2012

The Fit-HaNSA Demonstrates Reliability and Convergent Validity of Functional Performance in Patients with Shoulder Disorders

Prajyot Kumta

Joy C. MacDermid
McMaster University

Saurabh Mehta
Marshall University, mehtas@marshall.edu

Paul W. Stratford
McMaster University

Recommended Citation

Kumta, P., MacDermid, J. C., Mehta, S. P., & Stratford, P. W. (2012). The FIT-HaNSA demonstrates reliability and convergent validity of functional performance in patients with shoulder disorders. Journal of Orthopaedic & Sports Physical Therapy, 42(5), 455-464.

The Fit-HaNSA Demonstrates Reliability and Convergent Validity of Functional Performance in Patients with Shoulder Disorders

Prajyot Kumta PT, MSc[1]
Joy C. MacDermid PT, PhD[2]
Saurabh P. Mehta PT, MSc[3]
Paul W. Stratford PT, MSc[4]

[1]Physiotherapist, Sripathi Physiotherapy Inc, 1760 Main Street West, Hamilton, Ontario, Canada
[2]Professor, School of Rehabilitation Science, McMaster University, Hamilton, Ontario and Co-director Hand and Upper Limb Centre Clinical Research Laboratory, St. Joseph's Health Centre, 268 Grosvenor St., London, Ontario, Canada
[3]PhD Student and CIHR Doctoral Fellow, School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada
[4]Professor, School of Rehabilitation Science, and Associate Member, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

Ethics approval was obtained from the Research Ethics Board at McMaster University

Corresponding Author: Joy C. MacDermid, PhD , School of Rehabilitation Science, McMaster University, 1400 Main Street West - IAHS 402, Hamilton, Ontario L8S 1C7, Canada Email: macderj@mcmaster.ca Fax no.: 001 905 524 0069

The authors have no financial affiliation or involvement with any commercial organization that has a direct financial interest in any matter involved in this manuscript.

Study Design: Psychometric study design

Objectives: To assess the test-retest reliability and convergent validity of the Functional Impairment Test- Hand and Neck/Shoulder/Arm (FIT-HaNSA) in patients with shoulder disorders

Background: Performance tests that assess functional ability of patients with shoulder disorders can provide useful information for making clinical or return to activity decisions. No performance based shoulder test has yet demonstrated sufficient relevance or clinical measurement properties. The FIT-HaNSA examines upper extremity performance during repetitive tasks that emphasize shoulder reaching and static postures and therefore has greater relevance for assessing performance.

Methods: Thirty six patients with shoulder disorders and 65 healthy controls were recruited in the study. The FIT-HaNSA, Disabilities of the Arm, Shoulder and Hand (DASH), Shoulder Pain and Disability Index (SPADI), isometric shoulder strength, and shoulder range of motion (ROM) were assessed at baseline and repeated 7 days later. Test-retest reliability was described using intraclass correlation coefficient (ICC) and standard error of measurement. Pearson correlation coefficients were used to examine the level of association between the FIT-HaNSA scores and the other measures.

Results: The $ICCs_{2,1}$ for test retest reliability for the FIT-HaNSA ranged from 0.89-0.97 in the patient group and 0.79-0.91 in the control group. The FIT-HaNSA showed a high correlation with the DASH and the SPADI and moderate correlations with the shoulder ROM and muscle strength.

Conclusion: The FIT-HaNSA demonstrated high test-retest reliability and convergent validity with other related outcomes in patients with shoulder disorders. Further

longitudinal studies are required to evaluate the responsiveness of the FIT-HaNSA in patients with different upper extremity conditions.

Key Words: *performance measure, return to work, shoulder disability, psychometrics*

The burden of pain and disability due to shoulder problems is a common concern in the general population. Shoulder problems are the third most common of the musculoskeletal conditions that require consultation in primary care and the prevalence of self-reported shoulder pain is estimated to be between 16% and 26%.[15, 27] Upper extremity disorders are particularly problematic in working populations.[6, 18, 19] Workers who are required to perform forceful, repetitive movements of the arms and work with the arm in elevated positions report high prevalence of upper extremity symptoms.[22, 26] In Ontario, Canada, 7% of Workplace Safety and Insurance Board (WSIB) claims were due to shoulder problems (n= 4274).[28] Furthermore, services directed at the rehabilitation of work injuries affecting the shoulder require more treatments than other body areas.[24]

Physical therapists often face challenging questions from patients, insurance companies, and compensation providers regarding the extent of recovery and whether the patient is able to return safely to normal work duties. Self-reported measures of upper extremity functions are often used for gaining patients' perspective on their functional disability related to the shoulder condition.[3, 16, 23] However, self-reported measures are dependent upon patients' perception regarding their functional status and often do not correspond with assessment performed by their clinicians.[13] This can be particularly problematic when clinicians need to make decision whether a particular patient can return to work or not. Patients with shoulder injuries may not have been required to do activities that incorporate repetitive shoulder movements to assess return-to-work capacity when only self-report measures are used for assessing function. Self-report scales can also be influenced by language, culture, cognitive impairment,

depression, affective response to illness, education, environmental, and personal factors.[4, 5]

Performance tests that assess functional ability of patients with shoulder problems might provide useful information for making clinical or return to activity decisions. A limited number of performance tests that have been described to assess shoulder functions including the Simple Shoulder Endurance test (SSET),[12] the Functional Shoulder Elevation Test (FSET),[10] and function-related tests.[29] No test has yet demonstrated sufficient relevance or clinical measurement properties. The SSET has demonstrated fair test-retest reliability ($ICC_{2, 1}$ of 0.59) in healthy individuals with no shoulder pathology,[12] suggesting that performance may be too unstable for making decision about patients. Furthermore, the endurance of the affected shoulder is assessed in a single position, by flexing the shoulder at 45 degrees and elbow at 30 degrees. This position provides minimal potential for subacromial impingement; therefore it may not replicate functional movements that stress the shoulder.

The FSET is designed to assess pain intensity while lifting a resistance equivalent to 5% of an individual's body weight for 1 repetition in each of the sagittal, scapular, and coronal planes.[10] Therefore, the FSET does not assess shoulder functions in activities that require sustained use of shoulder muscles but rather is an indicator of irritability. Neither reliability nor the relationship for predicting functional outcomes have been determined for the FSET.

The function-related tests developed by Yang et al[29] assess the patient's ability to reach for objects located overhead, across the body, and in their back pocket. While this test showed high intrarater and interrater reliability (weighted κ between 0.83 -

0.90), the authors acknowledge that the test does not consider fatigue, endurance, and movement efficiency.[29] This test might be considered a screen for functional movement rather than a comprehensive indicator of function.

