

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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, Grahn, Birgitta; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg Hill, Jonathan; Keele University, Research Institute of Primary Care and Health Sciences Petersson, Ingemar; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Skåne University Hospital Post Sennehed, Charlotte; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg 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

SCHOLARONE™
Manuscripts

1
2
3 1 Can the STarT Back Tool predict health related quality of life and work
4
5
6 2 ability after an acute/subacute episode with back or neck pain? – a
7
8
9 3 prospective cohort study in primary care

10
11
12 4 M. Forsbrand^{1,2*}, B. Grahn^{1,3}, JC. Hill⁴, IF. Petersson^{1,6}, C. Post Sennehed^{1,3}, K. Stigmar^{5,6}

13
14
15 5 ¹Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund, Sweden,

16
17 6 ²Blekinge Centre of Competence, Landstinget Blekinge, Karlskrona, Sweden, ³Department of Research and

18
19 7 Development, Region Kronoberg, Växjö, Sweden, ⁴Research Institute of Primary Care and Health Sciences,

20
21 8 Keele University, Stoke-on-Trent, United Kingdom, ⁵Department of Health Sciences, Physiotherapy, Lund

22
23 9 University, Lund, Sweden, ⁶Skåne University Hospital, Lund, Sweden.

24
25
26 10
27
28 11 ***Corresponding author:**

29
30 12 Malin Forsbrand, RPT, PhD student.

31
32 13 Address: Blekinge Centre of Competence, SE-371 81 Karlskrona, Sweden.

33
34 14 E-mail: malin.forsbrand@med.lu.se

35
36 15 Telephone: +46455735616

37
38 16 **E-mails authors (not corresponding):** birgitta.grahn@kronoberg.se, j.hill@keele.ac.uk,

39
40 17 ingemar.petersson@skane.se, charlotte.sennehed@kronoberg.se, kjerstin.stigmar@med.lu.se

41
42 18 **Word count:** 3997 words

43
44 19 **Category:** Research article

45
46 20 **Funding sources:** This work was supported by The Scientific Committee of Blekinge County Council,

47
48 21 Sweden (Dnr 2017-00686) and the County Councils Region Skåne, Kronoberg and Blekinge, REHSAM

49
50 22 research, Sweden (RS2011/005).

51
52 23 **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.

1
2 48 **Keywords:** STarT Back Tool, health related quality of life, work ability, primary care, back pain, neck pain.
3
4 49

5
6
7 50 **Strengths and limitations of this study**
8

- 9
10 51 • This is the first study to evaluate the predictive validity of SBT of the outcomes HRQoL and work
11 ability at long-term follow-up in a population with acute/subacute back and/or neck pain.
12 52
13
14 53 • In this prospective study we have recruited patients from 35 different primary care centers, where
15 many physiotherapists were engaged.
16 54
17
18 55 • The predictive validity of the SBT was examined in different ways.
19
20
21 56 • Limited baseline information was available for one part of the cohort.
22
23 57 • Limitations of the study were the broad variation in time to follow-up.
24
25 58
26
27
28 59
29
30
31 60
32
33 61
34
35 62
36
37
38 63
39
40
41 64
42
43 65
44
45
46 66
47
48
49 67
50
51 68
52
53
54 69
55
56 70
57
58
59
60

71 Introduction

72 Musculoskeletal pain, especially back pain (BP) and neck pain (NP) are highly prevalent in the general
73 population^{1 2} causing disability for the individual and high costs for society³⁻⁵. Individuals with BP and NP
74 are mostly managed in primary care^{6 7} and patients presenting with these conditions are at risk of sickness
75 absence⁸ and poor health related quality of life (HRQoL)^{9 10}. Whilst most individuals with acute back pain
76 improve quickly and return to work¹¹, for some of them the pain is more severe and lasts for a longer period
77^{12 13}. In a Swedish cohort of individuals with BP and NP about half of the population reported pain and
78 disability 5 years after onset¹⁴. Evidence-based guidelines¹⁵ therefore, recommend that clinicians assess
79 patient prognosis using brief questionnaires to identify individuals at risk of poor outcomes in order to
80 achieve effective treatment allocation¹⁶ and to direct the limited healthcare resources available to those most
81 in need.

