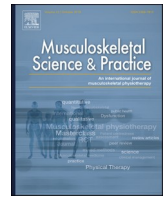




Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp

Original article

STarT back tool retained its predicting abilities in patients with acute and sub-acute low back pain patients after a transcultural adaptation and validation to Hebrew

Noa Ben Ami ^{a,*}, Asaf Weisman ^{b,c}, Tomer Yona ^d, Anat Shashua ^{c,e}, Jonathan Hill ^f, Tamar Pincus ^g

^a Department of Physiotherapy, Ariel University, Ariel, Israel

^b The Spinal Research Laboratory, Department of Physical Therapy, The Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel-Aviv University, Israel

^c Clalit Health Services, Department of Physiotherapy, Tel Aviv, Israel

^d The Israeli Physiotherapy Society, Tel-Aviv, Israel

^e Department of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

^f Department of Physiotherapy, Keele University, United Kingdom

^g Department of Psychology, Royal Holloway University of London, London, United Kingdom



ARTICLE INFO

Keywords:

Low back pain
Prognosis
STarT back screening tool
Translation
Validation

ABSTRACT

Background: The STarT Back Screening Tool (SBT) distributes low back pain (LBP) patients into three prognostic groups for stratified care. This approach has demonstrated beneficial clinical and cost-effectiveness.

Objectives: To translate and validate the SBT by investigating its psychometric properties among Israelis with acute and sub-acute LBP, and to evaluate its ability to predict disability after three months.

Design: Prospective study.

Method: The SBT was transcultural adapted into Hebrew using published guidelines. A total of 150 patients receiving physical therapy for acute or subacute LBP were administered the SBT. Clinical outcomes included the Roland-Morris Disability Questionnaire (RMDQ), the Hospital Anxiety and Depression Scale (HADS), the Fear-Avoidance Beliefs Questionnaire (FABQ) and a numerical pain rating scale (NPRS), collected by an independent interviewer by phone at the start of the physical therapy treatment and after three months.

Results: The test-retest reliability of the SBT total score and psychosocial subscale were excellent (intraclass correlation coefficient 0.89 and 0.82). Spearman's correlation coefficient between SBT total score and RMDQ was 0.82, HADS (Anxiety 0.66, Depression 0.76), FABQ (exercise 0.53), NPRS (severe pain 0.48, average pain 0.53). The SBT baseline score showed excellent predictive abilities in discriminating poor disability after three months (ROC curve = 0.825, $P < 0.001$, 95% CI 0.756–0.894).

Conclusion: The Israeli translation and cross-cultural adaptation of the SBT is a valid and reliable instrument. The SBT discriminated low, medium and high-risk groups, and predicts disability after three months.

1. Introduction

Low back pain (LBP) is the leading cause of disability worldwide, and its burden is growing due to increased population and life span (Buchbinder et al., 2018). Most cases of LBP are non-specific, meaning, without any identified spinal pathology (Foster et al., 2018). Poor understanding of etiology and prognosis complicates the selection of effective treatment for non-specific LBP. The STarT Back approach has

been suggested to improve clinical outcomes, where patients are selected for treatment by prognostic classification (Foster et al., 2014). The STarT Back is a cost-effective, evidence-based approach, utilizing a screening tool that has demonstrated clinical utility in predicting outcomes of patients with LBP (mainly tendency to chronicity) (Hill et al., 2011). The STarT Back screening tool (SBT) supports primary care decision making to allocate patients with acute and subacute non-specific LBP into three prognostic groups (low, medium and high risk) with

* Corresponding author.

E-mail address: noaba@ariel.ac.il (N. Ben Ami).

<https://doi.org/10.1016/j.msksp.2020.102134>

Received 18 November 2019; Received in revised form 23 January 2020; Accepted 19 February 2020

Available online 20 February 2020

2468-7812/© 2020 Elsevier Ltd. All rights reserved.

matched treatment pathways. The SBT has repeatedly proven to be a valid and reliable instrument and has been well accepted by both patients and clinicians (Hill et al., 2008). Recently, the NICE guidelines (de Campos, 2017) for LBP and sciatica assessment and management, recommended using risk assessment and risk stratification tools such as the STarT Back.

The original SBT was developed in English and been validated and translated into several languages. (Piironen et al., 2016; Bruyère et al., 2012; Morsø et al., 2011; Abedi et al., 2015; Pilz et al., 2014; Aebischer et al., 2015). The SBT was developed to screen patients with acute LBP at early stages in order to predict the transition to chronic LBP. Hence, we selected only patients in the early stages (acute and sub-acute) for our sample. This study aimed to translate and validate the SBT by investigating its psychometric properties among Israelis with acute and sub-acute LBP, and to evaluate its ability to predict disability after three months. We hypothesized that the SBT tool would yield high positive correlations with disability, depression, and anxiety, and moderately positively correlated with pain and fear avoidance of exercise.

