



Published in final edited form as:

Am J Ophthalmol. 2008 February ; 145(2): 381–385. doi:10.1016/j.ajo.2007.09.012.

Defining Real Change in Prism-Cover Test Measurements

Jonathan M. Holmes, David A. Leske, and George G. Hohberger

Department of Ophthalmology, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905

Abstract

Purpose—To describe the interobserver test-retest variability of both simultaneous prism and cover testing (SPCT) and alternate prism and cover testing (APCT) in horizontal deviations, and to calculate 95% limits of agreement that might be used to define real change.

Design—Prospective cohort study

Methods—23 patients with sixth nerve palsy and 3 controls were independently examined by two experienced strabismus surgeons. SPCT and APCT were performed at distance and near fixation. Test-retest variability and agreement between tests were evaluated using Bland-Altman plots and 95% limits of agreement were calculated.

Results—For SPCT, the 95% limits of agreement half-widths were 6.3 prism diopters (pd) at distance fixation and 6.9 pd at near. For APCT, the 95% limits of agreement half-widths were 10.2 pd at distance and 9.2 pd at near.

Conclusions—Based on 95% limits of agreement half-widths between two examiners, a change in strabismus measurements of less than 10 pd may be due to test-retest variability. Changes of 10 pd or more are likely to represent real change and might be used as the threshold for management decisions.

INTRODUCTION

Clinical decisions in strabismus management are made, in large part, on the basis of measuring the magnitude of the deviation using the simultaneous prism and cover test (SPCT) for the tropic component and the alternate prism and cover test (APCT) for the combined tropia and phoria. Management decisions, such as whether to order neuroimaging, and surgical decisions, such as whether to perform or defer surgery and what surgery to perform, are based partly on whether the angle of misalignment has changed. Nevertheless there are very few data to guide the clinician on the magnitude of change that likely represents a real change.

In order to provide data on real change in strabismus measurements we studied test-retest variability of the simultaneous prism and cover test and the alternate prism and cover test.

Corresponding author and reprint requests to: Dr. Jonathan M. Holmes, Department of Ophthalmology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, e-mail: holmes.jonathan@mayo.edu, Telephone 507-284-3760, Fax 507-284-8566.

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.

METHODS

Patients

As part of a study of photographic assessment of sixth nerve palsy and paresis,¹ we enrolled 23 patients with the clinical diagnosis of 6th nerve palsy and 3 controls (two with resolved sixth nerve palsy and one with a very small skew deviation). Controls were included to represent no deviation or very small angle deviations, so that a large range of angles of deviation could be studied. As reported previously, patient ages ranged from 16 to 81 years, 38% were male and 92% were white. All patients had an uncorrected visual acuity of 20/200 or better in each eye, to allow SPCT and APCT measurements. It was felt that visual acuity worse than 20/200 would make such measurements difficult to obtain..

Strabismus measurements

As described previously,¹ each patient was assessed independently by two fellowship-trained strabismus specialists. The second examiner was masked to the measurements of the first examiner and the second evaluation was made within one hour of the first evaluation. Each examiner recorded a SPCT measurement at distance fixation in the primary position using a 20/200 fixation letter at 3 meters and at near fixation using a 20/200 equivalent letter pasted on a tongue depressor held at 1/3 meter. This near distance was estimated by each examiner, and not formally measured, to mimic routine clinical testing performed by most strabismus surgeons. An APCT measurement in the primary position at 3 meters and 1/3 meter fixation was also recorded at the same fixation distances. A standard set of loose plastic prisms was used for all measurements. The individual prisms increased in power from 1 prism diopter (pd) to 10 pd in single pd increments, from 10 to 20 pd in 2 pd increments, and from 20 to 50 pd in 5 pd increments. As in common clinical practice, deviations were recorded as the value of the prism that came the closest to neutralizing the misalignment. In all deviations less than 50 pd, the primary deviation, rather than the secondary deviation, was measured by holding prism over the paretic eye. When the deviation exceeded 50 pd such that a single prism could not be used, prisms were split between eyes rather than stacked. The practice of the two examiners at the time of the current study was to leave the 50 pd prism over the paretic eye and add smaller prisms over the fellow eye. When prisms were split, the non-paretic eye was not technically in primary position, but there is no practical way to measure the deviation with the eye in primary position when the deviation exceeds 50 pd except by stacking prisms, which induces other artifacts. If prisms were split more equally between eyes, then the non-paretic eye would be even further from the primary position, and therefore the described method was used. Subjective methods, relying on patients responses of diplopia or no diplopia, were not used in this study. As in common clinical practice, head position was not formally controlled by using a bite bar or chin rest.

