

The Reliability and Validity of the Four Square Step Test in Patients with Hip Osteoarthritis Pre and Post Total Hip Replacement.

Martha Batting MRes, BSc (Hons) Physiotherapy ¹⁻²

Karen L Barker PhD, FCSP ^{1,3}

¹Nuffield Orthopaedic Centre, Physiotherapy Research Unit, Windmill Road, Headington, Oxford, OX3 7HE, UK; ²University of Southampton, Health Sciences, 12 University Road, Southampton, SO17 1BJ, UK. ³Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, University of Oxford, Oxford, OX3 7LD.

Twitter Handle: @PRUOxford

Corresponding author: Martha Batting, Physiotherapy Research Unit, Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Foundation Trust, Windmill Road, Headington, OX3 7HE. Tel: 01865 737526 Email - Martha.moore@ouh.nhs.uk

Manuscript word count: 3086

Funding

The corresponding author is in receipt of an NIHR funded masters in Clinical Research at the University of Southampton.

The work was also supported by the NIHR Biomedical Research Centre, Oxford (OxBRC).

28 **Acknowledgments**

29 We would like to acknowledge the contribution of Bronagh Walsh and Sean Ewings
30 from the University of Southampton for their general supervision and statistical
31 support.

32 We would also like to thank Erin Hannink from the Physiotherapy Research Unit for
33 her help in conducting the reliability assessments.

34

35 **Ethical approval**

36

37 Ethical approval was obtained for the study by the Office for Research Ethics
38 Committees Northern Ireland (Reference: 16/NI/0049).

39

Abstract

Objective: To determine the validity and inter-rater and intra-rater reliability of the Four Square Step Test (FSST) in assessing gait performance, balance and physical function for patients with hip osteoarthritis before and after total hip replacement (THR).

Design: Observational, repeated measures.

Setting: A specialist orthopaedic hospital.

Participants: Fifty-eight participants with Moderate to Severe Hip Osteoarthritis scheduled to receive primary hip replacement within four months from recruitment.

Main Outcome Measure: Time to complete the FSST, time and steps to complete the Figure of 8 Walk Test (F8W) and Berg Balance Scale score (BBS).

Results: Overall, on 95% of occasions, the difference between the FSST measurements for Intra-rater reliability pre THR was within 3.1s and post THR was 1.3s. For Inter-rater reliability the difference for the FSST was 1.6s on 95% of occasions. Concurrent validity was assessed and it was found that the FSST correlated highly with the F8W test ($r = .7$, $p < 0.001$) and moderately with the BBS ($r = .6$, $p < 0.001$). Only one participant was rated as moderate risk of falls on the BBS with the rest scoring low and again only one participant failed to complete the F8W test. This is in contrast to the FSST in which 21 people failed pre-operatively.

Conclusions: The FSST is a valid and reliable measure of multi-directional stepping speed and balance giving a more informative measure of gait performance over the F8W and BBS that is feasible for use in a clinical population of patients both before and after hip replacement.

Key words: Outcome Assessments, Gait, Balance, Osteoarthritis, Total Hip Replacement and Reproducibility of Results.

Contribution of the paper:

- The FSST is a valid and reliable measure when compared with the F8W and BBS
- The ability of the FSST to highlight specific limitations in balance and gait speed of participants over the reference standards make it a more appropriate measure for use in the total hip replacement population

The Reliability and Validity of the Four Square Step Test in Patients with Hip Osteoarthritis Pre and Post Total Hip Replacement.

Introduction

Our living communities are complex environments that continue to challenge our balance control and walking patterns to help avoid obstacles, change directions, carry loads, and plan a path to a destination based on prior cognitive maps¹. Most measures of walking ability and static balance however provide limited insight to everyday mobility, as they are constrained to situations of minimal environmental challenge². Patients with hip OA have been shown to have reduced toe clearance, impaired obstacle avoidance, and gait and balance disorders³. The same is true post operatively where patients with a total hip replacement (THR) are reported to have a slower gait speed and shorter stride length⁴ as well as reduced postural balance⁵.

One measure that does challenge obstacle avoidance and change of direction, is the Four Square Step Test (FSST). The test requires a person to step forwards, backwards and sideways over obstacles in a specified sequence.

Developed by Dite and Temple⁶ the FSST has been shown to provide a measure of dynamic standing balance and mobility. Both balance and mobility are the most consistently identified risk factors linked to falls^{7,8,9}. Most falls occur during movement itself, with trips and slips making up a large proportion of falls^{10,11,12} indicating that the ability to take a rapid step may help prevent some of these falls. The FSST could therefore be used to identify an increased risk of falls in those who

Four Square Step Test

have slower test times. Furthermore its' small testing space and low equipment requirements make the test an appealing outcome measure for the clinical setting. The FSST has been shown to discriminate between healthy and non-ambulatory populations ($P<.01$) displaying high interrater reliability (intraclass correlation coefficient [ICC] =0.99) and retest reliability (ICC=0.98)⁶. It has also been shown to be a valid and reliable measure to assess balance deficits in community dwelling older adults, patients with Parkinson's Disease, Huntington's Disease, Multiple Sclerosis, Vestibular disorders, post stroke, following transtibial amputation and knee pain¹³. The reliability of the FSST has also been gauged in participants with Hip Osteoarthritis (OA)¹⁴; however the validity for this population has yet to be determined.

It is therefore pertinent that the FSST be validated in this population so that effective rehabilitation to address any deficits can occur. To provide a measure of concurrent validity the FSST was evaluated against the Figure of 8 Walk test (F8W) and the Berg Balance scale (BBS), which are both measures of dynamic balance that are reliable and already used within the orthopaedic setting^{15,16}. It is also essential to determine the reliability of this population both pre and post THR as in clinical situations multiple assessors may need to use the test or the same assessor will need to repeat the FSST to measure change.

The objectives of this study were to explore the FSST's validity by comparing its agreement with the F8W ~~test~~ and BBS in patients with hip OA ~~both~~ before and after THR, and to establish the inter-rater and intra-rater reliability of the test by comparing the limits of agreement both within and between two separate assessors.

Methods

Participants

Eighty participants were recruited from a specialist orthopaedic hospital with a total of 58 participants completing all three study assessments. To ensure a sufficient sample size, the recommendations from the Consensus-based standards for the selection of health measurement instruments guidelines were followed¹⁷. Individuals were eligible if they had moderate to severe hip OA, were due to undergo a primary THR and were aged over 55. Individuals were not eligible if they were found to have severe cardiovascular or pulmonary disease, severe dementia or communication difficulties, Rheumatoid Arthritis or a neurological condition that would affect their ability to take part in the balance tests, registered with a visual impairment or were due to have further planned treatment on the hip within the next four weeks. We enrolled a consecutive series of participants onto the study. Ethical approval was obtained for the study ([Office for Research Ethics Committees, Northern Ireland](#)~~Reference: 16/NI/0049~~). All participants were informed about the purpose and

Four Square Step Test

procedures of the study and gave written consent. A flow diagram of participants throughout the study can be found in figure 1.

