BMJ Open

Can the STarT Back Tool predict health-related quality of life and work ability after an acute/subacute episode with back or neck pain? A psychometric validation study in primary care

Malin H Forsbrand,[1,2] Birgitta Grahn,[1,3] Jonathan C Hill,[4] Ingemar F Petersson,[1,5] Charlotte Post Sennehed,[1,3] Kjerstin Stigmar[5,6]

To cite: Forsbrand MH, Grahn B, Hill JC, *et al*. Can the STarT Back Tool predict health-related quality of life and work ability after an acute/subacute episode with back or neck pain? A psychometric validation study in primary care. *BMJ Open* 2018;**8**:e021748. doi:10.1136/bmjopen-2018-021748

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-021748).

Received 17 January 2018
Revised 20 April 2018
Accepted 28 June 2018

For numbered affiliations see end of article.

Correspondence to
Malin H Forsbrand;
malin.forsbrand@med.lu.se

ABSTRACT

Objectives The predictive ability of the STarT Back Tool (SBT) has not yet been examined among acute/subacute back and/or neck pain in a primary care setting in respect to health-related quality of life (HRQoL) and work ability outcomes. The aim of this study was to evaluate the SBT's predictive validity for HRQoL and work ability outcomes at long-term follow-up in a population with acute/subacute back and/or neck pain.

Setting Prospective data from 35 primary care centres in south Sweden during 2013.

Participants Patients (n=329) with acute/subacute back and/or neck pain, aged 18–67 years, not on sick leave or <60 days of sick leave completed the SBT when applying for physiotherapy treatment. Long-term follow-up measures (median 13 months, range 11–27 months) of HRQoL (EQ-5D) and work ability (Work Ability Score) was completed by 238 patients (72%).

Outcomes The predictive ability of the SBT for HRQoL and work ability outcomes was examined using Kruskal-Wallis test, logistic regression and area under the curve (AUC).

Results Based on SBT risk group stratification, 103 (43%), 107 (45%) and 28 (12%) patients were considered as low, medium and at high risk, respectively. There were statistically significant differences in HRQoL (p<0.001) and work ability (p<0.001) at follow-up between all three SBT risk groups. Patients in the high risk group had a significantly increased risk of having poor HRQoL (OR 6.16, 95% CI 1.50 to 25.26) and poor work ability (OR 5.08, 95% CI 1.75 to 14.71) vs the low risk group at follow-up. The AUC was 0.73 (95% CI 0.61 to 0.84) for HRQoL and 0.68 (95% CI 0.61 to 0.76) for work ability.

Conclusions The SBT is an appropriate tool for identifying patients with a poor long-term HRQoL and/or work ability outcome in a population with acute/subacute back and/or neck pain, and maybe a useful adjunct to primary care physiotherapy assessment and practice.

Trial registration number NCT02609750; Results.

Strengths and limitations of this study

- This is the first study to evaluate the predictive validity of STarT Back Tool (SBT) of the outcomes health-related quality of life and work ability at long-term follow-up in a population with acute/subacute back and/or neck pain.
- In this prospective study, we have recruited patients from 35 different primary care centres, where many physiotherapists were engaged.
- The predictive validity of the SBT was examined in different ways.
- Limited baseline data were available for one part of the study population.
- Limitations of the study were the broad variation in time to follow-up.

INTRODUCTION

Musculoskeletal pain, especially back pain (BP) and neck pain (NP) are highly prevalent in the general population[1 2] causing disability for the individual and high costs for society.[3–5] Individuals with BP and NP are mostly managed in primary care[6 7] and patients presenting with these conditions are at risk of sickness absence[8] and poor health-related quality of life (HRQoL).[9 10] To have concurrent BP and NP is also common[11] and increases the risk of work disability further in the long-term.[12] While most individuals with acute BP improve quickly and return to work,[13] for some of them the pain is more severe and lasts for a longer period.[14 15] In a Swedish cohort of individuals with BP and NP, about half of the population reported pain and disability 5 years after onset.[16] There are recommendations for the use of screening methods in healthcare to identify patients in early stages with the purpose to guide them to the best treatment,[17–19] to support staying at work or for enhancing return to work.[20 21] The UK National Institute for Health and Care Excellence guidance recommend

using brief questionnaires to identify individuals of poor outcomes and stratify care,[22] but there is a lack of such tools that can be used in primary care. The widely used STarT Back Tool (SBT)[23] is a brief risk stratification tool that includes nine questions on predictors for long-term disabling BP, in order to match individuals to appropriate targeted treatments, according to their prognostic profile. Using the SBT together with targeted treatment pathways has shown improved efficiency regarding patients' clinical outcomes and reduced healthcare costs in the UK.[24] The SBT is cross-culturally adapted and validated in Swedish[25] and recently also for a population with both BP and NP in primary care.[26] The SBT is developed and validated to predict future disability due to low BP of any duration,[23 27–30] but it has not yet been studied for the outcomes of HRQoL and work ability for a population with acute/subacute BP and NP in primary care. The aim of this study was therefore to evaluate the predictive validity of SBT of the outcomes HRQoL and work ability at long-term follow-up in a population with acute/subacute BP and/or NP.

METHODS

Design

We conducted a prospective psychometric validation study with long-term follow-up. The sample was identified in connection to a randomised clinical trial (RCT) in a primary care setting (ClinicalTrials.gov ID: NCT02609750).