Analysis

For one patient, the deviation was so large that only one of the examiners felt it could be quantified (85 pd by APCT), and since there was no corresponding pair of SPCT or APCT measurements from the other examiner, this patient was excluded from analysis of SPCT and APCT measurements. For one additional patient, one of the examiners recorded that angle by SPCT as “greater than 50 pd.” A specific SPCT value was not recorded for this patient and the patient was excluded from analysis of SPCT measurements. Therefore, we had 24 pairs of SPCT measurements and 25 pairs of APCT measurements for analysis.

The difference between each pair of measurements (Examiner 1 – Examiner 2) was calculated for each of the SPCT measurements (distance and near) and for each of the APCT measurements (distance and near). The 95% limits of agreement half-widths were calculated by multiplying the standard deviation of the differences by 1.96. The 95% limit of agreement

is a commonly used statistic in evaluating reliability, and represents the magnitude of a change where one would be 95% certain that the change is within test-retest variability. Bland-Altman plots were also constructed.² All analyses were conducted using SAS software version 9.1.3. Intraclass correlation coefficients were also calculated. The intraclass correlation coefficient accounts for any potential systematic difference between pairs of values, unlike a Pearson or Spearman correlation coefficient, which do not account for such differences.

RESULTS

We have previously reported¹ excellent intraclass correlation coefficients of SPCT and APCT in these patients ranging from 0.94 to 0.96.

Bland-Altman plots for SPCT at distance and near are shown in Figure 1 and Figure 2 respectively, and for ACT at distance and near in Figure 3 and Figure 4 respectively. The calculated 95% limits of agreement half-widths were 6.9 pd and 6.3 pd for SPCT at distance and near fixation and 10.2 pd and 9.2 pd for APCT at distance and near fixation (Table) 95% confidence intervals around the limits of agreement are shown in the Table.

Variability did not appear to change with magnitude of strabismus (Figure 1 to Figure 4); the smaller misalignments had similar variability to larger misalignments.

DISCUSSION

The 95% limits of agreement half-widths for interobserver test-retest variability were approximately 7 pd for SPCT at both distance and near fixation and approximately 10 pd for APCT at both distance and near fixation. Given the pre-determined steps in standard prism sets, 10 pd might be considered a practical threshold for defining real change in measurement with 95% certainty. The next smallest prism step for moderate and large deviations is commonly 5 pd, and this magnitude of apparent change falls well within test-retest variability. For smaller angles, when using smaller prism steps, the value of 7 pd could be used as a threshold for real change in SPCT.

Previous studies of test-retest variability in prism and cover tests have been limited to normal volunteers with pure phorias. For example Johns et al³ studied 72 volunteers with phorias ranging from 24 exophoria to 14 esophoria, and found 95% limit of agreement of ± 4 pd using alternate prism and cover testing. An earlier study by Rainey et al⁴ with a smaller range of phorias found a similar 95% limit of agreement of ± 3.6 pd. It is not surprising that we found wider limits of agreement of up to 10 pd in a clinical population of patients with sixth nerve palsy who had deviations of up to 65 pd.

The difference between SPCT and APCT is somewhat surprising because the SPCT is considered somewhat more difficult to perform and quantify. Nevertheless, the APCT might be confounded by differing periods of dissociation, revealing more or less of an underlying phoric component. Such variability in technique might lead to greater variability in measured deviations by APCT.

There are several limitations to our study. First, we only studied patients with a single condition, sixth nerve palsy or paresis. Sixth nerve palsy or paresis is by definition an example of parietic and incomitant strabismus. It is possible that test-retest variability might be less in non-parietic or more comitant types of strabismus since head position might be less critical. Second, we had a relatively small sample size, which resulted in fairly wide confidence intervals around our estimated limits of agreement. Third, although we had a wide range of deviations from zero to over 50 pd, we did not have a large amount of data for patients within subgroups of small, medium, and large deviations, limiting our ability to detect a difference in variability in

smaller versus larger angles. Fourth, we studied only adult patients, and the variability of strabismus measurements in children might be expected to be greater than those in adults.

The limits of agreement were similar for each of the tests at each of distance and near fixation. One might speculate that the absence of standardization of the testing distance for near fixation (for example by using a tape measure), would be an additional source of measurement error, but we did not find a marked difference between test-retest variability of SPCT and APCT measurements at distance and near fixation. Interestingly, there was a small bias towards larger angles by examiner 2 at near but not at distance (3 pd by SPCT and 4 pd by APCT; Figure 2 and Figure 4), which most likely reflects differences in fixation distance at near.

