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 prospective cohort study in primary care



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2018-021748 |
| Article Type: | Research |
| Date Submitted by the Author: | 17-Jan-2018 |
| Complete List of Authors: | Forsbrand, Malin; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Landstinget Blekinge Blekinge kompetenscentrum,<br>Grahn, Birgitta; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg<br>Hill, Jonathan; Keele University, Research Institute of Primary Care and Health Sciences<br>Petersson, Ingemar; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Skåne University Hospital<br>Post Sennehed, Charlotte; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg<br>Stigmar, Kjerstin; Department of Health Sciences, Physiotherapy, Lund University; Skåne University Hospital |
| Keywords: | STarT Back Tool, health related quality of life, work ability, PRIMARY CARE, neck pain, Back pain < ORTHOPAEDIC & TRAUMA SURGERY |
| | |

# 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 prospective cohort study in primary care

M. Forsbrand[1,2*], B. Grahn[1,3], JC. Hill[4], IF. Petersson[1,6], C. Post Sennehed[1,3], K. Stigmar[5,6]

[1]Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, 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, United Kingdom, [5]Department of Health Sciences, Physiotherapy, Lund University, Lund, Sweden, [6]Skåne University Hospital, Lund, Sweden.

***Corresponding author:**

Malin Forsbrand, RPT, PhD student.

Address: Blekinge Centre of Competence, SE-371 81 Karlskrona, Sweden.

E-mail: malin.forsbrand@med.lu.se

Telephone: +46455735616

**E-mails authors (not corresponding):** birgitta.grahn@kronoberg.se, j.hill@keele.ac.uk, ingemar.petersson@skane.se, charlotte.sennehed@kronoberg.se, kjerstin.stigmar@med.lu.se

**Word count:** 3997 words

**Category:** Research article

**Funding sources:** 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).

**No conflicts of interest declared**

# 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 centers in south Sweden during 2013.

**Participants:** Patients (n=329) with acute/subacute back and/or neck pain, aged 18-67, 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=0.000) and work ability (p=0.000) 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 1.50-25.26) and poor work ability (OR 5.08, 95 % CI 1.75-14.71) vs the low risk group at follow-up. The AUC was 0.73 (CI 0.61-0.84) for HRQoL and 0.68 (CI 0.61-0.76) 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.

**Keywords:** STarT Back Tool, health related quality of life, work ability, primary care, back pain, neck pain.

# Strengths and limitations of this study

- This is the first study 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 back and/or neck pain.
- In this prospective study we have recruited patients from 35 different primary care centers, where many physiotherapists were engaged.
- The predictive validity of the SBT was examined in different ways.
- Limited baseline information was available for one part of the cohort.
- 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]. Whilst most individuals with acute back pain improve quickly and return to work [11], for some of them the pain is more severe and lasts for a longer period [12 13]. In a Swedish cohort of individuals with BP and NP about half of the population reported pain and disability 5 years after onset [14]. Evidence-based guidelines [15] therefore, recommend that clinicians assess patient prognosis using brief questionnaires to identify individuals at risk of poor outcomes in order to achieve effective treatment allocation [16] and to direct the limited healthcare resources available to those most in need.

The widely used STarT Back Tool (SBT) [17], is a brief risk stratification tool that includes nine questions on predictors for long-term disabling back pain, 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 health care costs in the United Kingdom [18]. The SBT is developed and validated to predict future disability due to low back pain of any duration [17 19-23], but it has not yet been studied for the outcomes of HRQoL and work ability for a population with acute/subacute back and neck pain 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 back and/or neck pain. We separately evaluated the SBT specific risk groups and also the SBT overall score.

# Methods

## Design

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

## Participants and procedure

Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care centers in the southern parts of Sweden. All patients that applied for physiotherapy treatment by direct access due to an episode of acute or subacute (<12 weeks) non-specific BP and/or NP and who were not currently on sick leave or had been on sick leave for less than 60 days, were asked to complete the SBT questionnaire (n=329) at their first physiotherapy session. Patients that were older than 67 years or younger than 18 years (n=3) or did not accept to participate (n=4) were excluded. The broad inclusion criteria were chosen to identify a cohort representative for clinical practice. The SBT was completed at baseline and thereafter not actively used by the physiotherapist or any other professionals.

All patients were followed up with self-reported questionnaires including items on work ability and HRQoL. Patients with any missing item on the SBT (n=11) and those who were lost to follow-up (n=73) were excluded. The final study cohort included 238 participants. 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 study cohort consisted of patients that had been included in the RCT (RCT intervention group, n=61 and RCT control group, n=99) and patients that had not been included in the RCT (n=78). The reason we included patients who had been excluded from the RCT was to ensure we had as broad a sample as possible for this SBT predictive validity study. RCT patients (n=160) received either structured physiotherapy treatment (including examination, assessment, diagnosis, evidence-based treatment and follow-up) with a workplace intervention (RCT intervention group) or structured physiotherapy without a workplace intervention (RCT

control group) (ClinicalTrials.gov ID: NCT02609750) and were followed up at 12-months (median 12 months, range 11-19). Excluded RCT patients received usual primary care and were followed up around 18-24 months (median 22 months, range 16-27). Data from all questionnaires were manually entered into a SPSS 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 group, RCT control group or usual primary care) and self-reports of SBT, age and gender.

### STarT Back Tool

The STarT Back Tool (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 [17]. 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 catastrophizing, depressive mood and overall impact from their BP. Items 1–8 have a dichotomous response option; “disagree” (0p) or “agree” (1p). 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 [17].

## Long-term follow-up data

### *Health related quality of life*

Health-related quality of life (HRQoL) was measured by the EuroQol five-dimension (EQ-5D, 3L) questionnaire [24] which is a generic, health-related quality of life instrument [25 26]. The EQ-5D comprises the EQ descriptive system which has 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The digits for the 5 dimensions are combined in a 5-digit number describing the respondent´s health state [27]. The 5-digit number is given a value between -0.59 and 1.0 according to the UK tariff [28], where 1 corresponds to full health and lower EQ-5D values reflect lower HRQoL. Health Related Quality of Life was dichotomized 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 back and neck pain [29].

### *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) [30 31]. This first item in the WAI is known as the "Work Ability score" (WAS) [32]. 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 [33 34]. Work ability was dichotomized using a previously published cut-off score [33] into "poor" work ability (WAS<8 points) and "good" work ability (WAS ≥8 points).

## Statistical analyses

SPSS 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 cohort and for each SBT risk group.

## 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 poor or good HRQoL and work ability, respectively. Mann Whitney U-test and Chi-squared test for trend was used to confirm potential differences.

Second, we calculated the odds ratios (95% confidence intervals) for poor outcome on HRQoL (EQ-5D<0.6) and work ability (WAS<8) for SBT risk groups using binary logistic regression. Independent variables age, sex, treatment group or time to follow-up (months) 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 group n=61, RCT control group n=99, Not RCT group 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 [35]. The strength of discrimination was set according to the following descriptors: 0.7-<0.8 acceptable discrimination, 0.8-<0.9 excellent discrimination, and ≥0.9 outstanding discrimination [36].

In addition, the predictive validity of the SBT risk group cutoffs (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 cutoffs (low/medium and medium/high) were used in line with the original study [17]. 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. Likelihood ratios above 5 or below 0.2 are generally seen as supporting a strong test, whereas values close to 1 indicate poor test performance [37].

# Ethics

The study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2012/497, 2013/426, Dnr 2015/214). 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 flowchart (Figure 1).

**INSERT FIG 1 here**

The final sample consisted of 238/329 patients (72%) including 160 (67%) females and 78 (33%) males. Baseline characteristics of the study population are summarized 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.

**Table 1.** Baseline characteristics of the study population – total cohort and stratified by SBT risk groups.

| Variable | Total population | SBT risk group | | |
|---|---|---|---|---|
| | | Low | Medium | High |
| | *n*=238 | *n*=103 (43%) | *n*=107 (45%) | *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[a] | | | | |
| BP[b], *n* (%) | 91 (38) | 42 (41) | 41 (38) | 8 (29) |
| NP + BP[c], *n* (%) | 147 (62) | 61 (59) | 66 (62) | 20 (71) |
| SBT total score 0-9, median (range) | 4 (0-9) | 2 (0-3) | 5 (4-7) | 7 (6-9) |

SBT, STarT Back Tool

[a]Area of pain Based on question number 2 (neck or shoulder pain) on SBT

[b]BP Back pain

[c]NP + BP Patients with neck or shoulder pain (NP) with or without back pain

## 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=0.000) 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=0.000) (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 long-term follow-up (p=0.000) 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=0.000) (Table 2).