Participants and procedure

Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care centres in the southern parts of Sweden, as part of an RCT.[31] Patients who all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) BP and/or NP, who were not currently on sick leave or had been on sick leave for <60 days and who had been working ≥4 consecutive weeks last year were asked to participate. It could be either a first episode or a recurrent episode of BP and/or NP after a period of at least 3 months of no substantial pain. Patients who were pregnant, had severe pathology ('red flags')[32] or were not able to understand the Swedish language were not eligible to participate. At baseline, patients completed the 'ÖMPSQ-short',[33] which was used for screening for inclusion to the RCT (≥40 points)[31] and the SBT, which was administered only for the purpose of psychometric testing. Thereafter, the SBT was not actively used by the physiotherapists or any other professionals. In all, 329 patients completed the SBT questionnaire and formed the population of this psychometric study. Patients who were older than 67 years or younger than 18 years (n=3), declined participation (n=4), had any missing item on the SBT (n=11) or those who were lost to follow-up (n=73) were excluded. The final study population (n=238) consisted of patients included in the RCT (RCT intervention, n=61 and RCT control, n=99) and patients not included in the RCT (n=78). The analyses were restricted to those who had complete data for work ability (n=235) and HRQoL (n=238) outcomes at long-term follow-up. The reason we included both RCT and not RCT patients was to ensure as broad a sample as possible for this SBT predictive validity study. RCT patients received either structured physiotherapy treatment with a workplace intervention (RCT intervention) or structured physiotherapy without a workplace intervention (RCT control)[31] and were followed up at the planned 12 months follow-up. Not RCT patients received usual primary care and were followed up by postal questionnaires. Data from all questionnaires were manually entered into a SPSS V.22.0 database and were thoroughly checked and validated. All questionnaires were scored, and missing items handled, according to the methods specified by the instrument developers.

Baseline data

Baseline questionnaire data included type of treatment received (RCT intervention, RCT control or usual primary care) and self-reports of SBT, age and gender.

STarT Back Tool

The SBT is a 9-item questionnaire with questions relating to modifiable physical (item 1–4) and psychosocial (item 5–9) risk factors for long-term disabling BP, designed to support clinicians in directing individuals to different levels of care.[23] The SBT has three risk subgroups which classifies patients into low, medium or high risk for poor disability outcomes. The SBT overall score ranges between 0 and 9. Item 1–4 is about referred leg pain, neck or shoulder pain, difficulties in walking and difficulties in dressing. Item 5–9 form the psychosocial subscale, which screen for fear of physical activity, anxiety, pain catastrophising, depressive mood and overall impact from their BP. Items 1–8 have a dichotomous response option; 'disagree' (0 p) or 'agree' (1 p). Item 9 uses a 5-point Likert scale from 'not at all' to 'extremely', where responses 'very much' or 'extremely' are counted as one point and the other responses as zero. A total score of ≤3 points indicates low risk, a total score ≥4 points in combination with <4 points on the psychosocial subscale (item 5–9) are medium risk and a psychosocial subscale score of ≥4 points indicates high risk for poor disability outcomes.[23]

Long-term follow-up data

Health-related quality of life

HRQoL was measured by the EuroQol five-dimension (EQ-5D, 3 L) questionnaire,[34] which is a generic, HRQoL instrument.[35 36] The EQ-5D comprises the EQ descriptive system which has five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The digits for the five dimensions are combined in a 5-digit number describing the respondent's health state.[37] The 5-digit number is given a value between −0.59 and 1.0 according to the UK tariff,[38] where 1 corresponds to

full health and lower EQ-5D values reflect lower HRQoL. HRQoL was also dichotomised into 'poor' HRQoL (EQ-5D<0.6) and 'good' HRQoL (EQ-5D≥0.6), based on a proposed cut-off for having sufficient capacity to be able to work for a population with BP and NP.[39]

Work ability

Work ability was measured by self-reports on the single item question ('current work ability compared with the lifetime best') from the Work Ability Index (WAI).[40 41] This first item in the WAI is known as the 'Work Ability Score' (WAS).[42] It consists of a scale from 0 representing 'cannot work at all right now' to 10 representing 'my work ability as at its best right now' and has been proposed to be used as a simple indicator for assessing the status and progress of work ability.[43 44] Work ability was also dichotomised using a previously published cut-off score[43] into 'poor' work ability (WAS <8 points) and 'good' work ability (WAS ≥8 points).

Statistical analyses

SPSS V.22.0 was used for all analyses. We used a non-parametric approach which was chosen based on the distribution of the data. Descriptive data on the study population was presented for the total population and for each SBT risk group. We separately evaluated the SBT-specific risk groups and also the SBT overall score.

Predictive performance of the SBT

First, cross-tabulations were used to describe the proportion of participants in each SBT risk group that had poor outcome in long-term follow-up for each outcome. The Kruskal-Wallis test was used to study if there were any differences between the SBT risk groups on follow-up data on HRQoL and work ability (median), respectively. Potential differences were confirmed with Mann-Whitney U test. X^2 test for trend was used to confirm potential differences concerning poor or good HRQoL and work ability.

Second, we calculated the ORs (95% CIs) for SBT risk groups to predict poor HRQoL (EQ-5D<0.6) and poor work ability (WAS<8) using binary logistic regression. Independent variables age, sex, treatment group and time to follow-up were also included in the analysis. We built a multiple logistic model where all independent variables were entered together with the SBT risk groups. For SBT, we used the SBT low risk group as the reference group and for treatment groups (RCT intervention n=61, RCT control n=99, not RCT n=78), we used the 'not RCT group' as the reference group. The significance level was set at 5%.

