Research Protocol

Determining the consistency of scores across two measurements of the visual system: Test-retest reliability

Principal investigator

Ian Shrier, MD, PhD – McGill University

Co-investigators

Stephanie Long (MSc candidate – McGill University)

David Tinjust, PhD - APEXK

Suzanne Leclerc, MD, PhD – INS – Université de Montréal

Study Location

Institut National du Sport (INS)

Date

May 15, 2017

ABREVIATIONS LIST

INS – Institut National du Sport

PTVS – Post Traumatic Visual Syndrome

Rationale and study purpose

Concussion is a mild form of traumatic brain injury that affects 1.6 to 3.8 million people each year in the United States [1]. However, the exact incidence is difficult to determine due to underreporting by patients [2]. Concussion is defined as a set of pathological reactions leading to direct damage of the brain, which may occur from a direct or indirect blow to the head [3]. Common symptoms fall in three domains: (1) physical signs and symptoms (e.g. headache, dizziness, nausea, balance problems, fatigue, light and noise sensitivity), (2) cognitive deficits (e.g. difficulty concentrating, decreased reaction time, memory problems), and (3) emotional reactions (e.g. depression, anxiety) [4, 5]. Visual complaints, such as blurred vision, eye fatigue, and difficulties sustaining attention, are particularly common in adult concussions [4, 6]. Recently, we have witnessed a shift in the focus of some research on concussion to include more detailed investigation of vestibular and visual symptoms [2].

One common visual deficiency that can mimic symptoms of concussion is convergence insufficiency. Convergence insufficiency (CI) is a common binocular vision disorder [7] often associated with a wide range of symptoms including headaches, poor concentration, slow reading, blurred vision, diplopia, and eyestrain [8-10]. Various treatment options are commonly prescribed such as home-based therapy, pencil push-ups, office-based vision therapy, and orthoptic treatment [11]. Orthoptic exercises have been in use for over 70 years and represent an established therapy for CI [12]. Despite frequent use, the effectiveness of these treatments had not been examined until recently by the Convergence Insufficiency Treatment Trial (CITT) group which compared the effectiveness of different treatment methods on CI. It was determined that orthoptic exercises resulted in significantly greater improvement of symptoms in children aged 9 to 17 years as compared to other common treatments [11].

To date, tests for CI and other orthoptic visual dysfunctions have yet to be validated or tested for reliability in healthy adults. Reliability is an important measurement property for all measures of health [13]. The Consensus-based Standards for the Selection of Health Measurements Instruments (COSMIN) has defined reliability as “the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions”. There are several types of reliability that are of interest in measurement of health outcomes, such as inter-rater reliability, intra-rater reliability, internal consistency, and test-retest reliability. In terms of health measures, such as the orthoptic tests, they must be valid and sensitive to the system of interest (i.e. visual system) in order to be used clinically to guide patient management. Presently, it is difficult to quantify the validity of these orthoptic measures due to a lack of a “gold-standard” in this domain [14], thus, test-retest reliability must be established [15]. Therefore, the purpose of this research is to determine the test-retest reliability of seven orthoptic measures across two independent measurements. As these tests have applicability in many health conditions (e.g. concussion, etc), this study has the potential to contribute to improved management of symptoms related to the visual system.

Study design and methods

The overall study design is a test-retest reliability study of seven different orthoptic tests in healthy participants aged 16 to 35 years. In order to determine if there is clinically useful agreement and consistency between the results when the orthoptic tests are repeated, each participant will be assessed at two time points.

At each time point, participants will be evaluated with seven orthoptic measures:

- (T1) at participant's earliest convenience
- (T2) 1 week following their initial visit

Each of the seven orthoptic measures will be administered to each participant by the same clinician (DT). To minimize the risk that the clinician will remember the scores of any participant, we will set up each testing day so that the clinician sees up to five participants in a random order. At their next and final visit, all the same participants will return to be evaluated, but their order will be altered from their initial visit. To further decrease the likelihood of clinician memory affecting the responses, a research assistant will record the test results as reported verbally by the clinician. After a participant has completed his/her two visits, a research assistant will enter the data into the proprietary software which will provide the clinician with their official "chart" that is required as part of good management practice.

