STarT MSK tool: Translation, adaptation and validation in Hebrew

Noa Ben Ami, PT, PhD[1], Jonathan Hill, PT, PhD[2], Tamar Pincus, PhD[3]

1. Department of Physiotherapy, Ariel University, Ariel, Israel.

2. Department of Physiotherapy, Keele University, United Kingdom.

3. Department of Psychology, Royal Holloway University of London, London, United Kingdom.

Abstract

Background: The Keele STarT MSK Tool divides musculoskeletal patients into three prognostic groups for risk-stratified care. It has shown good predictive and discriminative ability in development and validation samples.

Objectives: This study aimed to translate and validate the STarT MSK in a Hebrew version, among Israeli people living with musculoskeletal (MSK) pain.

Method: A Cross-sectional study, with nested prospective sub-sample. The STarT MSK was translated into Hebrew using published guidelines. A total of 153 adults (18+) who reported living with MSK pain were administered the STarT MSK. Clinical measures included for validity testing included the 12-Item Short-Form Health Survey (SF-12), the Hospital Anxiety and Depression Scale (HADS), the Fear-Avoidance Beliefs Questionnaire (FABQ) and a numerical pain rating scale (NPRS).

Results: The STarT MSK was forward and backward translated, with minor changes to ensure cultural adaptation. The test-retest reliability of the STarT MSK total score was excellent (intraclass correlation coefficient 0.92). Internal consistency for the MSK scale was (α=0.612). The Spearman's correlation coefficients between STarT MSK total score and the validation measures confirmed the hypotheses and were significant.

Conclusion: The Israeli translation and validation of the STarT MSK suggest that it is a valid and reliable instrument. The STarT MSK discriminated low, medium, and high-risk groups.

Keywords: Musculoskeletal pain; Keele STarT MSK screening tool; Translation; Validation.

1 INTRODUCTION

Chronic musculoskeletal (MSK) pain is a widespread medical, social, and economic problem, characterized by pain and loss of function. (Woolf and Pfleger 2003). The prevalence of persistent MSK pain in the world population ranges from 25% to 32% (Wijnhoven, de Vet, and Picavet 2006). Evidence from the global burden of diseases studies suggests that low back pain (LBP) (the most prevalent MSK pain site) is one of the leading causes of years lived with disability in Western Europe and Australia (Murray et al. 2012). It is therefore imperative to reduce this burden by preventing the transition from early stages of MSK pain to persistent disability.

Clinical guidelines for LBP, which is the most common MSK problem, (de Campos 2017) recommend an assessment using a risk stratification approach that links to appropriate matched treatment options. This approach has been shown to improve primary care decision-making (Jull 2017; Hill et al. 2020). The potential of risk stratification approaches is that they can help to maximize the benefit of treatment, whilst reducing the potential for harm and increasing health care efficiency (Campbell et al. 2016).

The success in screening and matching for LBP led to the development and validation of the Keele STarT MSK Tool (STarT MSK), for use in patients with the five most common MSK pain areas seen in primary care (back, neck, shoulder, knee, and multisite pain) which has shown a moderate to good predictive and discriminative abilities (Dunn et al. 2017; Campbell et al. 2016; Dunn et al. 2021).

Objectives: In this study, we aimed to translate and validate the STarT MSK into Hebrew, according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) taxonomy (Mokkink et al. 2010). The psychometric properties assessed for reliability include test-retest and internal consistency. For construct validity, the hypotheses propose significant associations between the newly-translated Hebrew StarT MSK with measures of pain and disability as well as depression, anxiety, and fear-avoidance of exercise. We hypothesized that the STarT MSK would yield strong positive correlations with disability and pain, and a moderately positive correlation with depression, anxiety, and fear-avoidance of exercise.