Measures

Standardised protocols were developed according to the procedures stated in each measures original validation study. Background information about an individual's relevant medical history and physical characteristics, including height, weight and any previous lower limb surgery were collected. Participants also completed the Oxford Hip Score (OHS), a 12-item questionnaire that measures patient reported outcomes of hip function¹⁸ and the Activities Specific Balance Confidence Scale (ABC) which is a subjective measure of confidence in performing various ambulatory activities¹⁹.

Index Test - Four Square Step Test

Four 90cm canes were placed in a square resting flat on the floor. Each square is numbered one to four. The bottom left square is labelled number one and the participant was asked to stand in this facing forwards. The patient was required to step into each square in the following sequence: 2, 3, 4, 1, 4, 3, 2, and 1 (See figure 2). The patient was instructed to try to complete the sequence as fast as possible without touching the sticks. Both feet had to make contact with the floor in each

Four Square Step Test

square and the participant must face forward during the entire sequence. One practice trial was carried out followed by two trials of the test and the best time was taken.

Reference Standard - Figure of 8 Walk Test

Participants were instructed to stand mid-way between two cones (1.52m apart) and to face towards one of the cones. An unmarked 0.6m boundary around the testing area was noted by the assessor to check for walking accuracy (See figure 3). Participants were asked to walk at a comfortable, self-selected speed and direction following a figure-of-8 path around the cones, stopping when they had returned to their starting position. The timer began when the participant moved to take their first step and the timer stopped once both feet were back at the starting position. The time and number of steps taken were recorded. The assessor noted whether the participant kept within the unmarked 0.6m boundary and also rated the participant on a three-item walking smoothness score which evaluated stopping, change of pace or hesitation.

Reference Standard - Berg Balance Scale

The participant was asked to complete 14 different balance tests including sit-to-stand, stand-to-sit, unsupported standing, unsupported sitting, transfers, standing with eyes closed, standing with feet together, reaching forward, picking up an object

Four Square Step Test

from the floor, turning to look behind, 360 degree turn, step onto stool, tandem stance and single leg stance. Each task was marked from zero to four with four being the best score. The test was totalled out of 56 with higher scores indicating lower falls risk.

Procedures

The FSST was conducted a total of five times over three separate visits for each participant both before they had their surgery and then again post-operatively. At the initial assessment all three outcome measures were completed once and carried out in a pre-set sequence picked at random. To establish inter-rater reliability, a second assessment was conducted by two separate assessors; assessor-A, who carried out the initial assessment, and assessor-B. Both assessor-A and assessor-B carried out the FSST test one time and allowed the patient a minute's rest in between the tests. Both assessors were blinded to each other's result. The second assessment was scheduled 7 to 35 days after the first assessment in order to minimise learning effects. The participants all then had their surgery and the final assessment was conducted by assessor-A six months following THR. The FSST was carried out twice with a minutes rest in between each test to establish intra-rater reliability after THR.

Statistical Analysis

Descriptive statistics were used to characterize the study population. Means and standard deviations (SD) were calculated for continuous variables; frequencies and

Four Square Step Test

percentages for ~~continuous or~~ binary variables. An analysis of intra-rater and inter-rater reliability for the FSST was performed for all three assessment time points using SPSS statistical package version 24 and based on the fastest FSST for each assessment. Bland and Altman plots with 95% limits of agreement were produced to provide a clinically meaningful picture of the size and range of the raters scores²⁰. In the plots, the differences between each pair of measurements are plotted against the mean of each pair of measurements. If the differences follow a standard normal distribution, then 95% of the differences will lie between 2SD. ~~Intra-class correlation coefficients (ICCs) with 95% Confidence Intervals (CI) were calculated using SPSS statistical package version 24 and based on the fastest FSST test time in each assessment using an absolute-agreement model to determine reliability of the FSST. Bland and Altman plots were produced to provide a more clinically meaningful picture of the size and range of rater differences²⁰. It is expected in a reliable assessment that the mean difference between measures will be close to zero and that at least 95% of the individual differences should be within a range of two standard deviations from the mean difference.~~

To investigate concurrent validity between the FSST, F8W and BBS a Pearson's correlation coefficient was calculated. Scatter plots were also produced to show the relationship between each measure. The outcomes produced in each of the three tests are different. Consequently, when we measured limits of agreement between differences and averages for the three measures, we standardized the raw measures of time and points by creating z-scores based on the sample's distribution. Gait/step speed and balance are the outcomes of interest, which will be compared between measures.

Four Square Step Test

Results

A total of 58 participants were included in the final review for the study from the initial eighty recruited (See figure 1). The participant characteristics are reported in table 1.

On average participants were assessed 18 days (± 8.83 SD) between the first and second appointments and 6 months (± 0.90 SD) between surgery and their final assessment. A summary of the outcome measures scores is provided in table 2.

Reasons and frequency of test fails are also noted in table 3.

Intra-rater and Inter-rater reliability

Overall, on 95% of occasions, the difference between the measurements for Intra-rater reliability before THR was within 6s. This was notably lower when measured after THR with a difference of 1.3s on 95% of occasions. Finally for Inter-rater reliability the difference was also found to be narrower at 1.6s on 95% of occasions (Figure 4).

There were three notable outliers presented within figure 4A which when removed produced upper limits of agreement to 3.5, lower limits of agreement to -3.2 and a mean of 0.14. When we recalculated the difference between measurements for intra-rater reliability before THR with these outliers removed we found a narrower agreement of 3.1s on 95% of occasions.

Four Square Step Test

~~Good intra-rater reliability was found between assessment one and two (ICC= .874) with a 95% CI (0.787, 0.926). Additionally, excellent inter-rater reliability at assessment two was found (ICC= .976) with 95% CI (0.943, 0.988). Likewise, excellent intra-rater reliability at the 6 month follow up assessment was shown (ICC= .934) with 95% CI (0.884, 0.962).~~

~~Bland and Altman plots indicate good intra-rater reliability for assessor A at assessments one and two (mean difference = 0.142, limits of agreement = 3.456 to -3.172; Figure 4A) and at assessment three post-surgery (mean difference = 0.240, limits of agreement = 2.015 to -1.535; Figure 4C). Similarly, good inter-rater reliability was found for assessors A and B (mean difference = 0.374, limits of agreement = 2.393 to -1.645; Figure 4B).~~

Concurrent Validity

The FSST and F8W test times were strongly correlated $r=0.729$, $p < 0.01$. Moderate negative correlations were also found between the FSST and BBS $r=-0.620$, $p < 0.01$. Scatter plots produced show a strong-positive linear relationship between the FSST and F8W (Figure 5A). A strong-negative relationship was also seen between the FSST and BBS (Figure 5B).

