

Intraexaminer repeatability and agreement in stereoacuity measurements made in young adults

Beatriz Antona, Ana Barrio, Isabel Sanchez, Enrique Gonzalez, Guadalupe Gonzalez

Department of Optics II (Optometry and Vision), Faculty of Optics and Optometry, Complutense University of Madrid, Madrid 28037, Spain

Correspondence to: Beatriz Antona. Faculty of Optics and Optometry, Complutense University of Madrid, C/Arcos de Jalón, n° 118, Madrid 28037, Spain. bantona@opt.ucm.es

Received: 2013-12-28 Accepted: 2014-12-31

Abstract

• **AIM:** To determine the repeatability and agreement of stereoacuity measurements made using some of the most widely used clinical tests: Frisby, TNO, Randot and Titmus.

• **METHODS:** Stereoacuity was measured in two different sessions separated by a time interval of at least 24h but no longer than 1wk in 74 subjects of mean age 20.6y using the four methods. The study participants were divided into two groups: subjects with normal binocular vision and subjects with abnormal binocular vision.

• **RESULTS:** Best repeatability was shown by the Frisby and Titmus [coefficient of repeatability (COR): ± 13 and ± 12 s arc respectively] in the subjects with normal binocular vision though a clear ceiling effect was noted. In the subjects with abnormal binocular vision, best repeatability was shown by the Frisby (COR: ± 69 s arc) and Randot (COR: ± 72 s arc). In both groups, the TNO test showed poorest agreement with the other tests.

• **CONCLUSION:** The repeatability of stereoacuity measures was low in subjects with poor binocular vision yet fairly good in subjects with normal binocular vision with the exception of the TNO test. The reduced agreement detected between the tests indicates they cannot be used interchangeably.

• **KEYWORDS:** stereoacuity, repeatability, agreement, stereopsis

DOI:10.3980/j.issn.2222-3959.2015.02.29

Antona B, Barrio A, Sanchez I, Gonzalez E, Gonzalez G. Intraexaminer repeatability and agreement in stereoacuity measurements made in young adults. *Int J Ophthalmol* 2015;8(2):374-381

INTRODUCTION

Stereoscopic vision is an essential characteristic of human binocular vision^[1]. The demonstration of normal

stereopsis for age indicates correct development of sensory and motor functions^[2-4]. Near stereoacuity tests are commonly used in the clinical setting to identify abnormalities in binocular function such as heterotropia and amblyopia and to manage patients with binocular vision disorders^[3,5-8].

There are presently several tests available to assess stereoacuity. In general, all commonly used stereoacuity tests are easy to administer and each has its own characteristics^[9]. Moreover, wide variation in stereoacuity thresholds has been observed in individual subjects between stereotests^[10,11].

Random dot stereotests that measure global stereopsis are more recommended at detecting strabismus, amblyopia, low visual acuity (VA) or anisometropia than the linear or contour stereotests (for example Titmus and Randot), which measure local stereopsis. This is because contour stereotests have monocular, contour or lateral displacement cues, which are not present in the random-dot stereotest (for example TNO test)^[3,12-14]. As a result, many subjects may be inaccurately assigned values of stereoacuity reflecting rescued or restored binocular vision, when only coarse or no stereopsis is present.

Real depth stereotests such as the Howard Dolman and the Frisby test have the advantage of not requiring dissociative glasses and are therefore less artificial^[9]. The manufacturers of the Frisby indicate that the only monocular depth cue present is motion parallax, and this is easily controlled by restricting relative movement between patient and test plate^[15]. This has in effect been confirmed in studies conducted both in children and adults^[16].

It has been reported that most observers with no ocular abnormalities can discriminate depth differences produced by a relative disparity of as little as 10s arc^[17]. However, this report failed to describe the types of test in which these stereoacuties are attainable^[18,19]. Given that the lowest disparity available in clinical tests is considerably greater, many subjects are able to identify the lowest disparity offered by the tests, which gives rise to a ceiling effect^[8,10,20,21].

When selecting a test for clinical use, its repeatability needs to be known to correctly interpret if a change in the measures made using a given stereotest is sufficient to be clinically significant or may be attributed to variation to the technique^[22].

To date, no study has compared at the same time the reliability of these four frequently used stereotests. There have been some reports on these tests^[8,10,20,22], but in most

Table 1 Description of the stereoacuity tests

| Test | Characteristics of the tests | Attainable stereoacuity scores (s arc) | | | Test distance (cm) |
|--------|---|--|------|--------|--------------------|
| | | 6 mm | 3 mm | 1.5 mm | |
| Frisby | 3 sheets of different thickness: 6, 3 and 1.5 mm (random elements) | 600 | 300 | 150 | 30 |
| | | 340 | 170 | 85 | 40 |
| | No filters (real depth test) | 215 | 110 | 55 | 50 |
| | | 150 | 75 | 40 | 60 |
| | | 110 | 55 | 30 | 70 |
| TNO | 3 sheets-6 degrees of disparity (random points) Red/green glasses | 85 | 40 | 20 | 80 |
| | | 480, 240, 120, 60, 30, 5 | | | Always at 40 |
| Randot | 10 groups of circles (contour design with background of random points) Polarized glasses | 400, 200, 140, 100, 70, 50, 40, 30, 25, 20 | | | Always at 40 |
| Titmus | 9 groups of circles (contour design) | 800, 400, 200, 140, 100, 80, 60, 50, 40 | | | Always at 40 |
| | Polarized glasses | | | | |

cases, the statistical analysis of data has not been adequate. Owing to the different working mechanisms of stereo vision tests (real depth, random points or contour design) and to the different disparities evaluated by each test (Table 1), their reliability is likely to differ. Given that stereotests use coarse steps at low stereo levels, we would expect lower reliability within tests and poorer agreement between tests in subjects with abnormal binocular vision than in those with normal binocular vision.