2 METHODS

2.1 STarT MSK Tool

The STarT MSK is a prognostic questionnaire with ten-item assessing physical and psychological risk in MSK patients with back, neck, shoulder, knee, multisite pain (Campbell et al. 2016; Hill et al. 2020). Items include pain severity, anxiety, bothersomeness, disability, comorbid pain, catastrophizing, health issues, depression, fear of movement, and chronicity. The response to each item is a binary yes or no, except for the question on the severity of pain, which is scores on an 11 point scale in which 0=no pain, and 10=the worse pain. This scale is coded into 0-3. Responses are added to create the total score (range 0-12), which categorizes the patient's risk level: low risk = 0 to 4, medium risk=5 to 8, and high risk= 9 to 12 (Campbell et al. 2016).

2.2 Translation procedure

The translation was carried out in Israel with the permission of the original developers and followed the recommendations for best practices in questionnaire translation. (Beaton et al. 2000; "WHO | Process of Translation and Adaptation of Instruments" n.d.) The translation steps were: Three translators, who were fluent in both languages, had Hebrew as their mother tongue, translated the questionnaire from English to Hebrew, and compared the translations, to avoid ambiguous wording and discrepancies, which were resolved through discussion. Two translators were aware of the concepts being tested in the questionnaire, and one translator, representing the lay-person perspective, was unaware of the concepts.

Next, a translation back to English was performed by two new translators. Both translators had English as their mother tongue and were unaware of the concepts behind the questionnaire and did not understand the field.

An expert committee (including two physiotherapy students in their final year, a physical therapist, an MSK researcher, a psychologist specializing in research in pain) compared the content of the original, translated, and back-translated questionnaire. The advice was also provided by one of the original authors of the questionnaire. Discrepancies were resolved through discussion, and a pre-final Hebrew language version was developed.

The pre-final version was then discussed with ten Israeli participants with MSK pain, in order to check the understandings, comprehensiveness, and readability of the

translated version. No difficulties in comprehension were noted at this stage and a final version was produced.

2.3 Procedure for recruitment for the validation cohort

Individuals with MSK pain were recruited through opportunity sampling between November 2019 and April 2020. Inclusion criteria were adults (18+), fluency in Hebrew, and back pain, neck, shoulders, knees, and /or multisite MSK pain. Following participant consent, individuals were asked to complete a set of printed baseline questionnaires. For the test-retest assessment, the STarT MSK was completed again within a 1-week period.(Marx et al. 2003) The ethical review board of Ariel University in Israel approved the study (Number AU-HEA-NBA-2019104).

2.4 Test-retest reliability

Forty-five participants were included in the test-retest investigation. Participants were asked whether they had improved or not over the past week and were included only if symptoms had not changed.

2.5 Baseline Questionnaires

The correlations of the following measures with the STarT MSK was tested: 1) The 12-Item Short-Form Health Survey (SF-12) (Ware, Kosinski, and Keller 1996) as a measure of disability, 2) Numerical pain rating scale (NPRS)(Ferreira-Valente, Pais-Ribeiro, and Jensen 2011) for the most severe and average pain intensity (0= no pain and 10 =the worst pain), 3) Hospital Anxiety and Depression Scale (HADS)(Bjelland et al. 2002) as a measure of anxiety and depressive symptoms and 4) Part 1 of the Fear-Avoidance Beliefs Questionnaire (FABQ) which measures fear-avoidance of exercise. These questionnaires are frequently used in MSK research and formed the basis for validity testing in the original development and testing of the questionnaire (Campbell et al. 2016; Dunn et al. 2021). All the questionnaires are reliable and valid in Hebrew.("התכונות הפסיכומטריות של שאלון להערכת חרדה ודיכאון (HADS – Hospital Anxiety and Depression Scale) האגודה הישראלית לגרונטולוגיה | המתורגם לעברית" n.d.; Jacob et al. 2001) In addition, age, sex, weight, height, smoking habits, occupation,

and employment status (employed, unemployed, on sick leave, retired) were measured.