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

| Follow-up measure | Total population *n*=238 | SBT risk group Low *n*=103 | Medium *n*=107 | High *n*=28 | *p*-value |
|---|---|---|---|---|---|
| **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) | |
| EQ-5D[a] <0.6, n (%) | 26 (11) | 4 (4) | 12 (11) | 10 (36) | p=0.000[d] |
| **Work ability**[b]; median (range) | 8 (0-10) | 9 (0-10) | 8 (1-10) | 7 (0-10) | |
| WAS[c] <8, n (%) | 78 (33) | 23 (22) | 38 (35) | 17 (68) | p=0.000[d] |

SBT, STarT Back Tool; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score
[a]EQ-5D scores, range -0.59-1
[b]3 missing from the high risk group (total cohort: n=235 and n=25 for the high risk group)
[c]Where 0 equates to "completely unable to work" and 10 equates to "work ability at its best"
[d]Chi square test for trend

The regression analysis showed that the SBT high risk group significantly predicted poor HRQoL (OR 6.16, CI 1.50-25.26, *B*=1.82, *p*=0.012) and poor work ability (OR 5.08, CI 1.75-14.71, *B*=1.62, *p*=0.003) at long-term follow-up. None of the variables age, sex, treatment or time to follow-up had a significant influence on the ability of the SBT to predict HRQoL or work ability. Our regression model was well adapted to the data material (for HRQoL; $\chi^2$-test=5.41, df 8, p=0.71 and for work ability; $\chi^2$-test=5.27, df 8, p=0.73) as a non-significant p-value >0.05 indicates that the model is good [38]. For HRQoL, the Cox-Snell $R^2$ was 0.12 and Nagelkerke $R^2$ was 0.21 and for work ability, the Cox-Snell $R^2$ was 0.11 and Nagelkerke $R^2$ was 0.16.

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 area under the curve (AUC) was 0.73 (CI 0.61-0.84) which was

'acceptable' (≥0.7) (Fig. 2). For work ability, the area under the curve (AUC) was 0.68 (CI 0.61-0.76) which was just below the limit (≥7) for acceptable discrimination (Fig. 3).

**INSERT FIG 2 and FIG 3 here**

The sensitivity, specificity, PPV, NPV and likelihood ratios for the SBT risk groups for HRQoL and work ability are presented in Table 3. The LR+s were higher and the LR-s were lower for HRQoL outcomes compared to work ability outcomes which indicate better discrimination of the SBT for poor HRQoL compared to poor work ability (Table 3).

**Table 3**. Discriminative ability of the SBT risk group cutoffs (low/medium and medium/high) 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 (1.29-1.95) | 0.33 (0.13-0.82) |
| L/M vs. H | 38.5 | 91.5 | 35.7 | 92.4 | 4.53 (2.35-8.74) | 0.67 (0.49-0.91) |
| **Work ability (WAS <8)** | | | | | | |
| L vs. M/H | 70.5 | 51.0 | 41.7 | 77.7 | 1.44 (1.16-1.78) | 0.58 (0.40-0.84) |
| L/M vs. H | 21.8 | 94.9 | 68.0 | 71.0 | 4.28 (1.93-9.47) | 0.82 (0.73-0.93) |

SBT, STarT Back Tool; HRQoL, Health related quality of life; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

# 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 back and/or neck pain. 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 back and/or neck pain 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 to individuals classified as SBT low risk.

Strengths of this study include the prospective design of a well characterized group of individuals from 35 different primary care centers. The SBT was used and administered by many different physiotherapists which makes this setting real and clinically relevant. 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 back pain and including neck pain. As might be expected, the distribution between the SBT risk groups at baseline differed compared to the UK development population [17]. In our study population, the percentage of individuals at high risk were lower (12%) compared to the original UK sample (15%) [17] 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. As the majority of the patients in this study (n=160/238) were included in an RCT (ClinicalTrials.gov ID: NCT02609750) we have access to information about tentative confounding factors and we investigated several of these factors (age, sex, type of treatment and time to follow-up) that may have potentially influenced the prognostic ability of the SBT. This study showed that age, sex, type of treatment or time to follow-up did not significantly influence the ability of the SBT to predict HRQoL and work ability outcomes at long-term follow-up. 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 [39] 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). Another strength is that we analyzed 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 analyzed the outcomes HRQoL and work ability both on the continuous scale (Kruskal-Wallis) and as dichotomized (logistic regression).

A weakness of this study is that we had limited access to information about patients not included in the RCT (n=78/238) compared to patients included in the RCT (n=160/238). For patients not included in the RCT, we did not have access to registered diagnoses, pain duration (acute or subacute) or self-reported HRQoL and work ability questionnaires at baseline. For that reason, we were not able to do comparative analyzes on baseline and follow-up data. Another weakness might be the variation in time to follow-up between patients which may have influenced the results. For patients included in the RCT, median time to follow-up was 12 months (range 11-19) and for patients not included in the RCT, it was 22 months (range 16-27). Therefore we used the follow-up time as one of the independent variables in the regression analysis. All data in this study is self-reported. However, self-reported data on sickness absence among employees in Sweden has been reported at least as valuable as register data [40].

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). In a recent systematic review, Karran et al [41] investigated how well prognostic screening instruments for BP, including the SBT, discriminate between patients who develop a poor outcome and those who do not [41]. Prognostic screening tools tend to perform poorly at assigning higher risk scores to individuals who develop chronic pain compared to those who do not and they also tend to predict disability outcomes better than most other outcomes [41]. 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) [41] and the SBT 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 back pain and the primary outcomes were disability at 3 months follow-up [17 42], 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 dichotomization 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 dichotomized (low vs medium/high or low/medium vs high). Similar differences in discrimination were also found in the original study for disability outcomes[17]. But regardless of which cutoff 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 [18] and also detrimental in patients with acute back pain [43].

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 back and neck pain [29]. Another cut-off has been used in a study of patients with musculoskeletal pain taking part in a national rehabilitation program in Sweden where ≥0.5 on EQ-5D at start showed reduced sick leave days after the rehabilitation [44]. 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)[45]. The fact that our sample included patients at an early stage of their pain (acute or 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 [31]. 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 categorization as for the whole WAI [32]. The WAS has shown to be a good alternative to the whole WAI [46] even though the whole WAI is superior compared to its individual items [47].

There are recommendations for the use of screening methods in health care to identify patients in early stages with the purpose to guide them to the best treatment [48-50] and also for enhancing return to work [51 52]. SBTs concurrent validity has earlier been studied for patients with back and/or neck pain [53] and a modified SBT have been tested to predict physical health outcome, using the SF-36 [54] 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 back and neck pain. This study showed that the SBT can identify acute/subacute back and neck pain patients with a poor long-term HRQoL and work ability outcome. Therefore this study widens the usefulness of the SBT compared to earlier studies [17 55-58]. More research is needed to find appropriate treatments for patients with nonspecific acute/subacute back and/or neck pain [59]. The SBT is primarily designed as a "stratified care tool" which involves targeting treatment to subgroups of patients based on their key characteristics [60]. Future studies are required to investigate whether the implementation of screening together

with matched treatment pathways improves HRQoL and work ability outcomes for these patients. The results of this study suggest that the SBT may help clinicians in primary care to pay more attention to work related factors at an early stage which is a priority in preventing chronicity [61] and essential for a successful rehabilitation process [62].

## Authors Contributions

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.

## Acknowledgements

Senior statistician Mikael Åström for statistical support, datamanager 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. All physiotherapists and patients involved in this project, including the WorkUp project, REHSAM, in Region Skåne, Region Kronoberg and Landstinget Blekinge, Sweden, for help with data collection.