The purpose of this study was to assess intra-observer repeatability of some of the tests most widely used in clinical practice to obtain a threshold stereoacuity (Frisby, TNO, Randot circles and Titmus circles tests). Also examined was the level of agreement between these four tests in subjects with and without normal binocular vision. Our final goal was to generate information to help clinicians to accordingly interpret stereoacuity measurements.

SUBJECTS AND METHODS

Study Population The study population was comprised of 74 subjects aged 18 to 32y (mean 20.6, SD 2.8y) recruited from the first year students attending the School of Optics, Universidad Complutense de Madrid (Madrid, Spain). Having recently been admitted, the subjects selected were unaccustomed to the type of tests performed. The results of this study could therefore be extrapolated to a healthy population of this age group with similar near work demands. The project was approved by the Ethics Committee of Clinical Studies of the University Hospital Ramon & Cajal (Madrid, Spain). The study protocol fulfilled the tenets of the Declaration of Helsinki and the subjects gave their consent to participate after the nature of the study had been explained to them.

The subjects were first required to complete a questionnaire to record their age, sex and eye history. Next, VA and binocular vision characteristics of each subject were determined with his subjective correction.

Table 2 Cut-off values for the initial screening tests

| Test | Method | Cut-off for inclusion |
|--------------------------------|------------------------|-------------------------|
| Amplitude of accommodation | Push-up test | ≥6 D |
| Near point of convergence | Accommodation test | Break point ≤7.5 |
| | | Recovery point ≤10 m |
| Deviation at distance and near | Cover test | Distance: ortho-3BI |
| | Von Graefe technique | Near: ortho-6BI |
| Central suppression | 4Δ Base-out prism test | Negative |
| Step vergence testing | Prism bar | Far vision PFV ≥4/5 Δ |
| | | NFV ≥4/2 Δ |
| | | Near vision PFV ≥10/7 Δ |
| | | NFV ≥7/5 Δ |

The study participants were divided into two groups according to the results of the preliminary tests. The first group (n=54) had normal binocular vision and fulfilled the inclusion criteria: 1) a corrected VA greater or equal to 0.9 Snellen decimal VA in each eye at distance and near; 2) no history of eye disease, refractive surgery, strabismus, nystagmus or amblyopia. No manifest deviation using the cover test and a negative 4 Δ base-out prism test; 3) no medication or disease that could affect accommodation, fusional vergences or ocular motility; 4) asymptomatic subjects without accommodative or vergence alterations. Cut-off values for the screening tests are provided in Table 2.

The second group was made up of 20 subjects with the binocular disorders: amblyopia (n=15), constant strabismus (n=3) or intermittent strabismus (n=2). All the subjects in this group had no history of eye disease or medication that could affect accommodation, vergence or ocular motility, and were required to fulfil one or both of the criteria: 1) a corrected VA lower or equal to 0.8 Snellen decimal VA in one or both eyes at distance and near; 2) manifest deviation using the cover test and/or a positive four base-out prism test.

Stereoacuity Measurements Stereoacuity was quantified in subjects using four different tests: Frisby Stereotest (Clement Clarke, Harlow, UK), TNO (Lameris Instrumenten, Groenekan, Netherlands), Randot circles (Stereo Optical

Table 3 Mean stereopsis±SD (s arc)

| Group | Frisby | TNO | Randot | Titmus | ¹ P |
|---------------------------|--------|---------|--------|--------|----------------|
| Normal binocular vision | 21±3 | 52±25 | 29±10 | 41±5 | <0.0001 |
| Abnormal binocular vision | 52±44 | 158±149 | 59±36 | 91±53 | |

¹Mann Whitney U-test.

Company, Chicago, IL, USA) and Titmus circles (Stereo Optical Company, Chicago, IL, USA). In Table 1, we provide a short description of the tests.

According to Bland and Altman ^[23], the best way of assessing the repeatability of an instrument is to take several measurements in a series of subjects. Thus, measurements were taken on two occasions separated by a time interval of at least 24h but no longer than 1wk to avoid visual function changes that could affect repeatability. All the measurements were taken between 12:30 and 15:30. At the first visit, the objectives of the study were briefly explained to each subject. The order of the tests for each subject was randomly selected to balance out variables such as fatigue and practice. All the stereoacuity tests in the two sessions were undertaken by the same examiner to minimise inter-examiner variability. Neither the examiner nor the subject was aware of the group they had been allocated to (normal or abnormal binocular vision). The results of the first set of measurements were not visible during second session, to avoid any possible influence of these on the examiner. Each subject underwent all the tests for one session on the same day. Each test was administered in the same way to each subject according to the manufacturers' guidelines. The test charts were always placed in primary gaze position and the examiner checked that the subject avoided head movements. When needed, subjects wore their correction as well as the special filters required for some of the tests. Lighting for the tests included an additional light above the subject's shoulder for uniform illumination of the stereotest (approximately 90 cd/m²).