A test that provides information about upper extremity performance during repetitive tasks that emphasize shoulder reaching and static postures would potentially have more relevance to patients with shoulder disorders than the tests currently reported in the literature. Such a test that assesses shoulder endurance should have feasibility for use in clinical practice. The Functional Impairment Test- Hand and Neck/Shoulder/Arm (FIT-HaNSA) was recently developed to assess performance during tasks that represent different aspects of upper extremity gross motor functions (reaching/placing objects at different heights, sustained overhead work).[17] Each task is repeated at a designated pace for up to 5 minutes (300 seconds). Preliminary research that included patients with mild shoulder impingement and controls has demonstrated excellent test-retest reliability (ICC = 0.98) and discriminative validity of the FIT-HaNSA.[17] However, the developmental study reported on only a small sample of subjects and thus the precision of these estimates is questionable.

The objectives of this current study were to examine selected properties of the FIT-HaNSA. First, the percentage of patients and controls who could perform all 3 subtasks by reaching a maximum score of 300 seconds were examined. Second, the test-retest reliability of the FIT-HaNSA in patients with shoulder disorders was assessed. Correlations (convergent validity) of the FIT-HaNSA to 2 upper extremity self-report measures (the Disabilities of the arm, shoulder and hand (DASH) and Shoulder pain and disability index (SPADI)) and to 2 measures of impairment (isometric strength

and range of motion (ROM)) were also examined. Lastly, the known group differences in the scores of the FIT-HaNSA between patients with shoulder pathology versus healthy controls and also for patients who are waitlisted for surgery versus patients with mild to moderate shoulder problems were examined.

METHODS

Participants

Ethics approval for the study was obtained from Hamilton Health Sciences/McMaster University Research Ethics Board. Participants in both groups were provided with an information sheet about the study and written informed consent was obtained.

Participants were recruited from September 2007 to March 2008 for 2 separate categories: patients or healthy controls. The patient subgroup was further subdivided into those managed conservatively and those waiting for shoulder surgery. Participants were included if they were above 18 years of age and had good proficiency in writing and speaking English. Potential participants in the patient and control groups were excluded from the study if they had unstable cardio-respiratory condition, history of stroke, or lower extremity impairments such as peripheral vascular diseases which could limit their ability to stand for 20 minutes to perform the full test. Participants with recent unstable fractures of upper extremity, symptoms of thoracic outlet syndrome, carpal tunnel syndrome or any peripheral nerve entrapment, or chronic systemic inflammatory conditions such as rheumatoid arthritis were also excluded from the study. A self-administered co-morbidity questionnaire[25] was used before the recruitment process to screen the participants for these inclusion/exclusion criteria.

The conservatively managed patients were recruited from local private practice physical therapy clinics, 1 hospital-based physical therapy department, and 1 university-based orthopedic/sport physical therapy clinic. Patients who were waitlisted for shoulder surgery were recruited from a single shoulder surgeon's clinic. Participants in the patient group were diagnosed with 1 or more shoulder conditions such as impingement, partial or full thickness rotator cuff tear, periarthritis, labral tear, shoulder joint arthritis, patients with shoulder arthritis wait listed for shoulder joint replacements and surgical repair of rotator cuff tear. In patients with bilateral involvement, the shoulder with greater involvement was used for testing.

The participants in the control group were recruited through public advertisement and flyers posted around the university offering participation in the study. They were tested for the FIT-HaNSA on their dominant side. The participants in the control group were excluded if they had pre-existing shoulder or neck conditions.

Outcome Measures

FIT-HaNSA

The FIT-HaNSA was conducted as described[17] using JobSim (JTech Medical, Salt Lake City, USA) for staging the tasks. The FIT-HANSA is a test battery of 3 tasks that simulate activities of lifting and sustained overhead work. The attribute of interest is sustained activity involving repeated lifting and overhead shoulder movements. In the first task, a shelf is placed at the participant's waist level and another shelf is placed 25 cm above the first one. Whereas in the second task, the shelf height is adjusted such that the top shelf is at the participant's eye level and the bottom shelf is 25 cm below it. For both tasks, 3 jars weighing 1 kg each are placed 10 cm apart from each other on

the lower shelf. The participant is required to use the affected arm to lift the jars 1 at a time from one shelf to the other shelf at a standard speed of 60 beats per minute as monitored with a metronome. The assessor measures the amount of time that the participant is able to perform each task using a stop watch. The first task is based on a reaching activity done at waist level, whereas the second task represents reaching activity done at shoulder level. In the third task, a shelf is placed at the participant's eye level. Attached perpendicular to the shelf is a plate projecting outward. The participant is instructed to use both his/her arms in a sustained overhead position to screw and unscrew bolts and transfer them between 3 holes in the attachment plate. If the participant drops 1 of the bolts he/she is instructed to maintain the overhead position of the arms while a spare bolt is given to him/her by the assessor - so the task maintains sustained overhead work. This third task represents sustained overhead activity. Approximately 30 seconds of rest is provided between each task during which the shelf heights are adjusted. Each task is continued for a maximum of 300 seconds or until any of the following stopping criteria is reached. The task is stopped if the participant reports extreme pain or fatigue, if the examiner observes the participant producing substitute trunk/body movements that could occur due to fatigue, there is a concern of injury to the participant, or if the participant is severely off pacing to the extent that he/she is unable to complete 1 repetition of the movement within 2 beats of the metronome. An overall summary score is calculated by averaging the time for the 3 tasks.

Self-report measures

The DASH is a self-reported measure of upper extremity disability.[11] The DASH has 30 questions, of which 21 are related to physical function, 6 are related to

symptoms, and 3 are related to social/role function. Each item on the DASH has a response option ranging from 1-5. A total score on the DASH can range between 0 to 100 with higher scores indicating greater disability.

The SPADI is a self-reported outcome measure developed primarily to assess pain and disability associated with shoulder problems.[21] The SPADI consists of 13 items in 2 subscales: pain (5 items) and disability (8 items). The participants respond on the numeric rating scale of 0-10 with higher score indicating greater pain/disability.

Physical impairment measures

A hand held dynamometer (HHD) (Lafayette Instrument Company, Lafayette, Indiana) was used to measure isometric muscle strength of the shoulder flexors, abductors, and rotators. Strength testing was performed with the participant sitting upright on a chair with back supported and feet touching the floor. The detailed procedures have been described previously.[8] For testing the flexors, the evaluator stood beside the participant and the participant was asked to hold the shoulder in a position of 90° flexion. For testing the abductors, the evaluator stood behind the participant and the participants was asked to hold the shoulder in a position of 90° of abduction. Strength was defined as the value (Kg) at which the participant could no longer match the force applied by the tester (break test). The test was repeated for 3 times and an average of the 3 scores used for analysis.

Active range of motion (ROM) for the shoulder flexion, abduction, internal rotation, and external rotation were measured using a full circle plastic goniometer. All measurements were made with participant sitting upright in a chair with back supported and feet touching the floor. For the external and internal rotation, the arm was

positioned in 0° glenohumeral abduction, 90° elbow flexion and in neutral supination/pronation. Participant was asked to move the shoulder actively through the available range in the limit of pain tolerance. One repetition was performed for each movement and recorded for analysis. These test procedures have been described previously and considered to have comparable reliability with other methods of ROM measurements for shoulder.[9]

Study Protocol

Two assessors were used and a single training session was attended by both the assessors to learn the test procedures.