2.6 Sample size

The sample size was calculated with G*Power 3.1.9.4 using the z-test family to detect the correlation between two measures. The input parameters were as follows: for a two-tailed test, assuming a medium Cohen's *q* of 0.5, α=0.05, and β=0.95, the total sample size recommended was 117 participants, assuming a dropout rate of 25% we would like to recruit 150 participants.

2.7 Statistical analysis

Data analysis was performed using IBM SPSS Statistics version 25. Characteristics of the sample were described using frequencies and means with standard deviations, and standard error measurement. Normality was evaluated by looking at each variable's skewenss and kurtosis. The depression variable was right-skewed and a logarithmic transformation was applied to correct it. SF-12 was left-skewed and corrected with a squared transformation. The equal variances assumption was examined by the Levene test, which was insignificant for each variable examined except physical component score (PCS) and mental component score (MCS). Internal consistency was measured by calculating Cronbach's alpha for the STarT MSK scale. Test-retest reliability of STarT MSK scale, between the baseline and 1-week follow-up, was evaluated by calculating the intra-class correlation coefficient (ICC) for the total score, and corresponding risk groups (i.e., low, medium or high risk) and each question individually. For reliability, we carried out an Intraclass correlation coefficient (ICC), on each item between the first measurement and the second measurement, within one week, in 45 participants. A two-way mixed effect test-retest absolute agreement ICC was used (Koo and Li 2016). ICC values interpreted as follows: poor < 0.40, fair 0.40-0.59, good 0.60-0.74, and excellent 0.75-1.00 (Cicchetti 1994). Finally, the standard error of measurement (SEM) values was calculated based on the differences between times of measurement, as conducted in previous studies (Geerinck et al. 2019).

Construct validity was assessed by analyzing the correlations between the STarT MSK total score and reference standards including the NPRS, SF12, HADS, FABQ,

using Spearman’s correlation coefficients. The criteria for correlation values used was: weak <0.30, moderate 0.30-0.59, strong ≥0.60. (Fritz, Beneciuk, and George 2011).

3 RESULTS

3.1 Linguistic translation

During the forward and backward translations, we found minor linguistic differences in the following items: item 2 (“manage your pain”), item 5 (“joint or muscle pain”), and item 9 ("unsafe"). Translation of item 2 was challenging because the concept of pain management in Hebrew is not commonly used, and the term "deal with" was substituted. The combination of " joint or muscle pain" in item 5 was not clear to a person with no medical background, and was replaced by "joint or muscle problems”.

3.2 Participants

Characteristics of the Israeli participants who completed the set of questionnaires (n = 153), stratified by STarT MSK risk groups are described in Table 1. Sixty-seven participants (43.8%) were in the low-risk group, 72 (47.1%) in the medium risk, and 14 (9.2%) at high risk. In the categorization of pain areas, back pain was the most common pain, followed by multi-site pain. A positive correlation was found between participants' age and risk groups. Older participants were more likely to belong to the higher risk group. The body mass index was higher in the high-risk group. ANOVA revealed that the mean scores of SF12, NPRS, FABQ, and HADS were significantly different across STarT MSK risk groups (Table 1). Post-hoc analyses were used to determine significant differences between every two risk groups.

Table 1 around hear

3.3 Test-retest reliability and internal consistency

Teat-retest reliability was carried out among 45 patients. The STarT MSK indicated excellent test-retest reliability, with an ICC total score of 0.92 (95 % CI 0.85 -0.95). The values for the individual items also demonstrated good test-retest reliability (Table 2). SEM values are also presented in table 2.

Internal consistency was measured by calculating Cronbach's alpha for the MSK scale (α=0.612).

Table 2 around hear

3.4 Validity

The STarT MSK correlated moderately with the reference scales, with Spearman correlations ranging from 0.382 to 0.639 (Table 3). Strong correlation was found between STarT MSK total score and disability measures (SF 12) (r_s = 0.611, p = 0.001), physical component score (r_s = 0.639, p = 0.001), and Average pain (r_s = 0.618, p = 0.001).