Data Analysis The Bland-Altman method was used to determine the repeatability and agreement of the tests^[23]. The advantage of this method is that agreement among tests is expressed in the same units of measure as the test itself and allows the clinician to establish his own criterion as to whether or not a difference is significant, since a small difference could be statistically significant yet not clinically significant.

The variables determined were the mean difference (MD), the standard deviation of differences (SD), the coefficient of repeatability (COR=±1.96×SD) and the limits of agreement at the 95% level (MD±COR). We also determined coefficients of agreement (COA) among the tests.

Differences between scores obtained in session 1 and 2 for a given test were plotted against their means to establish the 95% limits of agreement and obtain a better idea of the repeatability of the measures. The agreement interval

represents a threshold for the differences in successive measures that has to be surpassed if the difference indicates that a change in the value has in effect occurred and cannot simply be explained by natural variation among measurements.

We also determined agreement between the scores of two tests by plotting the difference between the test scores obtained by each participant against the mean of his or her 2 test scores. For this analysis, the data recorded in both sessions for each test were averaged. This is known as a Bland-Altman plot and it allows assessment of whether test-retest reliability depends on the level of stereoacuity. From these plots we can establish the 95% limits of agreement and determine the repeatability of the measurements.

The significance of the difference between the stereoacuity result was calculated using a mixed ANOVA model with one between-subjects factor [group (normal vs abnormal binocular vision)] and two within-subjects factors [method (Frisby, TNO, Randot or Titmus)] and session (first vs second).

The normal distribution of data was determined using the Shapiro-Wilks test. To establish the significance of the differences observed, a Student's *t*-test was used for normally distributed data. For non-parametric data, we used the Mann-Whitney *U* test for independent samples or Wilcoxon test for related samples. A *P*-value of less than 0.05 was taken to denote statistical significance.

Data analysis was performed using the Analyse-it program for Microsoft Excel (Leeds, UK. See <http://www.analyse-it.com>) and SPSS statistics 19 for Windows (SPSS Inc., IBM, Somers, New York, USA).

RESULTS

Table 3 shows the means and standard deviations for each of the tests recorded. Mean stereopsis was better in the normal binocular vision group than the abnormal binocular vision group.

Repeatability Figure 1 shows the repeatability results obtained for the four stereoacuity tests in the subjects with normal binocular vision. Best repeatability was shown by the Frisby and Titmus tests since these returned the narrowest 95% agreement intervals (COR: ±13 and ±12s arc respectively). In clinical terms, this reflects little variation between the two sets of test scores. Figure 2 shows the repeatability results obtained for the four stereoacuity tests in the subjects with a binocular disorder. Best repeatability was