First, demographic data such as age, sex, occupation, and side dominance were collected for all participants. Then, shoulder muscle strength and ROM measurements were performed. Subsequently, the DASH, SPADI, and FIT-HaNSA were administered to the participants. This order of testing allowed the participants to have about 15 minutes rest between strength testing and the FIT-HaNSA. This was considered sufficient to recover from any fatigue related to strength testing.

Participants involved in the test retest reliability assessment of the study repeated testing on the FIT-HaNSA within 7 days of the first visit. An attempt was made to ensure that the same rater performed the retest assessment for each patient.

Data Analysis

Data entry and visual screening were performed by the first author. SPSS version 16.0 (SPSS, Inc., Chicago IL) was used for all the analyses. Descriptive statistics (means, standard deviations, standard errors of mean, and 95% confidence intervals) were computed for the FIT-HaNSA, self report, and physical impairment

measures in patients and healthy controls. The differences in age, gender, FIT-HaNSA scores, and comparative measures for the 2 subgroups were examined by using independent t-tests. Histograms were plotted for the FIT-HaNSA scores for both subgroups. The percentage of patients and controls who were able to complete all 3 substasks by reaching a maximum score of 300 seconds was calculated.

Test retest reliability was described using intraclass correlation coefficient ($ICC_{2,1}$) and standard error of measurement (SEM). [20] The test retest reliability is considered to be high if the calculated value of the $ICC_{2,1}$ is greater than 0.75. The Bland and Altman's plot was used to examine the agreement between the scores obtained with the FIT-HaNSA administered on 2 occasions. [2] A graph of the differences between the summary scores of the FIT-HaNSA on 2 occasions for each participant was plotted against the grand mean of the FIT-HaNSA scores. The 95% limits of agreement (LOA) between the scores obtained on both occasions were calculated. This graph demonstrates the extent to which the scores of the FIT-HaNSA obtained on both occasions agreed with each other. The unit on both axes of the Bland and Altman plot is in seconds, which is the unit of measurements of the FIT-HaNSA.

Pearson correlation coefficients were used to assess the level of association between the FIT-HaNSA scores and the other measures (DASH, SPADI, the strength and ROM ratios of the affected to the unaffected side). We expected a negative correlation between the FIT-HaNSA and the self-report measures and a positive correlation between the FIT-HaNSA and the physical impairment measures. Correlations were classified as high ($r > 0.70$), moderate ($r > 0.40$ but < 0.70) and low ($r < 0.40$). [7] The known group validity examines the ability of the measure to discriminate

among groups with and without pathologies, in this study: control group, patients with pathologies being actively managed conservatively and patients with pathologies wait-listed for surgery). This was examined by assessing the differences in the summary scores of the FIT-HaNSA obtained on the first testing occasion across these 3 subgroups using an ANOVA.

RESULTS

A total of thirty six patients (15 females, 21 males) were recruited (mean ± SD age, 42 ± 16 years), of which 26 were conservatively managed and 10 were waitlisted for shoulder surgery. Of 36 patients, 34 were tested on 2 occasions. Two patients did not return for their second testing for unknown reasons. Thirty four patients had unilateral shoulder involvement, whereas the remaining 2 had bilateral shoulder problems. Nine patients had involvement of their non dominant side. Sixty five healthy participants (45 females, 20 males) were recruited in the control group (age, 29 ± 12). Six participants in the control group did not come for the second testing for unknown reasons. The characteristics of the participants are shown in TABLE 1.

The majority of participants in both groups were male. The 2 groups were significantly different in age ($p < 0.0001$) and gender distribution ($p= 0.008$). The DASH (mean ± SD, 21.4 ± 18.1) and the SPADI scores (27.98 ± 23.55) for the patient group were significantly higher as compared to the control group (DASH score of 2.5 ± 3.2 and SPADI score of 2.2 ± 4.9) which indicated that they had greater functional disability. The scores on the FIT-HaNSA were significantly lower ($p < 0.0001$) for the patient versus the control group on both testing occasions (182 ± 77 versus 273 ± 39 on occasion 1; 185 ± 83 versus 277 ± 41 on occasion 2) (TABLE 2). Both the groups were different in their

With the exception of shoulder strength for internal and external rotation, which was similar for both groups, shoulder strength and ROM values for the control group were significantly greater than for the patient group (TABLE 3). Figure 1 illustrates the histograms for the FIT-HaNSA scores for the patient and control subgroups with normal distribution curves for the first assessment. Of the 59 controls, 24 (41%) were able to complete all 3 subtasks to achieve the maximum score of 300 seconds on both testing occasions. Of 34 patients, only 1 (3%) was able to achieve the maximum score of 300 seconds on both occasions.

The $ICC_{2,1}$ for test retest reliability ranged from 0.89 to 0.95 in patients and 0.79 to 0.91 for controls for the 3 subtasks (TABLE 4). The $ICC_{2,1}$ values for the summary scores of the FIT-HaNSA were 0.97 for the patients and 0.91 for the controls. The test was designed to have a target for task completion by reaching a maximum score of 300 seconds. Because 24 participants in the control group completed the 3 subtasks and achieved the maximum score of 300 seconds on both testing occasions, the data had the potential to bias reliability estimation. This is because there was an artificial agreement (and a lack of score variation) across the 2 occasions. Therefore, the test-retest reliability analysis was also performed considering only the remaining 35 controls who did not reach the score of 300 seconds. In this group of controls, the $ICC_{2,1}$ was 0.88 for the summary FIT-HaNSA score. The lowest reliability was observed for task 2 in the patient group ($ICC_{2,1} = 0.89$) and the subset of the control groups who did not reach the maximum score of 300 seconds ($ICC_{2,1} = 0.72$). The estimated SEM for the patient group was 13.3 seconds (95% CI: 10.7, 17.5) compared to 12.2 seconds (95% CI: 10.3, 14.9) for the control group.

Figures 2 and 3 show the Bland Altman plot for the level of agreement between the FIT-HaNSA scores performed on 2 separate occasions. The mean difference between the 2 occasions was 2.9 ± 18.9 seconds for the patient group and 3.9 ± 17.1 seconds for the controls indicating that no systematic differences occurred between testing occasions.

The results of correlations between the subtasks of FIT-HaNSA and other measures in the patient group are shown in Table 5. The summary score of FIT-HaNSA showed a high level of association with the DASH (r = 0.76; 95% CI -0.87, -0.58) and the SPADI (r = -0.71; 95% CI -0.84, -0.50) and a moderate level of association with shoulder ROM (r between 0.45-0.64) and strength (r between 0.49-0.66). The results of correlations between the subtasks of FIT-HaNSA and other measures in the control group are shown in Table 6. No significant correlations were observed between the FIT-HaNSA and the comparative measures in the control group with an exception of the SPADI where a moderate level of association (r = -0.5; 95% CI -0.66, -0.29) was observed.