Figure 6 depicts Bland and Altman plots for the sample showing limits of agreement between the FSST, F8W and BBS z-scores. It confirms the correlation findings to show good agreement between the FSST, F8W and BBS with a mean difference close to zero and almost all observations spread within narrow limits of agreement

Four Square Step Test

~~for the F8W (1.6 to -0.8 z scores) and BBS(3.9 to -2 z scores). When plotting z scores between both measures the mean difference was approximately zero, variability fairly constant and almost all individual observations are spread within the narrow limits of agreement for F8W (1.18 to -1.18 z-scores) and BBS (2.844 to -2.844 z-scores).~~

Discussion

This has been the first study to assess the validity and reliability of the FSST before and after hip replacement surgery. The study found the FSST to be a valid and reliable measure of balance for this population. Both intra-rater and inter-rater reliability have been demonstrated with ~~excellent to good ICC values being reported~~ and relatively small variation in scoring both between and within assessors A and B. The FSST was found to strongly correlate to the F8W test and moderate negative correlations were reported between the FSST and BBS.

It is important to be able to determine a patient's stepping speed and ability to help assess their risk of falls and provide a clinically meaningful measure of balance to ensure appropriate and timely rehabilitation. This is particularly true in patients who have hip OA as they are known to have reduced stepping speed and obstacle clearance²¹. The ability to step at speed in multi-directions is required in everyday walking when responding to forward, backward and lateral perturbations, for example when walking in a busy street. It was noted by the assessors that pre-operatively participants struggled to step sideways in particular due to the known reduction in hip abductor strength²². Following surgery many patients reported

Four Square Step Test

finding it easier to take a side step, which may in part explain the reduction in mean FSST step test time with a mean difference of 2.9 seconds post operatively. This correlates with the known increase in gait speed after hip replacement surgery²³ ~~Excellent Intra-rater and Inter-rater reliability have been found for the FSST in other test populations that correspond to our findings including in community dwelling adults⁷ (ICC = .98-99), post-stroke²⁴ (ICC = .94-99), in Parkinson's disease²⁵ (ICC = .74-88) and those with hip osteoarthritis¹⁵ (ICC = .83-86).~~ We reported a slower mean test time for the FSST (13.73s first assessment, 14.2 second assessment and 11.13 third assessment) compared to the FSST in patients with hip OA¹⁵ (8.97-8.56s). These differences may in part be explained by the mean age of their participants, who were significantly younger. Our study also had a larger sample size with a greater age range of participants which is more representative of the current THR population. However, there exists a dearth of literature assessing the FSST in musculoskeletal conditions making it difficult to confidently conclude that our findings are reflective of the population as a whole.

The Bland and Altman plots produced reinforce our findings that the FSST is a reliable measure showing relatively narrow limits of agreement. Three significant outliers were highlighted when analysing the agreement within assessor A as they had a difference of more than 7 seconds between the two measurement times. If these outliers were not removed a variation of up to 6 seconds is shown for 95% of occasions. ~~removed from the pre-operative sample of 58 participants as their scores between assessments were more than 10 seconds apart which was notably larger than the rest of the data. Had these outliers not been removed the upper and lower limits of agreement would have shown a variation of approximately six seconds for~~

Four Square Step Test

~~intra-rater reliability and four seconds for inter-rater reliability.~~ Clinically, variations of up to six seconds would not be a reliable predictor of someone's falls risk, given that previous studies have found cut off scores as low as 9.68s²⁴⁶ for the FSST.

Additionally, when looking at two of the outliers we see that their second assessment time is markedly slower than the first for both assessors A and B which may indicate that their mobility had declined between assessment 1 and 2 and it is not due to measurement error alone. One of the outliers was seen to produce faster scores at assessment 2 for both assessors and this could also indicate a potential learning effect that we have not accounted for within the study.

It is also important to discuss the length of time between testing and how this may account for variations in intra-rater reliability pre and post operatively. The pre-operative re-test times were on average 18 days apart, however at assessment three the two re-test scores were measured just one minute apart. This would account for the 1.8s difference within rater scores from before to after THR as there are fewer confounding variables if the measures are taken within a minute of each other compared to multiple days.

When comparing the results of the FSST to the BBS scores we see that those who have a higher overall score are quicker at completing the FSST. For the BBS those at low risk of falls are scored from 41-56, moderate risk of falls 21-40 and high risk of falls 0-20²⁵⁷. However only one participant in our cohort was rated as moderate risk of falls scoring 37 on the BBS where 16 of the participants had reported a fall in the past year, indicating that perhaps the BBS was not sufficiently challenging enough and therefore sensitive to changes in this population of patients.

Four Square Step Test

299

300 Furthermore, the FSST is seen to correlate well with F8W scores, with those who
301 have quicker F8W test times scoring faster FSST times. Our mean F8W test time of
302 9.99s and step count of 15.8 was also similar to that shown in community dwelling
303 older adults of 10.49s and 17.51 steps, with a similar mean age of participants
304 sampled¹⁶. Again however the F8W only showed that one person was unable to stay
305 within the boundary of the test, which indicates a fail. This is in contrast to the FSST
306 which showed 21 people failed pre-operatively would lead to a conclusion that the
307 FSST is more receptive to limitations in participants balance and gait speed. Some
308 may argue these findings may demonstrate that the FSST is not suitable in this
309 population due to the failure rate in completing the test, however the improvement of
310 failure rate post-operatively to just 4 participants indicates that it is accurately
311 measuring participants balance and giving a greater indication of improvement over
312 the other measures.

313

314 In our cohort only five participants carried out the FSST with an aid pre-operatively,
315 dropping to just three requiring an aid to complete the test post-operatively. Given
316 the relatively small number of walking aid users it remains difficult to draw
317 conclusions as to whether the use of a walking aid would impact on the FSST time.

318

319 The FSST is a valid and reliable measure when compared with the F8W and BBS.
320 Furthermore the challenging nature of the FSST highlights those participants who
321 have poor balance or gait speed and can identify the potential causes of this such as
322 poor obstacle clearance or reduced motor planning. This provides a clear advantage

Four Square Step Test

for the use of the test over the reference standards making it a more appealing and accurate measure for use in the hip replacement population.