82 The widely used STarT Back Tool (SBT)¹⁷, is a brief risk stratification tool that includes nine questions on
83 predictors for long-term disabling back pain, in order to match individuals to appropriate targeted
84 treatments, according to their prognostic profile. Using the SBT together with targeted treatment pathways
85 has shown improved efficiency regarding patients' clinical outcomes and reduced health care costs in the
86 United Kingdom¹⁸. The SBT is developed and validated to predict future disability due to low back pain of
87 any duration^{17 19-23}, but it has not yet been studied for the outcomes of HRQoL and work ability for a
88 population with acute/subacute back and neck pain in primary care. The aim of this study was therefore to
89 evaluate the predictive validity of SBT of the outcomes HRQoL and work ability at long-term follow-up in a
90 population with acute/subacute back and/or neck pain. We separately evaluated the SBT specific risk groups
91 and also the SBT overall score.

92

93

94

95 **Methods**

96 **Design**

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

100 **Participants and procedure**

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

109 All patients were followed up with self-reported questionnaires including items on work ability and HRQoL.
110 Patients with any missing item on the SBT (n=11) and those who were lost to follow-up (n=73) were
111 excluded. The final study cohort included 238 participants. The analyses were restricted to those who had
112 complete data for work ability (n=235) and HRQoL (n=238) outcomes at long-term follow-up. The study
113 cohort consisted of patients that had been included in the RCT (RCT intervention group, n=61 and RCT
114 control group, n=99) and patients that had not been included in the RCT (n=78). The reason we included
115 patients who had been excluded from the RCT was to ensure we had as broad a sample as possible for this
116 SBT predictive validity study. RCT patients (n=160) received either structured physiotherapy treatment
117 (including examination, assessment, diagnosis, evidence-based treatment and follow-up) with a workplace
118 intervention (RCT intervention group) or structured physiotherapy without a workplace intervention (RCT

1
2 119 control group) (ClinicalTrials.gov ID: NCT02609750) and were followed up at 12-months (median 12
3
4 120 months, range 11-19). Excluded RCT patients received usual primary care and were followed up around 18-
5
6 121 24 months (median 22 months, range 16-27). Data from all questionnaires were manually entered into a
7
8 122 SPSS 22.0 database and were thoroughly checked and validated. All questionnaires were scored, and
9
10 123 missing items handled, according to the methods specified by the instrument developers.
11
12
13
14 124

15 125 **Baseline data**

16 125
17
18
19 126 Baseline questionnaire data included type of treatment received (RCT intervention group, RCT control
20
21 127 group or usual primary care) and self-reports of SBT, age and gender.
22
23
24 128

25 128 26 27 129 **STarT Back Tool**

28
29 130 The STarT Back Tool (SBT) is a 9-item questionnaire with questions relating to modifiable physical (item
30
31 131 1–4) and psychosocial (item 5–9) risk factors for long-term disabling BP, designed to support clinicians in
32
33 132 directing individuals to different levels of care¹⁷. The SBT has three risk subgroups which classifies patients
34
35 133 into low, medium or high risk for poor disability outcomes. The SBT overall score ranges between 0 and 9.
36
37 134 Item 1–4 is about referred leg pain, neck or shoulder pain, difficulties in walking and difficulties in dressing.
38
39 135 Item 5–9 form the psychosocial subscale which screen for fear of physical activity, anxiety, pain
40
41 136 catastrophizing, depressive mood and overall impact from their BP. Items 1–8 have a dichotomous response
42
43 137 option; “disagree” (0p) or “agree” (1p). Item 9 uses a 5-point Likert Scale from “not at all” to “extremely”,
44
45 138 where responses “very much” or “extremely” are counted as one point and the other responses as zero. A
46
47 139 total score of ≤ 3 points indicates low risk, a total score ≥ 4 points in combination with < 4 points on the
48
49 140 psychosocial subscale (item 5–9) are medium risk and a psychosocial subscale score of ≥ 4 points indicates
50
51 141 high risk for poor disability outcomes¹⁷.
52
53
54
55 142