An additional weakness is that we did not study intra-observer test-retest variability. This might be important since clinicians often compare their own measurements from visit to visit. Nevertheless, it may be impossible to design a study that truly examines intra-observer test-retest reliability, since an examiner is very likely to remember the last measurement, and is likely to be influenced by that measurement. In addition, attempting to tape or otherwise mask prisms or prism bar may be thwarted by familiarity of the examiners with the relative shapes and sizes of the prisms. We believe that our present study of inter-observer test-retest variability between two experienced strabismus specialists, closely models what might be expected between measurements made by the same individual. An additional weakness of our study might be that measurements were performed within one hour of each other, and we are applying our results to measurements made weeks and months apart. To truly study interobserver test-retest variability over weeks or months, one would have to assume that the underlying condition had not changed, and such an assumption is unreasonable, particularly in cases of sixth nerve palsy, where spontaneous recovery would be expected in a proportion of cases.^{5, 6} The use of interobserver test-retest studies in defining real change has been accepted as a valid methodology and has been applied to visual acuity⁷ and to testing of stereoacuity.⁸

The available steps within the prism set limits our ability to more finely quantify strabismus, but finer gradation may give a false impression of accuracy, since intrinsic variability of testing may be a more important factor. Finer versus coarser steps have been discussed in the context of visual acuity testing,⁷ where decreasing scale coarseness had a modest effect on reducing measurement error. Nevertheless, applying measurement error principles to discrete steps requires choosing one step magnitude as the closest actual step that corresponds to the calculated measurement error. In this context, using a standard prism set, we determined that a change of 10 pd most likely represented real change.

It is of interest that variability among small deviations appears similar in magnitude to that found when measuring larger deviations (Figure 1–Figure 4). This may reflect the difficulty of determining the magnitude of small tropias and the difficulty of determining whether a small deviation even exists. For example, in the present study, two patients were judged to be orthotropic by the first examiner and judged by SPCT to have a 2 pd esotropia and 4 pd esotropia respectively by the second examiner, while a third patient was judged to have a 4 pd esotropia by the first examiner and judged to be orthotropic by the second examiner. It is also possible that these small tropias might have been intermittent and truly present for one examiner and absent for the other. Eye tracker technology applied to quantifying tropias may prove to be useful in both clinical and research settings as a method of objectively quantifying small angle misalignment (Leske DA et al, IOVS 2007;48:ARVO E-Abstract 900).

Based on our study, if a second prism-cover measurement of strabismus, made days or months later, is 5 pd more or less than the last measurement, then the difference is well within test-retest variability. A change of 10 pd from one visit to the next is likely to represent a real

change, and 10 pd might reasonably be used as a threshold for defining improvement, deterioration or instability of ocular misalignment.

ACKNOWLEDGEMENTS / DISCLOSURE

Funding / Support: National Institutes of Health Grants EY015799 (JMH), Research to Prevent Blindness, Inc., 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

Financial disclosures: No conflicting financial relationships exist

Contributions of authors: Design of study (JMH, DAL, GGH); Conduct of study (JMH, DAL, GGH); Collection (JMH, GGH); Management (JMH, DAL, GGH); Analysis (JMH, DAL); Interpretation of data (JMH, DAL, GGH); Preparation / review of manuscript (JMH, DAL, GGH)

Institutional Review Board/Ethics Committee approval was obtained and all patients gave written informed consent. All experiments and data collection were conducted in a manner compliant with the Health Insurance Portability and Accountability Act.

REFERENCES

1. Holmes JM, Hohberger GG, Leske DA. Photographic and clinical techniques for outcome assessment in sixth nerve palsy. *Ophthalmology* 2001;108:1300–1307. [PubMed: 11425691]
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310. [PubMed: 2868172]
3. Johns HA, Manny REF K, Hu YS. The intraexaminer and interexaminer repeatability of the alternate cover test using different prism neutralization endpoints. *Optom Vis Sci* 2004;81:939–946. [PubMed: 15592119]
4. Rainey BB, Schroeder TL, Goss DA, Grosvenor TP. Inter-examiner repeatability of heterophoria tests. *Optom Vis Sci* 1998;75:719–726. [PubMed: 9798211]
5. Holmes JM, Droste PJ, Beck RW. The natural history of acute traumatic sixth nerve palsy or paresis. *J AAPOS* 1998;2:265–268. [PubMed: 10646746]
6. Holmes JM, Leske DA, Christiansen SP. Initial treatment outcomes in chronic sixth nerve palsy. *J AAPOS* 2001;5:370–376. [PubMed: 11753258]
7. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci* 1993;34:120–129. [PubMed: 8425819]
8. Fawcett SL, Birch EE. Interobserver test-retest reliability of the Randot preschool stereoacuity test. *J AAPOS* 2000;4:354–358. [PubMed: 11124670]

