Journal of Physiotherapy xxx (2018) xxx–xxx

Journal of PHYSIOTHERAPY

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

Research

The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study

Michelle Kendell [a], Darren Beales [a], Peter O'Sullivan [a], Martin Rabey [a], Jonathan Hill [b], Anne Smith [a]

[a] School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia; [b] Institute of Primary Care and Health Sciences, Keele University, Staffordshire, United Kingdom

KEY WORDS

Low pack pain
Decision support techniques
Prognosis
Validation studies

ABSTRACT

Questions: In people with chronic non-specific low back pain (LBP), what is the predictive and discriminative validity of the STarT Back Tool (SBT) for pain intensity, self-reported LBP-related disability, and global self-perceived change at 1-year follow-up? What is the profile of the SBT risk subgroups with respect to demographic variables, pain intensity, self-reported LBP-related disability, and psychological measures? **Design**: Prospective cohort study. **Participants**: A total of 290 adults with dominant axial LBP of $\geq$ 3 months' duration recruited from the general community, and private physiotherapy, psychology, and pain-management clinics in Western Australia. **Outcome measures**: The 1-year follow-up measures were pain intensity, LBP-related disability, and global self-perceived change. **Results**: Outcomes were collected on 264 participants. The SBT categorised 82 participants (28%) as low risk, 116 (40%) as medium risk, and 92 (32%) as high risk. The risk subgroups differed significantly ($p < 0.05$) on baseline pain, disability, and psychological scores. The SBT's predictive ability was strongest for disability: RR was 2.30 (95% CI 1.28 to 4.10) in the medium-risk group and 2.86 (95% CI 1.60 to 5.11) in the high-risk group. The SBT's predictive ability was weaker for pain: RR was 1.25 (95% CI 1.04 to 1.51) in the medium-risk group and 1.26 (95% CI 1.03 to 1.52) in the high-risk group. For the SBT total score, the AUC was 0.71 (95% CI 0.64 to 0.77) for disability and 0.63 (95% CI 0.55 to 0.71) for pain. **Conclusion**: This was the first large study to investigate the SBT in a population exclusively with chronic LBP. The SBT provided an acceptable indication of 1-year disability, had poor predictive and discriminative ability for future pain, and was unable to predict or discriminate global perceived change. In this cohort with chronic non-specific LBP, the SBT's predictive and discriminative abilities were restricted to disability at 1 year. **[Kendell M, Beales D, O'Sullivan P, Rabey M, Hill J, Smith A (2018) The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *Journal of Physiotherapy* XX: XX–XX]**

Introduction

Chronic low back pain (LBP) is problematic for many people, and has significant personal, social, and economic impact. Worldwide, LBP is the leading cause of years lived with disability.[1] The economic burden of LBP in Australia is substantial.[2] Healthcare practices require more effective and affordable strategies to better manage the rising burden of LBP[3] and direct resources to those most in need.

Evidence-based guidelines[4,5] highlight the need to screen for indicators of poor prognosis and/or stratify patients with LBP based on risk. The STarT Back Tool (SBT) (http://www.keele.ac.uk/sbst/startbacktool/)[6] was developed to allow primary care/first-contact practitioners to subgroup patients with non-specific LBP into low risk, medium risk, and high risk of future disability, with the purpose of matching each subgroup to a care pathway. The nine-item SBT includes treatment-modifiable domains such as spread of pain, disability, and psychological factors.[7] A randomised trial demonstrated that a risk stratification approach based on the SBT resulted in better clinical outcomes and reduced costs compared to usual care in a primary care cohort in the United Kingdom.[8]

Initial development and validation of the SBT was undertaken in a United Kingdom primary care, general practice setting where participants had non-specific LBP of 'variable' episode duration (ie, acute/subacute/chronic LBP).[7] There is a growing body of evidence supporting the SBT's psychometric properties and predictive and discriminative ability particularly in populations with LBP of variable episode duration.[7,9–17] To date, only two studies have evaluated the SBT in a population with exclusively[18] or predominantly[19] chronic LBP. Studies have shown that the SBT's performance has differed depending on the population in which it was evaluated, with clinical setting, cultural context, LBP episode duration, treatment provided, and outcome measure evaluated all influencing the tool's predictive performance.[9,10,12,16–22] Authors have recommended that the SBT be evaluated in different populations[15,16,20] and that LBP episode duration be considered.[21]

The SBT risk subgroups have not been profiled nor has the tool's predictive and discriminative ability been adequately investigated

https://doi.org/10.1016/j.jphys.2018.02.009

in a population exclusively with chronic LBP. While the SBT is commonly used in Australia, it is yet to be evaluated in an Australian context. If the SBT is to be used with confidence in Australia for patients presenting with chronic LBP, the tool's performance must be evaluated in this group.

Therefore, the research questions for this prospective cohort study were:

1. In people with chronic non-specific LBP, what is the predictive and discriminative validity of the SBT for pain intensity, self-reported LBP-related disability, and global self-perceived change at 1-year follow-up?
2. What is the profile of the SBT risk subgroups with respect to demographic variables, pain intensity, self-reported LBP-related disability, and psychological measures?

Method

Design

This prospective cohort study had a 1-year follow-up. Data for this study were obtained from a prospective cohort study that evaluated the presence of multidimensional subgroups and reported on multidimensional prognostic modelling in a population with chronic non-specific LBP.[23–26] These prior publications did not consider the SBT risk subgroups and did not use the SBT data in any analyses.