## 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.[Erratum appears in Lancet. 2013 Feb 23;381(9867):628 Note: AlMazroa, Mohammad A [added]; Memish, Ziad A [added]]. *Lancet (London, England)* 2012;380(9859):2163-96. doi: http://dx.doi.org/10.1016/S0140-6736(12)61729-2
2. Bergman S, Herrstrom P, Hogstrom K, et al. Chronic musculoskeletal pain, prevalence rates, and sociodemographic associations in a Swedish population study. *J Rheumatol* 2001;28(6):1369-77. [published Online First: 2001/06/21]
3. Gerdle B, Bjork J, Coster L, et al. Prevalence of widespread pain and associations with work status: a population study. *BMC musculoskeletal disorders* 2008;9:102. doi: 10.1186/1471-2474-9-102 [published Online First: 2008/07/17]
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(2):289-99. doi: 10.1016/j.ejpain.2011.07.006 [published Online First: 2012/02/11]
5. Alexanderson K, Norlund A. . Aim, background, key concepts, regulations, and current statistics. *Scandinavian journal of 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 musculoskeletal disorders* 2015;16:75. doi: 10.1186/s12891-015-0536-z [published Online First: 2015/04/19]
7. Jordan KP, Joud 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(1):212-8. doi: 10.1136/annrheumdis-2012-202634 [published Online First: 2013/01/25]
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(1):34-41.
9. Nolet PS, Cote 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 Journal: Official Journal of the North American Spine Society* 2015;15(4):675-84. doi: https://dx.doi.org/10.1016/j.spinee.2014.12.009
10. Nolet PS, Kristman VL, Cote P, et al. Is low back pain associated with worse health-related quality of life 6 months later? *European Spine Journal* 2015;24(3):458-66. doi: https://dx.doi.org/10.1007/s00586-014-3649-4
11. Pengel LH, Herbert RD, Maher CG, et al. Acute low back pain: systematic review of its prognosis. *BMJ (Clinical research ed)* 2003;327(7410):323.
12. Dunn KM, Hestbaek L, Cassidy JD. Low back pain across the life course. *Best practice & research Clinical rheumatology* 2013;27(5):591-600. doi: 10.1016/j.berh.2013.09.007 [published Online First: 2013/12/10]
13. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2003;12(2):149-65. doi: 10.1007/s00586-002-0508-5 [published Online First: 2003/04/24]
14. 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 (Phila Pa 1976)* 2004;29(21):2458-65. [published Online First: 2004/10/28]
15. Bernstein IA, Malik Q, Carville S, et al. Low back pain and sciatica: summary of NICE guidance. *BMJ (Clinical research ed)* 2017;356:i6748. doi: 10.1136/bmj.i6748 [published Online First: 2017/01/08]
16. Viniol A, Jegan N, Hirsch O, et al. Chronic low back pain patient groups in primary care--a cross sectional cluster analysis. *BMC musculoskeletal disorders* 2013;14:294. doi: 10.1186/1471-2474-14-294 [published Online First: 2013/10/18]
17. 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(5):632-41. doi: http://dx.doi.org/10.1002/art.23563
18. Hill JC, Whitehurst DGT, 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 (London, England)* 2011;378(9802):1560-71. doi: http://dx.doi.org/10.1016/S0140-6736(11)60937-9

19. 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. *The Journal of orthopaedic and sports physical therapy* 2014;44(9):656-64. doi: http://dx.doi.org/10.2519/jospt.2014.5178
20. Morso L, Kent P, Albert HB, et al. The predictive and external validity of the STarT Back Tool in Danish primary care. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2013;22(8):1859-67.
21. Bruyere 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 (Phila Pa 1976)* 2014;39(2):E123-8. doi: 10.1097/brs.0000000000000062 [published Online First: 2013/10/11]
22. Betten C, Sandell C, Hill JC, et al. Cross-cultural adaptation and validation of the Swedish STarT Back Screening Tool. *European Journal of Physiotherapy* 2015;17(1):29-36. doi: http://dx.doi.org/10.3109/21679169.2014.1003962
23. Bier JD, Ostelo RW, 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 doi: 10.1093/ptj/pzx023 [published Online First: 2017/03/25]
24. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33(5):337-43.
25. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208. [published Online First: 1990/11/05]
26. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37(1):53-72. [published Online First: 1996/06/06]
27. Greiner W, Weijnen T, Nieuwenhuizen M, et al. A single European currency for EQ-5D health states. *The European Journal of Health Economics, formerly: HEPAC* 2003;4(3):222-31.
28. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108. [published Online First: 1997/11/21]
29. Hansson E, Hansson T, Jonsson R. Predictors for work ability and disability in men and women with low-back or neck problems. *European Spine Journal* 2006;15(6):780-93.
30. Tuomi K, Ilmarinen J, Jahkola A, et al. Work ability index, Finnish Institute of Occupational Health. 2: Helsinki, 1998.
31. Ilmarinen J, Tuomi K, Seitsamo J. New dimensions of work ability. *International Congress Series* 2005;1280:3-7. doi: http://dx.doi.org/10.1016/j.ics.2005.02.060
32. Gould R, Ilmarinen J, Järvisalo J, et al. Dimensions of work ability. 2008
33. 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(5):404-12. [published Online First: 2010/04/08]
34. 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. doi: 10.1186/1471-2458-13-305 [published Online First: 2013/04/10]
35. Kirkwood B, Sterne J. Medical Statistics. 2nd: Oxford, England.: Blackwell Publishing, 2003.
36. Hosmer Jr D, Lemeshow S, Sturdivant RX. Introduction to the Logistic Regression Model. *Applied Logistic Regression* 2000
37. 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(3):439-46. doi: 10.1002/ejp.615 [published Online First: 2014/11/11]
38. Cole T. Applied logistic regression. DW Hosmer and S. Lemeshow, Wiley, New York, 1989. : Wiley Online Library 1991.
39. Morso L, Kongsted A, Hestbaek L, et al. The prognostic ability of the STarT Back Tool was affected by episode duration.[Erratum appears in Eur Spine J. 2016 Aug;25(8):2677-8; PMID: 27160824]. *European Spine Journal* 2016;25(3):936-44. doi: https://dx.doi.org/10.1007/s00586-015-3915-0
40. Voss M, Stark S, Alfredsson L, et al. Comparisons of self-reported and register data on sickness absence among public employees in Sweden. *Occup Environ Med* 2008;65(1):61-7. doi: 10.1136/oem.2006.031427 [published Online First: 2007/08/21]
41. 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(1):13. doi: 10.1186/s12916-016-0774-4 [published Online First: 2017/01/20]

42. Morso L, Kent P, Albert HB, et al. The predictive and external validity of the STarT Back Tool in Danish primary care. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2013;22(8):1859-67. doi: 10.1007/s00586-013-2690-z [published Online First: 2013/02/12]
43. 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 (Phila Pa 1976)* 2003;28(23):2551-60. doi: 10.1097/01.brs.0000096677.84605.a2 [published Online First: 2003/12/04]
44. Stigmar KG, Petersson IF, Joud 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 musculoskeletal disorders* 2013;14:57. doi: 10.1186/1471-2474-14-57 [published Online First: 2013/02/07]
45. Burstrom K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the EQ-5D. *Qual Life Res* 2001;10(7):621-35. [published Online First: 2002/02/02]
46. 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* 2010:404-12.
47. 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(4):350-56. doi: 10.1177/1403494817702759 [published Online First: 2017/04/08]
48. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ (Clinical research ed)* 2013;346:e5595. doi: 10.1136/bmj.e5595 [published Online First: 2013/02/07]
49. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ (Clinical research ed)* 2013;346:e5793. doi: 10.1136/bmj.e5793 [published Online First: 2013/02/07]
50. 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(2):102-11. doi: 10.1370/afm.1625 [published Online First: 2014/03/13]
51. Kuoppala J, Lamminpaa A. Rehabilitation and work ability: a systematic literature review. *J Rehabil Med* 2008;40(10):796-804. doi: 10.2340/16501977-0270 [published Online First: 2009/02/27]
52. 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(8):607-21. doi: 10.3109/09638280903186301 [published Online First: 2010/03/09]
53. Forsbrand M, Grahn B, Hill JC, et al. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. *BMC musculoskeletal disorders* 2017;18(1):89. doi: 10.1186/s12891-017-1449-9
54. 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(10):e012445. doi: 10.1136/bmjopen-2016-012445 [published Online First: 2016/10/16]
55. 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(1):83-9. doi: http://dx.doi.org/10.1016/j.ejpain.2009.01.003
56. 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(11):2182-91.
57. 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. *Journal of Pain* 2015;16(1):19-30. doi: http://dx.doi.org/10.1016/j.jpain.2014.10.004
58. 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. *Physical therapy* 2011;91(5):722-32.
59. Åsenlöf P, Michaelson P, Grahn B, et al. Acute neck and back pain: preventive interventions - Effects of physical training, manual treatment and cognitive behavioural interventions. : Swedish Agency for Health technology Assessment and Assessment of Social Services (SBU) 2016.