Figure 4 shows the mean scores of the FIT-HaNSA for the subgroups recruited in this study based on their level of severity of shoulder problems (patients with mild to moderate shoulder condition, patients waiting for shoulder surgery, and control group). The patients waiting for shoulder surgery had significantly lower scores on the FIT-HaNSA (mean ± SD, 90.0 ± 23.5 seconds) than those with mild to moderate shoulder problems (221.0 ± 55.3 seconds) and healthy control subjects (273.0 ± 39.1 seconds). These differences were statistically significant ($p < 0.001$) and in the anticipated direction with those with the more severe conditions having a lower score. These results

supported our hypothesis of known group validity and suggested that the FIT-HaNSA discriminated between these subgroups.

DISCUSSION

This study examined the reliability and validity of the FIT-HaNSA as well as its ability to discriminate between the known subgroups of patients with different severities of shoulder problems. Our study extends support to the previous study that examined the validity of the FIT-HaNSA[17] by indicating similar findings in a larger sample of patients with shoulder pathologies and healthy controls.

ICCs indicate relative reliability and can be influenced by variability of the sample. Our patient group sample included patients with varying severity of shoulder problems, which may have contributed to higher between subject variance. However, wide variation in the severity of disability is common within clinical studies and our data reflect the type of shoulder pathology commonly seen in clinic. Further, data collected on subjects without pathology also showed high reliability.

When comparing individual patients over time, absolute reliability is more relevant as it tells more about the stability of a measure and highlights the error associated with a measurement. Because this is the first study on reliability estimation of the FIT-HaNSA on patients with shoulder pathology, we do not have any previous studies with which we can compare our data. This study did not deal with an error associated with a changed score. However, based on the point estimate for the SEM of 13.3 seconds for a single measurement, 68% of patients would be expected to display random fluctuations within 18.8 seconds (i.e., 1 x SEM x $\sqrt{2}$) between measurements

taken on 2 occasions and 90% of patients would be expected to display fluctuations within 31.1 seconds (i.e., 1.65 x SEM x √2).

Subtask 2 of the FIT-HaNSA showed high test retest reliability with an ICC of 0.89 (0.80-0.95 CI) but it was slightly lower than the other 2 subtasks. Subtask 2 is more challenging in nature requiring repeated movements into a position of relative shoulder impingement during a reach and grasp task. Therefore it is possible that performance might be different on 2 separate occasions. All 3 subtasks have different movement patterns which place stress on the shoulder joint and provide different neuromuscular challenges. It appears that each subtask is affected by shoulder pathology and hence its content validity is high. We know that performance of each subtask may be influenced by the underlying shoulder problem and therefore different subtasks are important for making the test applicable across broader clinical populations.

Twenty out of 34 patients and 45 out of the 59 controls had slightly higher summary scores for the FIT-HaNSA on testing occasion 2 compared to occasion 1. However, this change was not significant ($p > 0.05$). There could be a few reasons for this small increase in the score on occasion 2. First, there is a possibility of a learning effect and increased familiarity with the test on the second occasion. Second, the strength testing was performed only on occasion 1 and not on occasion 2. Although we gave 15 minutes between strength testing and performing the FIT-HaNSA to minimize the impact of fatigue, it is possible that this amount of time was not sufficient for some participants.

We observed a high level of association between the FIT-HaNSA and self-reported functional measures in the patient group, which support the validity of these

measures. There was some variation in the strength of this relationship across different subtasks of the FIT-HaNSA with subtask 2 demonstrating low correlation as compared to subtasks 1 and 3. Because the overall score on subtask 2 was lower, we believe that the task was more difficult to complete.

We found a moderate level of association between the scores on the FIT-HaNSA and shoulder strength for flexion ($r = 0.66$; 95% CI 0.42, 0.81) and abduction ($r = 0.55$; 95% CI 0.27, 0.74). In a previous study, the correlation values between these measures were low ($r < 0.29$).[17] This finding can be attributed to the difference in the study population. The previous study had 36 participants which included healthy controls and patients with only mild shoulder problems whereas the patient population in our study included patients with a broad spectrum of severity of shoulder problems.

We observed a moderate level of association between shoulder ROM and scores on the FIT-HaNSA in the patient group. These findings are consistent with previous studies[1, 14] in which a similar correlation was observed between scapular movements and self-reported functional limitation while performing activities similar to the FIT-HaNSA.

The level of association between the FIT-HaNSA and the impairment measures was low in the control group. Though the participants in the control group did not have any shoulder conditions, it is very likely that the FITHaNSA was relatively more difficult for many compared to the single task performance such as the muscle strength and ROM.

Twenty four healthy controls (41%) and 1 (3%) patient were able to complete the FIT-HaNSA with a summary score of 300 seconds on testing occasion 1. In the

previous study, most of those in the control group (95%) were able to complete the FIT-HaNSA.[17] The purpose of this study was not to establish “normal data” but this is an important consideration. The 300 seconds limit can also cause a “ceiling effect” in which the participant scores at the top of the scale with no further potential for improvement on the scale. However, time limits to the test are also important for feasibility because a test that is of very long duration may prove to be less practical to be administered in a busy outpatient clinic. Ideally the majority of young healthy people should be able to complete the task and it should be sensitive to picking up differences in less healthy or compromised individuals. However, the 300 seconds restriction may also limit the ability to detect change in certain patients who can achieve the maximum scare but still have deficits precluding full return to activities. However, for the purpose of standardization of the test and feasibility in clinical practice, the developers assumed that limiting each task to 300 seconds would be sufficient to assess the performance of most patients with shoulder or neck problems. This was substantiated in this study as only 1 participant in the patient group scored the maximum 300 seconds. No floor effect was seen for the FIT-HaNSA. A wide range of scores were observed and no apparent clustering was evident, suggesting that for most patients the FIT-HaNSA provides the ability to detect clinical changes of >13 seconds (SEM for the FIT-HaNSA). A low incidence of ceiling or floor effect for the FIT-HaNSA in patients makes this test more responsive to measure clinical change in a population with low level of functional ability. However our study did not deal with this aspect of the test and further study on patients with shoulder problems with repeated measurement before and after treatment would be useful to evaluate the responsiveness of the FIT-HaNSA.

Results of this study support the hypotheses about known group validity of FIT-HaNSA as significant differences in scores were observed between participant groups depending on their level of severity of shoulder problem. As expected, those patients who were waitlisted for surgery demonstrated poorest performance, those managed conservatively had limited performance, and those without pathology performed the best on the FIT-HaNSA.