Participants

Participants were recruited from the general community via multimedia advertisements in metropolitan and rural Western Australia, as well as private physiotherapy, psychology and pain-management clinics in Perth, Western Australia. The eligibility criteria have been published in detail previously.[23–26] In brief, participants were required to: be aged 18 to 70 years; have dominant axial non-specific LBP with the ratio of back:leg pain $\geq$ 60%; have had the pain for $\geq$ 3 months; report an average baseline pain intensity in the previous week of $\geq$ 2 on an 11-point Numerical Rating Scale; and report LBP-related disability of $\geq$ 5 on the Roland-Morris Disability Questionnaire. Exclusion criteria were: inability to understand English, pregnancy, a diagnosed neurological condition, serious spinal pathology, or spinal surgery in the previous 6 months.

Baseline measures

Eligible participants completed a standardised self-report questionnaire pack, which included questions related to demographic variables (age, gender, education level, employment status, occupation, compensation status, pain duration, and overall general health), the SBT, and psychological questionnaires, as shown in Table 1 and Table 2.

Follow-up measures

At 1-year follow-up, participants completed an online or paper questionnaire that included: average pain intensity in the previous week measured on an 11-point Numerical Rating Scale where 0 indicated 'no pain' and 10 indicated 'the worst pain imaginable'; LBP-related disability measured by the Roland-Morris Disability Questionnaire; and global perceived change measured by a single self-report item on a 7-point Global Rating of Change Scale. For global perceived change, participants were asked: 'With respect to your low back pain, how would you describe yourself now compared to 1 year ago when we examined you for the research project (laboratory session at Curtin University)?'[27] Responses ranged from −3 indicating 'very much worse' to 3 indicating 'very much improved'.

Follow-up measures were dichotomised into 'recovered' and 'not recovered'. Not recovered for pain was defined as a score of $\geq$ 3 on the Numerical Rating Scale.[20] Not recovered for disability was defined as a score of $\geq$ 7 on the Roland-Morris Disability Questionnaire as per prior investigations of the SBT.[7,16,19,20] Not recovered/not improved for global perceived change was defined as a score of $\leq$ 0 on the Global Rating of Change Scale (ie, no change or worse).

Data analysis

Questionnaire missing data were managed as suggested in original manuscripts, where described. Otherwise, the average of the relevant scale or subscale was used when one item was

Table 1
Baseline demographic and clinical characteristics of the total study cohort and stratified by STarT Back Tool risk subgroup.

Characteristic	Total cohort (n = 290)	Risk subgroup			*P*-value[a]
		Low (n = 82, 28%)	Medium (n = 116, 40%)	High (n = 92, 32%)	
Age *(yr)*, median (IQR)	51 (37 to 60)	49 (40 to 61)	52 (41 to 58)	50 (32 to 60)	0.556
Gender, n (%) female	166 (57)	49 (60)	73 (63)	44 (48)	0.079
Education level *(yr)*, mean (SD)[b]	15 (4)	16 (4)	15 (3)	14 (4)	0.012[b]
Employed, n (%)	223 (77)	65 (79)	89 (77)	69 (75)	0.799
Occupation, n (%)[c]					≤ 0.001
manual	72 (26)	13 (16)	24 (22)	35 (41)	
sedentary	174 (63)	61 (75)	76 (70)	37 (43)	
not working	29 (11)	7 (9)	8 (7)	14 (16)	
Compensated, n (%)[d]	45 (16)	15 (19)	15 (13)	15 (17)	0.712
Pain duration *(months)*, median (IQR)[e]	120 (42 to 240)	120 (42 to 240)	120 (48 to 300)	120 (36 to 192)	0.481
Overall general health *(0 to 5)*, median (IQR)[f]	3.0 (2.0 to 4.0)	2.0 (2.0 to 3.0)	3.0 (2.0 to 3.5)	3.0 (3.0 to 4.0)	≤ 0.001
Pain intensity on NRS *(0 to 10)*, mean (SD)	5.8 (1.9)	4.7 (1.8)	6.0 (1.8)	6.5 (1.6)	≤ 0.001
RMDQ *(0 to 24)*, median (IQR)	9.0 (6.0 to 13.0)	6.0 (5.0 to 8.0)	9.0 (7.0 to 13.0)	12.0 (8.5 to 15.0)	≤ 0.001
Pain location *(ratio back:leg)*, n (%)					0.091
100:0	145 (50)	51 (62)	49 (42)	45 (49)	
80:20	110 (38)	23 (28)	50 (43)	37 (40)	
60:40	35 (12)	8 (10)	17 (15)	10 (11)	

NRS = Numerical Rating Scale; RMDQ = Roland-Morris Disability Questionnaire.
[a] Evaluation of subgroup differences were performed using a one-way Analysis of Variance (ANOVA), Kruskal-Wallis test, or chi-squared test, depending on data type and distribution. *P*-values < 0.05 indicate significant pair-wise differences between all three subgroups unless otherwise indicated.
[b] Significant difference between the low-risk and high-risk groups only. Missing data points: 6 low, 1 medium, 7 high.
[c] Missing data points: 1 low, 8 medium, 6 high.
[d] Missing data points: 1 low, 2 medium, 3 high.
[e] Missing data points: 2 medium, 2 high.
[f] COOP-WONCA charts, where lower scores indicate better health.

Please cite this article in press as: Kendell M, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* (2018), https://doi.org/10.1016/j.jphys.2018.02.009

Table 2
Median (IQR) baseline psychological measures of the total study cohort and stratified by STarT Back Tool risk subgroup.