2. Methods

2.1. STarT Back Screening Tool

The SBT is a 9-item questionnaire that classifies patients into three risk categories, according to the presence of physical and psychosocial risk factors for persistent LBP symptoms (Hill et al., 2008). The nine items divided into physical and psychosocial subscales. The physical subscale includes four items: referred leg pain, disability (2 items), and comorbid pain. The psychosocial subscale includes five items: bothersomeness, catastrophizing, fear, anxiety, and depression. Each item scored as positive or negative. All points are added together for the total score (range 0–9). Items 5 to 9 form the psychosocial subscale (range 0–5). The Subscale scores are used to categorize the patient's risk level, low risk if the total score from both subscales is 0–3, high risk if the psychosocial subscale score is 4 or 5, and a medium risk for all others (Hill et al., 2008).

2.2. Translation procedure

The translation was done with the permission of Keele university researchers, who developed the SBT. The translation process followed the recommendations for best practice in questionnaire translation. (Beaton et al., 2000; "WHO | Process of Translation and Adaptation of Instruments" n.d.). The SBT was initially translated forwards to Hebrew and then back to English, by two independent pairs of translators fluent in both English and Hebrew. The original translators (English to Hebrew) spoke English fluently and had Hebrew as their mother tongue. They were all aware of the concepts behind the questionnaire. Discrepancies were resolved through mutual discussion. The second stage translators (Hebrew to English), had English as their mother tongue, were fluent in Hebrew and were unaware of the concepts behind the questionnaire. The expert committee consisted of two physiotherapists who are specialists both in LBP and in LBP research and one more psychologist specializing in research in psychological aspects (and their measurement) in pain populations. After comparing the content of the original and backward translated version, the observed differences were discussed, and a pre-final Hebrew language version was developed. The pre-final version was then discussed with ten patients with LBP, who commented on the burden, ease of understanding, comprehensiveness, and readability of the translated version. No difficulties in comprehension were noted at this stage and a final version was produced.

2.3. Baseline questionnaires

We tested the correlations of the following measures with the SBT: 1) the Roland-Morris Disability Questionnaire (RMDQ) (Chiarotto et al.,

2016) as a measure of back disability, 2) Numerical pain rating scale (NPRS) (Ferreira-Valente et al., 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) Fear-Avoidance Beliefs Questionnaire (FABQ) which measures fear-avoidance of exercise and work. We have excluded the section of fear-avoidance of work from the FABQ questionnaire as it is not available on the SBT (there is no question about work in SBT). These questionnaires were selected, as they are considered appropriate for studying the SBT's construct validity and are frequently used in LBP research (Abedi et al., 2015; Hill et al., 2008). Also, age, sex, weight, height, smoking habits, occupation, and employment status (employed, unemployed, on sick leave, retired) were included in the baseline data collection questionnaire.

2.4. Procedure for recruitment

Patients with LBP were recruited between March 2018 and June 2019, from three large outpatient physical therapy clinics of Clalit Health Services. Inclusion criteria were patients experiencing acute (less than six weeks) or subacute (6–12 weeks) LBP, with or without radicular pain, age of at least 18 years, and the ability to understand the Hebrew language. Exclusion criteria were chronic pain (more than 12 weeks of pain) and suspected red flags.

Their treating physiotherapist recruited patients at their initial assessment. Potential participants were asked for their permission to undergo a telephone interview. After patients gave their consent, a researcher called them by telephone to interview and fill out their baseline questionnaires. For the test-retest assessment, the same researcher called within one week to fill out the SBT questionnaire. Only patients that reported no change in their condition participated in the test-retest assessment. Three months after the first contact, the same researcher called each patient and filled out all the questionnaires again. All in all, the entire validation process for all versions of the questionnaire was conducted by telephone. The ethical review board of Clalit Health Services approved the study (No. 0157-17-COM2).

2.5. Test-retest reliability

Fifty patients were included in the test-retest investigation. Patients were asked whether they had improved or not over the past week and were included only if they reported 'no change' in their symptoms.

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 total SBT and Disability (which is the primary outcome on the SBT trial) (Hill et al., 2011). The input parameters were as follows: for a two-tailed test, assuming a medium effect size of 0.5, $\alpha = 0.05$ and $\beta = 0.95$, the total sample size recommended was 147 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 skewness and kurtosis. The equal variance was examined by the Levene test, which was insignificant for each variable examined. Internal consistency was measured by calculating Cronbach's alpha for the SBT scale. Test-retest reliability of SBT, between the baseline and 1-week follow-up, was evaluated by calculating the intra-class correlation coefficient (ICC) for the total score, psychosocial subscale, and corresponding risk groups (i. e., low, medium or high risk; and low or high psychosocial score) 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 50 patients. A two-way mixed effect test-retest absolute agreement ICC was used (Koo and Mae 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, standard error of measurement (SEM) values were 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 SBT (total score and psychosocial subscale) and reference questionnaires (NPRS, RMDQ, 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 et al., 2011). The difference between risk groups was measured by a two-way repeated measure analysis of variance (ANOVA) to verify change from baseline to 3-month clinical outcomes. We considered the factors: time (baseline and post three months) and risk category (low, medium and high), according to the baseline risk category.

