) for angles 20^{Δ} and 12.8^{Δ} (95% CI, 5.3^{Δ} – 20.3^{Δ}) for angles $>20^{\Delta}$.

Conclusions—Test–retest reliability data in this study provide thresholds to help determine clinically significant change in angle of strabismus in children with intermittent exotropia. These data should facilitate evidence-based assessment of long-term change in intermittent exotropia over time.

In children with intermittent exotropia, clinicians use a variety of measures to evaluate change in severity of the condition, although change in angle of exodeviation is the most commonly reported method.^{1–3} Because angle of deviation is routinely measured at follow-up examinations, tracking its change over time is appealing; yet the magnitude of angle change that is clinically significant has not been established.

A difference in angle of deviation measurements over time is often assumed to represent change in the severity of the condition, but change may in fact be attributable to other factors such as intrinsic variability of the condition itself, measurement error, change in tenacious fusion, changes in testing conditions, or a combination of these factors. Previous studies^{1,2} of longitudinal changes in angle in intermittent exotropia have used arbitrary thresholds, such as 10^{Δ} , to signify a change in severity, but the true threshold for a clinically significant change in measured angle of deviation is unknown. The purpose of this

© 2012 American Association for Pediatric Ophthalmology and Strabismus. Published by Mosby, Inc. All rights reserved.

Correspondence: Dr. Jonathan M. Holmes, Ophthalmology W7, Mayo Clinic, Rochester, MN, 55905 (holmes.jonathan@mayo.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

None of the authors have any proprietary or financial interests to disclose.

study was to derive test–retest thresholds and thereby quantify variability attributable both to measurement error and short-term intrinsic variability of the condition. Such data can then enable more reliable assessment of clinically significant change in angle of deviation over longer periods of time in children with intermittent exotropia.

Patients and Methods

Approval of the institutional review board of the Mayo Clinic, Rochester, was obtained and all participants provided informed consent. All procedures and data collection were conducted in a manner compliant with the Health Insurance Portability and Accountability Act.

Children (aged 1–17 years) with either basic, pseudo-divergence excess or true divergence excess types of intermittent exotropia of at least 10^Δ at distance were eligible for inclusion. Due to the burden of testing, patients noted to have poor cooperation were not included. Patients with sensory exotropia, paralytic exotropia, or coexisting developmental delay were excluded. In addition, patients with convergence insufficiency type intermittent exotropia (near angle more than 10^Δ greater than distance) were not included since we were interested in studying the more frequently encountered basic and distance types and convergence insufficiency type may behave somewhat differently. Best-corrected visual acuity (where measurable) was 20/40 or better in each eye. All children underwent repeat measures of angle of deviation over a day, as part of a larger study analyzing repeat measures of control, stereoacuity, angle, and motor fusion: some control, stereoacuity, and fusion data from these patients have been reported previously (Liebermann L, Hatt SR, Leske DA, Holmes JM. Variability of fusional convergence in childhood intermittent exotropia Invest Ophthalmol Vis Sci 2011;52:ARVO E-Abstract 6379).^{4–6}

Angle measurements were obtained using the prism and alternate cover test, with patients wearing habitual refractive correction if prescribed. The patient viewed an accommodative target at distance fixation (3 meters) and then at near fixation (1/3 meter). Base-in prisms were placed over one eye and dissociation continued, with gradually increasing prism strength, until the angle of deviation was overcorrected. Prism strength was then decreased to identify the largest prism strength at which the exodeviation was neutralized. The largest prism magnitude that neutralized the exodeviation was recorded as the angle of deviation.

To limit the influence of other potential sources of variability, individual patients underwent repeat assessments under the same conditions and in the same room. The same examiner performed repeat measurements whenever possible. To reduce the influence of examiner memory, previous measurement(s) were not reviewed when repeating the assessment.

Since each subject had more than two measurements and therefore more than one test–retest difference, we used the 95% repeatability coefficient,⁷ which is equivalent to the 95% limits of agreement in a two-measure test–retest study but accounts for more than one test–retest difference per patient, that is, when three or more measurements per patient are analyzed. The 95% repeatability coefficient was calculated by multiplying the within subject standard deviation (square root of the mean squares) of the measures by 2.77 (ie, $1.96 \sqrt{2} s_w$, where s_w indicates “within subject” standard deviation).⁷

We also calculated the 95% limits of agreement on individual measurements,⁸ which provide an estimate of precision for a measurement, that is, a defined range that would contain the real measurement 95% of the time: individual measurements were calculated by multiplying the within-subject standard deviation of the individual measures by 1.96.