143 **Long-term follow-up data**

145 *Health related quality of life*

146 Health-related quality of life (HRQoL) was measured by the EuroQol five-dimension (EQ-5D, 3L)
147 questionnaire²⁴ which is a generic, health-related quality of life instrument^{25 26}. The EQ-5D comprises the
148 EQ descriptive system which has 5 dimensions: mobility, self-care, usual activities, pain/discomfort and
149 anxiety/depression. The digits for the 5 dimensions are combined in a 5-digit number describing the
150 respondent's health state²⁷. The 5-digit number is given a value between -0.59 and 1.0 according to the UK
151 tariff²⁸, where 1 corresponds to full health and lower EQ-5D values reflect lower HRQoL. Health Related
152 Quality of Life was dichotomized into "poor" HRQoL (EQ-5D <0.6) and "good" HRQoL (EQ-5D ≥0.6),
153 based on a proposed cut-off for having sufficient capacity to be able to work for a population with back and
154 neck pain²⁹.

156 *Work ability*

157 Work ability was measured by self-reports on the single item question ("current work ability compared with
158 the lifetime best") from the Work Ability Index (WAI)^{30 31}. This first item in the WAI is known as the
159 "Work Ability score" (WAS)³². It consists of a scale from 0 representing "cannot work at all right now" to
160 10 representing "my work ability as at its best right now" and has been proposed to be used as a simple
161 indicator for assessing the status and progress of work ability^{33 34}. Work ability was dichotomized using a
162 previously published cut-off score³³ into "poor" work ability (WAS <8 points) and "good" work ability
163 (WAS ≥8 points).

167 **Statistical analyses**

168 SPSS 22.0 was used for all analyses. We used a non-parametric approach which was chosen based on the
169 distribution of the data. Descriptive data on the study population was presented for the total cohort and for
170 each SBT risk group.

172 **Predictive performance of the SBT**

173 First, cross tabulations were used to describe the proportion of participants in each SBT risk group that had
174 poor outcome in long-term follow-up for each outcome. The Kruskal Wallis test was used to study if there
175 were any differences between the SBT risk groups on follow-up data on poor or good HRQoL and work
176 ability, respectively. Mann Whitney U-test and Chi-squared test for trend was used to confirm potential
177 differences.

178 Second, we calculated the odds ratios (95% confidence intervals) for poor outcome on HRQoL (EQ-5D<0.6)
179 and work ability (WAS<8) for SBT risk groups using binary logistic regression. Independent variables age,
180 sex, treatment group or time to follow-up (months) were also included in the analysis. We built a multiple
181 logistic model where all independent variables were entered together with the SBT risk groups. For SBT, we
182 used the SBT low risk group as the reference group and for treatment groups (RCT intervention group n=61,
183 RCT control group n=99, Not RCT group n=78), we used the “Not RCT group” as the reference group. The
184 significance level was set at 5%.

185 Third, we evaluated the ability of the SBT overall scores (0-9 points) to discriminate between individuals
186 with poor or good HRQoL/work ability in long-term follow-up. For that purpose, we used the area under the
187 curve (AUC) statistics from receiver operating characteristic (ROC) curves³⁵. The strength of discrimination
188 was set according to the following descriptors: 0.7-<0.8 acceptable discrimination, 0.8-<0.9 excellent
189 discrimination, and ≥ 0.9 outstanding discrimination³⁶.

1
2 190 In addition, the predictive validity of the SBT risk group cutoffs (low/medium and medium/high) was
3
4 191 assessed by calculating sensitivity, specificity, positive predictive values (PPV), negative predictive values
5
6 192 (NPV) and positive and negative likelihood ratios (LRs) against long-term HRQoL and work ability
7
8 193 outcomes. The SBT risk group cutoffs (low/medium and medium/high) were used in line with the original
9
10 194 study¹⁷. The PPV is the probability that a poor outcome is present when the test is positive and the NPV is
11
12 195 the probability that a good outcome is present when the test is negative. Higher positive LR and lower
13
14 196 negative LR indicate better discrimination. Likelihood ratios above 5 or below 0.2 are generally seen as
15
16 197 supporting a strong test, whereas values close to 1 indicate poor test performance³⁷.
17
18
19
20
21 198

199 **Ethics**

200 The study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2012/497, 2013/426,
201 Dnr 2015/214). Prior to inclusion, all patients obtained written information about the purpose of the study
202 and each individual gave informed consent to participate in the study (opt-out). The principles of the
203 Declarations of Helsinki were followed.
204
205