60. Foster NE, Hill JC, O'Sullivan P, et al. Stratified models of care. *Best practice & research Clinical rheumatology* 2013;27(5):649-61. doi: 10.1016/j.berh.2013.10.005 [published Online First: 2013/12/10]
61. Rozenberg S, Foltz V, Fautrel B. Treatment strategy for chronic low back pain. *Joint Bone Spine* 2012;79(6):555-59. doi: http://dx.doi.org/10.1016/j.jbspin.2012.09.003
62. Wahlin C, Ekberg K, Persson J, et al. Evaluation of self-reported work ability and usefulness of interventions among sick-listed patients. *Journal of occupational rehabilitation* 2013;23(1):32-43. doi: 10.1007/s10926-012-9376-y [published Online First: 2012/07/05]

# Figure legends

**Fig. 1 Flowchart of inclusion and exclusion of participants.** [1]Start Back Screening Tool. [2]EuroQol five-dimension. [3]Work Ability Score.

**Fig. 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.** AUC, area under the receiving operation curve; ROC, receiver operation characteristic; EQ-5D, Euroqol 5-dimension questionnaire. **Note:** The area under the ROC curve was 0.73.

**Fig. 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.** AUC, area under the receiving operation curve; ROC, receiver operation characteristic; WAS, work ability score. **Note:** The area under the ROC curve was 0.68.

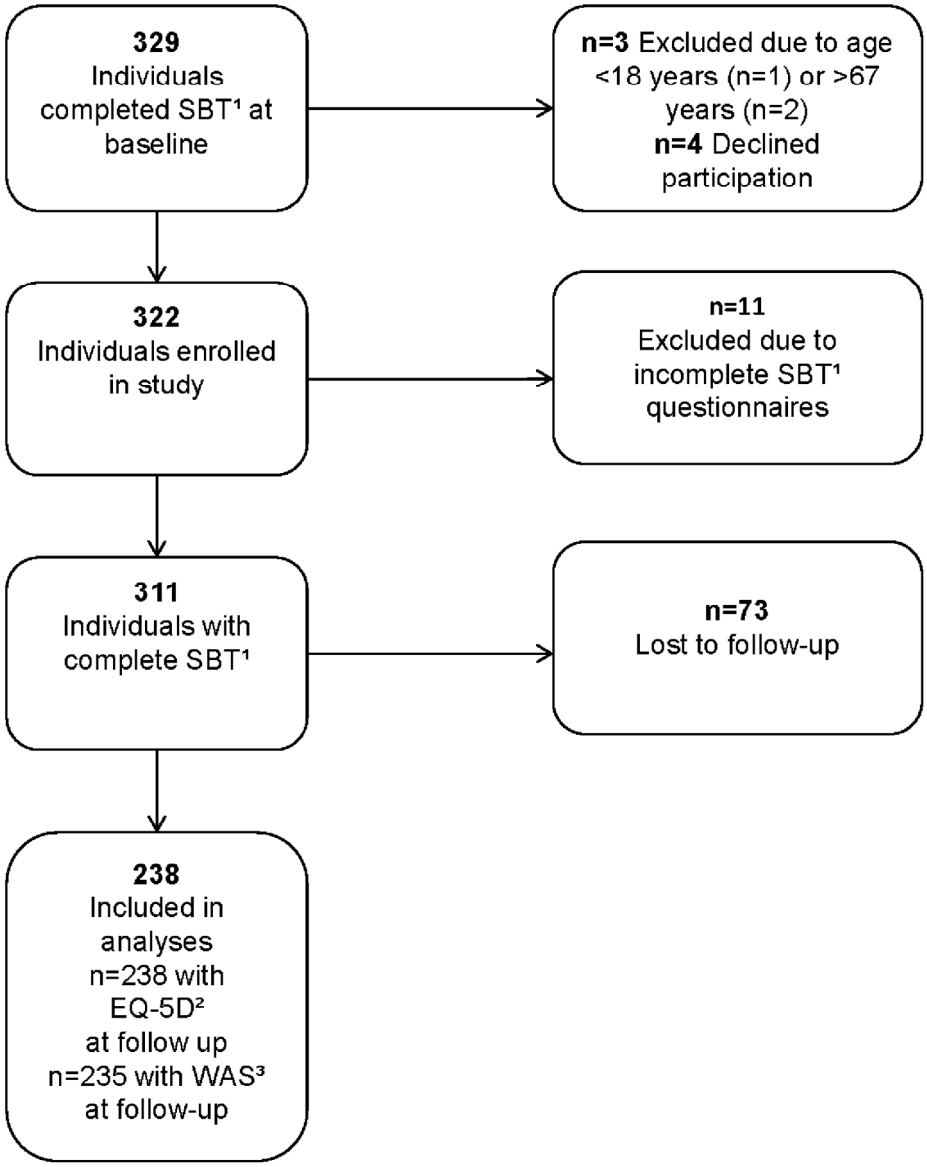


Fig.1. Flowchart of inclusion and exclusion of participants. ¹STarT Back Tool. ²EuroQol five-dimension. ³Work Ability Score.

151x181mm (300 x 300 DPI)

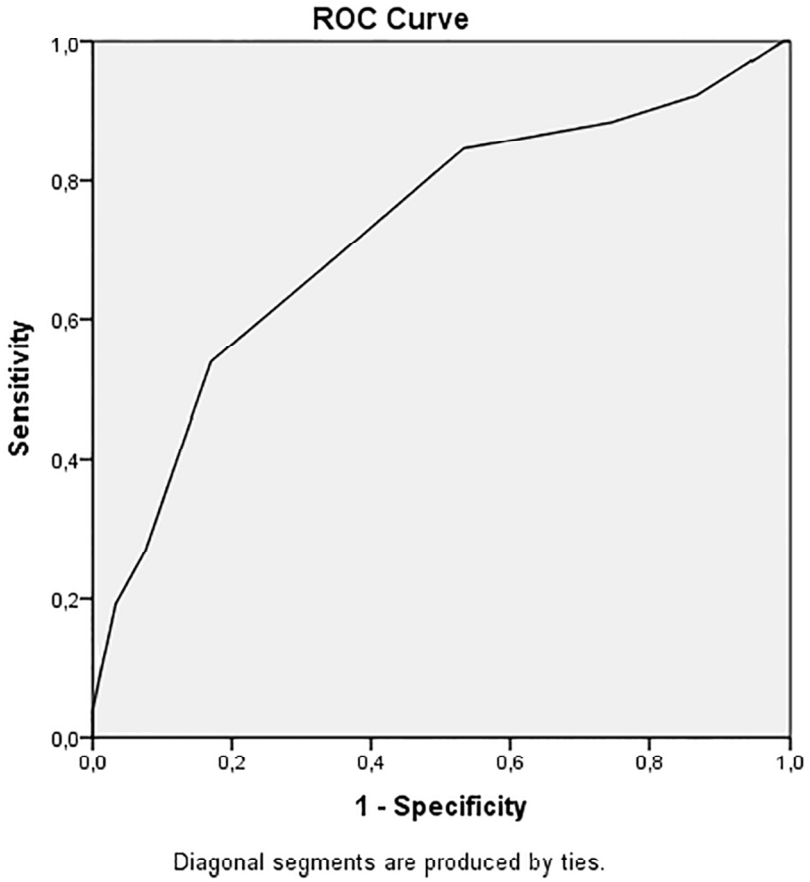


Fig. 2 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor health related quality of life (EQ-5D <0.6) in long-term follow up. Each point on the ROC curve has a corresponding cut-off value. AUC, area under the receiving operation curve; ROC, receiver operation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

111x113mm (300 x 300 DPI)