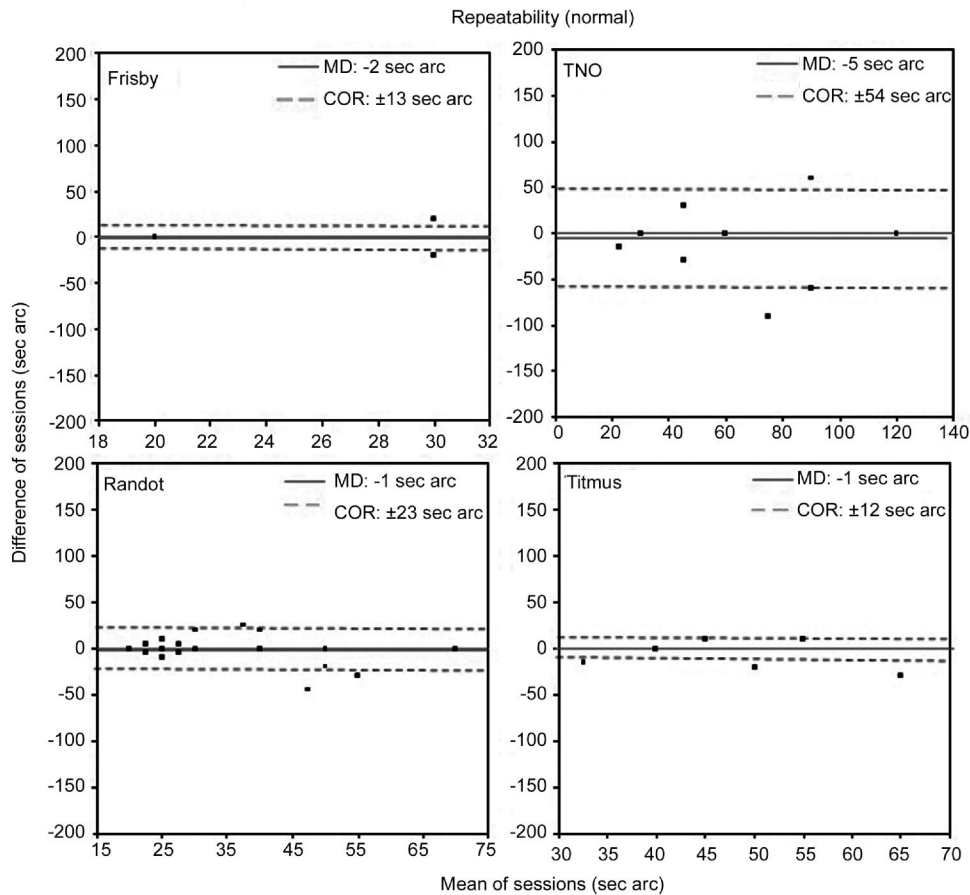


Figure 1 Bland–Altman plots of the repeatability of stereoacuity measurements taken in subjects with normal binocular vision
 The plot for the Frisby had only three points: (20, 0) 48 subjects; (30, 20) 1 subject; and (30, -20) 5 subjects. The solid line represents the averaged difference between final session and initial session measures (MD, mean difference). Dotted lines indicate the lower and the upper 95% limits of agreement (coefficient of repeatability, $COR=\pm 1.96\times SD$) (Wilcoxon test $P>0.05$).

shown by the Frisby test (MD=-13s arc and $COR=\pm 69$ s arc) and Randot test (MD=-14s arc and $COR=\pm 72$ s arc). In both subject groups, differences in the scores obtained in the two sessions for each of the methods (MD) were not significant ($P>0.05$).

Agreement Figure 3 shows the agreement values obtained among the different stereoacuity tests in the subjects with normal binocular vision. The TNO test showed the poorest agreement with the other tests, as indicated by the wide agreement interval ($COA=\pm 48$ s arc). Greatest agreement was shown between the Frisby and Titmus with a COA of ± 10 s arc. MDs were always significant ($P<0.001$).

Figure 4 shows the agreement values obtained among the different stereoacuity tests for the deficient binocular vision group. The TNO test showed poorest agreement with the other tests, as indicated by the wide agreement interval ($COA>\pm 225$ s arc). Agreement among the remaining tests was similar and COAs were close to ± 80 s arc. With the exception of Frisby versus Randot, differences in MDs were always significant ($P<0.05$).

Analysis of the Variance Mixed ANOVA revealed a main significant difference effect according to the method used ($F=43.34$; $P<0.0001$). However, no significant difference

emerged between the two measurement sessions ($F=3.70$; $P=0.06$). An interaction effect was detected between method and group ($F=14.70$; $P<0.0001$; Figure 5), which was significant for the Titmus ($P<0.001$) and TNO ($P=0.001$). The interactions session \times group ($F=0.98$; $P=0.3$), method \times session ($F=1.88$; $P=0.1$), and method \times session group ($F=2.04$; $P=0.1$) were all non-significant.

DISCUSSION

Repeatability The results of our study confirm our working hypothesis that the repeatability of each test was different and that for each test, repeatability was worse when tested in subjects with abnormal binocular vision. In those with normal vision, the repeatability values obtained in each test except the TNO can be considered good. However, the high COR shown in Figure 2 indicates that none of the 4 tests (especially the TNO) would be adequate for the follow up of patients with a binocular anomaly, since any change produced would be masked by the low reliability shown by the tests in this group of subjects. Although, the number of subjects could be a limitation of our results and it will be interesting to repeat the study with a higher sample size, especially in the group with binocular anomalies. The small number of subjects with strabismus ($n=5$) does not allow us