206 **Results**

207 **Study population**

208 The inclusion and exclusion of participants in the study is presented in a flowchart (Figure 1).
209
210

211 **INSERT FIG 1 here**

1
2 210 The final sample consisted of 238/329 patients (72%) including 160 (67%) females and 78 (33%) males.
3
4 211 Baseline characteristics of the study population are summarized in Table 1. The patient sample included 103
5
6 212 (43%) patients at low risk, 107 (45%) patients at medium risk, and 28 (12%) patients at high risk. The
7
8 213 median time to long-term follow-up was 13 (range 11-27) months.
9
10
11
12 214

13
14 215 **Table 1.** Baseline characteristics of the study population – total cohort and stratified by SBT risk groups.

Variable	Total population <i>n</i> =238	SBT risk group		
		Low <i>n</i> =103 (43%)	Medium <i>n</i> =107 (45%)	High <i>n</i> =28 (12%)
Age, median (range)	46 (19-67)	45 (22-64)	47 (21-67)	38 (19-63)
Sex, <i>n</i> (%) female	160 (67)	73 (71)	72 (67)	15 (54)
Area of pain ^a				
BP ^b , <i>n</i> (%)	91 (38)	42 (41)	41 (38)	8 (29)
NP + BP ^c , <i>n</i> (%)	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)

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 216 SBT, STarT Back Tool

33 217 ^aArea of pain Based on question number 2 (neck or shoulder pain) on SBT

34 218 ^bBP Back pain

35 219 ^cNP + BP Patients with neck or shoulder pain (NP) with or without back pain
36
37
38
39
40

41 220 **Predictive performance of the SBT**

42
43
44 221 There were statistically significant differences in the distribution of HRQoL scores (*n*=238) between the
45
46 222 SBT low, medium and high risk groups at long-term follow-up (*p*=0.000) and the proportion of patients with
47
48 223 poor HRQoL (EQ-5D<0.6) was significantly higher in higher risk groups (low risk 4%, medium risk 11%,
49
50 224 high risk 36%) (*p*=0.000) (Table 2). We also found differences in the distribution of work ability (WAS)
51
52 225 scores (*n*=235) between the SBT low, medium and high risk groups at long-term follow-up (*p*=0.000) and
53
54
55
56
57
58
59
60

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 <i>n</i> =238	SBT risk group			<i>p</i> -value
		Low <i>n</i> =103	Medium <i>n</i> =107	High <i>n</i> =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)	
EQ-5D ^a <0.6, <i>n</i> (%)	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, <i>n</i> (%)	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

^aEQ-5D scores, range -0.59-1

^b3 missing from the high risk group (total cohort: $n=235$ and $n=25$ for the high risk group)

^cWhere 0 equates to "completely unable to work" and 10 equates to "work ability at its best"

^dChi 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; χ^2 -test=5.41, df 8, $p=0.71$ and for work ability; χ^2 -test=5.27, df 8, $p=0.73$) as a non-significant p -value >0.05 indicates that the model is good³⁸. 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

1
2 246 'acceptable' (≥ 0.7) (Fig. 2). For work ability, the area under the curve (AUC) was 0.68 (CI 0.61-0.76) which
3
4 247 was just below the limit (≥ 0.7) for acceptable discrimination (Fig. 3).

5
6
7 248

8
9
10 249 **INSERT FIG 2 and FIG 3 here**

11
12
13 250

14
15
16 251 The sensitivity, specificity, PPV, NPV and likelihood ratios for the SBT risk groups for HRQoL and work
17
18 252 ability are presented in Table 3. The LR+s were higher and the LR-s were lower for HRQoL outcomes
19
20 253 compared to work ability outcomes which indicate better discrimination of the SBT for poor HRQoL
21
22 254 compared to poor work ability (Table 3).