For the ability to predict patients with poor disability outcomes after three months, we used the receiver operating characteristic curve (ROC curve). This was performed by calculating the ROC curve for the tool’s overall scores against a reference standard cut point for poor disability after three months (RMDQ ≥7). The ROC Curve values interpreted as follows: not acceptable < 0.5, acceptable 0.7–0.8, excellent 0.8–0.9, an outstanding 0.9–1.00(Mandrekar, 2010).

3. Results

3.1. Linguistic translation

During the forward and backward translations, we found minor linguistic differences in the following items: item 1 (“spread down my legs”), item 1 and 2 (“at some time”), item 5 (“it’s not really safe”), item 6 (“worrying thoughts”) and item 9 (“extremely”). Translation of item 1 was challenging because not all people understand the meaning of “radiate” down the leg in Hebrew. The term ‘Radiate’ was noted by the research team a-priori as a problematic term, that did not translate well. It was removed before the questionnaire was tested with patients. Therefore, it was agreed to use the term “spread down my legs.” Additionally, the term “at some time” required changing in Hebrew. In the translation of item 5, “it’s not really safe” there was much consideration. After consulting with the developer of SBT, who recommended emphasizing the severity of the feeling, it was translated as “dangerous.” The same considerations applied regarding “worrying thoughts” and “extremely.”

3.2. Participants

Characteristics of the patients who completed the first set of questionnaires (n = 150) with the results stratified by SBT risk groups are described in Table 1. Forty-eight patients (32.5%) were allocated to the low-risk group, 75 (49.7%) in the medium risk, and 27 (17.9%) at the high risk. ANOVA revealed that the mean scores of NPRS, RMDQ, FABQ, and HADS were significantly different across SBT risk groups (Table 1) at baseline. Post-hoc analyses were used to determine significant differences between every two risk groups.

3.3. Test-retest reliability and psychometric properties

Test-retest reliability was carried out among 50 patients. The SBT indicated excellent test-retest reliability, with an ICC total score of 0.89 (95% CI 0.84–0.93) and for the psychosocial subscale of 0.82 (95% CI 0.735–0.886). The values for the individual items also demonstrated excellent test-retest reliability (Table 2). SEM values are also presented in Table 2. Internal consistency was measured by calculating Cronbach’s

Table 1

Descriptive baseline characteristics according to risk groups in STaRT Back Tool.

Variable	All patients N = 150	Low risk N = 48, 32%	Medium risk N = 75, 50%	High risk N = 27, 18%	P
Age	53.5 ± 19.5	51.7 ± 21.4	55.7 ± 18.1	50.7 ± 19.5	0.381
Gender, n (%)	87 (58%)	25 (51%)	46 (61.3%)	16 (59.3%)	0.597
women					
Symptom duration (week)	6.5 ± 3.6	7.3 ± 3.6 ^a	6.4 ± 3.5	5.1 ± 3.6 ^a	0.03 ^a
BMI	26.0 ± 4.4	24.5 ± 3.8 ^a	27.1 ± 4.4 ^a	25.6 ± 4.6	0.004 ^a
Smoking, n (%)	32 (21.3%)	6 (12.2%) ^a	16 (21.3%)	10 (37%) ^a	0.04 ^a
Work status					0.253
Employed, n (%)	79 (52.7%)	32 (65.3%)	35 (46.7%)	12 (44.4%)	
Unemployed, n (%)	9 (6%)	1 (2%)	5 (6.7%)	3 (11.1%)	
Sick leave, n (%)	18 (12%)	2 (4.1%)	11 (14.7%)	5 (18.5%)	
Retired, n (%)	44 (29.3%)	13 (26.5%)	24 (32%)	7 (25.9%)	
Disability RMDQ (0–23)	13.05 ± 5.6	7.0 ± 3.8 ^{a, b}	15.4 ± 3.8 ^a	17.2 ± 3.5 ^{a, b}	0.001 ^a
Severe pain (0–10)	7.5 ± 2.3	6.3 ± 2.2 ^{a, b}	7.9 ± 2.1 ^a	8.7 ± 2.1 ^{a, b}	0.001 ^a
Average pain (0–10)	6.0 ± 2.1	4.8 ± 1.8 ^{a, b}	6.3 ± 1.9 ^a	7.1 ± 2.1 ^{a, b}	0.001 ^a
Fear avoidance exercise (0–24)	9.6 ± 6.3	5.5 ± 4.8 ^a	10.3 ± 5.9 ^a	15.2 ± 5.0 ^a	0.001 ^a
Anxiety (0–21)	6.1 ± 5.1	2.2 ± 2.9 ^a	7.2 ± 5.0 ^a	10.0 ± 4.4 ^a	0.001 ^a
Depression (0–21)	5.7 ± 4.7	1.8 ± 1.7 ^a	6.6 ± 4.5 ^a	9.9 ± 4.1 ^a	0.001 ^a

^a Values represent means±standard deviations unless otherwise indicated. BMI = body mass index, RMDQ = Roland-Morris Disability Questionnaire. Super-script letters denote significant differences between every two risk groups.

alpha for the SBT scale (α = 0.698).