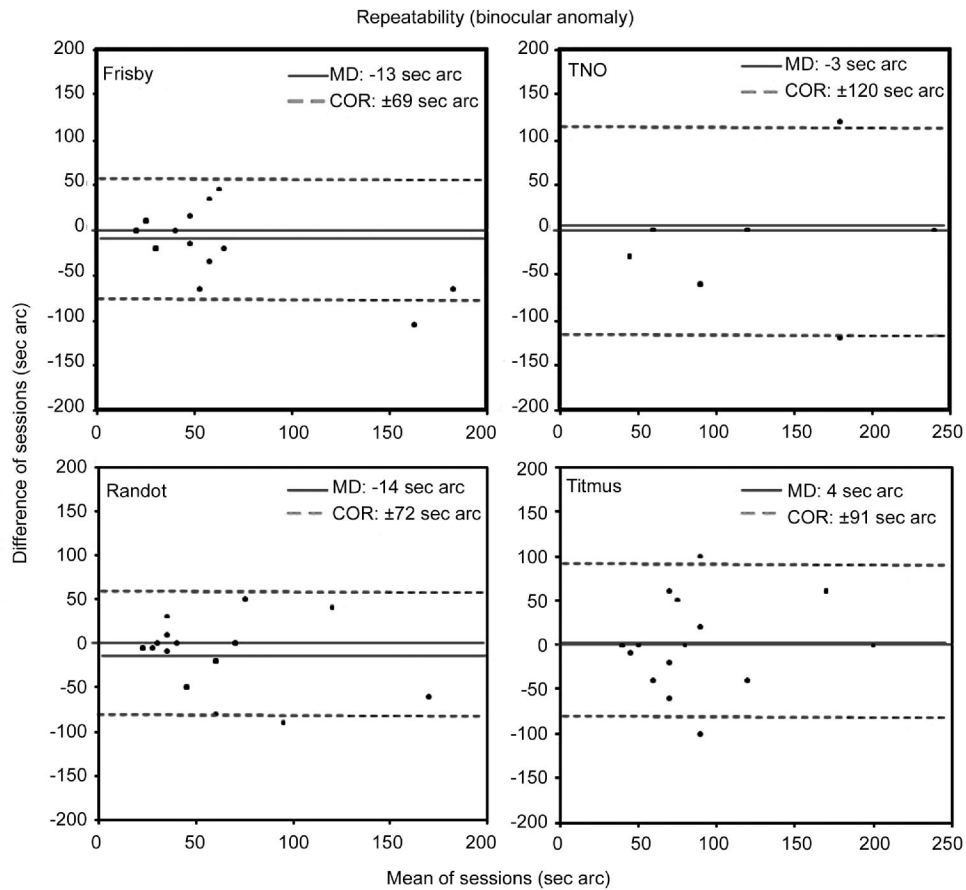


Figure 2 Bland–Altman plots of the repeatability of stereoacuity measurements taken in subjects with abnormal binocular vision
 The solid line represents the averaged difference between final session and initial session measures (MD, mean difference). Dotted lines indicate the lower and upper 95% limits of agreement (coefficient of repeatability, $COR = \pm 1.96 \times SD$) (Wilcoxon test $P > 0.05$).

to study the influence of the etiology of binocular anomaly over repeatability. But with a higher sample size, it could be interesting to study if the repeatability of the measurements could be different in severe amblyopia, light amblyopia, constant strabismus or intermittent squints.

MD between the 2 test scores was not significantly different from zero both in statistical and clinical terms, so it may be concluded that fatigue or learning effect did not influence the test scores.

Repeatability can be strongly affected by the step size of the tests (Table 3) as was observed here for the TNO test in both subject groups. Mean stereoacuity in subjects with normal binocular vision was close to 60s arc yet step size at this level is greater which explains the reduction in reliability. Moreover, these larger step sizes at higher thresholds could be the main reason why the COR was larger for subjects with abnormal binocularity in all the tests. Using several stereotests it could be easier to detect if a high stereoacuity change could correspond to a true change or in contrast, it can be attributed to the normal variability of tests associate to their lower repeatability in subjects with poor binocular vision. So in this kind of patients, we recommend the clinicians to measure the level of stereopsis using several tests to verify if there is the same tendency in the different

measurements.

Repeatability may also be affected by the ceiling effect [10]. In the normal binocular function group, stereoacuity values for the Frisby and Titmus tests were biased since this ceiling effect was produced because most subjects (89%) attained the lowest disparity measurable by the test in both sets of measurements. This threshold effect determined that many subjects showed zero difference between sessions. In contrast, in the Randot test, only 18% of subjects attained the maximum value in both sessions and in the TNO, no subject achieved the maximum value in both sessions. In the abnormal binocular vision group, no ceiling effect was produced and very few subjects attained the maximum value in both measurement sessions [Frisby 3 subjects (15%), TNO (0%), Randot (0%), Titmus 4 subjects (20%)].

Agreement Figures 3 and 4 illustrate the low level of agreement among tests (worse in subjects with a binocular abnormality), especially between the TNO and remaining tests. Marsh *et al* [20] proposed that differences between tests are determined by factors such as: a) the configuration and size of the disparity areas; b) shape of the figures; c) testing distances; d) red-green versus polarized glasses; e) size of figures; f) the ability of subjects to identify the disparate area. The circles tests have strong monocular cues of lateral