Third, we evaluated the ability of the SBT overall scores (0–9 points) to discriminate between individuals with poor or good HRQoL/work ability in long-term follow-up. For that purpose, we used the area under the curve (AUC) statistics from receiver operating characteristic (ROC) curves.[45] The strength of discrimination was set according to the following descriptors: 0.7 to <0.8 acceptable discrimination, 0.8 to <0.9 excellent discrimination, and ≥0.9 outstanding discrimination.[46]

In addition, the predictive validity of the SBT risk group cut-offs (low/medium and medium/high) was assessed by calculating sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV) and positive and negative likelihood ratios (LRs) against long-term HRQoL and work ability outcomes. The SBT risk group cut-offs (low/medium and medium/high) were used in line with the original study.[23] The PPV is the probability that a poor outcome is present when the test is positive and the NPV is the probability that a good outcome is present when the test is negative. Higher positive LRs and lower negative LRs indicate better discrimination. LRs >5 or <0.2 are generally seen as supporting a strong test, whereas values close to 1 indicate poor test performance.[47]

Patient and public involvement

Relevant patient organisations were involved in the development and design of the RCT, where this study was embedded. For this psychometric study, no patients were involved. The results of this study will be disseminated to study participants by the use of SBT in primary care.

Ethics

Prior to inclusion, all patients obtained written information about the purpose of the study and each individual gave informed consent to participate in the study (opt-out). The principles of the Declarations of Helsinki were followed.

RESULTS

Study population

The inclusion and exclusion of participants in the study is presented in a flow chart (figure 1).

The final sample consisted of 238/329 patients (72%) including 160 (67%) females and 78 (33%) males. Baseline characteristics of the study population are summarised in table 1. The patient sample included 103 (43%) patients at low risk, 107 (45%) patients at medium risk and 28 (12%) patients at high risk. The median time to long-term follow-up was 13 (range 11–27) months. For RCT patients, the median time to follow-up was 12 months (range 11–19) and for not RCT patients, the median time was 22 months (range 16–27).

Predictive performance of the SBT

There were statistically significant differences in the distribution of HRQoL scores (n=238) between the SBT low, medium and high risk groups at long-term follow-up (p<0.001) and the proportion of patients with poor HRQoL (EQ-5D<0.6) was significantly higher in higher risk groups (low risk 4%, medium risk 11%, high risk 36%) (p<0.001) (table 2). We also found differences in the distribution of work ability (WAS) scores (n=235) between the SBT low, medium and high risk groups at

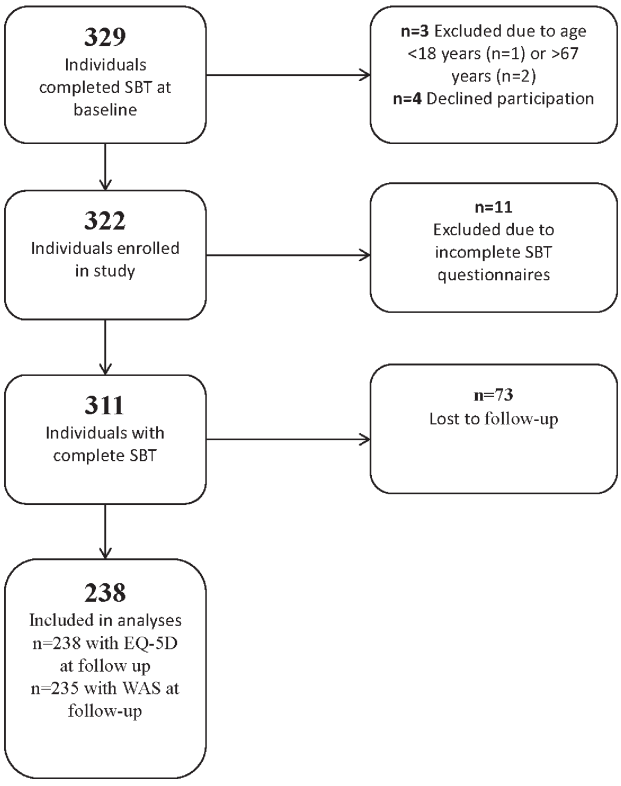

Figure 1 Flow chart of inclusion and exclusion of participants. EQ-5D, EuroQol five-dimension; SBT, STarT Back Tool; WAS, Work Ability Score.

long-term follow-up (p<0.001) and the proportion of patients with poor work ability (WAS <8) was significantly higher in higher risk groups (low risk 22%, medium risk 35%, high risk 68%) (p<0.001) (table 2).

The regression analysis showed that the SBT high risk group could significantly predict poor HRQoL (OR 6.16, 95% CI 1.50 to 25.26, B=1.82, p=0.012) and poor work ability (OR 5.08, 95% CI 1.75 to 14.71, B=1.62, p=0.003) at long-term follow-up also after adjusting for age, sex, treatment and time to follow-up (table 3). Our regression model was well adapted to the data material as a non-significant p value >0.05 of the test by Hosmer and Lemeshow indicates that the model is good[48] (table 3).

Regarding the ability of the SBT total scores (0–9 points) to discriminate between individuals with poor or good HRQoL at long-term follow-up, the AUC was 0.73 (95% CI 0.61 to 0.84), which was 'acceptable' (≥0.7) (figure 2). For work ability, the AUC was 0.68 (95% CI 0.61 to 0.76), which was just below the limit (≥7) for acceptable discrimination (figure 3).

The sensitivity, specificity, PPV, NPV and LRs for the SBT risk groups for HRQoL and work ability are presented in table 4. The LRs+ were higher and the LRs− were lower for HRQoL outcomes compared with work ability outcomes which indicate better discrimination of the SBT for poor HRQoL compared with poor work ability (table 3).