23
24
25 255

26
27
28 256 **Table 3.** Discriminative ability of the SBT risk group cutoffs (low/medium and medium/high) to predict
29
30 257 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)

31
32
33
34
35
36
37
38
39
40
41
42
43
44 258 SBT, STarT Back Tool; HRQoL, Health related quality of life; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score;
45 259 PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

46 260

47
48
49 261

50
51
52
53 262

263 Discussion and Conclusions

264 This is the first study to evaluate the predictive validity of SBT for HRQoL and work ability outcomes at
265 long-term follow-up in a population with acute/subacute back and/or neck pain. The findings of this study
266 support the ability of the SBT risk groups to predict future poor HRQoL or poor work ability, for patients
267 presenting with an episode of acute/subacute back and/or neck pain in primary care. Individuals classified as
268 SBT high risk had a significantly increased risk of having poor HRQoL (OR 6.2) and poor work ability (OR
269 5.1) in the long-term compared to individuals classified as SBT low risk.

270
271 Strengths of this study include the prospective design of a well characterized group of individuals from 35
272 different primary care centers. The SBT was used and administered by many different physiotherapists
273 which makes this setting real and clinically relevant. The population studied was relatively homogenous
274 including only patients with acute or subacute pain, not individuals with chronic pain. This study population
275 differs from the original UK development population for SBT by excluding chronic back pain and including
276 neck pain. As might be expected, the distribution between the SBT risk groups at baseline differed compared
277 to the UK development population¹⁷. In our study population, the percentage of individuals at high risk
278 were lower (12%) compared to the original UK sample (15%)¹⁷ which may be due to our sample including
279 patients with acute/subacute pain. However, there is still a clear and statistically significant difference in
280 HRQoL and work ability outcomes between the three risk groups in the expected direction in our Swedish
281 sample. As the majority of the patients in this study (n=160/238) were included in an RCT
282 (ClinicalTrials.gov ID: NCT02609750) we have access to information about tentative confounding factors
283 and we investigated several of these factors (age, sex, type of treatment and time to follow-up) that may
284 have potentially influenced the prognostic ability of the SBT. This study showed that age, sex, type of
285 treatment or time to follow-up did not significantly influence the ability of the SBT to predict HRQoL and
286 work ability outcomes at long-term follow-up. In another SBT non-stratified primary care setting where they

1
2 287 studied different influences (care setting, episode duration and time to follow-up) on the prognostic ability of
3
4 288 the SBT for disability outcomes³⁹ they found that the only factor that modified the prognostic ability of the
5
6 289 SBT risk groups was episode duration with SBT being less predictive in very acute patients (<2 weeks
7
8 290 duration). Another strength is that we analyzed the predictive validity in different ways, for example we
9
10 291 studied both the established SBT risk groups and the SBT overall score to predict the outcomes of HRQoL
11
12 292 and work ability. We also analyzed the outcomes HRQoL and work ability both on the continuous scale
13
14 293 (Kruskal-Wallis) and as dichotomized (logistic regression).
15
16
17
18 294

19
20
21 295 A weakness of this study is that we had limited access to information about patients not included in the RCT
22
23 296 (n=78/238) compared to patients included in the RCT (n=160/238). For patients not included in the RCT, we
24
25 297 did not have access to registered diagnoses, pain duration (acute or subacute) or self-reported HRQoL and
26
27 298 work ability questionnaires at baseline. For that reason, we were not able to do comparative analyzes on
28
29 299 baseline and follow-up data. Another weakness might be the variation in time to follow-up between patients
30
31 300 which may have influenced the results. For patients included in the RCT, median time to follow-up was 12
32
33 301 months (range 11-19) and for patients not included in the RCT, it was 22 months (range 16-27). Therefore
34
35 302 we used the follow-up time as one of the independent variables in the regression analysis. All data in this
36
37 303 study is self-reported. However, self-reported data on sickness absence among employees in Sweden has
38
39 304 been reported at least as valuable as register data⁴⁰.
40
41
42

43 305 The ability of the SBT overall score to discriminate between patients with poor or good HRQoL and work
44
45 306 ability differed slightly between the two outcomes with a slightly better discrimination for HRQoL (0.73)
46
47 307 than for work ability (0.68). In a recent systematic review, Karran et al⁴¹ investigated how well prognostic
48
49 308 screening instruments for BP, including the SBT, discriminate between patients who develop a poor
50
51 309 outcome and those who do not⁴¹. Prognostic screening tools tend to perform poorly at assigning higher risk
52
53 310 scores to individuals who develop chronic pain compared to those who do not and they also tend to predict
54
55 311 disability outcomes better than most other outcomes⁴¹. The discriminative performance of SBT for work
56
57
58
59
60