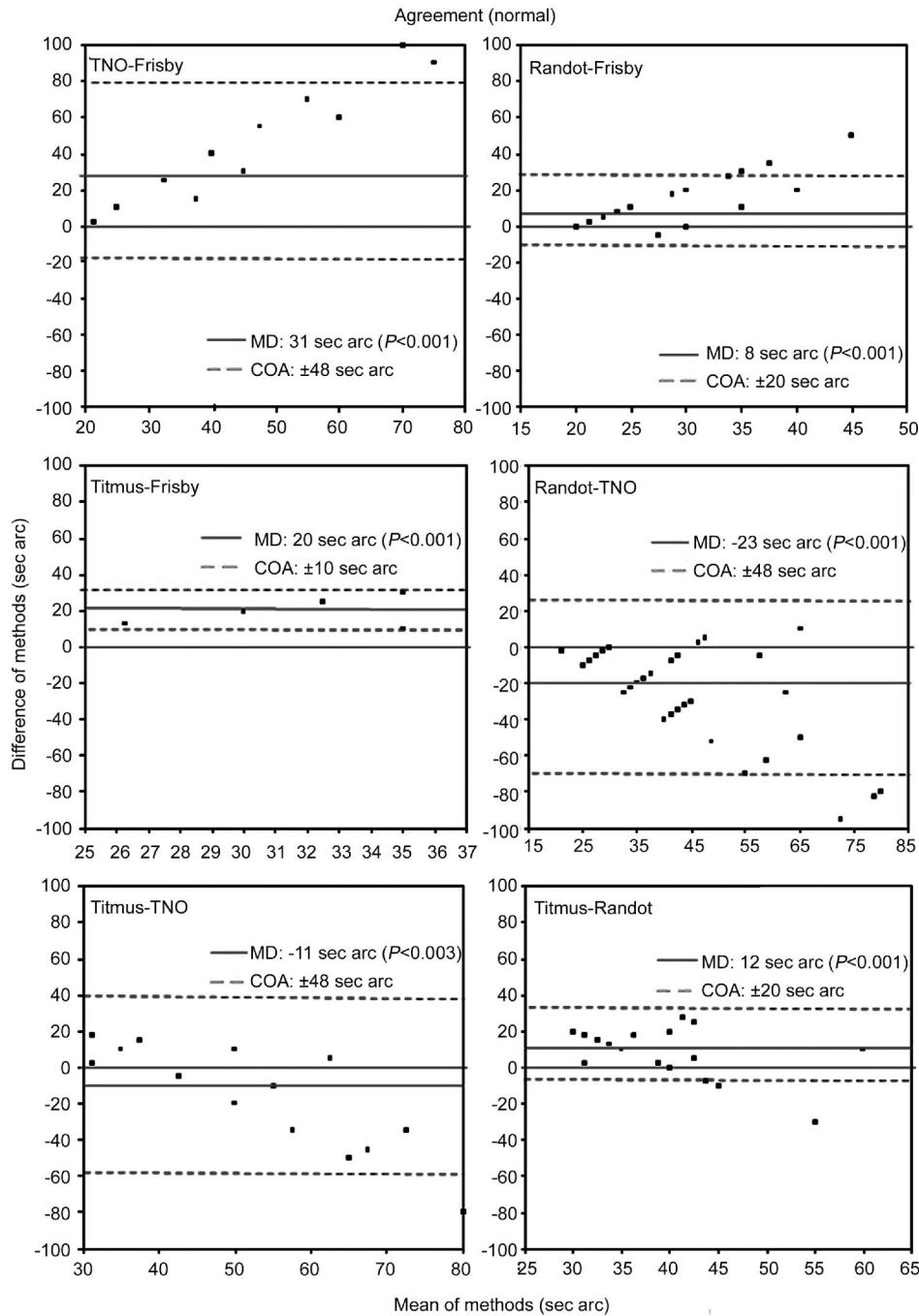


Figure 3 Bland–Altman plots of agreement between the tests used to measure stereoacuity in subjects with normal binocular vision The solid line represents the averaged difference between final session and initial session measures (MD, mean difference). Dotted lines indicate the lower and the upper 95% limits of agreement (coefficient of agreement, $COA = \pm 1.96 \times SD$) (Wilcoxon test P value).

displacement but only for the first 3 or 4 circles in the Titmus (1 or 2 in the Randot), so these monocular cues should not affect those subjects with thresholds of 100s arc or better in the Titmus test and 140s arc in the Randot test [12-14]. Therefore, the difference between the Titmus or Randot tests compared to TNO in subjects with normal binocular vision is unlikely due to the presence or absence of monocular cues. The results provided in Table 3 indicate that the TNO test was the most difficult to correctly complete with a mean stereoacuity threshold of 52s arc in subjects with normal binocular vision, and 158s arc in subjects with deficient binocular vision. The design of the TNO itself could explain

this greater difficulty [10]. It could be, however, that the random dot stereograms in TNO are more difficult to perceive because a global stereopsis target may be processed differently (neurally) than a contoured or a real depth stereopsis target [24-26]. Other contributing factors to the low agreement between the TNO and remaining stereotests are the worse repeatability of the TNO and the fact that it is the only test used here that requires the use of red/green filters. In anaglyph tests, investigators have found that luminous transmittance and contrast may differ significantly between the red and green filters. Such differences between the two eyes during testing may affect suppression tendencies and