Study limitations

There are some limitations to this study, which may influence how these results are interpreted. We had initially recruited 80 patients to our baseline cohort; there was a relatively high dropout rate of participants after the first assessment at 16.25%, which although not outside the expected range for observational studies, when combined with the further 11.25% who were unable to complete all three assessments reduced our study sample to just 58 participants. This can significantly affect the generalisability of our findings given that those participants who were unable to complete the assessments usually differ to the other subjects in the sample. Furthermore, it may have been of clinical interest to retest those subjects who were unable to perform the FSST pre-operatively after their THR to see if there were any improvements in their ability to carry out the test.

There was also a large variation in time from assessment one to two and only a one minute rest between re-testing at assessments 2 and 3. The unpredictability of surgery dates and fitting around participant schedules meant that it was not possible to see each patient at the same three week time point. This resulted in some patients being seen much sooner for the retest than others and may affect intra-rater ~~and inter-rater~~ reliability findings. The shorter time period for re-testing at assessment 2 and 3 also makes it difficult to conclude that there were no learning or fatigue effects from the inter-rater and post-operative intra-rater reliability. As well as this the results

Four Square Step Test

may not be generalizable to current clinical practice where typically testing intervals will be longer.

Finally, the purpose of this study was not to assess the predictive value of the FSST in measuring falls risk, nor was it to assess the responsiveness of the FSST to change. Further studies could look to provide information about the cut off scores for falls risk in this population and the minimally clinically important difference so that the test can effectively be used as a measure both pre and post THR.

Conclusions

The FSST is a valid and reliable measure of multi-directional stepping speed and balance giving a measure of gait performance that is feasible for use in a clinical population of patients both pre and post THR. The FSST also provides more informative and clinically useful data on balance and mobility when compared to the F8W test and BBS. The study is limited in generalising the pre THR inter-rater and post THR intra-rater reliability findings due to the shorter timing interval between measurements which may not account for learning or fatigue effects nor reflect current practice of retesting periods. Further research needs to be conducted to assess the predictive validity of the test in this population and it's responsiveness to change if it is to provide meaningful outcomes of participant's physical function and falls risk.

Four Square Step Test

373

374 **Ethical approval**

375

376 Ethical approval was obtained for the study by the Office for Research Ethics

377 Committees Northern Ireland (Reference: 16/NI/0049).

378

379 **Funding**

380

381 See title page.

382

383 **Conflicts of interest**

384 The authors declare they have no conflict of interests.

385

386

387

388

389

390

391

392

393

394

395

396 |

397

398

399

References

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

1. A Patla. Mobility in complex environments: implications for clinical assessment and rehabilitation. *Neurol Rep* 2001;25:82–90
2. Frank J and Patla A. Balance and mobility challenges in older adults: implications for preserving community mobility. *American Journal of preventative medicine* 2003;25(3):157-163
3. Ng C and Tan M. Osteoarthritis and falls in the older person. *Age and Ageing* 2013;42(5):1-6
4. Sliwinski MM, Sisto SA, Batavia M, Chen B, Forrest GF. Dynamic stability during walking following unilateral total hip arthroplasty. *Gait Posture* 2004;19:141–147
5. Nantel J, Termoz N, Centomo H, Lavigne M, Vendittoli PA, Prince F. Postural balance during quiet standing in patients with total hip arthroplasty and surface replacement arthroplasty. *Clin Biomech* 2008;23:402–407
6. Dite W and Temple V A clinical test of stepping and change of direction to identify multiple falling older adults. *Archives of Physical Medicine and Rehabilitation* 2002;83(11):1566-1571
7. Studenski S, Duncan PW, Chandler J et al. Predicting Falls: The Role of Mobility and Nonphysical Factors. *Journal of the American Geriatrics Society* 1994;42:297-302.
8. Anacker SL and Di Fabio, RP. Influence of Sensory Inputs on Standing Balance in Community-Dwelling Elders with a Recent History of Falling. *Phys Ther* 1992;72: 575-581
9. Altug F, Isik E and Cavallak U. Reliability and Validity of Four Step Square Test in Older Adults. *Turkish Journal of Geriatrics* 2015;18:151-155.
10. Berg WP, Alessio HM, Mills EM and Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. *Age and Ageing* 1997;26(4):261-8
11. Blake AJ, Morgan K, Bendall MJ et al. Falls By elderly people at home: Prevalence and associated factors. *Age and Ageing* 1988;17(6):365-72
12. Hill K, Schwarz J, Flicker L and Carroll S. Falls among healthy, community-dwelling, older women: a prospective study of frequency, circumstances, consequences and prediction accuracy. *Australian and New Zealand Journal of Public Health* 1999;23(1):41-8
13. Moore M and Barker K. The validity and reliability of the four square step test in different adult populations: a systematic review. *BMC Systematic Reviews* 2017;6:187
14. Choi YM, Dobson F, Martin J, Bennell KL and Hinman RS. Interrater and Intrarater reliability of common clinical standing balance tests for people with hip osteoarthritis. *Phys Ther* 2014;94(5):696–704
15. Hess R, Brach J, Piva A and VanSwearingen J. Walking Skill can be assessed in Older Adults: Validity of the Figure-of-8 Walk Test. *Journal of Physical Therapy* 2010;90:89-98

Four Square Step Test

16. Downs S, Marquez J and Chiarelli P. The Berg Balance Scale has high intra and inter-rater reliability but absolute reliability varies across the scale: A systematic Review. *Journal of Physiotherapy* 2013;59:93-99
17. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM and de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* 2012;21(4):651-7
18. Dawson J, Fitzpatrick R, Carr A and Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78(2):185-90
19. Powell LE and Myers AM. "The Activities-specific Balance Confidence (ABC) Scale." *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 1995;50(1):28-34
20. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology* 2011;64:96-106
21. Watelain E, Dujardin F, Babier F, Dubois D and Allard P. Pelvic and lower limb compensatory actions of subjects in an early stage of hip osteoarthritis. *Arch Phys Med Rehabil* 2001;82(12):1705–1711
22. Zacharias A, Pizzari T, English DJ, Kapakoulakis T and Green RA. Hip abductor muscle volume in hip osteoarthritis and matched controls. *Osteoarthritis Cartilage* 2016;24:1727–1735
23. Van den Akker-Scheek I, Stevens M, Bulstra SK, Groothoff JW, van Horn JR and Zijlstra W. Recovery of gait after short-stay total hip arthroplasty. *Arch Phys Med Rehabil* 2007;88(3):361–367
- ~~24. Blennerhassett JM and Jayalath VM. The four square step test is a feasible and valid clinical test of dynamic standing balance for use in ambulant people poststroke. *Arch Phys Med Rehabil* 2008;89(11):2156–61~~
- ~~25. McKee KE and Hackney ME. The four square step test in individuals with Parkinson's disease: association with executive function and comparison with older adults. *Neuro Rehabil*. 2014;35(2):279–89~~
- 26-24. Duncan RP and Earhart GM. Four square step test performance in people with Parkinson disease. *J Neurol Phys Ther* 2013;37(1):2–8
- 27-25. Berg K, Wood-Dauphinee S, Williams JI and Maki, B. Measuring balance in the elderly: Validation of an instrument. *Can. J. Pub Health* 1992;2:7-11

Four Square Step Test

Table 1. Characteristics of participants

	n	mean	SD	(min – max)
Age (years)	58	70.62	7.08	56 – 94
BMI (kg/m ²)	58	28.3	4.90	20.2 – 40.6
ABC Scale score ¹	55	69.82	20.9	17-99
OHS questionnaire ²	56	21.38	8.15	5 – 40

¹Excludes 3 participants because of incomplete ABC forms (missing data).