Distance- and near-angle data were initially analyzed separately with a plan to combine data if variability at distance and near was similar. Agreement between angle measurements was represented using Bland-Altman plots, showing the magnitude of difference between each sequential pair of measurements. Regression analyses were performed to determine whether or not there was greater variability with larger deviations. If the slope was significant, test-retest differences would be analyzed separately for large angles ($>20^\Delta$) and for small angles ($\leq 20^\Delta$). The threshold of 20^Δ was chosen a priori because the prism increment changes from 2^Δ to 5^Δ at 20^Δ , and this threshold has been used in an analogous study of esotropia.⁸ For this analysis, angle measurements were calculated as the average of a patient's measurements over the whole day (up to four per patient).

Results

A total of 26 children (median age, 7 years; range, 1–13) were recruited. Of these, 8 (31%) wore habitual refractive correction for angle measurements. Eighteen (69%) were female; 23 (88%) reported their race as white. The median angle of deviation for the 26 children in the present study was 25^Δ (range, 10^Δ – 40^Δ) at distance (3 meters) and 14.5^Δ (range, 5^Δ – 45^Δ) at near by prism and alternate cover test, based on the first measurement for each child. Prism and alternate cover test angle measurements were obtained during either four examinations (22 patients) or three examinations (4 patients) at least 2 hours apart, over the course of one day. The total number of included measurements in the 26 children was 100 at distance and 93 at near. The same examiner performed the test and retest measurements for 67 of 74 comparisons at distance and 64 of 69 comparisons at near.

Test-retest differences are represented in Bland-Altman plots in Figure 1A (distance), and Figure 1B (near). There was significantly greater variability with larger versus smaller deviations at near ($P < 0.0001$) but not at distance ($P = 0.09$). Despite the borderline difference in variability with larger angles at distance, we elected to separate larger and smaller angles for analysis because we had few patients with small-distance angles (minimum 10^Δ required for inclusion in the study), somewhat limiting our ability to find an effect if there was one. Variability was somewhat greater for near fixation than distance fixation, so analysis of distance and near data was performed separately.

The 95% repeatability coefficients (equivalent to the 95% limits of agreement for the difference between two measures) are shown in Table 1. Based on these data, a change of at least 4^Δ would be consistent with a clinically significant change in distance angle of deviation if the reference measurement was $\leq 20^\Delta$, and a change of at least 8^Δ would be consistent with a clinically significant change if the reference measurement was $>20^\Delta$. At near, a change of at least 7^Δ would be consistent with a clinically significant change in angle of deviation if the reference measurement was $\leq 20^\Delta$, and a change of at least 13^Δ would be consistent with a clinically significant change if the reference measurement was $>20^\Delta$.

For individual measurements the 95% repeatability coefficient (equivalent to the 95% limits of agreement for an individual measurement) at distance was 2.4^Δ (95% CI, 0.5^Δ – 4.4^Δ) for deviations $\leq 20^\Delta$ and 5.1^Δ (95% CI, 3.1^Δ – 7.0^Δ) for deviations $>20^\Delta$. At near, the 95% repeatability coefficient for individual measurements was 4.7^Δ (95% CI, 2.6^Δ – 6.8^Δ) for deviations $\leq 20^\Delta$ and 9.0^Δ (95% CI, 3.7^Δ – 14.4^Δ) for deviations $>20^\Delta$.

Discussion

By analyzing test-retest differences of prism and alternate cover test measurements in children with intermittent exotropia, we have quantified certain elements of measurement variability. Derived thresholds enable more rigorous evaluation of clinically significant

change in the angle of deviation in intermittent exotropia over time. These thresholds for change incorporate both test–retest differences and differences due to short-term intrinsic variability of the condition over a period of hours.

In previously published, large cohort studies of intermittent exotropia, a threshold of 10^Δ is often used to signify clinically significant change in angle of deviation over time.^{1,2,9} Our data showed a threshold of as little as 4^Δ would indicate clinically significant change in the magnitude of the exodeviation at distance fixation (in small angles) and therefore applying 10^Δ could result in under-reporting of change in intermittent exotropia over time in some cases. Lower thresholds, such as $>5^\Delta$, have been used to assess change over time,³ but based on the results of the present study, thresholds for clinically significant change differ depending on the magnitude of the reference measurement. Indiscriminately applying one threshold, distance and near, regardless of magnitude may result in under or over-reporting of change in angle.