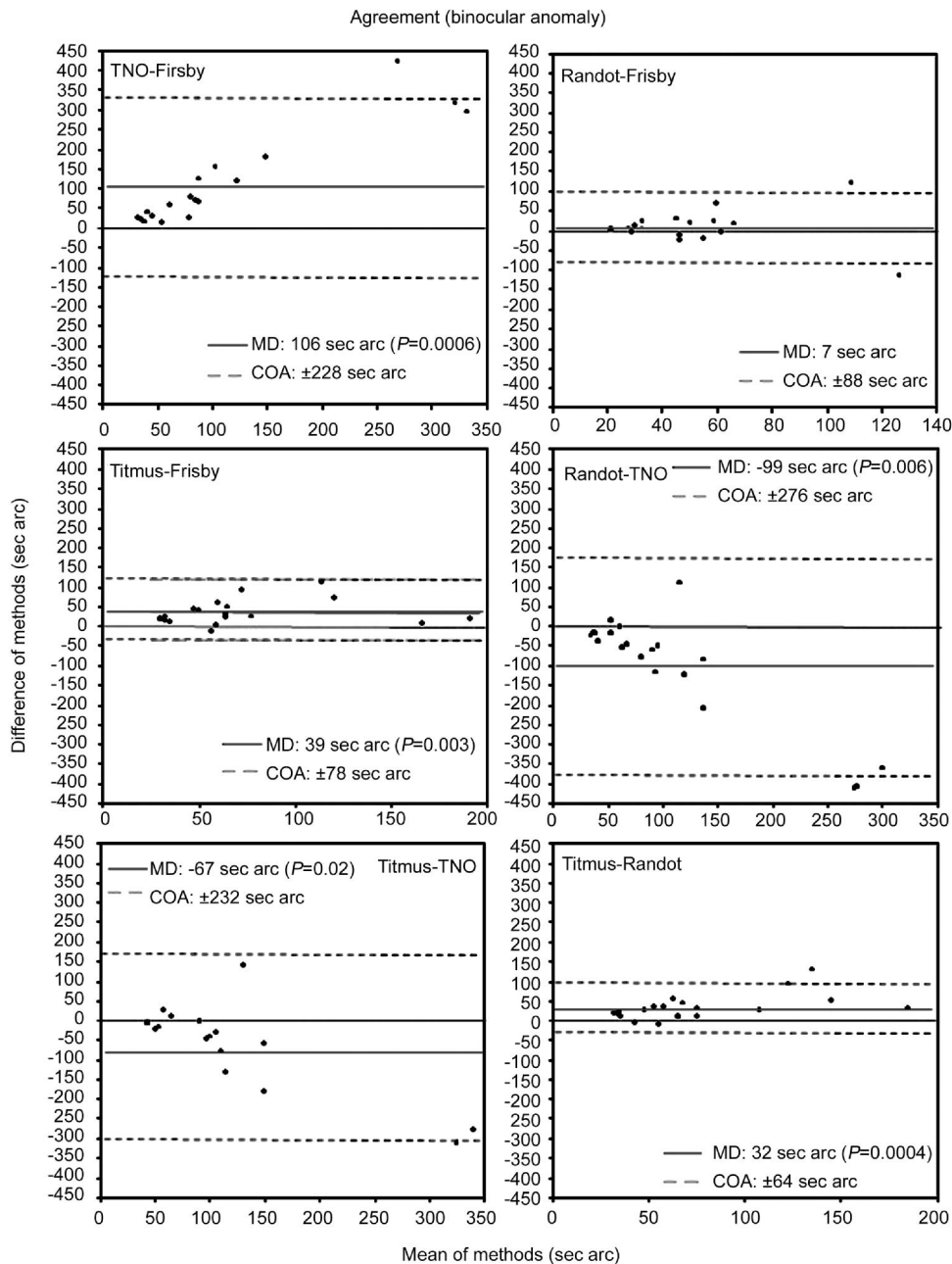


Figure 4 Bland–Altman plots of agreement between the tests used to measure stereoacuity in subjects with abnormal binocular vision. The solid line represents the averaged difference between final session and initial session measures (mean difference, MD). Dotted lines indicate the lower and the upper 95% limits of agreement (coefficient of agreement, COA=±1.96×SD) (Wilcoxon test *P* value).

may explain an overall reduced stereopsis in some subjects^[9,18,19].

Rosner and Clift^[27] compared stereoacuity measures obtained using the Frisby and TNO tests in 20 subjects aged 23 to 37y (mean 27.4y) with good binocular vision, and obtained a Pearson's correlation coefficient of $r=0.73$ ($P<0.001$); the TNO tended to yield lower stereoacuties than the Frisby. Hall^[10] measured the stereo-thresholds of 67 normal binocular subjects aged 18 to 24y using the tests Titmus, Frisby, TNO and two-needle test and found that there was low, but significant, correlation between the TNO and Frisby ($r=0.35$; $P<0.05$) and TNO and Titmus ($r=0.25$; $P<0.02$).

If we compare the tests from a practical perspective, the Frisby test needs no artificial means to assess stereoacuity,

while the TNO, Randot and Titmus rely on red-green or polarized glasses to produce binocular disparity, which could cause partial dissociation or introduce retinal rivalry^[8,9]. However, to change the degree of disparity in the Frisby test, impractical changes in the observation distance are needed, while the remaining tests are administered at a fixed distance of 40 cm. As an advantage of the Frisby test, expected replies cannot be memorized or learnt which could happen in the other tests, since their targets appear at a specific position. In summary, since it is presently not easy to select the best test for use in clinical practice, there is a need for improved designs that bring together all the desired features in a single test.

According to the findings of our study, we would recommend

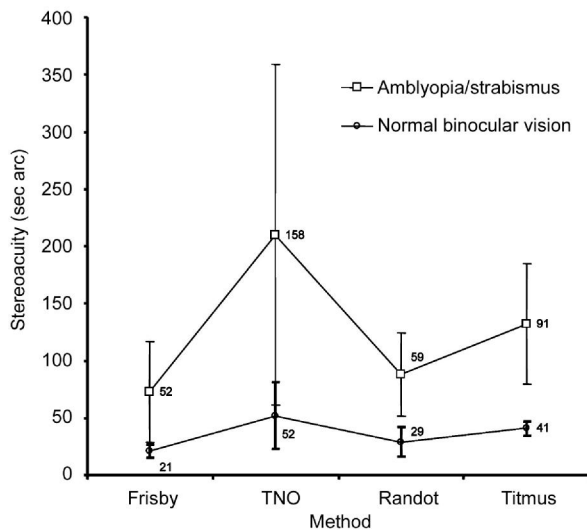


Figure 5 Mean stereoaucities recorded using the different tests in session 1 for the two groups of subjects (showing normal or abnormal binocular vision) Error bars represent 1 SD.

adding lower disparity levels to the Frisby and Titmus tests to reduce the ceiling effect. In addition, if intermediate disparity levels were introduced in the TNO, this would improve its repeatability and allow discrimination of the severity of a binocular disorder.