The SBT total score and the psychosocial subscale correlated moderately with the reference scales, with Spearman correlations ranging from 0.443 to 0.827 (Table 3). Strong correlation was found between SBT and disability (r_s = 0.827, p = 0.001), anxiety (r_s = 0.666, p = 0.001) and depression (r_s = 0.768, p = 0.001). However, severe pain, average pain and fear-avoidance from exercise were only moderately correlated with SBT (r_s = 0.485, p = 0.001; r_s = 0.53, p = 0.001; r_s = 0.536, p = 0.001, respectively).

3.4. Ability to predict clinical outcomes after three months

Comparisons between baseline and final mean scores of all questionnaires, among each risk group revealed significant differences (Table 4). The proportion of patients with disabling back pain at 3

Table 2

The intraclass correlation coefficient (ICC), and 95% confidence intervals (CI) for the test-retest reliability of translated STaRT Back Screening Tool (N = 50).

	ICC	95% CI	SEM
Total score	0.890	0.840–0.930	0.378
Psychosocial subscale	0.82	0.735–0.886	0.408
1: Referred leg pain	0.92	0.862–0.956	0.174
2: Comorbid pain	0.980	0.965–0.989	0.099
3: Difficulties in walking	0.941	0.895–0.966	0.139
4: Difficulties in dressing	0.955	0.921–0.975	0.139
5: Fear of physical activity	0.882	0.792–0.933	0.174
6: Anxiety	0.960	0.930–0.977	0.139
7: Pain catastrophizing	0.907	0.835–0.947	0.169
8: Depressive mood	1		0
9: Overall impact of pain	0.956	0.923–0.975	0.139

Table 3

Spearman correlation between STarT Back total score, subscale score and the other questionnaires (n = 150).

Variable	SBT Total score	SBT subscale score	P-value
Severe pain	0.485	0.443	0.001
Average pain	0.53	0.442	0.001
RMDQ	0.827	0.643	0.001
FABQ exercise	0.536	0.528	0.001
Anxiety	0.666	0.669	0.001
Depression	0.768	0.717	0.001

SBT – Start Back tool; RMDQ - Roland-Morris Disability Questionnaire; FABQ – Fear-avoidance beliefs questionnaire.

months follow-up was 8/47 (17%) in the low-risk group, 45/67 (67%) in the medium-risk group, and 21/25 (84%) in the high-risk group. The change score in the clinical outcomes after three months, according to the risk category at baseline, are displayed in Table 5. Significant differences were found between baseline and post three months and risk category (low, medium and high) by time. However, there were no significant differences by group (Table 5), except for the variables anxiety (P = 0.007) and depression (P = 0.002). There was a main effect for time on depression and anxiety. Further analysis revealed an interaction of time and risk group, revealing no significant change between time 1 and time 2 in the low risk, and significant changes in group 2 and 3 (Table 5).

ROC curve analysis demonstrated excellent discriminant abilities of the SBT baseline score for predicting poor disability at three months across the different risk groups (AUC = 0.825, 95% CI 0.756–0.894) (Fig. 1).

4. Discussion

The translation of the Israeli version of the SBT showed excellent repeatability and moderate to high correlation with other scales among Hebrew speakers and indicated an excellent predictive ability to detect disability at three months.

The reliability of the Hebrew version of the SBT is on par with other formal translated versions of the questionnaire. The translation process in our study was similar to the one used in other studies (Morsø et al., 2011; Bruyère et al., 2012; Piironen et al., 2016; Abedi et al., 2015; Pilz et al., 2014; Aebischer et al., 2015). The instruments we used in this study were similar to the original version except for FABQ. In the original version, the Tampa Scale for Kinesiophobia (TSK) for fear of movement avoidance was used (Hill et al., 2008). Previous studies have found similar results with different questionnaires testing fear avoidance. (Abedi et al., 2015; Aebischer et al., 2015). The correlations with RMDQ and HADS were similar to a previous study (Abedi et al., 2015). Previous studies reported similar results with different questionnaires testing fear avoidance and Pain. (Abedi et al., 2015; Aebischer et al., 2015).

Although the SBT has been adapted and translated to several different languages, its predictive abilities have been tested only for English (Hill et al., 2008; Fritz et al., 2011; Suri et al., 2018), Dutch (Bier

Table 4

Clinical outcomes at baseline and three months, according to risk category at baseline.