DISCUSSION AND CONCLUSIONS

This is the first study to evaluate the predictive validity of SBT for HRQoL and work ability outcomes at long-term follow-up in a population with acute/subacute BP and/or NP. The findings of this study support the ability of the SBT risk groups to predict future poor HRQoL or poor work ability, for patients presenting with an episode of acute/subacute BP and/or NP in primary care. Individuals classified as SBT high risk had a significantly increased risk of having poor HRQoL (OR 6.2) and poor work ability (OR 5.1) in the long-term compared with individuals classified as SBT low risk. The population studied was relatively homogenous including only patients with acute or subacute pain, not individuals with chronic pain. This study population differs from the original UK development population for SBT by excluding chronic BP and including NP. As might be expected,

Table 1 Baseline characteristics of the study population—total population and stratified by SBT risk groups

Variable	Total population n=238	SBT risk group		
		Low n=103 (43%)	Medium n=107 (45%)	High n=28 (12%)
Age, median (range)	46 (19–67)	45 (22–64)	47 (21–67)	38 (19–63)
Sex, *n* (%) female	160 (67)	73 (71)	72 (67)	15 (54)
Area of pain*				
BP, *n* (%)	91 (38)	42 (41)	41 (38)	8 (29)
NP+BP†, *n* (%)	147 (62)	61 (59)	66 (62)	20 (71)
Type of intervention, *n* (%)				
RCT control	99 (41)	21 (20)	60 (56)	18 (64)
RCT intervention	61 (26)	21 (20)	31 (29)	9 (32)
Not RCT	78 (33)	61 (60)	16 (15)	1 (4)

*Area of pain based on question number 2 (NP or shoulder pain) on SBT.
†NP+BP patients with neck or shoulder pain with or without back pain.
BP, back pain; NP, neck pain; RCT, randomised clinical trial; SBT, STarT Back Tool.

Table 2 Health-related quality of life and work ability at long-term follow-up—total population and stratified by SBT risk groups

		SBT risk group			
	Total population	Low	Medium	High	
Follow-up measure	n=238	n=103	n=107	n=28	P values
Health-related quality of life; median (range)	0.80 (–0.14–1)	0.80 (0.09–1)	0.76 (0.09–1)	0.67 (–0.14–1)	<0.001*
EQ-5D† <0.6, n (%)	26 (11)	4 (4)	12 (11)	10 (36)	<0.001§
Work ability‡; median (range)	8 (0–10)	9 (0–10)	8 (1–10)	7 (0–10)	<0.001*
WAS¶ <8, n (%)	78 (33)	23 (22)	38 (35)	17 (68)	<0.001§

*Kruskal-Wallis test.
†EQ-5D scores, range –0.59–1.
‡Three missing from the high risk group (total population: n=235 and n=25 for the high risk group).
§X^2 test for trend.
¶Where 0 equates to 'completely unable to work' and 10 equates to 'work ability at its best'.
EQ-5D, EuroQol five-dimension; SBT, STarT Back Tool; WAS, Work Ability Score.

the distribution between the SBT risk groups at baseline differed compared with the UK development population.[23] In our study population, the percentage of individuals at high risk was lower (12%) compared with the original UK sample (15%),[23] which may be due to our sample including patients with acute/subacute pain. However, there is still a clear and statistically significant difference in HRQoL and work ability outcomes between the three risk groups in the expected direction in our Swedish sample.

Strengths of this study include the prospective design of a well-characterised group of individuals from 35 different primary care centres. The SBT was used and administered by many different physiotherapists which makes this setting real and clinically relevant. Another strength is that we analysed the predictive validity in different ways, for example, we studied both the established SBT risk groups and the SBT overall score to predict the outcomes of HRQoL and work ability. We also analysed the outcomes HRQoL and work ability both on the continuous scale (Kruskal-Wallis) and as dichotomised (logistic regression).

A weakness of this study is that we had limited access to baseline data from patients not included in the RCT (n=78/238) compared with RCT patients (n=160/238). For not RCT patients, we did not have access to baseline data from HRQoL and work ability questionnaires. For that reason, we were not able to do comparative analyses

Table 3 The ability of the SBT risk groups to predict poor HRQoL* and poor work ability† at long-term follow-up

	HRQoL			Work ability		
Coefficient	OR	95% CI for OR	P values	OR	95% CI for OR	P values
SBT low risk group (ref)	1			1		
SBT medium risk group	1.814	0.506 to 6.509	0.361	1.361	0.684	0.380
SBT high risk group	6.160	1.502 to 25.264	0.012	5.075	1.751 to 14.705	0.003
Treatment not RCT (ref)	1			1		
Treatment RCT control	1.411	0.073 to 27.252	0.820	7.631	1.284 to 45.341	0.025
Treatment RCT intervention	2.932	0.183 to 47.073	0.448	8.156	1.485 to 44.803	0.016
Time to follow-up (months)	0.949	0.734 to 1.227	0.688	1.146	0.983 to 1.336	0.081
Age (years)	0.984	0.947 to 1.022	0.403	1.014	0.988 to 1.040	0.306
Sex, 0=female, 1=male (ref)	0.449	0.183 to 1.106	0.082	0.706	0.381 to 1.309	0.269
Test	**χ^2 test**	**P values**	**df**	**χ^2 test**	**P values**	**df**
Goodness-of-fit test Hosmer and Lemeshow test	5.41	0.71	8	5.27	0.73	8

*Poor HRQoL measured by EQ-5D questionnaire <0.6.
†Poor work ability measured by WAS<8.
HRQoL: Cox-Snell R^2=0.12. Nagelkerke R^2=0.21, n=238.
Work ability: Cox-Snell R^2=0.11. Nagelkerke R^2=0.16, n=235.
EQ-5D, EuroQol five-dimension; HRQoL, health-related quality of life; RCT, randomised clinical trial; SBT, StarT Back Tool; WAS, Work Ability Score.