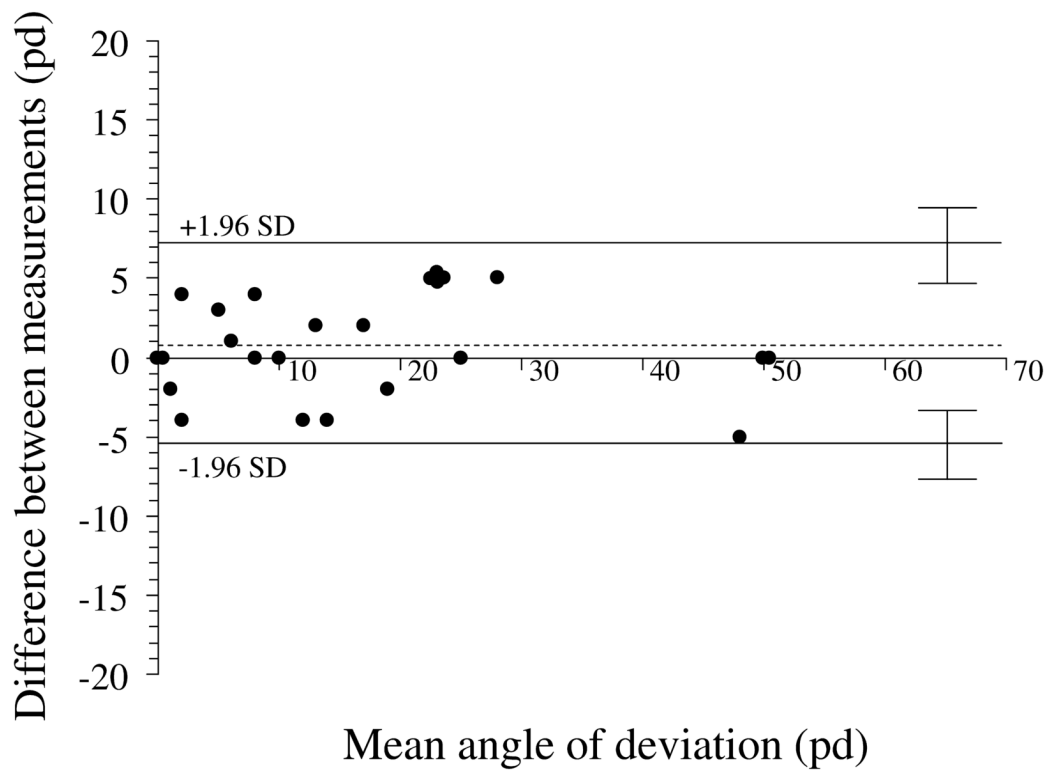


Figure 1. Test-retest variability of *simultaneous* prism and cover test at *distance* fixation. Bland-Altman plot, showing 95% limits of agreement half-widths ($\pm 1.96 \times$ standard deviation of the difference) to be ± 6.3 pd. Dotted line represents the mean of test-retest differences; $0.8 \text{ pd} \pm 3.2 \text{ pd}$. There appears to be no relationship between magnitude of deviation and test-retest variability. pd = prism diopters