Characteristic	Total cohort (n = 290)	Risk subgroup			*P*-value[a]
		Low (n = 82, 28%)	Medium (n = 116, 40%)	High (n = 92, 32%)	
DASS-21					
Depression *(0 to 42)*	6.0 (2.0 to 14.0)	4.0 (2.0 to 6.0)	6.0 (2.0 to 12.0)	14.0 (6.0 to 24.0)	≤ 0.001
Anxiety *(0 to 42)*	4.0 (2.0 to 8.0)	2.0 (0.0 to 6.0)	4.0 (2.0 to 8.0)	6.0 (2.0 to 11.0)	≤ 0.00[b]
Stress *(0 to 42)*	12.0 (6.0 to 20.0)	8.0 (4.0 to 14.0)	12.0 (6.0 to 18.0)	16.0 (10.0 to 24.0)	≤ 0.001
FABQ					
Physical activity *(0 to 24)*	15.0 (11.0 to 19.0)	13.0 (7.0 to 17.0)	13.5 (10.0 to 17.0)	18.0 (14.0 to 21.0)	≤ 0.001[c]
Work *(0 to 42)*[d]	16.0 (8.0 to 27.0)	13.0 (6.5 to 24.0)	15.0 (6.0 to 24.0)	20.0 (9.0 to 32.0)	0.010[d]
PCS[e]					
Total *(0 to 52)*	17.0 (9.0 to 27.0)	9.0 (4.0 to 15.0)	16.0 (10.0 to 23.0)	28.0 (20.5 to 36.0)	≤ 0.001
Rumination *(0 to 16)*	6.0 (3.0 to 10.0)	3.0 (1.0 to 7.0)	6.0 (3.0 to 9.0)	10.0 (5.0 to 13.5)	≤ 0.001
Magnification *(0 to 12)*	3.0 (1.0 to 5.0)	2.0 (1.0 to 4.0)	3.0 (1.0 to 5.0)	5.0 (4.0 to 7.0)	≤ 0.001
Helplessness *(0 to 24)*	8.0 (4.0 to 13.0)	4.0 (2.0 to 6.0)	7.0 (4.0 to 11.0)	13.0 (9.0 to 16.5)	≤ 0.001
Perceived risk of persistence *(0 to 10)*	9.0 (8.0 to 10.0)	8.0 (7.0 to 9.0)	9.0 (8.0 to 10.0)	9.0 (8.0 to 10.0)	≤ 0.001
PSEQ *(0 to 60)*[f]	42.0 (32.0 to 50.0)	49.0 (44.0 to 53.0)	43.0 (32.0 to 49.0)	34.0 (24.5 to 41.0)	≤ 0.001
CPAQ-8[g]					
Total *(0 to 48)*	26.0 (21.0 to 31.0)	30.0 (26.0 to 36.0)	27.0 (23.0 to 31.0)	21.0 (15.5 to 25.5)	≤ 0.001
Pain willingness *(0 to 24)*	9.0 (6.0 to 12.0)	11.0 (8.0 to 15.0)	9.0 (7.0 to 12.5)	6.0 (3.0 to 9.0)	≤ 0.001
Activity engagement *(0 to 24)*	18.0 (14.0 to 21.0)	20.0 (17.0 to 22.0)	18.0 (15.0 to 21.0)	15.0 (11.0 to 19.0)	≤ 0.001

DASS-21 = Depression Anxiety Stress Scale, FABQ = Fear Avoidance Beliefs Questionnaire; PCS = Pain Catastrophising Scale; PSEQ = Pain Self-efficacy Questionnaire; CPAQ-8 = Chronic Pain Acceptance Questionnaire.
[a] Kruskal-Wallis test. *P*-values < 0.05 indicate significant pair-wise differences between all three subgroups, unless otherwise indicated.
[b] No significant difference between the medium-risk and high-risk groups.
[c] No significant difference between the low-risk and medium-risk groups.
[d] No significant difference between the low-risk and medium-risk groups. Missing data points: 2 low, 2 medium, 5 high.
[e] Missing data point: 1 medium.
[f] Higher scores indicate greater confidence.
[g] Higher scores indicate greater acceptance.

missing. When two or more items were missing, the questionnaire score was coded as missing.

Baseline data analysis

Descriptive statistics were calculated for demographic variables, clinical measures of pain intensity and disability, and psychological measures, with respect to the total cohort and each SBT risk subgroup. Subgroup differences for continuous demographic variables, clinical measures, and psychological measures were examined using a one-way analysis of variance for normally distributed variables and the Kruskal-Wallis test for variables with skewed data. Subgroup differences for categorical variables were examined using the chi-squared test.

Follow-up data analysis

Comparisons between responders and non-responders at the 1-year follow-up were made for demographic variables, baseline pain intensity and disability, and SBT stratification. Depending on data type and distribution, an independent sample *t*-test, Mann-Whitney U, or chi-squared test was used to evaluate these differences.

Descriptive statistics were calculated for follow-up pain intensity, disability, and global perceived change for the total cohort and for each SBT risk subgroup. The proportion of participants not recovered at 1 year for each of the follow-up measures was calculated at a cohort level and by SBT risk subgroup.