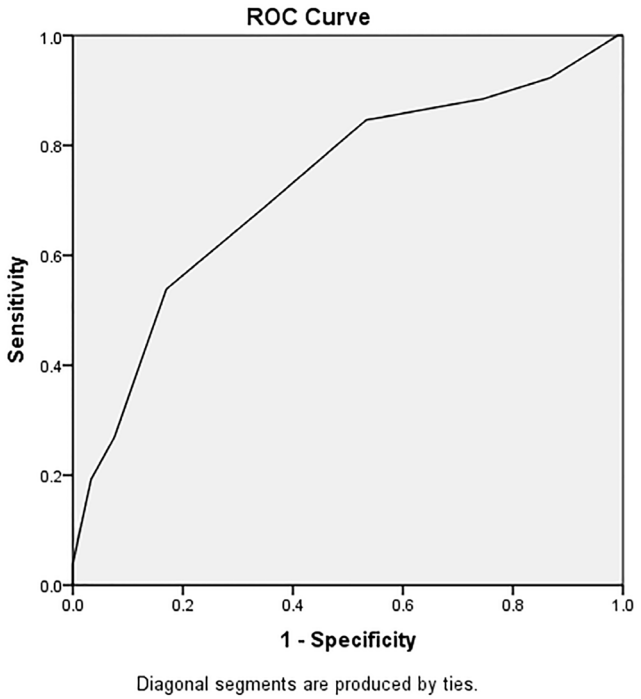

Figure 2 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor health-related quality of life (EQ-5D<0.6) in long-term follow-up. Each point on the ROC curve has a corresponding cut-off value. The area under the ROC curve was 0.73. AUC, area under the curve; EQ-5D, Euroqol 5-dimension questionnaire; ROC, receiver operation characteristic.

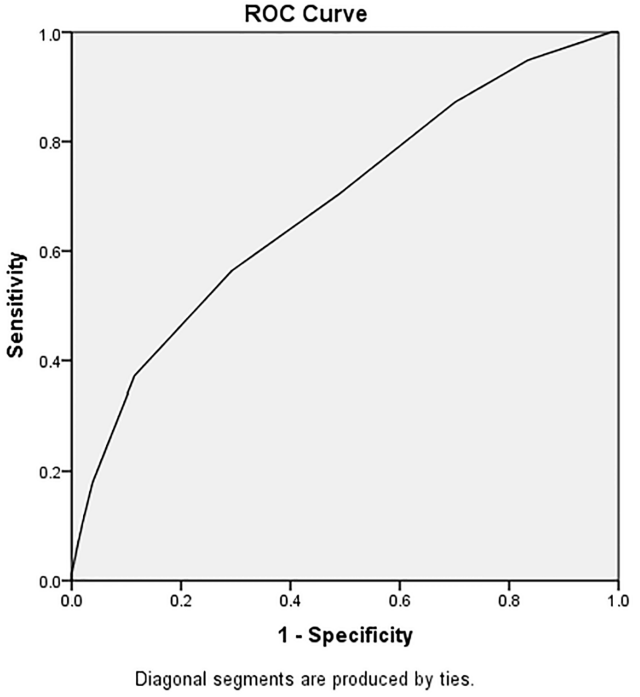

Figure 3 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor work ability (WAS<8) in long-term follow-up. Each point on the ROC curve has a corresponding cut-off value. The area under the ROC curve was 0.68. AUC, area under the curve; EQ-5D, Euroqol 5-dimension questionnaire; ROC, receiver operation characteristic.

on baseline and follow-up data. When recommending tools for use in primary care settings, preferably they should have been validated in large trials within this specific setting. However, as is the case with this study of the SBT, information from smaller studies is still of scientific value. We accept that our study population (n=329) is unlikely to be representative of all individuals consulting primary care for acute/subacute BP and/or NP. However, even if they are a selected group of participants, we do not think that this will have substantially affected the psychometric validation questions examined in this study.

The time to follow-up varied between patients in our study which may have influenced the results. The optimal time point for identifying patients at risk of developing persistent BP may vary and is a forum for discussion.[49] In our study, two-third of the study population (n=160) were in the RCT and were followed up at a planned physiotherapy visit at 12 months. For not RCT patients (n=78), the ambition was also to follow-up at 12 months but these patients were followed up with postal questionnaires and due to practical reasons there were a wider variation on the time for follow-up. This is of course a limitation, but did not have impact on the results in the regression analyses. However, we had access to information about tentative confounding factors and we investigated several of these factors (age, sex, treatment and time to follow-up) that may have potentially influenced the prognostic ability of the SBT. In this study, we included both patients with NP and BP. Since this group of patients often have concurrent pain from the back or neck,[11] we decided to not include this in the regression analysis. In another SBT non-stratified primary care setting where they studied different influences (care setting, episode duration and time to follow-up) on the prognostic ability of the SBT for disability outcomes[50] they found that the only factor that modified the prognostic ability of the SBT risk groups was episode duration with SBT being less predictive in very acute patients (<2 weeks duration).