² Excludes 2 participants because of incomplete OHS forms (missing data).

*ABC = Activites specific Balance Confidence Scale

OHS = Oxford Hip Score

	n	%
BMI (kg/m ²)		
Normal weight (18.00 – 24.99)	12	20.7
Overweight (25.00 – 29.99)	27	46.5
Obese (30 or more)	19	32.8
Gender		
Female	32	55.2
Male	26	44.8
Previous Lower Limb Surgery		
Yes	31	53.4
No	27	46.6
Other musculoskeletal conditions		
Yes	22	37.9
No	36	62.1
Falls history in previous year		
Yes	16	27.6%
No	42	72.4%

Four Square Step Test

508
509
510
511
512
513

Table 2. Outcome measurements summary

	n	Mean (SD)	95%CI
First assessment (only assessor A)			
Four Square Step Test (seconds)	58	13.73(5.55)	12.2 – 15.1
Figure of 8 walk test (seconds)	58	9.99 (3.41)	9.09 – 10.88
Figure of 8 walk test steps	58	15.8 (4.09)	14.80 – 16.95
Berg Balance Scale total score	58	52.62 (3.68)	51.71 – 53.64
Second assessment			
Assessor A			
Four Square Step Test (seconds)	57	14.20 (6.51)	12.07 – 15.93
Assessor B			
Four Square Step Test (seconds)	55	13.60 (5.62)	12.07 – 15.14
Follow up assessment (Only assessor A)			
Four Square Step Test (seconds) time 1	58	11.13 (3.19)	10.29 – 11.97
Four Square Step Test (seconds) time 2	58	10.76 (3.02)	9.97 – 11.56
	n	Percentage	
Four Square Step Test fails			
Assessment 1	7	12.1%	
Assessment 2	14	24.1%	
Assessment 3	4	6.89%	
Figure of 8 Walk Test boundary fails	n		
Assessment 1	1	1.7%	
Berg Balance Scale score breakdown			
Low risk of falls (41-56)	57	98.3%	
Medium risk of falls (scores 21-40)	1	1.7%	
High risk of falls (scores 0-20)	0	0%	

514
515
516
517
518
519
520
521
522
523
524

Four Square Step Test

Table 3. FSST reason for test fails

	First assessment	Second assessment; assessor A	Second Assessment; Assessor B	Follow up assessment; Trial 1	Follow up assessment; Trial 2
Failed test due to wrong sequence	4	4	1	1	1
Failed test due to touched sticks	3	5	3	2	1
Failed test due to loss of balance	0	1	0	0	0
Total fails	7	10	4	2	2

Four Square Step Test

555

Figure(s) 1

Figure 1.

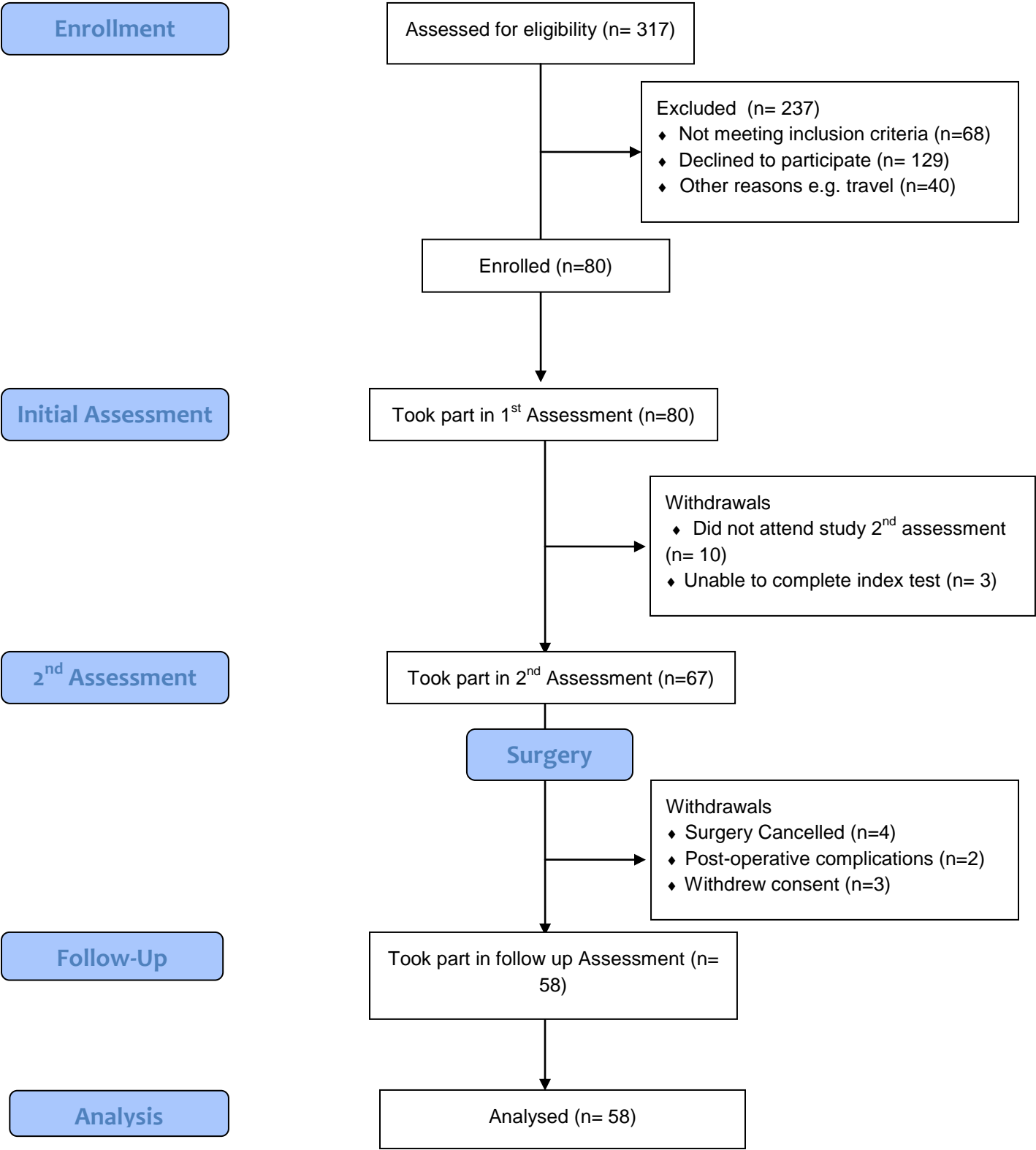


Figure 2.