The predictive ability of the SBT was evaluated by calculating the relative risk (RR) of non-recovery for participants classified by the SBT as medium-risk or high-risk, using the low-risk subgroup as the reference category. An RR of < 2.0 is unlikely to have much practical value, an RR of 3.0 can be considered a moderate effect, and an RR of 4.0 can be considered a strong effect.[28]

To evaluate the accuracy of the SBT baseline total score and psychological subscale score to discriminate between recovered and not recovered participants at follow-up, receiver operating characteristic (ROC) curves were constructed and the area under the curve (AUC) calculated. Where follow-up measures had significant AUC values, the positive likelihood ratio, negative likelihood ratio, sensitivity, specificity, and the diagnostic odds ratio (DOR) were calculated for: the low-risk group versus the medium/high-risk group; and the low/medium-risk group versus the high-risk group. The SBT risk subgroups were collapsed into low/medium and medium/high, as this reflected the risk subgroup cut-offs and facilitated comparison with previous studies.[7,18,20] An AUC of 0.50 suggests no discrimination, > 0.50 but < 0.70 poor discrimination, ≥ 0.70 but < 0.80 acceptable discrimination, ≥ 0.80 but < 0.90 excellent discrimination, and ≥ 0.90 outstanding discrimination.[29] A higher positive likelihood ratio and lower negative likelihood ratio indicate better discrimination.[7] Likelihood ratios > 5 or < 0.2 are generally seen as supporting a strong test, whereas values close to 1 indicate poor test performance.[20] The DOR ranges from 0 to infinity.[30] A higher DOR indicates better test discrimination, a value of 1 indicates that the test has no ability to discriminate, and a value < 1 indicates that the test classifies incorrectly.[30]

Because there have been a number of cut-off scores used in previous studies to indicate non-recovery with respect to pain, a sensitivity analysis was performed using a cut-off score of ≥ 6 on the Numerical Rating Scale. A sensitivity analysis was also performed using linear regression to assess and compare the proportion of variance explained by the SBT with pain, disability, and global perceived change as scale variables rather than dichotomised variables.

Results

Flow of participants through the study

The flow of participants through the study is shown in Figure 1. Of the 290 participants included at baseline, 228 (79%) were recruited from the general community, 59 (20%) from private physiotherapy clinics, and three (1%) from private psychology and pain management clinics. The 1-year follow-up data were available for 264 (91%) participants. There was no significant difference ($p > 0.05$) in age, gender, or baseline pain intensity between responders and non-responders. Non-responders had higher baseline disability than responders (Roland-Morris Disability Questionnaire median score 10.5 versus 8.0, respectively, $p = 0.034$). Non-responders also had higher risk status than responders (high-risk allocation 62% versus 29%, respectively, $p = 0.002$).

Please cite this article in press as: Kendell M, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* (2018), https://doi.org/10.1016/j.jphys.2018.02.009

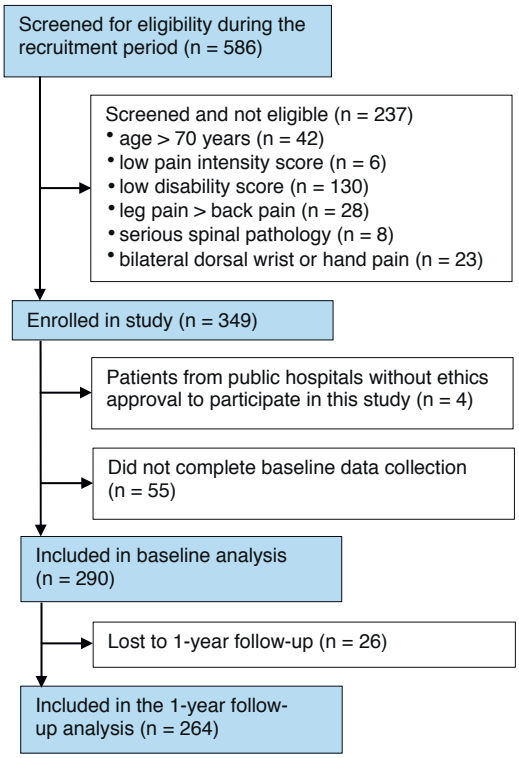

Figure 1. Flow of participants through the study.

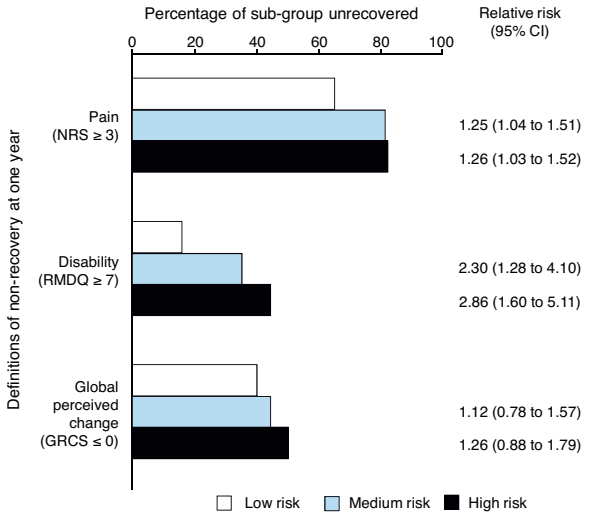

Figure 2. Relative risk of non-recovery at the 1-year follow-up by STarT Back Tool risk subgroup. Relative risks (95% CI) are calculated with the low-risk subgroup as the reference category. NRS = Numerical Rating Scale, RMDQ = Roland Morris Disability Questionnaire, GRCS = Global Rating of Change Scale.