Table 3 around hear

4 DISCUSSION

The Hebrew version of STarT MSK was found to be reliable and had good construct validity among Israeli participants with common MSK pain problems. The results indicated excellent test-retest reliability (0.92) and showed a good construct validity (0.382 to 0.639). This is the second translation and validation in a foreign language after the Dutch validation.(Broek et al. 2021).

The risk group distribution in our study was similar to the distribution in the STarT MSK pilot cluster randomized controlled trial (Hill et al. 2020) and pretty similar to the UK validation study (Dunn et al. 2017; 2021). Our cohort had a slight shift toward low risk at the expense of high risk as in the Dutch trial. (Broek et al. 2021) In our study, 67 participants (43%) were categorized as low risk, while the UK cohorts (respectively 38% and 30%). The medium risk was the same (respectively 47%, 52% and 51%), and our high risk was 9% compared to UK cohorts (respectively 10% and 19%). In the Dutch trial, only 4 patients (2.8%) were categorized as high risk. (Broek et al. 2021).

The use of this questionnaire can form part of the adaptation of the recommendations outlined in the Lancet series on back pain (Buchbinder et al. 2018; Foster et al. 2018; Hartvigsen et al. 2018), which include the development and implementation of strategies for early identification of patients who are at risk for the transition to chronic pain.

There are several limitations to the study. Of importance is the lack of a timeline to test whether the measure can predict the outcome later. In addition, participants were not necessarily seeking care for their pain. Finally, The study does not inform on the effectiveness of screening and matching strategies in MSK pain.

Conclusion

The Israeli translation and validation of the STarT MSK suggest that it is a valid and reliable instrument. The STarT MSK discriminated low, medium, and high-risk groups. Its predictive validity should now be examined within a clinical setting.

Conflict of Interest – None declared

Ethical Approval - This study received approval (number AU-HEA-NBA-2019104) from the ethical review board of Ariel University in Israel.

Funding – None declared

Acknowledgments - We thank the physiotherapist's students Maya Botbol, Danielle Fioriti, Shani Farkash, and Yael kapach, for the dedicated work in the interviewing and coding process. We thank Pavel Fridlin for the statistical analysis.

References

Beaton, D. E., C. Bombardier, F. Guillemin, and M. B. Ferraz. 2000. “Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures.” *Spine* 25 (24): 3186–91.

Bjelland, Ingvar, Alv A. Dahl, Tone Tangen Haug, and Dag Neckelmann. 2002. “The Validity of the Hospital Anxiety and Depression Scale. An Updated Literature Review.” *Journal of Psychosomatic Research* 52 (2): 69–77.

Broek, Anke G. van den, Corelien J. J. Kloek, Martijn F. Pisters, and Cindy Veenhof. 2021. “Validity and Reliability of the Dutch STarT MSK Tool in Patients with Musculoskeletal Pain in Primary Care Physiotherapy.” *PLOS ONE* 16 (3): e0248616. https://doi.org/10.1371/journal.pone.0248616.

Buchbinder, Rachelle, Maurits van Tulder, Birgitta Öberg, Lucíola Menezes Costa, Anthony Woolf, Mark Schoene, Peter Croft, et al. 2018. “Low Back Pain: A Call for Action.” *The Lancet* 391 (10137): 2384–88. https://doi.org/10.1016/S0140-6736(18)30488-4.

Campbell, Paul, Jonathan C. Hill, Joanne Protheroe, Ebenezer K. Afolabi, Martyn Lewis, Ruth Beardmore, Elaine M. Hay, et al. 2016. “Keele Aches and Pains Study Protocol: Validity, Acceptability, and Feasibility of the Keele STarT MSK Tool for Subgrouping Musculoskeletal Patients in Primary Care.” *Journal of Pain Research* 9: 807. https://doi.org/10.2147/JPR.S116614.