Variable	All patients N = 139		Low risk N = 47 (33.8%)		Medium risk N = 67 (48.2%)		High risk N = 25 (17.9%)	
	baseline	3 months	baseline	3 months	baseline	3 months	baseline	3 months
Severe pain	7.4 ± 2.3	5.0 ± 3.2	6.3 ± 2.2	3.9 ± 3.1	7.9 ± 2.1	5.0 ± 3.1	8.7 ± 2.1	7.0 ± 2.8
Average pain	6.0 ± 2.1	4.1 ± 3.0	4.8 ± 1.8	2.7 ± 2.5	6.3 ± 1.9	4.3 ± 2.9	7.1 ± 2.1	6.1 ± 2.5
RMDQ	12.9 ± 5.7	9.5 ± 6.6	7.0 ± 3.8	4.4 ± 4.3	15.4 ± 3.8	11.1 ± 6.3	17.2 ± 3.5	13.8 ± 5.5
FA exercise	9.6 ± 6.5	6.9 ± 6.1	5.5 ± 4.8	4.3 ± 5.5	10.3 ± 5.9	7.1 ± 5.8	15.2 ± 5.0	11.0 ± 5.5
Anxiety	5.8 ± 5.0	3.9 ± 4.4	2.2 ± 2.9	1.4 ± 2.6	7.2 ± 5.0	4.2 ± 4.3	10.0 ± 4.4	7.7 ± 4.2
Depression	5.5 ± 4.8	3.8 ± 4.1	1.8 ± 1.7	1.4 ± 2.2	6.6 ± 4.5	4.3 ± 4.2	9.9 ± 4.1	7.0 ± 4.3

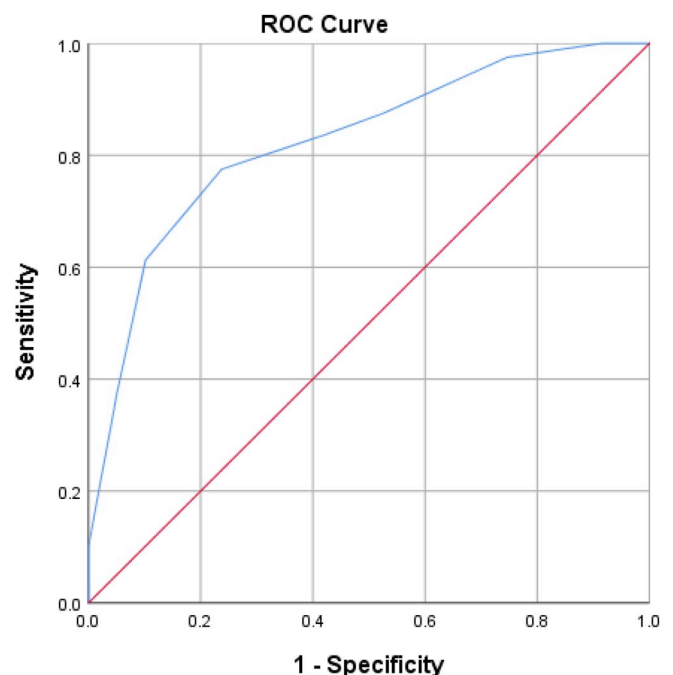
RMDQ - Roland Morris disability questionnaire; FA - Fear-avoidance.

Table 5

Change in 3-month clinical outcomes, according to baseline risk category.

Variable	Low risk N = 47	Medium risk N = 67	High risk N = 25	P By group	F (1,136) P By time
Severe pain	2.3 ± 3.0	2.6 ± 3.4	1.9 ± 2.5	0.56	63.7 P < 0.001
Average pain	2.1 ± 2.5	1.9 ± 3.1	1.0 ± 2.1	0.30	44.1 P < 0.001
RMDQ	2.5 ± 3.9	4.4 ± 5.1	3.2 ± 5.5	0.102	59.2 P < 0.001
FA exercise	1.0 ± 4.8	3.1 ± 7.1	4.6 ± 7.1	0.056	24.5 P < 0.001
Anxiety	0.7 ± 2.5	2.8 ± 3.7	1.9 ± 3.7	0.007*	34.7 P < 0.001
Depression	0.4 ± 2.1	2.2 ± 3.9	3.0 ± 2.7	0.002*	41.4 P < 0.001

RMDQ - Roland Morris disability questionnaire; FA - Fear avoidance.



Diagonal segments are produced by ties.

Fig. 1. Roc curve, predictive abilities in discriminating poor disability after three months.

et al., 2018), Japanese (Matsudaira et al., 2017), and German (Aebischer et al., 2015).

The proportion of patients with disabling back pain at three months follow up (RMDQ>7) in our study was higher than other predictive studies (Hill et al., 2008; Aebischer et al., 2015; Suri et al., 2018; Bier

et al., 2018), a possible explanation for this is the longer follow-up time of 6 months in some other studies (Hill et al., 2008; Suri et al., 2018). Another possible explanation is the higher baseline RMDQ score found in our study as compared to others (Hill et al., 2008; Aebischer et al., 2015; Suri et al., 2018; Bier et al., 2018).