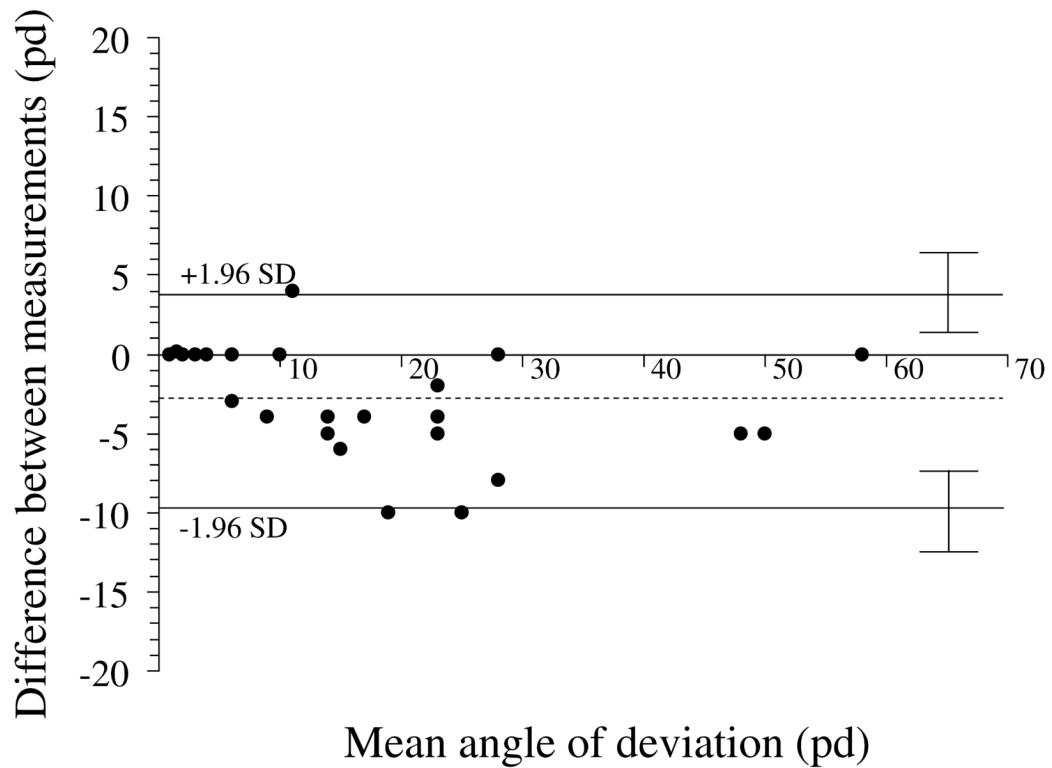


Figure 2. Test-retest variability of *simultaneous* prism and cover test at *near* fixation. Bland-Altman plot, showing 95% limits of agreement half-widths ($\pm 1.96 \times$ standard deviation of the difference) to be ± 6.9 pd. Dotted line represents the mean of test-retest differences; -3.0 pd ± 3.5 pd. There appears to be no relationship between magnitude of deviation and test-retest variability. pd = prism diopters

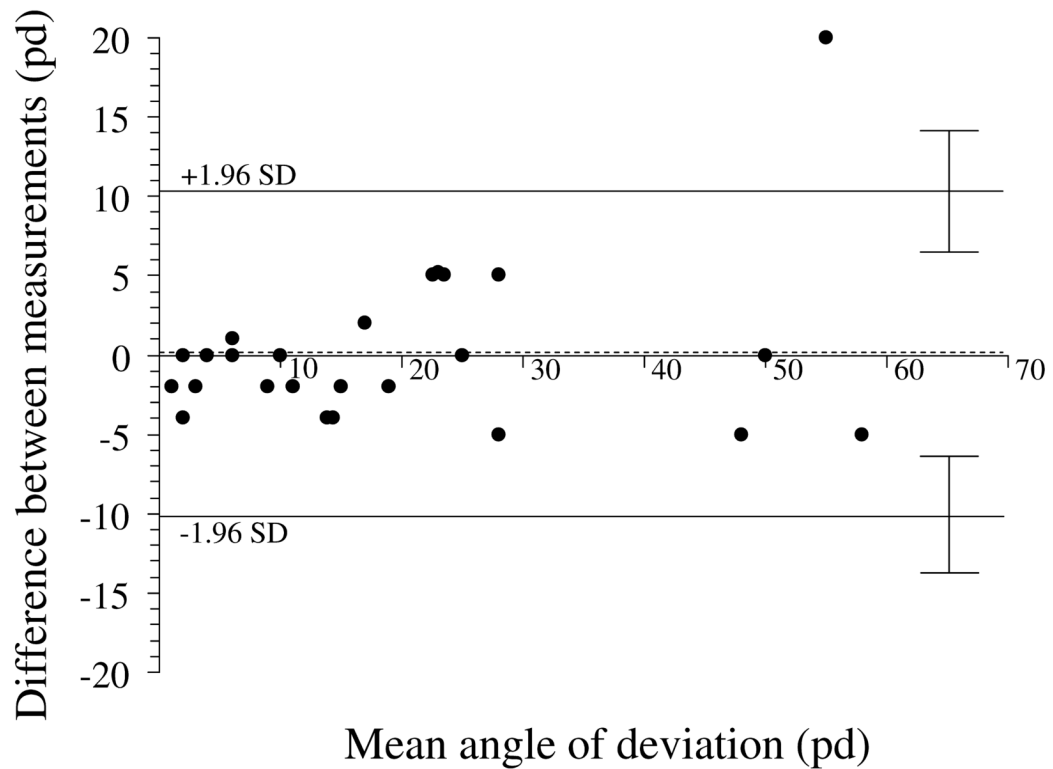


Figure 3.