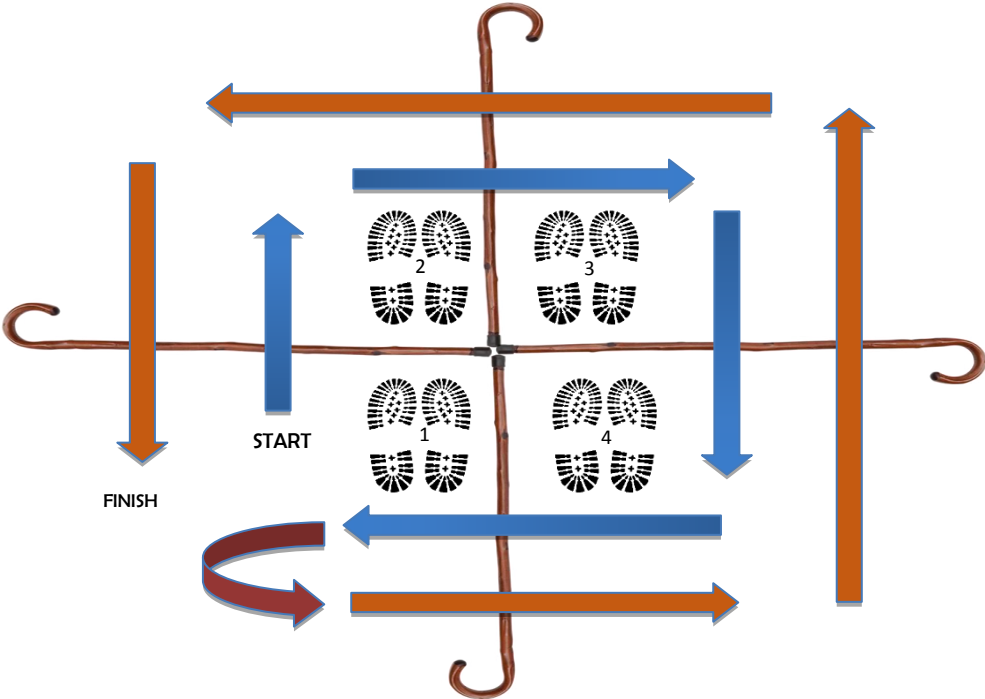
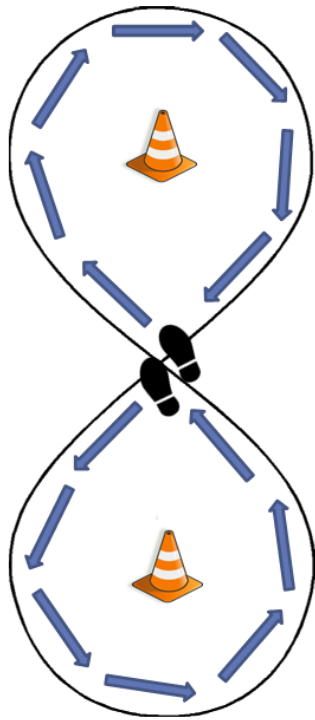


Figure 3.