Some studies did not report the percentages of patients with RMDQ > 7 in each group at follow-up (Matsudaira et al., 2017), while others did not report the RMDQ score for each group at baseline (Hill et al., 2008; Bier et al., 2018), and thus it is hard to compare the data to the full extent.

The stratified treatment model for LBP patients has been tested in a randomized controlled trial showing good results (Foster et al., 2014; Hill et al., 2011). Allocating patients into different prognostic groups and matching treatment resulted in better patient outcomes and economic benefits than usual treatments (Foster et al., 2014). Screening and matching were found to reduce time off work and to improve disability scores in patients in the UK health system (Foster et al., 2014). However, the approach has not been universally successful (Cherkin et al., 2018).

Health systems that have yet to implement screening and matching may be likely to over-treat or undertreat LBP patients (Bier et al., 2018). The SBT has also been found to have high responsiveness, and thus it is possible to use it to measure change for LBP patients (Medeiros et al., 2019). The Israeli health system is a public health system, similar to the one in the UK; therefore, this Hebrew translation and validity testing of SBT allows further investigation of the stratified treatment model among Israeli LBP patients.

4.1. Limitations

The following limitations of the current study are noteworthy:

- The analysis of measurement properties might have been improved by the use of the COSMIN approach.
- We recognize that the stronger methodology for translation is the recommendation by Beaton et al. Our sample size is relatively small for both the pre-final version and the reliability analysis. According to Beaton et al.'s (Beaton et al., 2000) recommendations, the pre-final version of the translation should be tested by at least 30 patients. In the current study, only 10 LBP patients were involved in that stage. Ideally, a larger sample should be tested in future research.
- Another limitation is that the translation did not assure that the second translator was completely naïve to the concepts being

quantified. In addition, the expert committee for the translation of the questionnaire omitted to include a language expert, as recommended by Beaton et al. (2000).

- Some of the participants in the study were already being treated by physiotherapists during data collection. So it is possible that their scores on the SBT may have been affected temporarily, possibly through reduced concern and therapeutic alliance.
- This version of the SBT was validated using a telephone interview to collect data. This might have influenced the results. Nevertheless, it seems that the questionnaire retained its predictive abilities despite this method.

5. Conclusion

The SBT was successfully translated into Hebrew, and the translated version was found to be a valid and reliable tool, among LBP patients in Israel. The ability to predict levels of disability at three months was confirmed. Future work should focus on developing and adapting matched interventions for each SBT risk category and testing their effectiveness in Israeli patients. Future studies should verify the responsiveness and minimal clinically important difference of the Hebrew version of the SBT.

Funding

None declared.

Ethical approval

This study received approval (number 0157-17-COM2) from the ethical review board of Clalit Health Services in Israel.

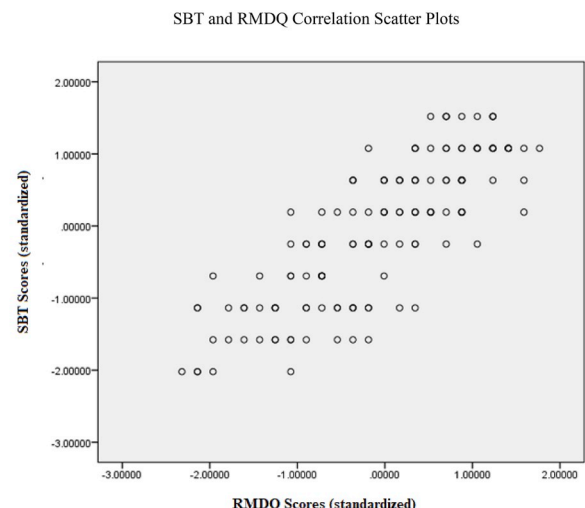
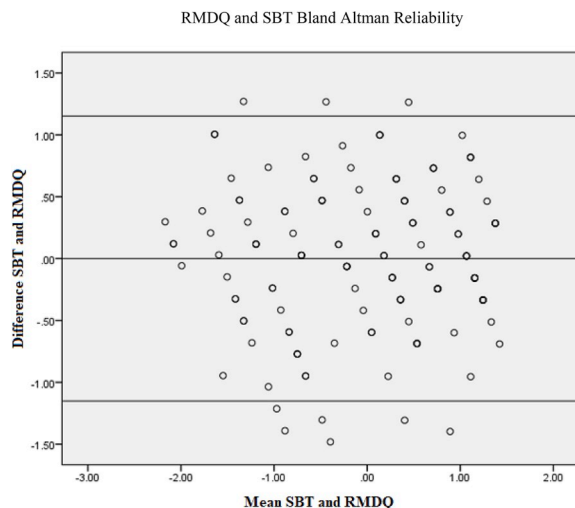
Declaration of competing interest

None declared.