Cross sectional profile of the STarT Back Tool risk subgroups

Participant baseline characteristics are presented in Table 1 and Table 2. The SBT categorised 82 participants (28%) as low risk, 116 (40%) as medium risk, and 92 (32%) as high risk. The SBT risk subgroups did not differ significantly for the majority of the demographic variables, including pain duration. In contrast, pain intensity ($p \leq 0.001$) and disability ($p \leq 0.001$) increased stepwise from the low-risk group through to the high-risk group. In addition, consistently greater negative psychological affect and cognitions, decreasing self-efficacy, and decreasing chronic pain acceptance were also seen from the low-risk group through to the high-risk group ($p \leq 0.001$).

Predictive and discriminative ability of the STarT Back Tool

Participant characteristics at the 1-year follow-up are presented in Table 3. Participants in both the medium-risk and high-risk groups had a 25% increased risk of not being recovered with respect to pain compared with the low-risk group (Figure 2). Participants in the medium-risk group had a 130% increased risk, and those in the high-risk group had a 186% increased risk of not being recovered with respect to disability compared with the low-risk group. Although a higher proportion of both the medium-risk group and the high-risk group perceived themselves as not improved compared with the low-risk group, the difference in risk was not statistically significant.

The ROC curves for the SBT total score and psychological subscale score are shown in Figure 3. The SBT's discriminative ability was highest for disability, lower for pain, and the tool was unable to discriminate global perceived change at 1 year (Table 4).

The likelihood ratios, sensitivity, specificity, and DORs for the SBT risk subgroups for pain and disability at 1-year follow-up are presented in Table 5. The positive likelihood ratios were higher and the negative likelihood ratios were lower for disability compared with those for pain.

The sensitivity analysis using a cut-off score of ≥ 6 instead of ≥ 3 for pain resulted in similar AUC values (0.62 versus 0.63, respectively). The sensitivity analysis using the follow-up measures at a scale level rather than dichotomised gave the same results, in that the SBT was significantly and most strongly predictive of disability ($r^2 = 0.09$), significantly but less predictive of pain ($r^2 = 0.04$), and not predictive of global perceived change ($r^2 = 0.00$).

Discussion

This was the first large study to investigate the SBT in a population exclusively with chronic non-specific LBP and in an Australian context.

At baseline, those with a higher SBT risk categorisation had significantly greater pain intensity, greater disability, higher scores

Table 3
Characteristics at the 1-year follow-up: total cohort and stratified by STarT Back Tool risk subgroup.

Follow-up measure	Total cohort (n = 264)	Risk subgroup[a]		
		Low (n = 79)	Medium (n = 106)	High (n = 76)
Pain *(0 to 10)*[b]				
mean (SD)	4.2 (2.1)	3.5 (2.0)	4.4 (2.1)	4.5 (2.2)
n (%) not recovered (≥ 3)	201 (76)	51 (65)	88 (81)	62 (82)
Disability *(0 to 24)*[c]				
median (IQR)	4.0 (2.0 to 8.0)	3.0 (1.0 to 5.0)	5.0 (2.0 to 8.0)	6.0 (3.0 to 10.5)
n (%) not recovered (≥ 7)	83 (31)	12 (15)	38 (35)	33 (43)
Global perceived change *(−3 to 3)*[d]				
median (IQR)	1.0 (0.0 to 2.0)	1.0 (0.0 to 2.0)	1.0 (0.0 to 2.0)	0.5 (−1.0 to 2.0)
n (%) not improved (≤0)	117 (45)	31 (40)	48 (44)	38 (50)

[a] Lost to follow-up by risk subgroup: 3 low, 7 medium and 16 high.
[b] Numerical Rating Scale.
[c] Roland-Morris Disability Questionnaire.
[d] Global Rating of Change Scale, where −3 equates to 'very much worse' and 3 equates to 'very much improved'. Missing data point: 1 low.

Please cite this article in press as: Kendell M, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* (2018), https://doi.org/10.1016/j.jphys.2018.02.009