Although these findings provide preliminary evidence regarding the reliability and validity of FIT-HaNSA in patients with shoulder problems, we recognize certain limitations in this study. Because many of the healthy controls were young university students, we used analysis of covariance to control for age effects where matching would have been preferable. Although all surgical patients had a standardized physical examination and imaging from a single shoulder surgeon and all patients were examined by a single physical therapist, we relied on a simple treatment-based classification to differentiate levels of severity. Additionally, we cannot confirm that the high level of relative reliability observed in this study would transfer to a more homogeneous sample such as a clinical subgroup with a specific athletic injury. This study was also limited by its cross sectional design. We were unable to determine whether the relationships observed between the FIT-HaNSA and other outcome measures were stable over time. Future studies also need to establish the age-adjusted normative values for the FIT-HaNSA which will enable clinicians to compare the results of their patients in context of expected values for their age.

CONCLUSION

This cross sectional study indicated that the FIT-HaNSA has excellent test-retest reliability in patients with shoulder disorders. The study also provided preliminary evidence regarding the expected relationships and convergent validity of the FIT-HaNSA with selected self-report measures and objective outcomes in these patients. The FIT-HaNSA was able to discriminate between different levels of severity of the shoulder pathologies. Further longitudinal studies are required to evaluate the responsiveness and predictive validity of the FIT-HaNSA in patients with different upper extremity conditions.

KEY POINTS

- Findings: The results of this study support the reliability and validity of the FIT-HaNSA for assessing performance of shoulder functions.
- Implication: Physical therapists can use the FIT-HaNSA in patients with shoulder pathologies for assessing shoulder endurance in different aspects of upper extremity gross motor functions (reaching/placing objects at different heights, sustained overhead work).
- Caution: Future work should focus on assessing longitudinal validity of the FIT-HaNSA and assess its utility across different upper extremity conditions.

Reference List

(1) Bekkering WP, ten CR, van Suijlekom-Smit LW, Mul D, van der Velde EA, van den Ende CH. The relationship between impairments in joint function and disabilities in independent function in children with systemic juvenile idiopathic arthritis. *J Rheumatol* 2001;28(5):1099-1105.

(2) Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007;17(4):571-582.

(3) Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004;63(4):335-341.

(4) Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. *Soc Sci Med* 1999;48(11):1549-1561.

(5) Daltroy LH, Phillips CB, Eaton HM et al. Objectively measuring physical ability in elderly persons: the Physical Capacity Evaluation. *Am J Public Health* 1995;85(4):558-560.

(6) De SE, Germeys F, De SL. Prevalence of work related upper limb disorders in hairdressers: a cross sectional study on the influence of working conditions and psychological, ergonomic and physical factors. *Work* 2009;34(3):325-330.

(7) Domholdt E. *Physical therapy research: Principle and applications*. 2nd ed. Philadelphia: WB Saunders Co.; 2000.

(8) Hayes K, Walton JR, Szomor ZL, Murrell GA. Reliability of 3 methods for assessing shoulder strength. *J Shoulder Elbow Surg* 2002;11(1):33-39.

(9) Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. *Aust J Physiother* 2001;47(4):289-294.

(10) Hollinshead RM, Mohtadi NG, Vande Guchte RA, Wadey VM. Two 6-year follow-up studies of large and massive rotator cuff tears: comparison of outcome measures. *J Shoulder Elbow Surg* 2000;9(5):373-381.

(11) Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med* 1996;29(6):602-608.

(12) Hughes RE, Johnson ME, Skow A, An KN, O'Driscoll SW. Reliability of a simple shoulder endurance test. *J Musculoskelet Res* 1999;3:195-200.

(13) Larmer PJ, McNair PJ, Smythe L, Williams M. Ankle sprains: patient perceptions of function and performance of physical tasks. A mixed methods approach. *Disabil Rehabil* 2011;33(23-24):2299-304. Epub 2011

(14) Lin JJ, Hanten WP, Olson SL et al. Shoulder dysfunction assessment: self-report and impaired scapular movements. *Phys Ther* 2006;86(8):1065-1074.

(15) Luime JJ, Koes BW, Hendriksen IJ et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33(2):73-81.

(16) MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. *J Shoulder Elbow Surg* 2006;15(4):407-414.

(17) MacDermid JC, Ghobrial M, Quirion KB et al. Validation of a new test that assesses functional performance of the upper extremity and neck (FIT-HaNSA) in patients with shoulder pathology. *BMC Musculoskelet Disord* 2007;8:42.

(18) Matsudaira K, Palmer KT, Reading I, Hirai M, Yoshimura N, Coggon D. Prevalence and correlates of regional pain and associated disability in Japanese workers. *Occup Environ Med* 2010;68(3):191-6. Epub 2010

(19) Osborne A, Blake C, McNamara J, Meredith D, Phelan J, Cunningham C. Musculoskeletal disorders among Irish farmers. *Occup Med (Lond)* 2010;60(8):598-603.

(20) Portney LG, Watkins MP. *Foundations of Clinical Research : Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2007.

(21) Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res* 1991;4(4):143-149.

(22) Roquelaure Y, Ha C, Rouillon C et al. Risk factors for upper-extremity musculoskeletal disorders in the working population. *Arthritis Rheum* 2009;61(10):1425-1434.

(23) Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum* 2009;61(5):623-632.

(24) Sadi J, MacDermid JC, Chesworth B, Birmingham T. A 13-year cohort study of musculoskeletal disorders treated in an autoplant, on-site physiotherapy clinic. *J Occup Rehabil* 2007;17(4):610-622.

(25) Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003;49(2):156-163.

(26) Stock SR. Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. *Am J Ind Med* 1991;19(1):87-107.

(27) Urwin M, Symmons D, Allison T et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis* 1998;57(11):649-655.

(28) Workplace Safety and Insurance Board. Statistical Supplement to the 2009 Annual Report. http://www.wsib.on.ca/files/Content/AnnualReports2009StatisticalSupplement/StatSupp09.pdf . 2009.

(29) Yang JL, Lin JJ. Reliability of function-related tests in patients with shoulder pathologies. *J Orthop Sports Phys Ther* 2006;36(8):572-576.

FIGURE 1. Histograms for the FIT-HaNSA summary scores for both the patients (A) and control (B) groups

One participant in the patient group and 24 participants in the control group scored the maximum score of 300 seconds.

FIGURE 2. The difference between the summary scores on the FIT-HaNSA for testing performed on 2 occasions plotted against the grand mean of FIT-HaNSA scores for the controls

Mean difference (d = 3.85 seconds) is the blue line, the limits of agreement (d ± 2SD = 37.95 and -30.25) are the red lines.

FIGURE 3. The difference between the summary scores on the FIT-HaNSA for testing on 2 occasions plotted against the grand mean of the FIT-HaNSA scores for the patients

Mean difference (d = 2.84 seconds) is the blue line, the limits of agreement (d ± 2SD = 40.54 and -34.86) are the red lines.

FIGURE 4. Summary scores of the FIT-HaNSA for the 3 groups

The box-plot summaries of FIT-HaNSA scores across the three subgroups recruited in this study (patients with mild to moderate shoulder condition, patients waiting for shoulder surgery, and control group) are illustrated in this figure. The dark line in each box represents the median, the upper and lower margins of the box depict the first and third quartiles, and the whisker length represents 1.5 times the interquartile range (IQR) provided a data point is located at that value, otherwise the whisker is drawn to the closest data point within 1.5 IQR. The dots show outliers. The mean scores (273 sec, 221 sec, and 90 sec) were significantly different among groups ($p < 0.0001$).