The ability of the SBT overall score to discriminate between patients with poor or good HRQoL and work ability differed slightly between the two outcomes with a slightly better discrimination for HRQoL (0.73) than for work ability (0.68). The AUC values are not very high, but still around 0.7, which is considered as acceptable.[46] In a recent systematic review, Karran *et al*[51] investigated how well prognostic screening instruments for BP, including the SBT, discriminate between patients who develop a poor outcome and those who do not.[51] Prognostic screening tools tend to perform poorly at assigning higher risk scores to individuals who develop chronic pain compared with those who do not and they also tend to predict disability outcomes better than most other outcomes.[51] The discriminative performance of SBT for work ability outcomes in this study (AUC 0.68) was higher than for other prognostic tool's reported abilities to discriminate pain outcomes (pooled AUC=0.59)[51] and the SBT

Table 4 Discriminative ability of the SBT risk group cut-offs (L/M and M/H) to predict poor HRQoL and poor work ability in long-term follow-up

Subgroups	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+ (95% CI)	LR– (95% CI)
HRQoL (EQ-5D<0.6)						
L vs M/H	84.6	46.7	16.3	96.1	1.59 (1.29 to 1.95)	0.33 (0.13 to 0.82)
L/M vs H	38.5	91.5	35.7	92.4	4.53 (2.35 to 8.74)	0.67 (0.49 to 0.91)
Work ability (WAS<8)						
L vs M/H	70.5	51.0	41.7	77.7	1.44 (1.16 to 1.78)	0.58 (0.40 to 0.84)
L/M vs H	21.8	94.9	68.0	71.0	4.28 (1.93 to 9.47)	0.82 (0.73 to 0.93)

EQ-5D, EuroQol five-dimension; H, high; HRQoL, health-related quality of life; L, low; LR+, positive likelihood ratio; LR–, negative likelihood ratio; M, medium; NPV, negative predictive value; PPV, positive predictive value; SBT, STarT Back Tool; WAS, Work Ability Score.

discriminative performance for HRQoL outcomes in this study (AUC 0.73) was in line with the pooled disability predictive performance (pooled AUC=0.74). In comparison to the original UK sample and a Danish sample in primary care, where participants had variable duration of BP and the primary outcomes were disability at 3 months follow-up,[23 28] the predictive ability of the SBT in our study was not as strong as in the UK population (AUC 0.81) but similar to the Danish population (AUC 0.71). In our study, as in the Danish study, the physiotherapy treatment was not targeted to SBT risk groups and treatment was therefore likely to be heterogeneous. A variation of values are expected as the AUC (derived from the ROC curve: sensitivity/1−specificity), depends on the characteristics of the population and possible explanations might be cultural and differences in treatment. Another possible explanation in variation of AUC values may be that a ROC curve analysis requires dichotomisation of outcomes and the definitions of poor outcome may also have affected the results. The discriminative ability of the SBT risk groups to predict poor HRQoL and work ability outcome was affected of how the three risk groups were merged and dichotomised (low vs medium/high or low/medium vs high). Similar differences in discrimination were also found in the original study for disability outcomes.[23] But regardless of which cut-off that was used, the results of the LRs indicate a slightly better discrimination of the SBT for poor HRQoL than for poor work ability and that the NPVs were consistently high for both outcomes which indicate a high probability that a good outcome is present when patients are classified as low risk. The proportion of patients with poor HRQoL and poor work ability was significantly higher in higher SBT risk groups at long-term follow-up, but not all patients were correctly classified. When patients are misclassified as low risk they may be undertreated and when patients are misclassified as high risk they may be overtreated. It is important for clinicians to be aware of the potential of misclassification as costs for misclassification and overtreatment of patients with a good prognosis can be high[24] and also detrimental in patients with acute BP.[52]

The EQ-5D was applied to measure HRQoL because it has been found to have good prediction of return to work and the cut-off ≥0.6 on EQ-5D has been proposed to be a limit for having sufficient capacity to work for patients with BP and NP.[39] Another cut-off has been used in a study of patients with musculoskeletal pain taking part in a national rehabilitation programme in Sweden where ≥0.5 on EQ-5D at start showed reduced sick leave days after the rehabilitation.[53] Our population had a median EQ-5D score of 0.80, which is just below the mean scores for a Swedish normal population (0.84).[54] The fact that our sample included patients at an early stage of their pain (acute/subacute) with no or short time of sick leave may have influenced the high level of HRQoL in our study sample. To measure work ability, we used the WAS which is the first item in the WAI, a widely used questionnaire for measuring the health and functional capacity dimension of work ability.[41] The cut-off (WAS<8/≥8) chosen in this study represents poor or moderate (poor) and good/excellent (good) work ability based on the same categorisation as for the whole WAI.[42] The WAS has shown to be a good alternative to the whole WAI,[43] even though the whole WAI is superior compared with its individual items.[55]

SBTs concurrent validity has earlier been studied for patients with BP and/or NP[26] and a modified SBT have been tested to predict physical health outcome, using the SF-36[56] but this was the first time the predictive validity of the SBT was studied for the outcomes of HRQoL and work ability for individuals with both BP and NP. Therefore, this study widens the usefulness of the SBT compared with earlier studies.[23 57–60] There is also need for short questionnaires that are easy to use in clinical to distribute and interpret, especially in primary care. The SBT is primarily designed as a 'stratified care tool', which involves targeting treatment to subgroups of patients based on their key characteristics[61] but in this study, we wanted to study if the SBT could predict the important outcomes HRQoL and work ability when applied in an RCT of NP and BP. In this study, the physiotherapists did not target treatment based on SBT. However, we accept that some of the constructs within the SBT may

have been addressed by the intervention provided, which may have affected SBTs ability to predict the above-mentioned outcomes. The results of this study suggest that the SBT can be used as a prognostic tool in primary care for subgroup identification of acute/subacute BP and/or NP patients at risk of poor long-term HRQoL and/or work ability outcome. This information about important risk factors may help clinicians in primary care to develop personalised treatment strategies which are a priority in research.[62] Future studies are required to investigate whether the implementation of screening together with matched treatment pathways have an effect on HRQoL and work ability outcomes for these patients.