Test–retest reliability of strabismus angle measurements has been evaluated in a previous study of childhood esotropia (minimum angle 10^Δ).⁸ Using interobserver test–retest measurements, the esotropia study also found different thresholds for clinically significant change at different magnitudes of esodeviation: the 95% limits of agreement on a difference between two measurements at distance was 10.4^Δ if the angle was $>20^\Delta$, and 5.8^Δ if the angle was 10^Δ to 20^Δ . Our test–retest data in intermittent exotropia also indicated slightly less variability if the angle was $<20^\Delta$ compared with angles $>20^\Delta$. It is likely that this increase in variability with larger deviations is to some extent a function of measurement method. Prism diopter increments in a standard prism set are 1^Δ to 2^Δ steps up to 20^Δ and then 5^Δ steps from 20^Δ to 50^Δ ; it is thus not common to measure a difference of $<5^\Delta$ if the angle of deviation is $>20^\Delta$, forcing greater variability between measurements. In the esotropia study different observers were used for test and retest measurements,⁸ whereas in the present study the same observer performed the retest examination in 67 of 74 cases (91%), potentially reducing variability compared with using different observers for test and retest. Another test–retest study¹⁰ evaluated angle measurements in patients with abducens nerve palsy (16–81 years of age), and found the 95% limits of agreement were 10.2^Δ at distance and 9.2^Δ at near using prism and alternate cover test,¹⁰ on average slightly more than the present study, but again using different observers.

Intermittent exotropia is by definition a condition that varies between periods of frank exotropia and periods of normal alignment with binocular single vision. It is therefore not surprising that some clinical tests may show greater variability in intermittent exotropia than in other types of strabismus. In previous studies of test–retest variability of Preschool Randot stereoacuity,^{11,12} a two-level change was found to be necessary to exceed test–retest variability in a range of strabismus types. Nevertheless, analyzing intermittent exotropia alone, test–retest variability was greater, requiring a three-level difference to exceed test–retest variability.¹¹ These thresholds incorporate changes due to test–retest variability and those due to intrinsic variability of the condition. In the present study, we expected angle of deviation to show greater variability in intermittent exotropia than in other types of strabismus but in fact found that variability was possibly somewhat less than that seen in esotropic strabismus.

The variable nature of intermittent exotropia makes it difficult to determine the specific cause of any change between measurements. Differences may be attributable to measurement error, intrinsic variability of the condition, change in testing conditions (lighting, etc), a clinically significant change in underlying severity, or a combination of these factors. We aimed to minimize rather than maximize potential sources of variability so that derived test–retest values most closely reflected variability due to measurement error

and intrinsic short-term variability of the condition. Our reported values therefore serve as a guide for the clinician. The potential influence of other factors that may affect variability (eg, a different observer, room, or lighting) should be considered when reviewing sequential angle measurements. Unfortunately, it is probably impractical to study the magnitude of variability attributable to each potential factor.

It is possible that clinically significant change in the underlying condition occurred between measurements in our study. In fact, we have previously reported change in control even from minute to minute,⁴ and so in the present study we evaluated changes that incorporate both test–retest variability and possible short-term change in the underlying condition. We aimed to establish thresholds that would allow determination of deterioration or improvement over longer periods of follow-up such as from one clinic visit to another over periods of weeks or months and we would expect such changes in the underlying severity to exceed the magnitude of change seen over one day. We also computed the reliability of individual measurements, providing a measure of precision for a given angle measurement in intermittent exotropia. There are few comparable data in strabismus populations, but a test–retest study in childhood esotropia⁸ reported reliability of individual measurements: at distance, 95% limits of agreement were 7.3^{Δ} for esotropic angles $>20^{\Delta}$ and 4.1^{Δ} for esotropic angles of 10^{Δ} to 20^{Δ} ,⁸ similar to values found in the present study of intermittent exotropia.

Due to the demands of repeat measurements in our study, only a small number of patients could be recruited. A larger patient cohort may have yielded slightly different thresholds. We also studied a wide range of ages but did not have sufficient numbers to analyze variability by age groups: the effect of age on variability is unknown, but it is possible that there may have been more variability in younger than in older children. Another potential limitation is that the same observer performed the repeat measurement of angle in 67 of 74 cases (91%) for distance testing and 64 of 69 cases (93%) for near testing: this may have resulted in closer agreement between measures than if we had used a different observer for repeat measurements. Nevertheless, it might be expected that there would be greater variability in intermittent exotropia than in other types of strabismus, potentially biasing the examiner towards finding differences in measures. In addition, care was taken to not review previous measurements. It is possible that the presence of tenacious proximal fusion may have influenced variability in angle measurements, although it is uncertain whether patients with tenacious proximal fusion would show increased or decreased variability. To overcome tenacious proximal fusion it would have been necessary to perform both test and retest measurements following prolonged periods of occlusion, which in the present study would have been impractical. In addition, the effects of repeated periods of occlusion on the underlying angle are unknown.