Test-retest reliability studies have used varying durations between testing intervals ranging from consecutive days to a few weeks [15, 16]. We have selected a seven-day interval between testing times because of scheduling efficiency and potential carry-over effects of the orthoptic tests. First, although these orthoptic measures are standard tests that have been used in the public, certain individuals may experience mild symptoms that typically resolve in minutes to a few hours. Therefore, if we repeated the test within a few hours, this may lead to inaccurate measurements due to residual effects of the test, itself. Although we could conduct the second test within 24 to 48 hours, we are concerned that scheduling might be difficult as participants may only have free time at particular times/days of the week (e.g. each Monday afternoon). Thus, we have chosen to have the tests repeated exactly seven days later. Finally, the time interval must be one in which it is unlikely that a participant may change. For instance, with an interval that is too long (e.g. one month) there is an increase likelihood of the participant having changed significantly from his/her initial measurement. We believe a one week interval is short enough that participants are very unlikely to change unless a major event occurs (e.g. car accident). Therefore, we believe a one-week interval represents the appropriate balance for avoiding a carry-over effect, probability of change in the participant, and feasibility.

Study population

Our target population is healthy adults aged 16 to 35 years. This study is being conducted as part of a larger project examining the correlation of the orthoptic measures with symptoms of concussions; therefore, this population has been selected to match that target population of young adults with concussion symptoms to ensure applicability of findings to the patient population in our larger study. We will exclude participants who have conditions or are taking treatments that could affect the results of the orthoptic tests.

Inclusion Criteria

- Participants must be 18 years or older
- Informed consent to participate in the study

Exclusion Criteria

- Participants with strabismus or a history of treated strabismus
- History of migraine or know neurologic disorders
- Use of myorelaxants, as they interfere with orthoptic testing
- We will exclude any participants who are currently taking medications for depression, anxiety or other psychological condition.

Although a random sample is usually the preferred sampling method, it is difficult to achieve in a timely manner and can be very costly [17]. Further, in the context of our study a random sample is not necessary because we are interested in examining the test-retest reliability across two measurements in the same participant. Therefore, we do not expect relevant heterogeneity across any known demographic measure that might be used as a criterion for representative samplings. In addition, as previously mentioned, this study is conducted in conjunction with another study examining these orthoptic tests in a target population of young adults; therefore, estimates derived from a similar sample (i.e. young adults) will be applicable to the group we apply it to.

Thus, we will use a convenience sampling of potential participants. Volunteers will be recruited from McGill University and the Montreal area through word-of-mouth and via personal contacts.

Measures

Demographic Variables

We will record demographic data relevant to our study in order to appropriately describe the population and explore if these factors modify the test-retest correlations. We will include the year of birth, sex, highest level of education achieved (secondary school, CEGEP, university), the use of corrective glasses for vision problems, and occupation.

Orthoptic Measures of the Visual System

Most clinicians understand that eye function can be tested for acuity, field of vision and convergence using standard tests available at any clinician's office. However, these tests can only detect gross deficiencies. Optometrists have access to specialized equipment to detect more subtle visual deficiencies. The field of orthoptics uses even more advanced equipment to measure very small deviations in several domains of the visual system. All orthotic measures will be conducted by a single clinician. A detailed description of the 7 orthoptic measures follows:

Gross stereoscopic acuity: (range 0-15 arc seconds)

Our binocular vision allows us to see in three dimensions (3D), or more simply, to see depth. In this test, seated participants wearing 3D glasses are shown images. Inability to see depth or 3D will cause images to appear as points instead of objects. The objects are presented in different stages, with each stage requiring them to discriminate different levels of depth perception. The test is scored in optical units, with a range of 0 to 15 arc-seconds. The maximum score corresponds to the level where the last object was identified.

Convergence measured by "motor punctum proximum": (cm)

When an object is moving towards our eyes, they symmetrically converge in order to maintain focus. However, there is a point at which our eyes no longer symmetrically converge (point of convergence or motor punctum proximum). This test measures the distance (cm) between the bridge of the nose and point of convergence in seated participants as an object is moved closer to the head.

Binocular fusion with convergence: (diopters, prism convergence units)

This test measures how well someone can adapt to challenges in focusing light on the retina. There are two almost identical tests. One test occurs with an object placed at 3m from the seated participant, and the other with an object at 30cm from the seated participant. Light from an image is passed through a prism. This is analogous to moving the image further away from the body. In response, the eyes must diverge (separate) to focus on the object, just as they would if the image actually moved away from the body. Different prisms are used to create increasing challenges. The score for these tests is simply the maximum amount of prism convergence (dioptres, noted on the prism, as one would note diopters on eye glasses) that the seated participant can accommodate at 3m and at 30cm.

Saccadic movements or oculomotor capacity: (Score = bad, medium, good)

A light appears on the screen and the participant move their eyes to fix on the object. While they eyes adjust, they will temporarily cover small distances until they achieve a fixed focus. These are called saccadic movements. Lights appear and disappear, in different locations on the screen, at a rate of 100 per minute, lasting 2 minutes. The test result is scored by the evaluator based on a global impression over the entire 2 minutes, with 3 separate sub scores on an ordinal scale for quality (bad, medium, good), for synchronization (bad, medium, good) and saccadic correction (many corrections, few corrections, no corrections). The three sub scores are combined into an overall score according to our industry partner's (Apexk) proprietary algorithm.