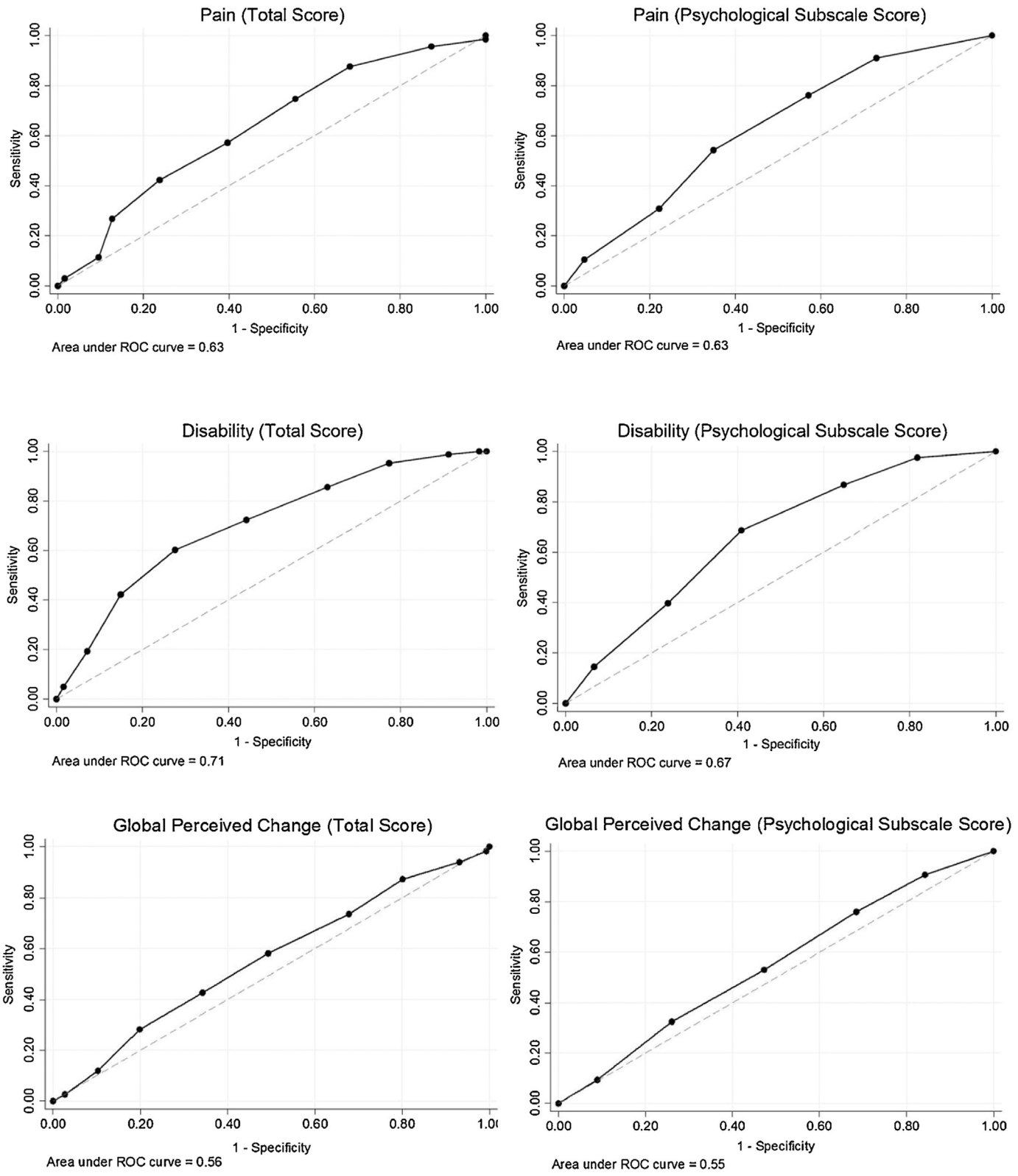

Figure 3. ROC curves for the STarT Back Tool baseline total score and psychological subscale score for pain (NRS $\geq$ 3), disability (RMDQ $\geq$ 7), and global perceived change (GRCS $\leq$ 0) at the 1-year follow-up. ROC = receiver operating characteristic; NRS = Numerical Rating Scale; RMDQ = Roland Morris Disability Questionnaire; GRCS = Global Rating of Change Scale.

on the negative psychological constructs, and lower scores on the positive psychological constructs. Previously published studies in physiotherapy and chiropractic clinical settings have also shown that the SBT risk subgroups and/or score were related to pain intensity,[12,13] disability,[11–13,18] depression,[10,11,15] fear avoidance beliefs,[10,11,15] catastrophising,[10,11,15] kinesiophobia,[10,11,18] and anxiety.[11] These results support that the SBT is able to distinguish elevated levels of pain, disability, and negative psychological affect and cognitions with reference to unidimensional questionnaires. The SBT may be an acceptable surrogate measure for multiple full-length unidimensional measures. Other authors have made similar suggestions.[11]

The prospective results demonstrated that the SBT had moderate predictive and acceptable discriminative ability for disability 1 year later. However, the SBT's predictive and discriminative ability was poor for pain, and the tool was unable to identify

Please cite this article in press as: Kendell M, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* (2018), https://doi.org/10.1016/j.jphys.2018.02.009

Table 4
Discriminative ability of the STarT Back Tool baseline total score and psychological subscale score.

Follow-up measure	Case definition	n (%)	STarT Back Tool AUC (95% CI)	
			Total score	Psychological subscale score
Pain	NRS ≥ 3	201 (76)	0.63 (0.55 to 0.71)	0.63 (0.55 to 0.71)
Disability	RMDQ ≥ 7	83 (31)	0.71 (0.64 to 0.77)	0.67 (0.60 to 0.73)
Global perceived change[a]	GRCS ≤ 0	117 (44)	0.56 (0.49 to 0.63)	0.55 (0.48 to 0.62)

AUC = area under the curve, NRS = Numerical Rating Scale, RMDQ = Roland-Morris Disability Questionnaire, GRCS = Global Rating of Change Scale.
[a] 1 missing from the low-risk group (n = 117/263).

Table 5
Discriminative ability of the STarT Back Tool risk subgroups.

Subgroups	LR+ (95% CI)	LR− (95% CI)	Sensitivity (%)	Specificity (%)	DOR
Pain (NRS ≥ 3)					
low versus medium/high	1.34 (1.06 to 1.70)	0.57 (0.40 to 0.82)	75	44	2.35
low/medium versus high	1.39 (0.84 to 2.30)	0.89 (0.76 to 1.05)	31	78	1.56
Disability (RMDQ ≥ 7)					
low versus medium/high	1.36 (1.18 to 1.57)	0.39 (0.22 to 0.68)	86	37	3.48
low/medium versus high	1.67 (1.15 to 2.43)	0.79 (0.65 to 0.96)	40	76	2.12

DOR = diagnostic odds ratio, LR+ = positive likelihood ratio, LR− = negative likelihood ratio, NRS = Numerical Rating Scale, RMDQ = Roland-Morris Disability Questionnaire.

participants who perceived themselves as improved versus not improved at follow-up. Previous studies have reported that the SBT's prognostic ability was better for disability than pain,[9,10,18,22] and that the SBT had little or no prognostic ability for global perceived change regardless of population, setting, or follow-up period.[12,17,18] The SBT was originally developed and validated to predict future disability,[7] so it is not surprising that the tool performed best with respect to this measure.