Using test–retest reliability data, we have calculated thresholds to be used as a guide for assessing clinically significant change in the angle of deviation using the prism and alternate cover test in children with intermittent exotropia. These rigorously derived data will allow a more evidence-based approach to assessing long-term change in angle of deviation in children with intermittent exotropia. Nevertheless, it would be entirely reasonable for other factors, such as control and stereoacuity, to be considered when evaluating change in overall severity of intermittent exotropia and when making treatment decisions, until a future, definitive study determines which, if any, of these parameters are truly important.

Acknowledgments

Supported by National Institutes of Health Grants EY015799 and EY018810 (JMH), Research to Prevent Blindness, New York, NY (JMH as Olga Keith Weiss Scholar and an unrestricted grant to the Department of Ophthalmology, Mayo Clinic), and Mayo Foundation, Rochester, MN.

References

1. Romanchuk KG, Dotchin SA, Zurevinsky J. The natural history of surgically untreated intermittent exotropia—looking into the distant future. *J AAPOS*. 2006; 10:225–231. [PubMed: 16814175]
2. Nusz KJ, Mohny BG, Diehl NN. The course of intermittent exotropia in a population-based cohort. *Ophthalmology*. 2006; 113:1154–1158. [PubMed: 16647128]
3. Chia A, Seenyen L, Long QB. A retrospective review of 287 consecutive children in Singapore presenting with intermittent exotropia. *J AAPOS*. 2005; 9:257–263. [PubMed: 15956946]
4. Hatt SR, Mohny BG, Leske DA, Holmes JM. Variability of control in intermittent exotropia. *Ophthalmology*. 2008; 115:371–376. [PubMed: 17629562]
5. Hatt SR, Mohny BG, Leske DA, Holmes JM. Variability of stereoacuity in intermittent exotropia. *Am J Ophthalmol*. 2008; 145:556–561. [PubMed: 18201680]
6. Hatt SR, Liebermann L, Leske DA, Mohny BG, Holmes JM. Improved Assessment of Control in Intermittent Exotropia Using Multiple Measures. *Am J Ophthalmol*. 2011; 152:872–876. [PubMed: 21843874]
7. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999; 8:135–160. [PubMed: 10501650]
8. Pediatric Eye Disease Investigator Group. Interobserver reliability of the prism and alternate cover test in children with esotropia. *Arch Ophthalmol*. 2009; 127:59–65. [PubMed: 19139339]
9. Hiles DA, Davies GT, Costenbader FD. Long-term observations on unoperated intermittent exotropia. *Arch Ophthalmol*. 1968; 80:436–442. [PubMed: 5674800]
10. Holmes JM, Leske DA, Hohberger GG. Defining real change in prism-cover test measurements. *Am J Ophthalmol*. 2008; 145:381–385. [PubMed: 18045567]
11. Adams WE, Leske DA, Hatt SR, Holmes JM. Defining real change in measures of stereoacuity. *Ophthalmology*. 2009; 116:281–285. [PubMed: 19091410]
12. Fawcett SL, Birch EE. Interobserver test–retest reliability of the Randot preschool stereoacuity test. *J AAPOS*. 2000; 4:354–358. [PubMed: 11124670]

FIG 1.

Test-retest variability of prism and alternate cover test measurements. A, at distance fixation. B, at near fixation. Bland-Altman plot showing 95% repeatability coefficient (equivalent to 95% limits of agreement for the difference between 2 measures). Middle dotted line represents the mean of test-retest differences (0). There was significantly greater variability with larger versus smaller deviations at near ($P < 0.0001$) with borderline greater variability with larger angles at distance ($P = 0.09$).

Table 1

Derived 95% repeatability coefficient^a values for angle of deviation test–retest differences, using the prism and alternate cover test in children with intermittent exotropia

	PACT angle 20 PD		PACT angle >20 PD	
	95% repeatability coefficient	95% CI	95% repeatability coefficient	95% CI
Distance (PD)	3.4	0.7 to 6.2	7.2	4.4 to 9.9
Near (PD)	6.6	3.7 to 9.6	12.8	5.3 to 20.3

CI, confidence interval; *PACT*, prism and alternate cover test; *PD*, prism diopters.

^aEquivalent to the 95% limits of agreement in a two-measure study.