Campos, Tarcisio Folly de. 2017. “Low Back Pain and Sciatica in over 16s: Assessment and Management NICE Guideline [NG59].” *Journal of Physiotherapy* 63 (2): 120. https://doi.org/10.1016/j.jphys.2017.02.012.

Cicchetti, Domenic V. 1994. “Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology.” *Psychological Assessment* 6 (4): 284–90. https://doi.org/10.1037/1040-3590.6.4.284.

Dunn, Kate M., Paul Campbell, Ebenezer K. Afolabi, Martyn Lewis, Danielle van der Windt, Jonathan C. Hill, Christian D. Mallen, Joanne Protheroe, Elaine M. Hay, and Nadine E. Foster. 2017. “176. REFINEMENT AND VALIDATION OF THE KEELE START MSK TOOL FOR MUSCULOSKELETAL PAIN IN PRIMARY CARE.” *Rheumatology* 56 (suppl_2). https://doi.org/10.1093/rheumatology/kex062.177.

Dunn, Kate M., Paul Campbell, Martyn Lewis, Jonathan C. Hill, Danielle A. van der Windt, Ebenezer Afolabi, Joanne Protheroe, et al. 2021. “Refinement and Validation of a Tool for Stratifying Patients with Musculoskeletal Pain.” *European Journal of Pain (London, England)* 25 (10): 2081–93. https://doi.org/10.1002/ejp.1821.

Ferreira-Valente, Maria Alexandra, José Luís Pais-Ribeiro, and Mark P. Jensen. 2011. “Validity of Four Pain Intensity Rating Scales.” *Pain* 152 (10): 2399–2404. https://doi.org/10.1016/j.pain.2011.07.005.

Foster, Nadine E, Johannes R Anema, Dan Cherkin, Roger Chou, Steven P Cohen, Douglas P Gross, Paulo H Ferreira, et al. 2018. “Prevention and Treatment of Low Back Pain: Evidence, Challenges, and Promising Directions.” *The Lancet* 391 (10137): 2368–83. https://doi.org/10.1016/S0140-6736(18)30489-6.

Fritz, Julie M., Jason M. Beneciuk, and Steven Z. George. 2011. “Relationship between Categorization with the STarT Back Screening Tool and Prognosis for People Receiving Physical Therapy for Low Back Pain.” *Physical Therapy* 91 (5): 722–32. https://doi.org/10.2522/ptj.20100109.

Geerinck, Anton, Vidmantas Alekna, Charlotte Beaudart, Ivan Bautmans, Cyrus Cooper, Fabiana De Souza Orlandi, Jerzy Konstantynowicz, et al. 2019. "Standard Error of Measurement and Smallest Detectable Change of the Sarcopenia Quality of Life (SarQoL) Questionnaire: An Analysis of Subjects from 9 Validation Studies." *PLOS ONE* 14 (4): e0216065. https://doi.org/10.1371/journal.pone.0216065.

Hartvigsen, Jan, Mark J. Hancock, Alice Kongsted, Quinette Louw, Manuela L. Ferreira, Stéphane Genevay, Damian Hoy, et al. 2018. "What Low Back Pain Is and Why We Need to Pay Attention." *Lancet (London, England)* 391 (10137): 2356–67. https://doi.org/10.1016/S0140-6736(18)30480-X.

Hill, J. C., S. Garvin, Y. Chen, V. Cooper, S. Wathall, B. Saunders, M. Lewis, et al. 2020. "Stratified Primary Care versus Non-Stratified Care for Musculoskeletal Pain: Findings from the STarT MSK Feasibility and Pilot Cluster Randomized Controlled Trial." *BMC Family Practice* 21 (1): 30. https://doi.org/10.1186/s12875-019-1074-9.

Jacob, T., M. Baras, A. Zeev, and L. Epstein. 2001. "Low Back Pain: Reliability of a Set of Pain Measurement Tools." *Archives of Physical Medicine and Rehabilitation* 82 (6): 735–42. https://doi.org/10.1053/apmr.2001.22623.