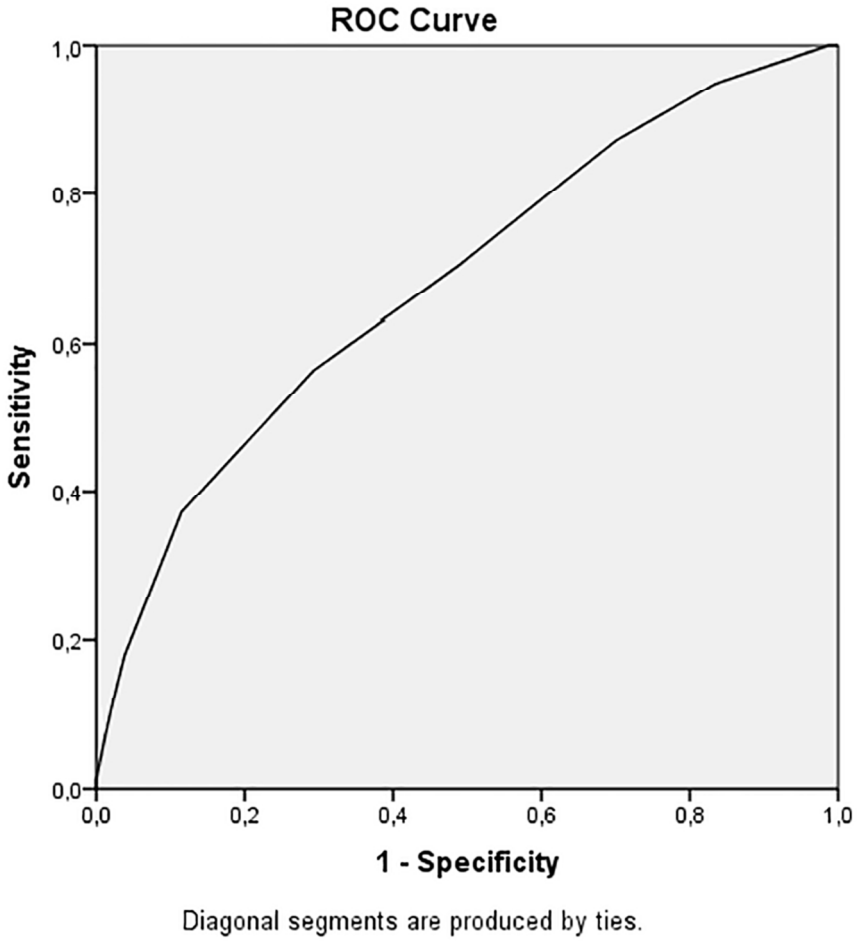


Fig. 3 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor work ability (WAS<8) in long-term follow up. Each point on the ROC curve has a corresponding cut-off value. AUC, area under the receiving operation curve; ROC, receiver operation characteristic;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

98x99mm (300 x 300 DPI)

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies***

| Section/Topic | Item # | Recommendation | Reported on page # |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 4 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 5 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 5-6 |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up | 5-6 |
| | | (*b*) For matched studies, give matching criteria and number of exposed and unexposed | Not relevant |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 5-8 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6-7 |
| Bias | 9 | Describe any efforts to address potential sources of bias | 6 |
| Study size | 10 | Explain how the study size was arrived at | 5 |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 8 |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 8-9 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | 8 |
| | | (*c*) Explain how missing data were addressed | 5,6 |
| | | (*d*) If applicable, explain how loss to follow-up was addressed | 6 |
| | | (*e*) Describe any sensitivity analyses | 9 |
| **Results** | | | |

| | | | |
|---|---|---|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | 5, 9 (flowchart) |
| | | (b) Give reasons for non-participation at each stage | 5,9 |
| | | (c) Consider use of a flow diagram | 9 (flowchart) |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 5,8 |
| | | (b) Indicate number of participants with missing data for each variable of interest | 5,9,11 (table 2) |
| | | (c) Summarise follow-up time (eg, average and total amount) | 6,10 |
| Outcome data | 15* | Report numbers of outcome events or summary measures over time | 5,9,11 (table 2) |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 11 |
| | | (*b*) Report category boundaries when continuous variables were categorized | 7 |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | Not relevant |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 12 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 13 |
| **Limitations** | | | |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 14,15 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 13,15,16 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 1 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

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



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2018-021748.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | 20-Apr-2018 |
| Complete List of Authors: | Forsbrand, Malin; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Landstinget Blekinge Blekinge kompetenscentrum,<br>Grahn, Birgitta; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg<br>Hill, Jonathan; Keele University, Research Institute of Primary Care and Health Sciences<br>Petersson, Ingemar; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Skåne University Hospital<br>Post Sennehed, Charlotte; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg<br>Stigmar, Kjerstin; Department of Health Sciences, Physiotherapy, Lund University; Skåne University Hospital |
| Primary Subject Heading: | Rehabilitation medicine |
| Secondary Subject Heading: | Patient-centred medicine |
| Keywords: | STarT Back Tool, health related quality of life, work ability, PRIMARY CARE, neck pain, Back pain < ORTHOPAEDIC & TRAUMA SURGERY |
| | |

# 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

M. Forsbrand[1,2*], B. Grahn[1,3], JC. Hill[4], IF. Petersson[1,6], C. Post Sennehed[1,3], K. Stigmar[5,6]

[1]Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, 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, United Kingdom, [5]Department of Health Sciences, Physiotherapy, Lund University, Lund, Sweden, [6]Skåne University Hospital, Lund, Sweden.

***Corresponding author:**

Malin Forsbrand, RPT, PhD student.

Address: Blekinge Centre of Competence, SE-371 81 Karlskrona, Sweden.

E-mail: malin.forsbrand@med.lu.se

Telephone: +46455735616

**E-mails authors (not corresponding):** birgitta.grahn@kronoberg.se, j.hill@keele.ac.uk, ingemar.petersson@skane.se, charlotte.sennehed@kronoberg.se, kjerstin.stigmar@med.lu.se

**Word count:** 4108 words

**Category:** Research article

**Funding sources:** 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).

**No conflicts of interest declared**

# 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 centers in south Sweden during 2013.

**Participants:** Patients (n=329) with acute/subacute back and/or neck pain, aged 18-67, 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 1.50-25.26) and poor work ability (OR 5.08, 95 % CI 1.75-14.71) vs the low risk group at follow-up. The AUC was 0.73 (CI 0.61-0.84) for HRQoL and 0.68 (CI 0.61-0.76) 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.

**Keywords:** STarT Back Tool, health related quality of life, work ability, primary care, back pain, neck pain.

# Strengths and limitations of this study

- This is the first study 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 back and/or neck pain.
- In this prospective study we have recruited patients from 35 different primary care centers, where many physiotherapists were engaged.
- The predictive validity of the SBT was examined in different ways.
- Limited baseline data was 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]. Whilst most individuals with acute back pain 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 health care 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 Nice 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 back pain, 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 health care costs in the United Kingdom [24]. The SBT is cross-culturally adapted and validated in Swedish[25] and recently also for a population with both back and neck pain in primary care [26]. The SBT is developed and validated to predict future disability due to low back pain 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 back and neck pain 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 back and/or neck pain.

# Methods

## Design

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

## Participants and procedure

Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care centers in the southern parts of Sweden, as part of an RCT [31]. Patients that all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 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 back and/or neck pain after a period of at least three months of no substantial pain. Patients that 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 that 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 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 STarT Back Tool (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 catastrophizing, depressive mood and overall impact from their BP. Items 1–8 have a dichotomous response option; “disagree” (0p) or “agree” (1p). 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*

Health-related quality of life (HRQoL) was measured by the EuroQol five-dimension (EQ-5D, 3L) questionnaire [34] which is a generic, health-related quality of life instrument [35 36]. The EQ-5D comprises the EQ descriptive system which has 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The digits for the 5 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. Health Related Quality of Life was also dichotomized 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 back and neck pain [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 dichotomized 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 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. Chi-squared test for trend was used to confirm potential differences concerning poor or good HRQoL and work ability.

Second, we calculated the odds ratios (95% confidence intervals) 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-<0.8 acceptable discrimination, 0.8-<0.9 excellent discrimination, and ≥0.9 outstanding discrimination [46].