Figure(s) 4

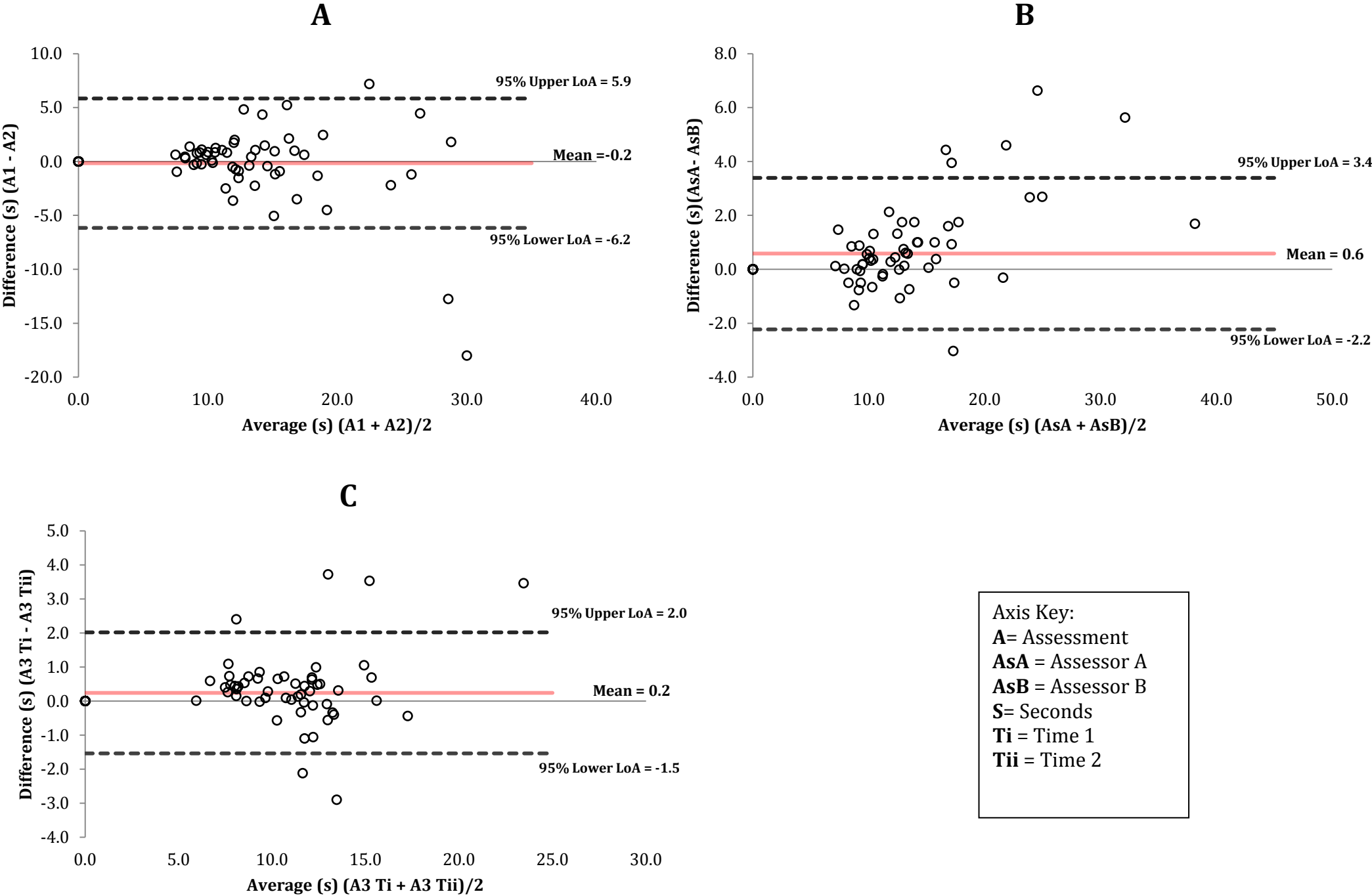
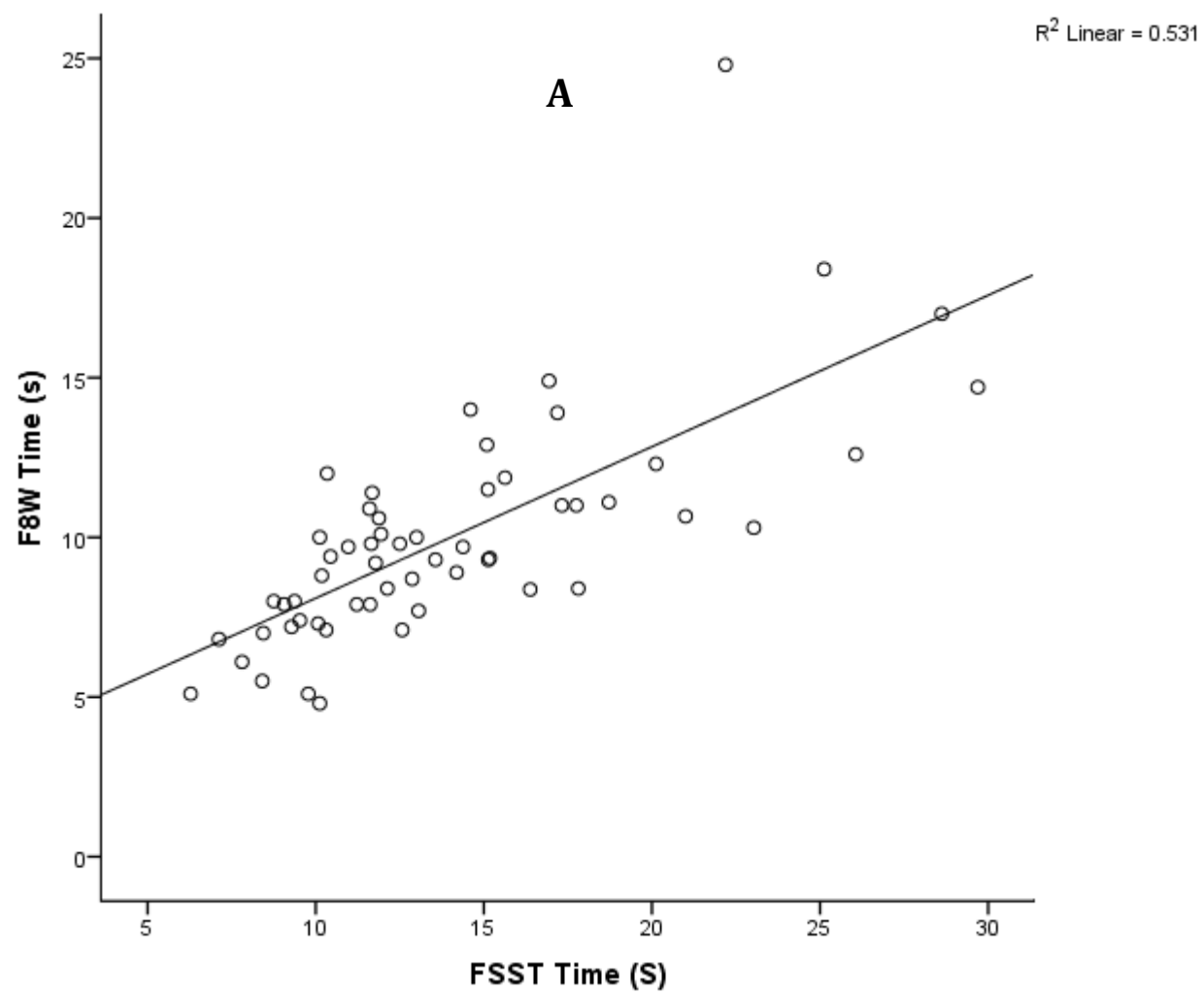


Figure 5.