Jull, Gwendolen. 2017. "Biopsychosocial Model of Disease: 40 Years on. Which Way Is the Pendulum Swinging?" *British Journal of Sports Medicine* 51 (16): 1187–88. https://doi.org/10.1136/bjsports-2016-097362.

Koo, Terry K., and Mae Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15 (2): 155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

Marx, Robert G., Alia Menezes, Lois Horovitz, Edward C. Jones, and Russell F. Warren. 2003. "A Comparison of Two Time Intervals for Test-Retest Reliability of Health Status Instruments." *Journal of Clinical Epidemiology* 56 (8): 730–35. https://doi.org/10.1016/s0895-4356(03)00084-2.

Mokkink, Lidwine B., Caroline B. Terwee, Dirk L. Knol, Paul W. Stratford, Jordi Alonso, Donald L. Patrick, Lex M. Bouter, and Henrica CW de Vet. 2010. "The COSMIN Checklist for Evaluating the Methodological Quality of Studies on Measurement Properties: A Clarification of Its Content." *BMC Medical Research Methodology* 10 (1): 22. https://doi.org/10.1186/1471-2288-10-22.

Murray, Christopher J. L., Theo Vos, Rafael Lozano, Mohsen Naghavi, Abraham D. Flaxman, Catherine Michaud, Majid Ezzati, et al. 2012. "Disability-Adjusted Life Years (DALYs) for 291 Diseases and Injuries in 21 Regions, 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010." *The Lancet* 380 (9859): 2197–2223. https://doi.org/10.1016/S0140-6736(12)61689-4.

Ware, J., M. Kosinski, and S. D. Keller. 1996. "A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity." *Medical Care* 34 (3): 220–33. https://doi.org/10.1097/00005650-199603000-00003.

"WHO | Process of Translation and Adaptation of Instruments." n.d. WHO. Accessed March 19, 2019. https://www.who.int/substance_abuse/research_tools/translation/en/?fbclid=IwAR3srUm1HjhDnmstrSwWQkNkftH-hCgkoDZpsh8Xe51lLaVx-dx4YIn49k8.

Wijnhoven, Hanneke A. H., Henrika C. W. de Vet, and H. Susan J. Picavet. 2006. "Prevalence of Musculoskeletal Disorders Is Systematically Higher in Women than in Men." *The Clinical Journal of Pain* 22 (8): 717–24. https://doi.org/10.1097/01.ajp.0000210912.95664.53.

Woolf, Anthony D., and Bruce Pfleger. 2003. "Burden of Major Musculoskeletal Conditions." *Bulletin of the World Health Organization* 81 (9): 646–56.

"התכונות הפסיכומטריות של שאלון להערכת חרדה ודיכאון (HADS – Hospital Anxiety and Depression Scale) האגודה הישראלית לגרונטולוגיה | המתורגם לעברית." n.d. Accessed November 15, 2021. http://www.gerontology.org.il/

Table 1. Descriptive baseline characteristics according to risk groups in STarT MSK Tool