In addition, the predictive validity of the SBT risk group cutoffs (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 cutoffs (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. Likelihood ratios above 5 or below 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 organizations 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

The study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2012/497, 2013/426, Dnr 2015/214). 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 flowchart (Figure 1).

**INSERT FIG 1 here**

The final sample consisted of 238/329 patients (72%) including 160 (67%) females and 78 (33%) males. Baseline characteristics of the study population are summarized 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 not RCT patients, the median time to follow-up was 12 months (range 11-19) and for RCT patients, the median time was 22 months (range 16-27).

**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[a] | | | | |
| BP[b], *n* (%) | 91 (38) | 42 (41) | 41 (38) | 8 (29) |
| NP + BP[c], *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) |

SBT, STarT Back Tool

[a]Area of pain Based on question number 2 (neck or shoulder pain) on SBT

[b]BP Back pain

[c]NP + BP Patients with neck or shoulder pain (NP) with or without back pain

## 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 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).

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

| Follow-up measure | Total population | SBT risk group | | | *p*-value |
|---|---|---|---|---|---|
| | | Low | Medium | High | |
| | *n*=238 | *n*=103 | *n*=107 | *n*=28 | |
| **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) | p<0.001 [d] |
| EQ-5D[a] <0.6, n (%) | 26 (11) | 4 (4) | 12 (11) | 10 (36) | p<0.001 [e] |
| **Work ability**[b]; median (range) | 8 (0-10) | 9 (0-10) | 8 (1-10) | 7 (0-10) | p<0.001 [d] |
| WAS[c] <8, n (%) | 78 (33) | 23 (22) | 38 (35) | 17 (68) | p<0.001 [e] |

SBT, STarT Back Tool; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score

[a]EQ-5D scores, range -0.59-1

[b]3 missing from the high risk group (total population: n=235 and n=25 for the high risk group)

[c]Where 0 equates to "completely unable to work" and 10 equates to "work ability at its best"

[d]Kruskal-Wallis test, [e]Chi square test for trend

The regression analysis showed that the SBT high risk group could significantly predict poor HRQoL (OR 6.16, CI 1.50-25.26, *B=1.82, p*=0.012) and poor work ability (OR 5.08, CI 1.75-14.71, *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 Hosmer and Lemeshow´s test indicates that the model is good [48] (Table 3).

**Table 3**. The ability of the SBT risk groups to predict poor health related quality of life[a] and poor work ability[b] at long-term follow-up.

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

SBT, StarT Back Tool; HRQoL, Health related quality of life; RCT, clinical trial

[a]Poor HRQoL measured by EuroQol five-dimension questionnaire (EQ-5D) <0.6

[b]Poor work ability measured by Work ability score (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.

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 area under the curve (AUC) was 0.73 (CI 0.61-0.84) which

was 'acceptable' (≥0.7) (Fig. 2). For work ability, the area under the curve (AUC) was 0.68 (CI 0.61-0.76) which was just below the limit (≥7) for acceptable discrimination (Fig. 3).

**INSERT FIG 2 and FIG 3 here**

The sensitivity, specificity, PPV, NPV and likelihood ratios for the SBT risk groups for HRQoL and work ability are presented in Table 4. The LR+s were higher and the LR-s were lower for HRQoL outcomes compared to work ability outcomes which indicate better discrimination of the SBT for poor HRQoL compared to poor work ability (Table 3).

**Table 4**. Discriminative ability of the SBT risk group cutoffs (low/medium and medium/high) 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 (1.29-1.95) | 0.33 (0.13-0.82) |
| L/M vs. H | 38.5 | 91.5 | 35.7 | 92.4 | 4.53 (2.35-8.74) | 0.67 (0.49-0.91) |
| **Work ability (WAS <8)** | | | | | | |
| L vs. M/H | 70.5 | 51.0 | 41.7 | 77.7 | 1.44 (1.16-1.78) | 0.58 (0.40-0.84) |
| L/M vs. H | 21.8 | 94.9 | 68.0 | 71.0 | 4.28 (1.93-9.47) | 0.82 (0.73-0.93) |

SBT, STarT Back Tool; HRQoL, Health related quality of life; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

# 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 back and/or neck pain. 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 back and/or neck pain 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 to 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 back pain and including neck pain. As might be expected, the distribution between the SBT risk groups at baseline differed compared to the UK development population [23]. In our study population, the percentage of individuals at high risk were lower (12%) compared to 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 characterized group of individuals from 35 different primary care centers. The SBT was used and administered by many different physiotherapists which makes this setting real and clinically relevant. Another strength is that we analyzed 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 analyzed the outcomes HRQoL and work ability both on the continuous scale (Kruskal-Wallis) and as dichotomized (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 to 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 analyzes 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 don´t 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 back pain 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 neck pain and back pain. 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 to 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 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 back pain 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 dichotomization 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 dichotomized (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 cutoff 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 back pain [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 back and neck pain [39]. Another cut-off has been used in a study of patients with musculoskeletal pain taking part in a national rehabilitation program 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 categorization as for the whole WAI [42]. The WAS has shown to be a good alternative to the whole WAI [55] even though the whole WAI is superior compared to its individual items [56].

SBTs concurrent validity has earlier been studied for patients with back and/or neck pain [26] and a modified SBT have been tested to predict physical health outcome, using the SF-36 [57] 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 back and neck pain. Therefore this study widens the usefulness of the SBT compared to earlier studies [23 58-61]. 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 [62] 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 neck and back pain. 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 back and/or neck pain 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 personalized treatment strategies which are a priority in research[63]. 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.

# Authors Contributions

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.

## Acknowledgements

Senior statistician Mikael Åström for statistical support, datamanager 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. All physiotherapists and patients involved in this project, including the WorkUp project, REHSAM, in Region Skåne, Region Kronoberg and Landstinget Blekinge, Sweden, for help with data collection.

## 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.[Erratum appears in Lancet. 2013 Feb 23;381(9867):628 Note: AlMazroa, Mohammad A [added]; Memish, Ziad A [added]]. *Lancet (London, England)* 2012;380(9859):2163-96. doi: http://dx.doi.org/10.1016/S0140-6736(12)61729-2
2. Bergman S, Herrstrom P, Hogstrom K, et al. Chronic musculoskeletal pain, prevalence rates, and sociodemographic associations in a Swedish population study. *J Rheumatol* 2001;28(6):1369-77. [published Online First: 2001/06/21]
3. Gerdle B, Bjork J, Coster L, et al. Prevalence of widespread pain and associations with work status: a population study. *BMC musculoskeletal disorders* 2008;9:102. doi: 10.1186/1471-2474-9-102 [published Online First: 2008/07/17]
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(2):289-99. doi: 10.1016/j.ejpain.2011.07.006 [published Online First: 2012/02/11]
5. Alexanderson K, Norlund A. . Aim, background, key concepts, regulations, and current statistics. *Scandinavian journal of 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 musculoskeletal disorders* 2015;16:75. doi: 10.1186/s12891-015-0536-z [published Online First: 2015/04/19]
7. Jordan KP, Joud 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(1):212-8. doi: 10.1136/annrheumdis-2012-202634 [published Online First: 2013/01/25]
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(1):34-41.
9. Nolet PS, Cote 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 Journal: Official Journal of the North American Spine Society* 2015;15(4):675-84. doi: https://dx.doi.org/10.1016/j.spinee.2014.12.009
10. Nolet PS, Kristman VL, Cote P, et al. Is low back pain associated with worse health-related quality of life 6 months later? *European Spine Journal* 2015;24(3):458-66. doi: https://dx.doi.org/10.1007/s00586-014-3649-4
11. Leijon O, Wahlstrom 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(17):1863-8. doi: http://dx.doi.org/10.1097/BRS.0b013e3181ab3397
12. Nyman T, Grooten WJ, Wiktorin C, et al. Sickness absence and concurrent low back and neck-shoulder pain: results from the MUSIC-Norrtalje study.[Erratum appears in Eur Spine J. 2007 May;16(5):639-40]. *European Spine Journal* 2007;16(5):631-8.
13. Pengel LH, Herbert RD, Maher CG, et al. Acute low back pain: systematic review of its prognosis. *BMJ (Clinical research ed)* 2003;327(7410):323.
14. Dunn KM, Hestbaek L, Cassidy JD. Low back pain across the life course. *Best practice & research Clinical rheumatology* 2013;27(5):591-600. doi: 10.1016/j.berh.2013.09.007 [published Online First: 2013/12/10]
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. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2003;12(2):149-65. doi: 10.1007/s00586-002-0508-5 [published Online First: 2003/04/24]