In comparison to a United Kingdom primary care setting, where participants reported variable duration LBP (where approximately 60% and 30% of the cohort, respectively, reported chronic LBP),[7,16] the SBT's performance was weaker in the current study. There are a number of possible explanations. The constructs included in the SBT were those shown to be the strongest independent predictors of disability in a United Kingdom primary care setting where participants reported LBP of variable duration.[7] Therefore, the SBT's constructs may not be as important for a cohort with chronic LBP. Given the complex, multidimensional nature of chronic LBP, it may be that the limited number of items and dichotomised response format of the SBT was unable to adequately capture prognostic risk in more complex patient presentations. Although the brevity and simplicity of the SBT is one of its strengths, it may also be a weakness when used in a population with chronic LBP. Importantly, there are a large number of factors that may influence prognosis prediction and future recovery that are unrelated to the content of the SBT.[26] The SBT was originally designed as a stratified care tool and only included clinically modifiable items, which may mean that important non-modifiable prognostic indicators were not included. The performance of a stratified care tool is best evaluated by an effect size obtained from a randomised, controlled trial; hence, caution is required if using the SBT for the sole purpose of predicting prognosis.[31]

Strengths of this study included: a low follow-up attrition rate; data collected on a diverse range of psychological measures, including three measures not previously investigated in relation to the SBT (perceived risk of pain persistence, self-efficacy, and chronic pain acceptance); patient-relevant follow-up measures that are in line with recommendations for research involving participants with chronic pain;[32] and a 1-year follow-up period, which makes this one of a few studies that has evaluated the SBT at a follow-up time point longer than 6 months.

There were a number of study limitations. Participants were self-selecting volunteers predominantly from the general community, which may have limited the generalisability of the results to specific clinical or healthcare-seeking populations. However, 213 of 260 (82%) participants received some form of intervention between baseline and 1-year follow-up, suggesting that this cohort was largely care-seeking and therefore broadly representative of patients with chronic LBP seeking care in a primary care setting. Participants had dominant axial LBP so the results may not be relevant to people with dominant low-back-related leg pain. The participants lost to follow-up had higher baseline disability and risk status, which may have biased the results. The follow-up measures were dichotomised, which facilitates reader understanding but may have resulted in a loss of information. Using different operational definitions for recovery could have influenced the results. Participants were free to pursue treatment throughout the study period. Robust data on treatment type and efficacy were not collected;[26] hence, it is unknown what influence, if any, treatment may have had on recovery or the SBT's performance. Nonetheless, the SBT was designed to guide decision-making[7] regardless of what treatment the patient may or may not have in the future. Finally, this study did not provide any information on the clinical or economic benefits of stratified care. However, if the SBT has relatively weak prognostic ability, its usefulness to stratify care may be limited in people with chronic LBP.[33]

This study provides valuable information for clinicians on the usefulness and limitations of the SBT for patients with chronic LBP. The SBT has value as a substitute for the administration of multiple full-length, unidimensional questionnaires to initially screen for high levels of pain, disability, and negative psychological affect and cognitions. The practicality of the SBT is derived from its multidimensional nature, and low responder and assessor burden facilitating its use in busy clinical practice. Although the SBT had moderate predictive and acceptable discriminative ability for disability 1 year later, clinicians need to be aware that patient-relevant outcomes extend beyond self-reported disability. The SBT appears limited in its ability to predict all aspects of a patient's outcome. The SBT should be used alongside the clinical examination and in conjunction with sound clinical reasoning when making care decisions.[34]

What is already known about this topic: The SBT was designed to stratify patients with non-specific LBP into low risk, medium risk, and high risk of poor disability outcome, with a matched care pathway for each subgroup. The predictive and discriminative ability of the SBT in populations with LBP of variable episode duration is supported in the literature, particularly for future disability among primary care populations. To date, there have been no large studies that have investigated

Please cite this article in press as: Kendell M, et al. The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *J Physiother.* (2018), https://doi.org/10.1016/j.jphys.2018.02.009

the SBT's predictive and discriminative ability in a population exclusively with chronic LBP.
What this study adds: This study adds to the literature on SBT's generalisability across populations. In a cohort with chronic non-specific LBP, those with higher SBT risk categorisation had significantly greater pain intensity, greater disability, higher scores on negative psychological constructs, and lower scores on positive psychological constructs at baseline. While the SBT provided an acceptable indication of future disability in this population, it performed poorly with respect to pain and global perceived change at the 1-year follow-up. Reliance on the SBT as the sole indicator of prognosis in chronic LBP is not recommended.

Ethics approval: The Curtin University Human Research Ethics Committee approved this study (Approval Number PT262/2013). All participants gave written informed consent before data collection began.

Competing interests: Nil.

Source(s) of support: Martin Rabey was supported during his Doctor of Philosophy degree by a Musculoskeletal Association of Chartered Physiotherapists Doctoral Award, United Kingdom; Chartered Society of Physiotherapy Charitable Trust, United Kingdom; Curtin University Postgraduate Scholarship, Perth, Western Australia; and an Australian Postgraduate Award, Australia. Darren Beales was supported by a National Health and Medical Research Council of Australia Early Career Research Fellowship, Australia.

Acknowledgements: Nil.

Provenance: Not invited. Peer reviewed.