In conclusion, the repeatability of stereoaucity measures obtained using four different tests was fairly good in subjects with normal binocular vision with the exception of the TNO. In subjects with poor binocular vision the repeatability of the four tests was low thus they are inadequate for detecting small changes in visual performance such as changes in an individual over time.

The reduced agreement detected between the tests indicates they cannot be used interchangeably. Poorest agreement was observed between the TNO and remaining tests.

ACKNOWLEDGEMENTS

Foundation: Supported by the Direction General of Universities and Research (DGUI) of the Community of Madrid (No.CCG10-UCM/BIO-4889)

Conflicts of Interest: Antona B, None; Barrio A, None; Sanchez I, None; Gonzalez E, None; Gonzalez G, None.

REFERENCES

- 1 Kuang T, Hsu W, Chou CK, Tsai SY, Chou P. Impact of stereopsis on quality of life. *Eye (Lond)* 2005;19(5):540-545
- 2 Fawcett SL, Wang YZ, Birch EE. The critical period for susceptibility of human stereopsis. *Invest Ophthalmol Vis Sci* 2005;46(2):521-525
- 3 Saladin JJ. Stereopsis from a performance perspective. *Optom Vis Sci* 2005;82(3):186-205
- 4 Richardson SR, Wright CM, Hrisos S, Buck D, Clarke MP. Stereoaucity in unilateral visual impairment detected at preschool screening: outcomes from a randomized controlled trial. *Invest Ophthalmol Vis Sci* 2005;46(1):150-154
- 5 Lee SJ, Kim WS. Factors influencing stereoaucity levels after surgery to correct unilateral developmental cataracts in children. *Int J Ophthalmol*

- 2013;6(3):331-336
- 6 Holmes JM, Leske DA, Hatt SR, Brodsky MC, Mohny BC. Stability of near stereoaucity in childhood intermittent exotropia. *J AAPOS* 2011;15(5):462-467
- 7 Stewart CE, Wallace MP, Stephens DA, Fielder AR, Moseley MJ; MOTAS Cooperative. The effect of amblyopia treatment on stereoaucity. *J AAPOS* 2013;17(2):166-173
- 8 Simons K. A comparison fo the Frisby, Random-Dot E, TNO, and Randot circles stereotests in screening and office use. *Arch Ophthalmol* 1981;99(3):446-452
- 9 Westheimer G. Clinical evaluation of stereopsis. *Vision Res* 2013;90:38-42
- 10 Hall C. The relationship between clinical stereotest. *Ophthalmic Physiol Opt* 1982;2(2):135-143
- 11 O'Connor AR, Birch EE, Anderson S, Draper H; FSOS Research Group. The functional significance of stereopsis. *Invest Ophthalmol Vis Sci* 2010;51(4):2019-2023
- 12 Cooper J, Warshowsky J. Lateral displacement as a response cue in the Titmus Stereo Test. *Am J Optom Physiol Opt* 1977;54(8):537-541
- 13 Fawcett SL, Birch EE. Validity of the Titmus and Randot circles tasks in children with known binocular vision disorders. *J AAPOS* 2003;7(5):333-338
- 14 Reinecke RD, Simons K. A new stereoscopic test for amblyopia screening. *Am J Ophthalmol* 1974;78(4):714-721
- 15 Fricke TR, Siderov J. Stereopsis, stereotests, and their relation to vision screening and clinical practice. *Clin Exp Optom* 1997;80(5):165-172
- 16 Manny RE, Martinez AT, Fern KD. Testing stereopsis in the preschool child: is it clinically useful? *J Pediatr Ophthalmol Strabismus* 1991;28(4):223-231
- 17 Martínez FM, Pons AM. Fundamentos de visión binocular. Valencia: Universitat de Valencia; 2004
- 18 Simons K, Elhatton K. Artifacts in fusion and stereopsis testing based on red/green dichoptic image separation. *J Pediatr Ophthalmol Strabismus* 1994;31(5):290-297
- 19 Yamada T, Scheiman M, Mitchell GL. A comparison of stereopsis testing between red/green targets and polarized targets in children with normal binocular vision. *Optometry* 2008;79(3):138-142
- 20 Marsh WR, Rawlings SC, Mumma JV. Evaluation of clinical stereoaucity test. *Ophthalmology* 1980;87(12):1265-1272
- 21 Webber AL, Wood JM, Gole GA, Brown B. The effect of amblyopia on fine motor skills in children. *Invest Ophthalmol Vis Sci* 2008;49(2):594-603
- 22 Adams WE, Leske DA, Hatt SR, Holmes JM. Defining real change in measures of stereoaucity. *Ophthalmology* 2009;116(2):281-285
- 23 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-310
- 24 Fawcett SL. An evaluation of the agreement between contour-based circles and random dot-based near stereoaucity tests. *J AAPOS* 2005;9(6):572-578
- 25 Gantz L. Are local and global stereograms processed by separate mechanisms? Houston: University of Houston; 2009
- 26 Tanaka H, Ohzawa I. Neural basis for stereopsis from second-order contrast cues. *J Neurosci* 2006;26(16):4370-4382
- 27 Rosner J, Clift GD. The validity of the Frisby stereotest as a measure of precise stereoaucity. *J Am Optom Assoc* 1984;55(7):505-506