Author affiliations
[1]Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden
[2]Blekinge Centre of Competence, Landstinget Blekinge, Karlskrona, Sweden
[3]Department of Research and Development, Region Kronoberg, Växjö, Sweden
[4]Research Institute of Primary Care and Health Sciences, Keele University, Stoke-on-Trent, UK
[5]Skåne University Hospital, Lund, Sweden
[6]Department of Health Sciences, Physiotherapy, Lund University, Lund, Sweden

Acknowledgements The authors would like to thank Senior statistician Mikael Åström for statistical support, data manager Chanchai Manuswin and the staff at Data analysis and Register Centre, Region Skåne, Lund, Sweden, the research team at the Institute of Primary Care and Health Sciences, Keele University, Stoke-on-Trent, UK and the staff at the Blekinge Centre of Competence, Karlskrona, Sweden for help with support. The authors would also like to thank all physiotherapists and patients involved in this project, including the WorkUp project, REHSAM, in Region Skåne, Region Kronoberg and Landstinget Blekinge, Sweden, for their help with data collection.

Contributors All authors discussed the results and commented on the manuscript. MF, IP, KS and BG were responsible for the study design, data analysis and interpretation. MF, BG and KS prepared and validated data. MF collected data and drafted the manuscript. JH and CPS took part in study design, data analysis and interpretation of data. All authors read and approved the final version of the manuscript.

Funding This work was supported by The Scientific Committee of Blekinge County Council, Sweden (Dnr 2017-00686) and the County Councils Region Skåne, Kronoberg and Blekinge, REHSAM research, Sweden (RS2011/005).

Competing interests None declared.

Patient consent Not required.