TABLE 1. Demographics

		Patients (N = 36)	Controls (N = 65)	P values
Age (Mean ± SD)		41.8 ± 15.7	29.1 ± 11.5	< 0.001
Sex	M	21	45	N/A
	F	15	20	
Dominant Side	R	33	61	N/A
	L	4	4	
Side tested for the the FIT-HaNSA	D	27	65	N/A
	ND	9	0	
DASH (Mean ± SD)		21.4 ± 18.1	2.5 ± 3.2	< 0.001
SPADI (Mean ± SD)		28 ± 23.5	2.2 ± 4.9	< 0.001

Abbreviations: D, Dominant side; F, Female; L, Left; M, Male; ND, Non Dominant; R, Right

TABLE 2. Descriptive statistics of FIT-HaNSA scores*

		Occasion 1	Occasion 2
		Mean ± SD (95% CI)	**Mean ± SD (95% CI)**
Patients N = 34	Task1	222 ± 88 (191,253)	223 ± 90 (192,253)
	Task2	124 ± 73 (98,149)	134 ± 82 (105,163)
	Task3	201 ± 93 (169,234)	199 ± 97 (165,233)
	Summary Score	182 ± 77 (155,209)	185 ± 83 (156,214)
Controls N = 59	Task1	294 ± 22 (289, 300)	292 ± 32 (284,300)
	Task2	239 ± 76 (219,259)	250 ± 74 (231,269)
	Task3	285 ± 40 (275, 295)	288 ± 35 (279,297)
	Summary Score	273 ± 39 (263, 283)	277 ± 41 (266,288)

* Data are mean ± SD (95% confidence interval) in seconds.

TABLE 3. Descriptive statistics for shoulder strength and range of motion*

Strength (kgs)	Patients (N=36) Affected side Mean ± SD (Min, Max)	Patients (N=36) Ratio** Mean ± SD (Min, Max)	Controls (N=65) Dominant side Mean ± SD (Min, Max)	Controls (N=65) Ratio*** Mean ± SD (Min, Max)	P-value****
Flexion	14.6 ± 6.4 (5.2,30)	0.35 ± 0.1 (0.18,0.55)	19 ± 4.3 (10.8,27.1)	0.48 ± 0.04 (0.37,0.57)	0.001
Abduction	13.3 ± 6.5 (1.8,26)	0.34 ± 0.12 (0.1,0.62)	18 ± 4.4 (9.3,29)	0.47 ± 0.05 (0.32,0.63)	< 0.001
Internal Rotation	13.1 ± 4.6 (2.5,26)	0.39 ± 0.1 (0.1,0.61)	14.4 ± 4 (6.7,26.7)	0.48 ± 0.06 (0.26,0.59)	0.15
External Rotation	12.2 ± 4.4 (4.4,24)	0.4 ± 0.1 (0.21,0.62)	12.4 ± 2.7 (8.3,22.3)	0.44 ± 0.06 (0.32,0.66)	0.75
Shoulder Range of Motion (Degrees)					
Flexion	151 ± 19.4 (95,172)	0.93 ± 0.1 (0.68,1.11)	167 ± 4.6 (157,180)	1.02 ± 0.03 (0.95,1.12)	< 0.001
Abduction	149 ± 24 (88,180)	0.90 ± 0.13 (0.64,1.07)	170 ± 4.8 (155,180)	1.01 ± 0.02 (0.96,1.01)	< 0.001
Internal Rotation	54 ± 14.8 (30,82)	0.84 ± 0.17 (0.46,1.12)	67 ± 8.1 (28,85)	1.01 ± 0.1 (0.67,1.45)	< 0.001
External Rotation	51 ± 14.7 (30,85)	0.88 ± 0.2 (0.53,1.42)	66 ± 10.1 (42,88)	1.07 ± 0.12 (0.82,1.35)	< 0.001

* Data are mean ± SD (min, max)

** Ratio of affected versus unaffected side

*** Ratio of dominant versus non-dominant side

**** Difference between groups for strength and range of motion values

TABLE 4. Reliability coefficients and SEM (seconds) and their 95% confidence intervals for the FIT-HaNSA

		Patients (N=34)	Controls (N=59)	Controls* (N=35)
Task 1	$ICC_{2,1}$	0.95 (0.91,0.97)	0.79 (0.67,0.87)	0.78 (0.62,0.88)
	SEM	18.5 (15,24.4)	12.6 (10.6,15.3)	16.3 (13.2,21.4)
Task 2	$ICC_{2,1}$	0.89 (0.80,0.95)	0.81 (0.71,0.88)	0.72 (0.51,0.84)
	SEM	25.3 (20.4,33.3)	32.1 (27.1,39.2)	41.2 (33.2,53.7)
Task 3	$ICC_{2,1}$	0.95 (0.91,0.97)	0.91 (0.85,0.94)	0.90 (0.81,0.94)
	SEM	20 (16.1,26.3)	11.2 (9.5,13.7)	14.5 (11.7,19)
Summary Score**	$ICC_{2,1}$	0.97 (0.95,0.98)	0.91 (0.85,0.94)	0.88 (0.77,0.94)
	SEM	13.3 (10.7,17.5)	12.2 (10.3,14.9)	15.6 (12.6,20.5)

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval; SEM, Standard Error of Measurement.

*Subset of the control participants who did not reach the maximum score of 300 seconds on all 3 tasks of the FIT-HaNSA.

**Mean of 3 tasks

TABLE 5. Correlations between the scores on the FIT-HaNSA and self-report measures and shoulder strength and range of motion in the patient group (N=36)*

	Task 1	**Task 2**	**Task 3**	**Summary Score****
DASH	-0.75 (-0.87,-0.57)	-0.49 (-0.70,-0.19)	-0.81 (-0.90,-0.66)	-0.76 (-0.87,-0.58)
SPADI	-0.72 (-0.84,-0.50)	-0.45 (-0.68,-0.14)	-0.75 (-0.86,-0.55)	-0.71 (-0.84,-0.50)
Strength				
Flexion	0.55 (0.26,0.74)	0.54 (0.25,0.73)	0.70 (0.48,0.83)	0.66 (0.42,0.81)
Abduction	0.55 (0.27,0.74)	0.38 (0.05,0.62)	0.56 (0.28,0.75)	0.55 (0.27,0.74)
Internal Rotation	0.46 (0.15,0.68)	0.46 (0.15,0.68)	0.54 (0.25,0.73)	0.54 (0.25,0.73)
External Rotation	0.55 (0.27,0.74)	0.37 (0.04,0.62)	0.41 (0.09,0.64)	0.49 (0.19,0.70)
Range of Motion				
Flexion	0.63 (0.38,0.79)	0.40 (0.08,0.64)	0.62 (0.35,0.78)	0.61 (0.35,0.78)
Abduction	0.68 (0.45,0.82)	0.39 (0.07,0.63)	0.64 (0.4,0.8)	0.64 (0.4,0.8)
Internal Rotation	0.35 (0.02,0.60)	0.34 (0.01,0.60)	0.54 (0.25,0.73)	0.45 (0.14,0.68)

External Rotation	0.49 (0.19,0.70)	0.49 (0.19,0.70)	0.57 (0.29,0.75)	0.57 (0.29,0.75)

*Pearson correlation coefficient (r) and its associated 95% confidence interval are given for each subtask and the FIT-HaNSA summary score.