Correspondence: Dr Michelle Kendell, School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia. Email: M.Kendell@curtin.edu.au

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet.* 2015;386(9995):743–800.
2. Walker B, Muller R, Grant W. Low back pain in Australian adults: the economic burden. *Asia Pac J Public Health.* 2003;15:79–87.
3. Vos T, Flaxman A, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2163–2196.
4. Koes B, van Tulder M, Lin C, Macedo L, McAuley J, Maher C. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *Eur Spine J.* 2010;19:2075–2094.
5. National Institute for Health and Care Excellence (NICE). *Low back pain and sciatica in over 16s: assessment and management.* London: NICE; 2016.
6. Keele University. STarT Back. http://www.keele.ac.uk/sbst/startbacktool/. (Accessed 21 October, 2015).
7. Hill J, Dunn K, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum.* 2008;59:632–641.
8. Hill J, Whitehurst D, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STartT Back): a randomised controlled trial. *Lancet.* 2011;378:1560–1571.
9. Beneciuck J, Fritz J, George S. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. *J Orthop Sports Phys Ther.* 2014;44:656–664.
10. Beneciuk J, Bishop M, Fritz J, Robinson ME, Asal NR, Nisenzon AN, et al. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther.* 2013;93:321–333.
11. Beneciuk J, Robinson M, George S. Subgrouping for patients with low back pain: a multidimensional approach incorporating cluster analysis and the STarT Back Screening Tool. *J Pain.* 2015;16:19–30.
12. Field J, Newell D. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropr Man Therap.* 2012;20:17.
13. Fritz J, Beneciuk J, George S, Hill J, Hay E. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Phys Ther.* 2011;91:722–736.
14. George S, Beneciuk J. Psychological predictors of recovery from low back pain: a prospective study. *BMC Musculoskelet Disord.* 2015;16:49.
15. Kongsted A, Johannesen E, Leboeuf-Yde C. Feasibility of the STarT Back Screening Tool in chiropractic clinics: a cross-sectional study of patients with low back pain. *Chiropr Man Therap.* 2011;19:10.
16. Morsø L, Kent P, Albert H, Hill J, Kongsted A, Manniche C. The predictive and external validity of the STarT Back Tool in Danish primary care. *Eur Spine J.* 2013;22:1859–1867.
17. Newell D, Field J, Pollard D. Using the STarT Back Tool: does timing of stratification matter? *Man Ther.* 2015;20:533–539.
18. Page I, Abboud J, O. Shaughnessy J, Laurencelle L, Descarreaux M. Chronic low back pain clinical outcomes present higher associations with the STarT Back Screening Tool than with physiologic measures: a 12-month cohort study. *BMC Musculoskelet Disord.* 2015;16:201.
19. Morsø L, Kent P, Manniche C, Albert H. The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *Eur Spine J.* 2014;23:120–128.
20. Mehling W, Avins A, Acree M, Carey T, Hecht F. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *Eur J Pain.* 2015;19:439–446.
21. Morsø L, Kongsted A, Hestbaek L, Kent P. The prognostic ability of the STarT Back Tool was affected by episode duration. *Eur Spine J.* 2016;25:936–944.
22. Kongsted A, Andersen C, Hansen M, Hestbaek L. Prediction of outcome in patients with low back pain – a prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Man Ther.* 2016;21:120–127.
23. Rabey M, Slater H, O'Sullivan P, Beales D, Smith A. Somatosensory nociceptive characteristics differentiate subgroups in people with chronic low back pain: a cluster analysis. *Pain.* 2015;156:1874–1884.
24. Rabey M, Smith A, Beales D, Slater H, O'Sullivan P. Differing psychologically derived clusters in people with chronic low back pain are associated with different multidimensional profiles. *Clin J Pain.* 2016;32:1015–1027.
25. Rabey M, Smith A, Beales D, Slater H, O'Sullivan P. Pain provocation following sagittal plane repeated movements in people with chronic low back pain: associations with pain sensitivity and psychological profiles. *Scand J Pain.* 2017;16:22–28.
26. Rabey M, Smith A, Beales D, Slater H, O'Sullivan P. Multidimensional prognostic modelling in people with chronic axial low back pain. *Clin J Pain.* 2017;33:877–891.
27. Kamper S, Maher C, Mackay G. Global Rating of Change Scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther.* 2009;17:163–170.
28. Ferguson C. An effect size primer: a guide for clinicians and researchers. *Prof Psychol Res Pr.* 2009;40:532–538.
29. Hosmer D, Lemeshow S, Sturdivant R. *Wiley series in probability and statistics: applied logistic regression (3rd edition).* New York, NY, USA: John Wiley & Sons; 2013.
30. Glas A, Lijmer J, Prins M, Bonsel G, Bossuyt P. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol.* 2003;56:1129–1135.
31. Karran E, McAuley J, Traeger A, Hillier SL, Grabherr L, Russek LN, et al. Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. *BMC Med.* 2017;15:13.
32. Dworkin R, Turk D, Farrar J, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain.* 2005;113:9–19.
33. Kent P, Morsø L, Kongsted A. Letter to the editor concerning 'Using the STarT Back Tool: does timing of stratification matter?'. *Man Ther.* 2015;20:e13.
34. Beales D, Kendell M, Chang R, Håmsø M, Gregory L, Richardson K, et al. Association between the 10 item Örebro Musculoskeletal Pain Screening Questionnaire and physiotherapists' perception of the contribution of biopsychosocial factors in patients with musculoskeletal pain. *Man Ther.* 2016;23:48–55.