Test-retest variability of *alternate* prism and cover test at *distance* fixation. Bland-Altman plot, showing 95% limits of agreement half-widths ($\pm 1.96 \times$ standard deviation of the difference) to be ± 10.2 pd. Dotted line represents the mean of test-retest differences; 0.2 pd ± 5.2 pd. There appears to be no relationship between magnitude of deviation and test-retest variability. pd = prism diopters

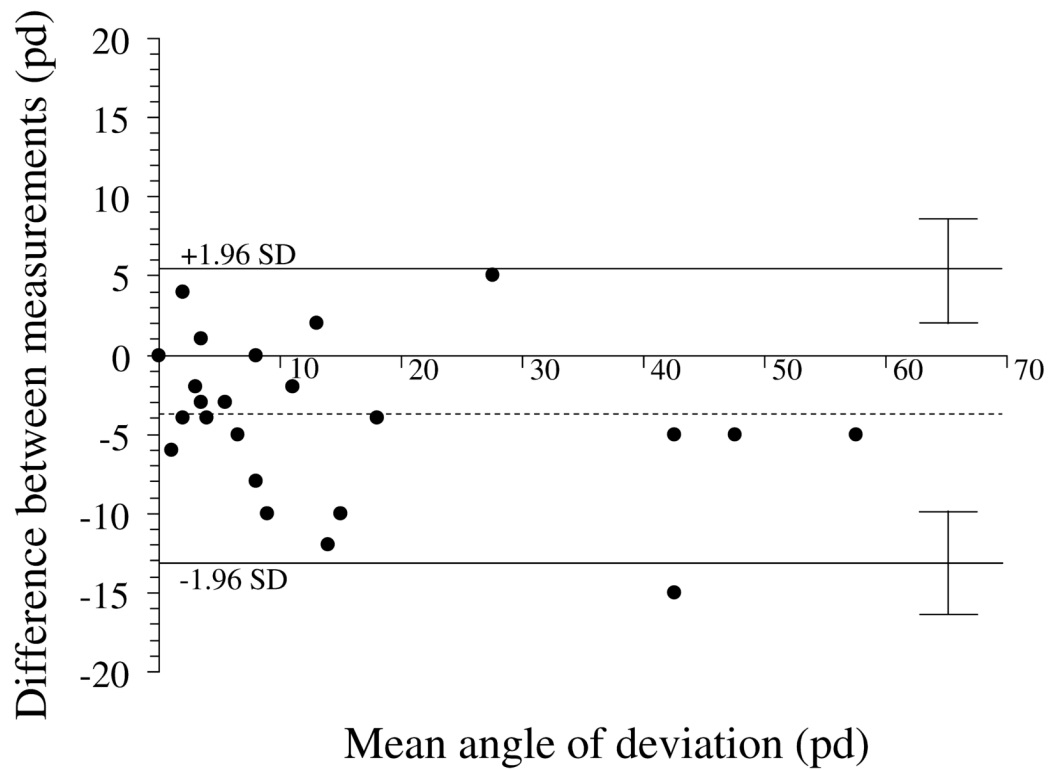


Figure 4. Test-retest variability of *alternate* prism and cover test at *near* fixation. Bland-Altman plot, showing 95% limits of agreement half-widths ($\pm 1.96 \times$ standard deviation of the difference) to be ± 9.2 pd. Dotted line represents the mean of test-retest differences; -3.9 pd ± 4.7 pd. There appears to be no relationship between magnitude of deviation and test-retest variability. pd = prism diopters

Table

95% Limits of Agreement for Measures of Strabismus

Type of measurement	95% limits of agreement half-widths (prism diopters)	95% CI on limits of agreement (prism diopters)
Simultaneous prism and cover test at <i>distance</i> fixation	± 6.3	± 3.7
Simultaneous prism and cover test at <i>near</i> fixation	± 6.9	± 3.3
Alternate prism and cover test at <i>distance</i> fixation	± 10.2	± 2.3
Alternate prism and cover test at <i>near</i> fixation	± 9.2	± 2.5