Anatomic oculomotor deviation: (diopters, prism convergence units)

This test measures the eyes' natural deviation (heterophoria) and also allows the detection of strabismus. In strabismus, anatomic deviation is evident and the person's dominant eye is looking at you, but the "lazy/deviated" eye is not. In heterophoria anatomic deviation is not visible to the naked eye and the deviation has to be triggered by covering in sequence, one eye at a time, to trigger the deviation. There are two identical tests: one occurs with an object placed at 3m from the seated participant (far vision), and the other with an object at 30cm from the seated participant (near vision). In this test, seated participants focus on an object. These movements can be seen by the orthoptist. The orthoptist covers/uncovers eyes to trigger movements and uses a prism to cancel the movement. The prism that achieves this cancellation is the measure of anatomic deviation. The rating of the prism that achieves this cancellation is considered the score for this test, with one score for the object placed at 3m and another score for the object placed at 30cm. Participants with strabismus are excluded in our study because strabismus is a contra-indication to post-concussion visual training, which is part of our larger study and thus, patients with strabismus do not represent our target population.

Convergence fusional proximum: (diopters, prism convergence units)

This test is similar to (2) above. When an object is moving closer to our head, the eyes symmetrically converge. When the object is moved beyond the participant's ability to converge, the participant will start to see two images (double vision). This test measures the distance between the bridge of the nose and point where double vision (cm) occurs in seated participants as an object is moved closer.

Binocular fusion with divergence: (diopters, prism convergence units)

This is the same test as (3), except that the prisms diverge the light and the participant has to converge their eyes to maintain focus. The score for these tests is simply the maximum amount of prism divergence (diopters, noted on the prism, as one would note diopters on glasses) that the seated participant can accommodate at 3m, and at 30cm.

Data collection and analysis

At the point of recruitment, a research assistant will describe what is needed of the participant in the study and then the participant will provide informed consent. At the first visit, a research assistant will collect demographic information of each participant. At each visit, the seven orthoptic tests will be administered to the participant by the single clinician (DT) as outlined above. Once each participant has completed their two testing periods, the data will be entered into the patient file by a research assistant. Following, the data will be exported into a comma-separated file for analysis using statistical software. The clinician conducting the orthoptic tests will not be involved in the data extraction or analysis phases.

Analysis plan

The main objective is to evaluate the agreement and consistency between measurements of the visual system taken at two different times to determine the test-retest reliability of the orthoptic measures. If the tests are perfectly reliable, we would expect the value of the first test to equal the value of the second test for each participant across the range of values for all individuals. The difference between the first and second test for each participant is called the within-group (each participant represents a group) variance. The difference across individuals is the between-group variance. We will assess reliability using the intra-class coefficient (ICC) which measures the between-group variance divided by the total variance (sum of between- and within-group variance). If the within-group variance is 0 (first and second measures are identical), then the ICC = 1 (between-group variance / (between-group variance + 0)). In addition to the ICC, limits of agreement will be estimated using the paired data and illustrated via Bland-Altman plots [18].

Sample Size Calculation

Sample size planning is based on the ICC which is our primary outcome. We consider a clinically relevant value in the study population to be ICC $\geq$ 0.75. Because the measured ICC will vary from sample to sample, we need to consider how precise the estimate of ICC should be. We consider the lower bound of clinical acceptability to be an ICC of 0.5 [19]. Therefore, if the true ICC is 0.75, our estimate precision (95% confidence interval width / 2) must be within 0.25 of the true estimate (0.75 to 0.5). If we include 20 participants, the precision of our ICC estimate

will be ± 0.20 [19]. This will be slightly more precise than is required, but allows us to account for some potentially missing data or minor technical problems that may occur during the study.

Safety, Confidentiality and Ethical considerations

Adverse Events

Participants will undergo the orthoptic visual tests. These tests stress the visual system and can lead to mild test-related symptoms. The clinical experience of our orthoptist is that these symptoms (namely dizziness, mild diplopia or mild headache) occur in a minority of patients, last minutes to hours and resolve by the next day. We consider this an acceptable risk given these tests are used as part of routine clinical management for many diseases, including concussion.

Any participant with persistent increase in symptoms following the test will be contacted the next day to ensure resolution, and will be referred to their treating physician if symptoms persist.