Variable	**All patients N=153**	**Low risk N=67, 43.8%**	**Medium risk N=72, 47.1%**	**High risk N=14, 9.2%**	**P**
Age	40±16.3	36.9±13.8[a]	40.4±16.4[b]	52.8±21.3[a,b]	0.04*
Gender, n (%) women	80 (52.3%)	28 (35%)	44 (55%)	8 (10%)	0.052
Gender, n (%) men	73 (47.7%)	39 (53.4%)	28 (38.4%)	6 (8.2%)	
Symptom duration (years)	6.2±8.0	6.0 ±7.9	5.5 ± 7.3	10.8 ±10.3	0.07
BMI	24.9±3.8	24.1±2.8[a]	25.0±4.2[b]	28.8±4.0[a,b]	0.001*
Smoking, n (%)	20 (13.1%)	9 (13.4%)	9 (12.5%)	2 (14.2%)	0.607
Work status n (%)					0.718
Employed	115 (75.2%)	53 (79.1%)	54 (75%)	8 (57.1%)	
Unemployed	17 (11.1%)	5 (7.4%)	8 (11%)	4 (28.5%)	
Sick leave	1 (0.7%)	1 (1%)	0	0	
Retired	8 (5.2%)	1	5	2 (0.1%)	
Other	12 (7.8%)	7 (10.4%)	5 (6.9%)	0	
Pain area, n (%)					0.058
Back	64 (41.8%)	29 (43.2%)	31 (40.2%)	4 (28.5%)	
Neck	9 (5.9%)	7 (10.4%)	2 (2.7%)	0	
Shoulder	15 (9.8%)	9 (13.4%)	6 (8.3%) 0	0	
Knee	14 (9.2%)	4 (5.9%)	8 (11.1%)	2 (14.2%)	
Multi-region	50 (32.7%)	17 (25.3%)	25 (34.7%)	8 (57.1%)	
MSK total	5.0±2.3	2.9±1.0[a]	6.1±1.0[a]	9.8±0.7[a]	0.001*
SF 12 (0-100)	66.1±22	77.6±13.8[a]	62.5±19.6[a]	29.7±20.2[a]	0.001*
PCS 12	66.2±26.1	81.3±14[a]	60.6±24[a]	22.6±20.5[a]	0.001*
MCS 12	66.0±21.1	73.8±17.9[a]	64.4±19.6[a]	36.9±21.6[a]	0.001*

Severe pain (0-10)	5.6±2.1	4.5±1.9[a]	6.1±1.7[a]	8.2±1.5[a]	0.001*
Average pain (0-10)	4.1±2.2	2.8±1.6[a]	4.7±1.8[a]	7.2±1.8[a]	0.001*
Fear-avoidance exercise (0-24)	11.9±6.6	9.3±5.3[a,b]	13.4±6.8[a]	16.4±6.9[b]	0.001*
Anxiety (0-21)	5.5±3.8	4.2±2.8[a]	6.0±3.7[a]	10.0±5.1[a]	0.001*
Depression (0-21)	3.4±3.2	2.0±2.1[a]	3.7±2.6[a]	9.1±3.8[a]	0.001*

* Values represent means±standard deviations unless otherwise indicated. BMI = body mass index, SF 12 - Short Form 12-item Survey ; PCS 12 - physical component score; MCS 12 - mental component score.

Table 2. The intraclass correlation coefficient (ICC), and 95 % confidence intervals (CI) for the test-retest reliability of translated STarT MSK Screening Tool (N=45)

	ICC	**95 % CI**	**SEM**
Total score	0.921	0.858 – 0.957	0.760
1: how intense was your pain	0.90	0.821 - 0.946	0.368
2: manage your pain condition	0.732	0.514 - 0.852	0.316
3: bothered a lot by your pain	0.654	0.375 – 0.813	0.302
4: walk short distances	0.872	0.771 - 0.937	0.213
5: troublesome joint or muscle	0.756	0.545 – 0.861	0.310
6: condition will last a long time	0.803	0.606 – 0.884	0.225
7: health problems	0.879	0.783 – 0.934	0.184
8: Depressive mood	0.826	0.686 – 0.905	0.259
9: Fear avoidence exrecise	0.722	0.491 – 0.884	0.234
10: pain problem for 6 months +	0.778	0.598 – 0.881	0.241

Table 3. Spearman correlation between STarT MSK total score, and the other questionnaires (n=153)

Variable	STarT MSK total score	P-value
Severe pain	0.577	0.001
Average pain	0.618	0.001
SF 12	0.611	0.001
PCS	0.639	0.001
MCS	0.462	0.001
FABQ exercise	0.389	0.001
Anxiety	0.382	0.001
Depression	0.507	0.001

MSK – Musculoskeletal; SF 12 - Short Form 12-item health survey (quality-of-life measures); PCS - physical component score; MCS - mental component score; FABQ – Fear-avoidance beliefs questionnaire