Acknowledgments

We thank Mr. Igal Levran, head of the physiotherapy in Clalit Health Services in Israel, and all the physiotherapists who were involved in the trial. We thank Michal Lev for the dedicated work in the interviewing and coding process.

Appendix 1



Appendix 2. Concerns and disagreements in the translation process

All concerns and disagreements were resolved by discussion. In the forward translation, concerns were raised about the following items (resolution was offered by the expert committee):

1. Translating the term 'at some time' (item 1) had three options in Hebrew.
2. In the same item, the term 'spread' had a more medical or a more commonly-used translation. The latter was selected.
3. Item 5: two options to translate 'be physically active', one focuses on exercise the other on physical activity. The latter was selected.
4. Item 5 if translated verbatim becomes complicated through negative words, and the word 'very' was added in the Hebrew translation, as there was a very high consensus from patients to the milder version, resulting in failure to differentiate.
5. Item 6, the term 'worrying' could be translated more in line with troublesome, or with worry. The latter was selected.
6. Item 7, the term absolutely terrible needed to be translated without the qualifier 'absolutely'.
7. Item 9, the term 'bothersome' could include shades of explicit worry, but the selected term is closer to the original English.
8. The response rate used in item 9 for 'extremely' utilized a stronger term.

The backward translation did not result in any concerns.