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 (Phila Pa 1976)* 2004;29(21):2458-65. [published Online First: 2004/10/28]
17. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ (Clinical research ed)* 2013;346:e5595. doi: 10.1136/bmj.e5595 [published Online First: 2013/02/07]
18. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ (Clinical research ed)* 2013;346:e5793. doi: 10.1136/bmj.e5793 [published Online First: 2013/02/07]
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(2):102-11. doi: 10.1370/afm.1625 [published Online First: 2014/03/13]
20. Kuoppala J, Lamminpaa A. Rehabilitation and work ability: a systematic literature review. *J Rehabil Med* 2008;40(10):796-804. doi: 10.2340/16501977-0270 [published Online First: 2009/02/27]
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(8):607-21. doi: 10.3109/09638280903186301 [published Online First: 2010/03/09]
22. Bernstein IA, Malik Q, Carville S, et al. Low back pain and sciatica: summary of NICE guidance. *BMJ (Clinical research ed)* 2017;356:i6748. doi: 10.1136/bmj.i6748 [published Online First: 2017/01/08]
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(5):632-41. doi: http://dx.doi.org/10.1002/art.23563
24. Hill JC, Whitehurst DGT, 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 (London, England)* 2011;378(9802):1560-71. doi: http://dx.doi.org/10.1016/S0140-6736(11)60937-9
25. Betten C, Sandell C, Hill JC, et al. Cross-cultural adaptation and validation of the Swedish STarT Back Screening Tool. *European Journal of Physiotherapy* 2015;17(1):29-36. doi: http://dx.doi.org/10.3109/21679169.2014.1003962
26. Forsbrand M, Grahn B, Hill JC, et al. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. *BMC musculoskeletal disorders* 2017;18(1):89. doi: 10.1186/s12891-017-1449-9
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. *The Journal of orthopaedic and sports physical therapy* 2014;44(9):656-64. doi: http://dx.doi.org/10.2519/jospt.2014.5178
28. Morso L, Kent P, Albert HB, et al. The predictive and external validity of the STarT Back Tool in Danish primary care. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2013;22(8):1859-67. doi: 10.1007/s00586-013-2690-z [published Online First: 2013/02/12]
29. Bruyere 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 (Phila Pa 1976)* 2014;39(2):E123-8. doi: 10.1097/brs.0000000000000062 [published Online First: 2013/10/11]
30. Bier JD, Ostelo RW, 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 doi: 10.1093/ptj/pzx023 [published Online First: 2017/03/25]
31. Sennehed CP, Holmberg S, Axen I, et al. Early Workplace Dialogue in Physiotherapy Practice Improved Work Ability at One-Year Follow-Up - Workup a Randomised Controlled Trial in Primary Care. *Pain* 2018 doi: 10.1097/j.pain.0000000000001216 [published Online First: 2018/03/20]
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. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2006;15 Suppl 2:S169-91. doi: 10.1007/s00586-006-1071-2 [published Online First: 2006/03/22]

33. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. *Spine* 2011;36(22):1891-5. doi: http://dx.doi.org/10.1097/BRS.0b013e3181f8f775
34. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33(5):337-43.
35. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208. [published Online First: 1990/11/05]
36. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37(1):53-72. [published Online First: 1996/06/06]
37. Greiner W, Weijnen T, Nieuwenhuizen M, et al. A single European currency for EQ-5D health states. *The European Journal of Health Economics, formerly: HEPAC* 2003;4(3):222-31.
38. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108. [published Online First: 1997/11/21]
39. Hansson E, Hansson T, Jonsson R. Predictors for work ability and disability in men and women with low-back or neck problems. *European Spine Journal* 2006;15(6):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. *International Congress Series* 2005;1280:3-7. doi: http://dx.doi.org/10.1016/j.ics.2005.02.060
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(5):404-12. [published Online First: 2010/04/08]
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. doi: 10.1186/1471-2458-13-305 [published Online First: 2013/04/10]
45. Kirkwood B, Sterne J. Medical Statistics. 2nd: Oxford, England.: 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(3):439-46. doi: 10.1002/ejp.615 [published Online First: 2014/11/11]
48. Cole T. Applied logistic regression. DW Hosmer and S. Lemeshow, Wiley, New York, 1989. : Wiley Online Library 1991.
49. Melloh M, Elfering A, Kaser A, et al. What is the best time point to identify patients at risk of developing persistent low back pain? *Journal of back and musculoskeletal rehabilitation* 2015;28(2):267-76. doi: 10.3233/bmr-140514 [published Online First: 2014/08/07]
50. Morso L, Kongsted A, Hestbaek L, et al. The prognostic ability of the STarT Back Tool was affected by episode duration.[Erratum appears in Eur Spine J. 2016 Aug;25(8):2677-8; PMID: 27160824]. *European Spine Journal* 2016;25(3):936-44. doi: https://dx.doi.org/10.1007/s00586-015-3915-0
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(1):13. doi: 10.1186/s12916-016-0774-4 [published Online First: 2017/01/20]
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 (Phila Pa 1976)* 2003;28(23):2551-60. doi: 10.1097/01.brs.0000096677.84605.a2 [published Online First: 2003/12/04]
53. Stigmar KG, Petersson IF, Joud 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 musculoskeletal disorders* 2013;14:57. doi: 10.1186/1471-2474-14-57 [published Online First: 2013/02/07]
54. Burstrom K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the EQ-5D. *Qual Life Res* 2001;10(7):621-35. [published Online First: 2002/02/02]

55. 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* 2010:404-12.
56. 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(4):350-56. doi: 10.1177/1403494817702759 [published Online First: 2017/04/08]
57. 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(10):e012445. doi: 10.1136/bmjopen-2016-012445 [published Online First: 2016/10/16]
58. 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(1):83-9. doi: http://dx.doi.org/10.1016/j.ejpain.2009.01.003
59. 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(11):2182-91.
60. 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. *Journal of Pain* 2015;16(1):19-30. doi: http://dx.doi.org/10.1016/j.jpain.2014.10.004
61. 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. *Physical therapy* 2011;91(5):722-32.
62. Foster NE, Hill JC, O'Sullivan P, et al. Stratified models of care. *Best practice & research Clinical rheumatology* 2013;27(5):649-61. doi: 10.1016/j.berh.2013.10.005 [published Online First: 2013/12/10]
63. Rozenberg S, Foltz V, Fautrel B. Treatment strategy for chronic low back pain. *Joint Bone Spine* 2012;79(6):555-59. doi: http://dx.doi.org/10.1016/j.jbspin.2012.09.003

# Figure legends

**Fig. 1 Flowchart of inclusion and exclusion of participants.** [1]STarT Back Tool. [2]EuroQol five-dimension. [3]Work Ability Score.

**Fig. 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.** AUC, area under the receiving operation curve; ROC, receiver operation characteristic; EQ-5D, Euroqol 5-dimension questionnaire. **Note:** The area under the ROC curve was 0.73.

**Fig. 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.** AUC, area under the receiving operation curve; ROC, receiver operation characteristic; WAS, work ability score. **Note:** The area under the ROC curve was 0.68.