Ethics approval The study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2012/497, 2013/426, Dnr 2015/214).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Vos T, Flaxman AD, Naghavi 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;380:628:2163–96. Note: AlMazroa, Mohammad A [added]; Memish, Ziad A [added]]. *Lancet (London, England)* 2012;380(9859):2163-96.
2. Bergman S, Herrström P, Högström K, *et al*. Chronic musculoskeletal pain, prevalence rates, and sociodemographic associations in a Swedish population study. *J Rheumatol* 2001;28:1369–77.
3. Gerdle B, Björk J, Cöster L, *et al*. Prevalence of widespread pain and associations with work status: a population study. *BMC Musculoskelet Disord* 2008;9:102.
4. Gustavsson A, Bjorkman J, Ljungcrantz C, *et al*. Socio-economic burden of patients with a diagnosis related to chronic pain--register data of 840,000 Swedish patients. *Eur J Pain* 2012;16:289–99.
5. Alexanderson K, Norlund A. Chapter 1. Aim, background, key concepts, regulations, and current statistics. *Scand J Public Health* 2004;32(63_suppl):12–30.
6. Kinge JM, Knudsen AK, Skirbekk V, *et al*. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. *BMC Musculoskelet Disord* 2015;16:75.
7. Jordan KP, Jöud A, Bergknut C, *et al*. International comparisons of the consultation prevalence of musculoskeletal conditions using population-based healthcare data from England and Sweden. *Ann Rheum Dis* 2014;73:212–8.
8. Holtermann A, Hansen JV, Burr H, *et al*. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. *Scand J Work Environ Health* 2010;36:34–41.
9. Nolet PS, Côté P, Kristman VL, *et al*. Is neck pain associated with worse health-related quality of life 6 months later? A population-based cohort study. *Spine J* 2015;15:675–84.
10. Nolet PS, Kristman VL, Côté P, *et al*. Is low back pain associated with worse health-related quality of life 6 months later? *Eur Spine J* 2015;24:458–66.
11. Leijon O, Wahlström J, Mulder M. Prevalence of self-reported neck-shoulder-arm pain and concurrent low back pain or psychological distress: time-trends in a general population, 1990-2006. *Spine* 2009;34:1863–8.
12. Nyman T, Grooten WJ, Wiktorin C, *et al*. Sickness absence and concurrent low back and neck-shoulder pain: results from the MUSIC-Norrtälje study. *Eur Spine J* 2007;16:639–40. *Eur Spine J* 2007;16(5):631-8.
13. Pengel LH, Herbert RD, Maher CG, *et al*. Acute low back pain: systematic review of its prognosis. *BMJ* 2003;327:323.
14. Dunn KM, Hestbaek L, Cassidy JD. Low back pain across the life course. *Best Pract Res Clin Rheumatol* 2013;27:591–600.
15. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *Eur Spine J* 2003;12:149–65.
16. Enthoven P, Skargren E, Oberg B. Clinical course in patients seeking primary care for back or neck pain: a prospective 5-year follow-up of outcome and health care consumption with subgroup analysis. *Spine* 2004;29:2458–65.
17. Hemingway H, Croft P, Perel P, *et al*. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
18. Hingorani AD, Windt DA, Riley RD, *et al*. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
19. Foster NE, Mullis R, Hill JC, *et al*. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. *Ann Fam Med* 2014;12:102–11.
20. Kuoppala J, Lamminpää A. Rehabilitation and work ability: a systematic literature review. *J Rehabil Med* 2008;40:796–804.
21. Carroll C, Rick J, Pilgrim H, *et al*. Workplace involvement improves return to work rates among employees with back pain on long-term sick leave: a systematic review of the effectiveness and cost-effectiveness of interventions. *Disabil Rehabil* 2010;32:607–21.
22. Bernstein IA, Malik Q, Carville S, *et al*. Low back pain and sciatica: summary of NICE guidance. *BMJ* 2017;356:i6748.
23. Hill JC, Dunn KM, Lewis M, *et al*. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;59:632–41.
24. Hill JC, Whitehurst DG, Lewis M, *et al*. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet* 2011;378:1560–71.
25. Betten C, Sandell C, Hill JC, *et al*. Cross-cultural adaptation and validation of the Swedish STarT Back Screening Tool. *Eur J Physiother* 2015;17:29–36.
26. Forsbrand M, Grahn B, Hill JC, *et al*. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Örebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. *BMC Musculoskelet Disord* 2017;18:89.
27. Beneciuk JM, Fritz JM, George SZ. 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;44:656–64.
28. Morsø L, Kent P, Albert HB, *et al*. The predictive and external validity of the STarT Back Tool in Danish primary care. *Eur Spine J* 2013;22:1859–67.
29. Bruyère O, Demoulin M, Beaudart C, *et al*. Validity and reliability of the French version of the STarT Back screening tool for patients with low back pain. *Spine* 2014;39:E123–8.
30. Bier JD, Ostelo R, van Hooff ML, *et al*. Validity and reproducibility of the STarT Back Tool (Dutch Version) in patients with low back pain in primary care settings. *Phys Ther* 2017;97:561–70.
31. Sennehed CP, Holmberg S, Axén I, *et al*. Early workplace dialogue in physiotherapy practice improved work ability at 1-year follow-up-WorkUp, a randomised controlled trial in primary care. *Pain* 2018;159:1456–64.
32. van Tulder M, Becker A, Bekkering T, *et al*. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 2006;15(Suppl 2):s169–91.
33. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Örebro Musculoskeletal Pain Screening Questionnaire. *Spine* 2011;36:1891–5.
34. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337–43.
35. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
36. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
37. Greiner W, Weijnen T, Nieuwenhuizen M, *et al*. A single European currency for EQ-5D health states. *The European Journal of Health Economics* 2003;4:222–31.
38. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095–108.
39. Hansson E, Hansson T, Jonsson R. Predictors for work ability and disability in men and women with low-back or neck problems. *Eur Spine J* 2006;15:780–93.
40. Tuomi K, Ilmarinen J, Jahkola A, *et al*. *Work ability index, Finnish Institute of Occupational Health. 2*. Helsinki, 1998.
41. Ilmarinen J, Tuomi K, Seitsamo J. New dimensions of work ability. *Int Congr Ser* 2005;1280:3–7.
42. Gould R, Ilmarinen J, Järvisalo J, *et al*. *Dimensions of work ability*, 2008.
43. Ahlstrom L, Grimby-Ekman A, Hagberg M, *et al*. The work ability index and single-item question: associations with sick leave, symptoms, and health--a prospective study of women on long-term sick leave. *Scand J Work Environ Health* 2010;36:404–12.
44. El Fassi M, Bocquet V, Majery N, *et al*. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. *BMC Public Health* 2013;13:305.
45. Kirkwood B, Sterne J. *Medical statistics*. 2nd edn. Oxford, UK: Blackwell Publishing, 2003.
46. Hosmer Jr D, Lemeshow S, Sturdivant RX. Introduction to the Logistic Regression Model. *Applied Logistic Regression* 2000.
47. Mehling WE, Avins AL, Acree MC, *et al*. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *Eur J Pain* 2015;19:439–46.
48. Cole T. In: Hosmer DW, Lemeshow S, *Applied logistic regression*. Wiley, New York: Wiley Online Library, 1991.
49. Melloh M, Elfering A, Käser A, *et al*. What is the best time point to identify patients at risk of developing persistent low back pain? *J Back Musculoskelet Rehabil* 2015;28:267–76.
50. Morso L, Kongsted A, Hestbaek L, *et al*. Erratum to: The prognostic ability of the STarT Back Tool was affected by episode duration. *Eur Spine J* 2016;25:2677–8.
51. Karran EL, McAuley JH, Traeger AC, *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.
52. George SZ, Fritz JM, Bialosky JE, *et al*. The effect of a fear-avoidance-based physical therapy intervention for patients with acute low back pain: results of a randomized clinical trial. *Spine* 2003;28:2551–60.
53. Stigmar KG, Petersson IF, Jöud A, *et al*. Promoting work ability in a structured national rehabilitation program in patients with musculoskeletal disorders: outcomes and predictors in a prospective cohort study. *BMC Musculoskelet Disord* 2013;14:57.
54. Burström K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the EQ-5D. *Qual Life Res* 2001;10:621–35.
55. Lundin A, Leijon O, Vaez M, *et al*. Predictive validity of the Work Ability Index and its individual items in the general population. *Scand J Public Health* 2017;45:350–6.
56. Hill JC, Afolabi EK, Lewis M, *et al*. Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data. *BMJ Open* 2016;6:e012445.
57. Hill JC, Dunn KM, Main CJ, *et al*. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 2010;14:83–9.
58. Wideman TH, Hill JC, Main CJ, *et al*. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain* 2012;153:2182–91.
59. Beneciuk JM, Robinson ME, George SZ. 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.
60. Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Phys Ther* 2011;91:722–32.
61. Foster NE, Hill JC, O'Sullivan P, *et al*. Stratified models of care. *Best Pract Res Clin Rheumatol* 2013;27:649–61.
62. Rozenberg S, Foltz V, Fautrel B. Treatment strategy for chronic low back pain. *Joint Bone Spine* 2012;79:555–9.