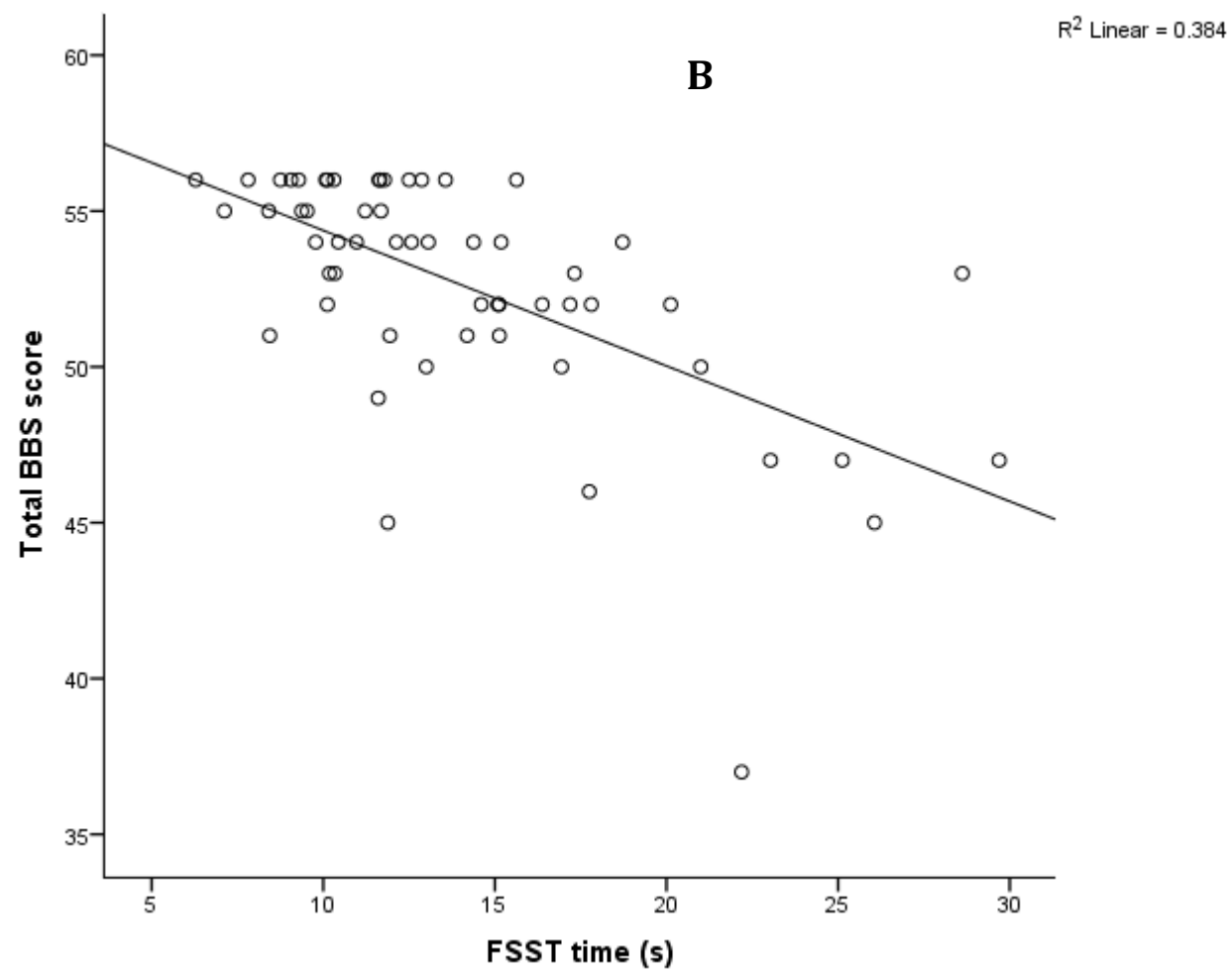


Figure 6.

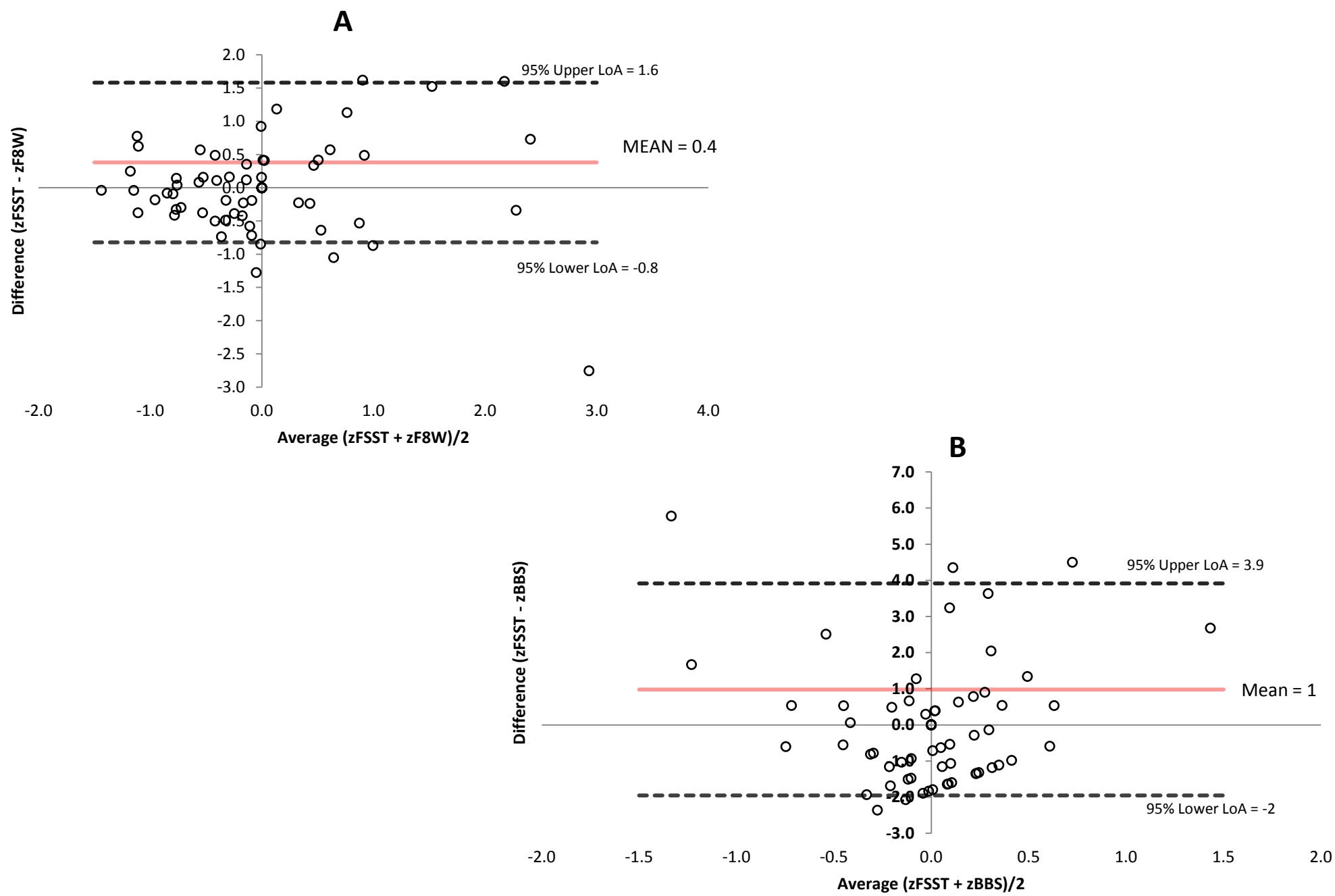


Figure 1. Flow of participants through study

Figure 2. Four Square Step Test Set-up

Figure 3. Figure of 8 Walk Test Diagram

Figure 4. Reliability Bland and Altman plots (4A shows Intra-rater reliability between the same assessor A for assessment 1 & 2, 4B shows Inter-rater reliability between assessor A and B at assessment 2 and 4C shows Intra-rater reliability for assessor A at the 6 month follow up assessment 3).

Figure 5. Scatter plots showing correlation between the Four Square Step test (FSST) and the Figure of 8 Walk test (F8W) (Figure 5A) and Berg Balance Scale (BBS) (Figure 5B) from assessment 1

Figure 6. Bland and Altman plots showing z-scores (6A shows difference in score between the Four Square Step Test (FSST) and Figure of 8 Walk Test (F8W), 6B shows difference in score between the FSST and Berg Balance Scale (BBS)).