**Mean of the 3 tasks

TABLE 6. Correlations between the scores on the FIT-HaNSA test and self-report measures and shoulder strength and range of motion in the control group (N=65)*

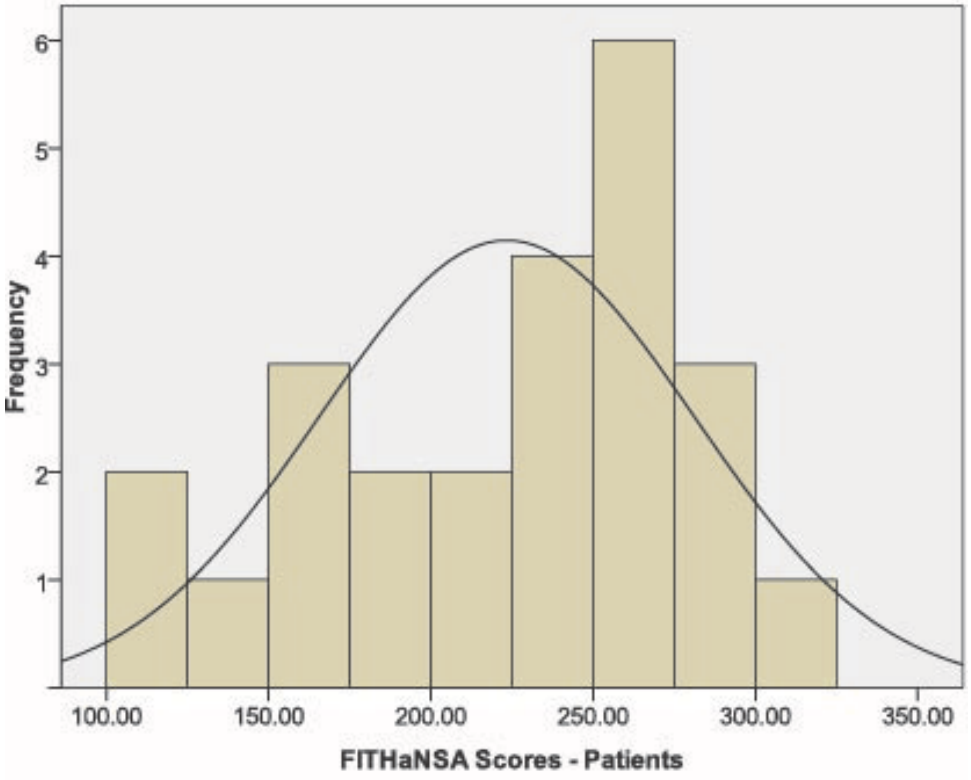
	Task 1	**Task 2**	**Task 3**	**Summary Score****
DASH	-0.23 (-0.44,0.01)	-0.09 (-0.32,0.16)	-0.24 (-0.46,-0.00)	-0.18 (-0.41,0.07)
SPADI	-0.52 (-0.68,-0.32)	-0.33 (-0.53,-0.09)	-0.57 (-0.71,-0.38)	-0.50 (-0.66,-0.29)
Strength				
Flexion	-0.18 (-0.41,0.06)	-0.03 (-0.27,0.22)	0.04 (-0.21,0.28)	-0.04 (-0.28,0.21)
Abduction	-0.12 (-0.35,0.13)	-0.11 (-0.34,0.14)	-0.04 (-0.28,0.21)	-0.10 (-0.33,0.15)
Internal Rotation	-0.18 (-0.41,0.06)	-0.09 (-0.33 0.16)	0.05 (-0.19,0.29)	-0.07 (-0.31,0.18)
External Rotation	0.02 (-0.22,0.26)	-0.02 (-0.26 0.22)	0.03 (-0.22,0.27)	0.00 (-0.24,0.24)
Range of motion				
Flexion	0.04 (-0.21,0.28)	-0.14 (-0.37,0.11)	0.08 (-0.17,0.38)	-0.06 (-0.29,0.18)
Abduction	-0.07 (-0.31,0.17)	-0.24 (-0.46,-0.00)	0.03 (-0.22,0.27)	-0.16 (-0.38,0.08)
Internal Rotation	0.09 (-0.16,0.32)	-0.01 (-0.25,0.23)	-0.03 (-0.27,0.21)	0.00 (-0.24,0.24)
External	-0.08	-0.17	0.02	-0.12

Rotation	(-0.32,0.17)	(-0.39,0.07)	(-0.23,0.26)	(-0.35,0.13)

*Pearson correlation coefficient (r) and its associated 95% confidence interval are given for each subtask and the FIT-HaNSA summary score.

**Mean of the 3 tasks

Figure 1. Histograms for the FIT-HaNSA for both the subgroups



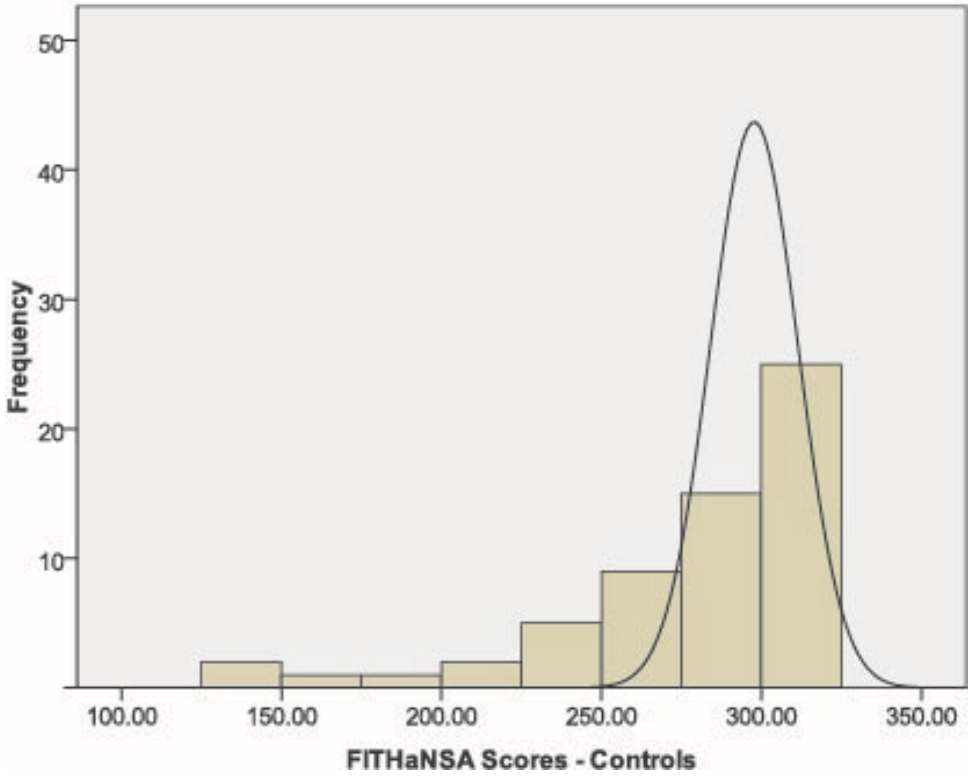


Figure 2. The difference between the summary scores of FIT-HaNSA on 2 occasions plotted against the grand mean of FIT-HaNSA scores for the controls