Serious Adverse Events

As we are only testing healthy adults at two time points, we do not expect any serious adverse events through participation in this study. We have excluded all cases of strabismus (by self-report, and by clinician examination before the tests begin) because it carries a risk of permanent diplopia (double vision) caused by visual orthoptic testing. No other major complications are reported for these tests.

Premature withdrawal from the study

Any consenting participant has the right to withdraw from the study at any time, for any reason he/she deems relevant. A premature withdrawal from the study will not result in harm to the patient, nor interfere with their normal course of treatment.

Confidentiality

All collected data will be held anonymously. Participants will be assigned numbers for identification, which will only be accessible to researchers directly involved in the project. Participant addresses and contact numbers will be stored separately from other data and will not be shared with industry partners or any third party. The electronic file will be password protected. Other study data recorded on paper will be in a locked filing cabinet. The information will be kept for a minimum of 7 years following publication of the results by the investigators.

Ethical considerations

The study will be conducted according to ethical principles stated in the declaration of Helsinki (2013) and will respect any directives determined by the IRB of McGill University. Ethics approval will be obtained before initiating the study. All participants will be able to satisfy all requirements for informed consent; consent forms will take into consideration the well being of all participants, with emphasis on their proper medical follow up and treatment regardless of research objectives, with respect for free-will, human rights and privacy.

Compensation

Participant will receive a financial compensation of 25$ per visit for their participation to cover transportation and miscellaneous costs.

References

1. Leddy, J.J., et al., *Rehabilitation of Concussion and Post-concussion Syndrome.* Sports Health, 2012. **4**(2): p. 147-54.
2. Ventura, R.E., L.J. Balcer, and S.L. Galetta, *The Concussion Toolbox: The Role of Vision in the Assessment of Concussion.* Semin Neurol, 2015. **35**(5): p. 599-606.
3. McCrory, P., et al., *Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport, Zurich, November 2012.* J Athl Train, 2013. **48**(4): p. 554-75.
4. Master, C.L., et al., *Vision Diagnoses Are Common After Concussion in Adolescents.* Clin Pediatr (Phila), 2016. **55**(3): p. 260-7.
5. Lynch, J.M., et al., *The gaming of concussions: a unique intervention in postconcussion syndrome.* J Athl Train, 2015. **50**(3): p. 270-6.
6. Ventura, R.E., et al., *Diagnostic tests for concussion: is vision part of the puzzle?* J Neuroophthalmol, 2015. **35**(1): p. 73-81.
7. Letourneau, J., N. Lapierre, and A. Lamont, *The relationship between convergence insufficiency and school achievement.* Am J Optom Physiol Opt, 1979. **56**: p. 18-22.
8. Scheiman, M., *Treatment of symptomatic convergence insufficiency in children with a home-based computer orthoptic exercise program.* J aapos, 2011. **15**(2): p. 123-4.
9. Lavrich, J.B., *Convergence insufficiency and its current treatment.* Curr Opin Ophthalmol, 2010. **21**(5): p. 356-60.
10. Huston, P.A. and D.L. Hoover, *Treatment of symptomatic convergence insufficiency with home-based computerized vergence system therapy in children.* J aapos, 2015. **19**(5): p. 417-21.
11. Convergence Insufficiency Treatment Trial Study, G., *Randomized clinical trial of treatments for symptomatic convergence insufficiency in children.* Arch Ophthalmol, 2008. **126**(10): p. 1336-49.
12. Horwood, A. and S. Toor, *Clinical test responses to different orthoptic exercise regimes in typical young adults.* Ophthalmic Physiol Opt, 2014. **34**(2): p. 250-62.
13. Polit, D.F., *Getting serious about test-retest reliability: a critique of retest research and some recommendations.* Qual Life Res, 2014. **23**(6): p. 1713-20.
14. Hachana, Y., et al., *Test-retest reliability, criterion-related validity, and minimal detectable change of the Illinois Agility Test in Male Team Sport Athletes.* Journal of Strength and Conditioning Research, 2013. **27**(10): p. 2752-2759.
15. Broglio, S.P., et al., *Test-retest reliability of computerized concussion assessment programs.* Journal of Athletic Training, 2007. **42**(4): p. 509-514.
16. Geber, C., et al., *Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study.* Pain, 2011. **152**(3): p. 548-56.
17. Endacott, R. and M. Botti, *Clinical research 3: Sample selection.* Accid Emerg Nurs, 2007. **15**(4): p. 234-8.
18. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet 1986. **1**(8476): p. 307-310.
19. Watson, P.F. and A. Petrie, *Method agreement analysis: a review of correct methodology.* Theriogenology, 2010. **73**(9): p. 1167-79.