References

- Abedi, Mohsen, Farideh Dehghan, Manshadi, Khalkhali, Mino, Seyed Javad, Mousavi, Alireza Akbarzadeh, Baghban, Ali, Montazeri, Parnianpour, Mohamad, 2015. Translation and validation of the Persian version of the STarT back screening tool in patients with nonspecific low back pain. *Man. Ther.* 20 (6), 850–854. <https://doi.org/10.1016/j.math.2015.04.006>.
- Aebischer, Bernhard, Hill, Jonathan C., Hilfiker, Roger, Karstens, Sven, 2015. German translation and cross-cultural adaptation of the STarT back screening tool. *PLoS One* 10 (7). <https://doi.org/10.1371/journal.pone.0132068> e0132068.
- Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B., 2000. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25 (24), 3186–3191.
- Bier, Jasper D., Janneke, J.W., Sandee-Geurts, Ostelo, Raymond W.J. G., Koes, Bart W., Verhagen, Arianne P., 2018. Can primary care for back and/or neck pain in The Netherlands benefit from stratification for risk groups according to the STarT back tool classification? *Arch. Phys. Med. Rehabil.* 99 (1), 65–71. <https://doi.org/10.1016/j.apmr.2017.06.011>.
- Bjelland, Ingvang, Dahl, Alv A., Haug, Tone Tangen, Neckelmann, Dag, 2002. The validity of the hospital anxiety and depression scale. An updated literature review. *J. Psychosom. Res.* 52 (2), 69–77.
- Bruyère, Olivier, Demoulin, Maryline, Brereton, Clara, Humblet, Fabienne, Flynn, Daniel, Hill, Jonathan C., Maquet, Didier, et al., 2012. Translation validation of a new back pain screening questionnaire (the STarT back screening tool) in French. *Arch. Publ. Health* 70 (1), 12. <https://doi.org/10.1186/0778-7367-70-12>.
- Buchbinder, Rachelle, Maurits van Tulder, Öberg, Birgitta, Costa, Luciola Menezes, Anthony, Woolf, Schoene, Mark, Croft, Peter, et al., 2018. Low back pain: a call for action. *Lancet* 391 (10137), 2384–2388. [https://doi.org/10.1016/S0140-6736\(18\)30488-4](https://doi.org/10.1016/S0140-6736(18)30488-4).
- de Campos, Tarcisio Folly, 2017. Low back pain and sciatica in over 16s: assessment and management NICE guideline [NG59]. *J. Physiother.* 63 (2), 120. <https://doi.org/10.1016/j.jphys.2017.02.012>.
- Cherkin, Dan, Balderson, Benjamin, Wellman, Rob, Hsu, Clarissa, Sherman, Karen J., Evers, Sarah C., Hawkes, Rene, et al., 2018. Effect of low back pain risk-stratification strategy on patient outcomes and care processes: the MATCH randomized trial in primary care. *J. Gen. Intern. Med.* 33 (8), 1324–1336. <https://doi.org/10.1007/s11606-018-4468-9>.
- Chiarotto, Alessandro, Maxwell, Lara J., Terwee, Caroline B., Wells, George A., Peter, Tugwell, Ostelo, Raymond W., 2016. Roland-morris disability questionnaire and Oswestry disability index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. *Phys. Ther.* 96 (10), 1620–1637. <https://doi.org/10.2522/ptj.20150420>.
- Cicchetti, Domenic V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6 (4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>.
- Ferreira-Valente, Alexandra, Maria, Pais-Ribeiro, José Luís, 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., Anema, Johannes R., Dan, Cherkin, Chou, Roger, Cohen, Steven P., Gross, Douglas P., Ferreira, Paulo H., et al., 2018. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet* 391 (10137), 2368–2383. [https://doi.org/10.1016/S0140-6736\(18\)30489-6](https://doi.org/10.1016/S0140-6736(18)30489-6).
- Foster, Nadine E., Mullis, Ricky, Hill, Jonathan C., Lewis, Martyn, Whitehurst, David G. T., Doyle, Carol, Konstantinou, Kika, et al., 2014. Effect of stratified care for low back pain in family practice (IMPACT back): a prospective population-based sequential comparison. *Ann. Fam. Med.* 12 (2), 102–111. <https://doi.org/10.1370/afm.1625>.
- Fritz, Julie M., Beneciuk, Jason M., George, Steven Z., 2011. Relationship between categorization with the STarT back screening tool and prognosis for people receiving physical therapy for low back pain. *Phys. Ther.* 91 (5), 722–732. <https://doi.org/10.2522/ptj.20100109>.
- Geerincx, Anton, Alekna, Vidmantas, Beaudart, Charlotte, Bautmans, Ivan, Cooper, Cyrus, De Souza Orlandi, Fabiana, Konstantynowicz, Jerzy, 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). <https://doi.org/10.1371/journal.pone.0216065> e0216065.
- Hill, Jonathan C., Dunn, Kate M., Lewis, Martyn, Mullis, Ricky, Main, Chris J., Foster, Nadine E., Hay, Elaine M., 2008. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum.* 59 (5), 632–641. <https://doi.org/10.1002/art.23563>.
- Hill, Jonathan C., Whitehurst, David G.T., Lewis, Martyn, Bryan, Stirling, Dunn, Kate M., Foster, Nadine E., Konstantinou, Kika, et al., 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT back): a randomised controlled trial. *Lancet (Lond., Engl.)* 378 (9802), 1560–1571. [https://doi.org/10.1016/S0140-6736\(11\)60937-9](https://doi.org/10.1016/S0140-6736(11)60937-9).
- Koo, Terry K., Mae, Y. Li, 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15 (2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- Mandrekar, Jayawant N., 2010. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.: Off. Publ. Int. Assoc. Stud. Lung Canc.* 5 (9), 1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>.
- Matsuda, Ko, Oka, Hiroyuki, Kikuchi, Norimasa, Haga, Yuri, Sawada, Takayuki, Tanaka, Sakae, 2017. The Japanese version of the STarT back tool predicts 6-month clinical outcomes of low back pain. *J. Orthop. Sci.* 22 (2), 224–229. <https://doi.org/10.1016/j.jos.2016.11.023>.
- Medeiros, Flávia Cordeiro, Costa, Leonardo Oliveira Pena, Soares Oliveira, Indira, Luciola da, Cunha Menezes Costa, 2019. A responsiveness analysis of the subgroups for targeted treatment (STarT) back screening tool in patients with nonspecific low back pain. *J. Orthop. Sports Phys. Ther.* 49 (10), 725–735. <https://doi.org/10.2519/jospt.2019.8776>.
- Morsø, Lars, Albert, Hanne, Kent, Peter, Manniche, Claus, Hill, Jonathan, 2011. Translation and discriminative validation of the STarT back screening tool into Danish. *Eur. Spine J.: Off. Publ. Eur. Spine Soc. Eur. Spinal Deformity Soc. Eur. Sect. Cervical Spine Res. Soc.* 20 (12), 2166–2173. <https://doi.org/10.1007/s00586-011-1911-6>.
- Piironen, Susanna, Paananen, Markus, Haapea, Marianne, Hupli, Markku, Zitting, Paavo, Ryyänen, Katja, Takala, Esa-Pekka, et al., 2016. Transcultural adaptation and psychometric properties of the STarT back screening tool among Finnish low back pain patients. *Eur. Spine J.: Off. Publ. Eur. Spine Soc. Eur. Spinal Deformity Soc. Eur. Sect. Cervical Spine Res. Soc.* 25 (1), 287–295. <https://doi.org/10.1007/s00586-015-3804-6>.
- Pilz, Bruna, Vasconcelos, Rodrigo A., Marcondes, Freddy B., Lodovichi, Samuel S., Wilson, Mello, Grossi, Débora B., Pilz, Bruna, et al., 2014. The Brazilian version of STarT back screening tool - translation, cross-cultural adaptation and reliability. *Braz. J. Phys. Ther.* 18 (5), 453–461. <https://doi.org/10.1590/bjpt-rbf.2014.0028>.
- Suri, Pradeep, Delaney, Kristin, Rundell, Sean D., Cherkin, Daniel C., 2018. Predictive validity of the STarT back tool for risk of persistent disabling back pain in a United States primary care setting. *Arch. Phys. Med. Rehabil.* 99 (8), 1533–1539. <https://doi.org/10.1016/j.apmr.2018.02.016> e2.
- 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-hCgkoDZpsh8Xe51LaVx-dx4Yn49k8.