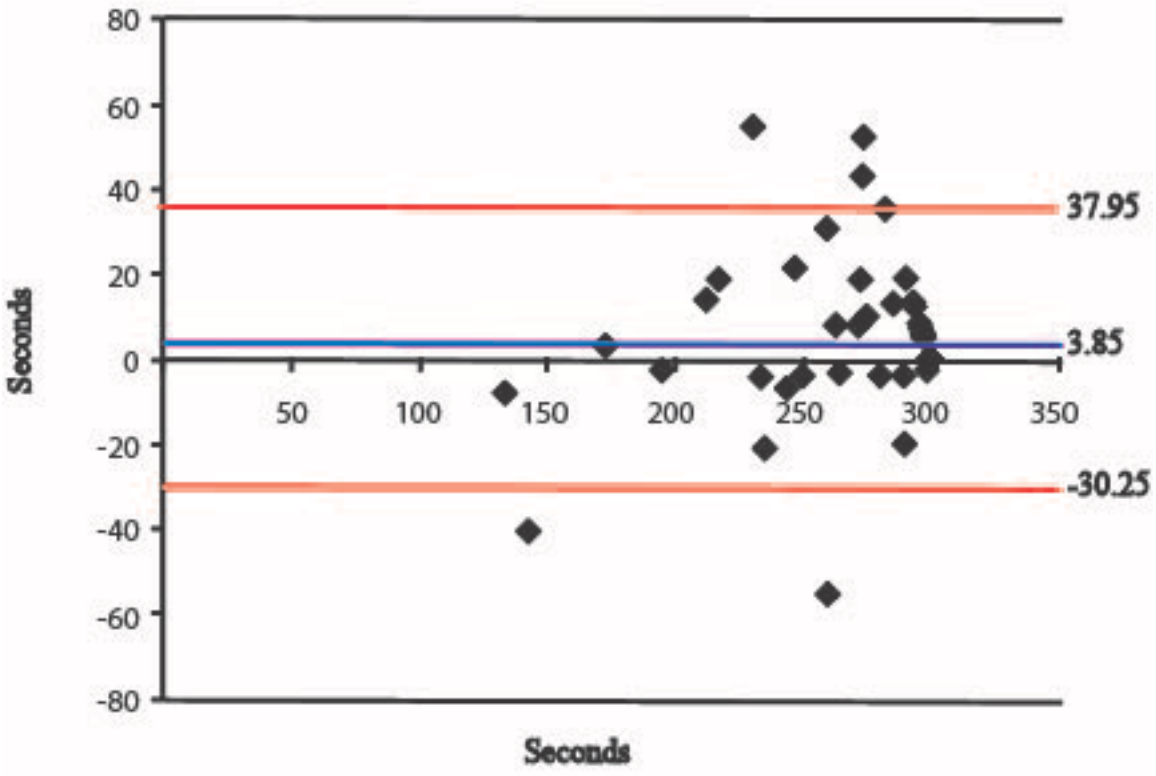


Figure 3. The difference between the summary scores of the FIT-HaNSA on 2 occasions plotted against the grand mean of the FIT-HaNSA scores for the patients

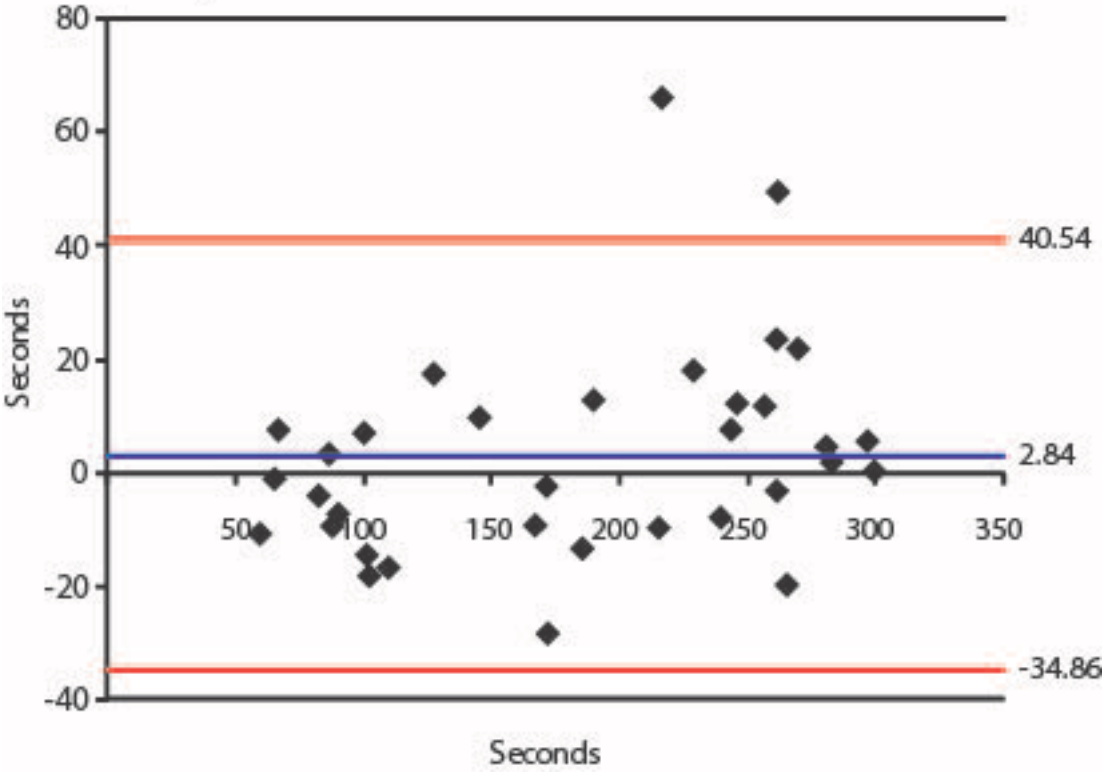


Figure 4. Summary scores of FIT-HaNSA across the 3 known groups