625

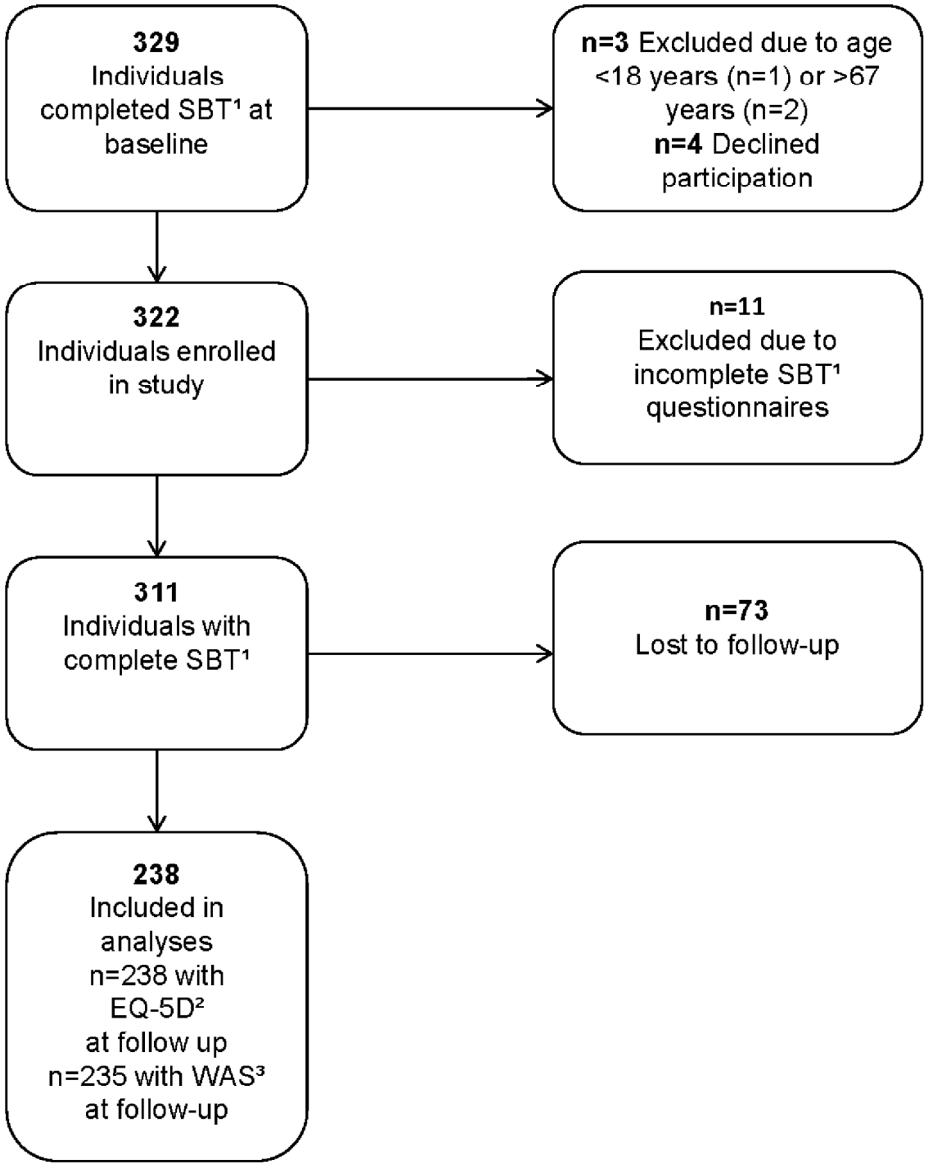


Fig.1. Flowchart of inclusion and exclusion of participants. ¹STarT Back Tool. ²EuroQol five-dimension. ³Work Ability Score.

151x181mm (300 x 300 DPI)

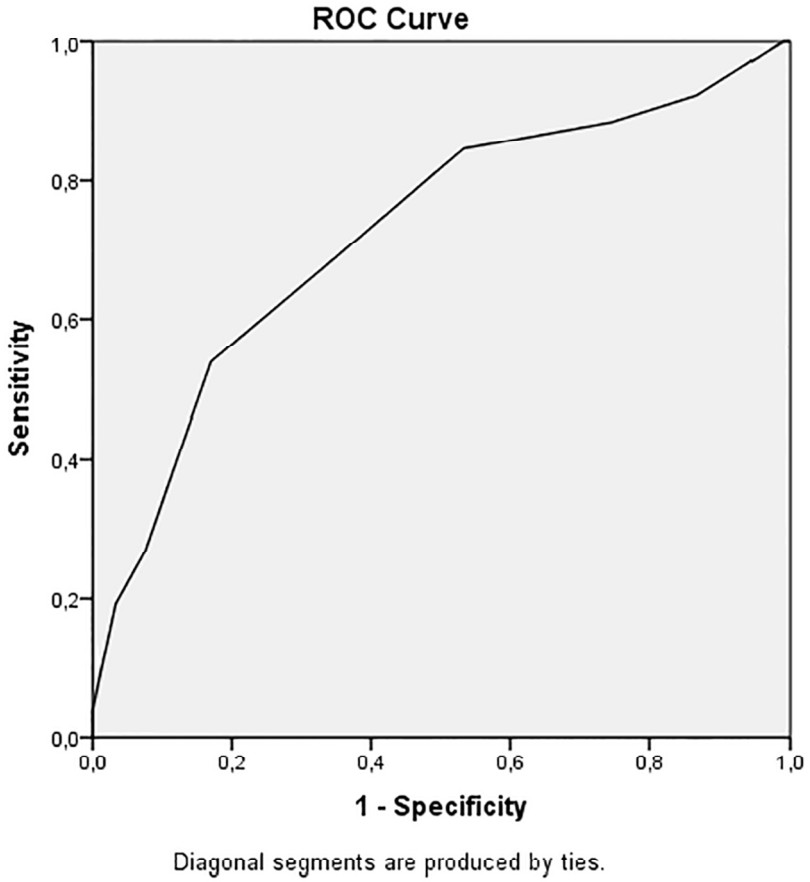


Fig. 2 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor health related quality of life (EQ-5D <0.6) in long-term follow up. Each point on the ROC curve has a corresponding cut-off value. AUC, area under the receiving operation curve; ROC, receiver operation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

111x113mm (300 x 300 DPI)

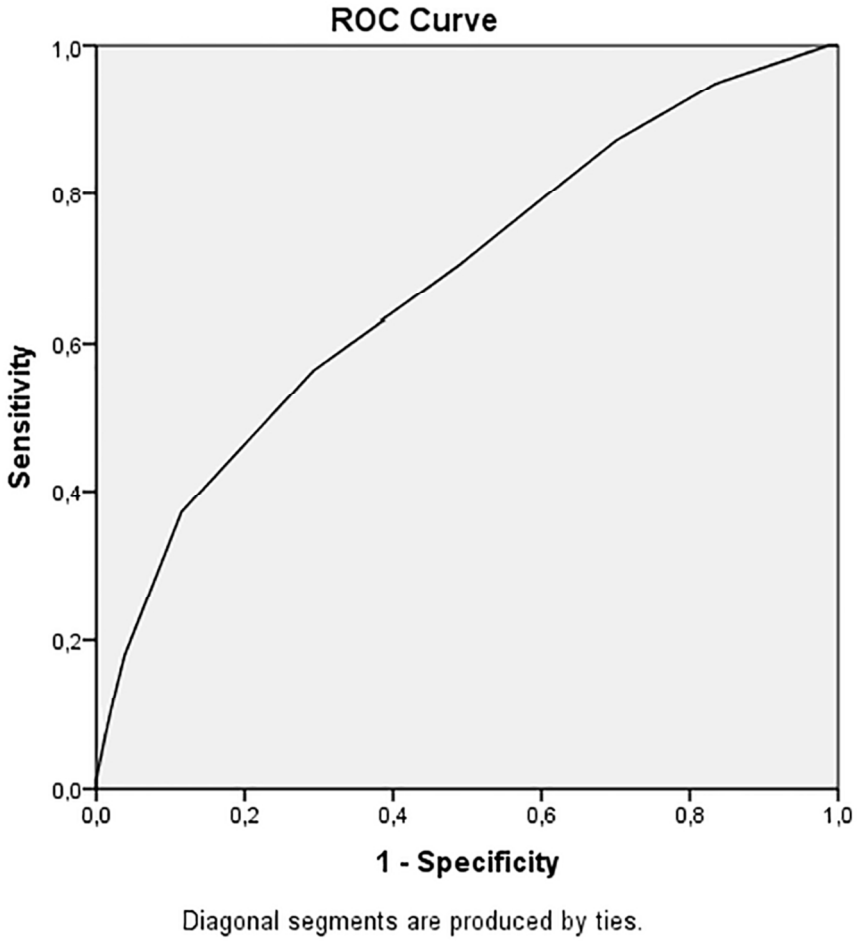


Fig. 3 AUC and ROC curve for overall STarT Back Tool scores to discriminate between individuals with poor work ability (WAS<8) in long-term follow up. Each point on the ROC curve has a corresponding cut-off value. AUC, area under the receiving operation curve; ROC, receiver operation characteristic;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

98x99mm (300 x 300 DPI)