1
2 312 ability outcomes in this study (AUC 0.68) was higher than for other prognostic tool's reported abilities to
3
4 313 discriminate pain outcomes (pooled AUC= 0.59)⁴¹ and the SBT discriminative performance for HRQoL
5
6 314 outcomes in this study (AUC 0.73) was in line with the pooled disability predictive performance (pooled
7
8 315 AUC=0.74). In comparison to the original UK sample and a Danish sample in primary care, where
9
10 316 participants had variable duration of back pain and the primary outcomes were disability at 3 months follow-
11
12 317 up^{17 42}, the predictive ability of the SBT in our study was not as strong as in the UK population (AUC 0.81)
13
14 318 but similar to the Danish population (AUC 0.71). In our study, as in the Danish study, the physiotherapy
15
16 319 treatment was not targeted to SBT risk groups and treatment was therefore likely to be heterogeneous. A
17
18 320 variation of values are expected as the AUC (derived from the ROC curve: sensitivity/1-specificity),
19
20 321 depends on the characteristics of the population and possible explanations might be cultural and differences
21
22 322 in treatment. Another possible explanation in variation of AUC values may be that a ROC curve analysis
23
24 323 requires dichotomization of outcomes and the definitions of poor outcome may also have affected the
25
26 324 results. The discriminative ability of the SBT risk groups to predict poor HRQoL and work ability outcome
27
28 325 was affected of how the three risk groups were merged and dichotomized (low vs medium/high or
29
30 326 low/medium vs high). Similar differences in discrimination were also found in the original study for
31
32 327 disability outcomes¹⁷. But regardless of which cutoff that was used, the results of the LRs indicate a slightly
33
34 328 better discrimination of the SBT for poor HRQoL than for poor work ability and that the NPVs were
35
36 329 consistently high for both outcomes which indicate a high probability that a good outcome is present when
37
38 330 patients are classified as low risk. The proportion of patients with poor HRQoL and poor work ability was
39
40 331 significantly higher in higher SBT risk groups at long-term follow-up, but not all patients were correctly
41
42 332 classified. When patients are misclassified as low risk they may be undertreated and when patients are
43
44 333 misclassified as high risk they may be overtreated. It is important for clinicians to be aware of the potential
45
46 334 of misclassification as costs for misclassification and overtreatment of patients with a good prognosis can be
47
48 335 high¹⁸ and also detrimental in patients with acute back pain⁴³.

336

1
2 337 The EQ-5D was applied to measure HRQoL because it has been found to have good prediction of return to
3
4 338 work and the cut-off ≥ 0.6 on EQ-5D has been proposed to be a limit for having sufficient capacity to work
5
6 339 for patients with back and neck pain²⁹. Another cut-off has been used in a study of patients with
7
8 340 musculoskeletal pain taking part in a national rehabilitation program in Sweden where ≥ 0.5 on EQ-5D at
9
10 341 start showed reduced sick leave days after the rehabilitation⁴⁴. Our population had a median EQ-5D score of
11
12 342 0.80 which is just below the mean scores for a Swedish normal population (0.84)⁴⁵. The fact that our sample
13
14 343 included patients at an early stage of their pain (acute or subacute) with no or short time of sick leave may
15
16 344 have influenced the high level of HRQoL in our study sample. To measure work ability, we used the WAS
17
18 345 which is the first item in the WAI, a widely used questionnaire for measuring the health and functional
19
20 346 capacity dimension of work ability³¹. The cut-off (WAS $< 8 / \geq 8$) chosen in this study represents poor or
21
22 347 moderate (poor) and good/excellent (good) work ability based on the same categorization as for the whole
23
24 348 WAI³². The WAS has shown to be a good alternative to the whole WAI⁴⁶ even though the whole WAI is
25
26 349 superior compared to its individual items⁴⁷.

27
28
29
30
31 350
32
33
34 351 There are recommendations for the use of screening methods in health care to identify patients in early
35
36 352 stages with the purpose to guide them to the best treatment⁴⁸⁻⁵⁰ and also for enhancing return to work^{51 52}.
37
38 353 SBTs concurrent validity has earlier been studied for patients with back and/or neck pain⁵³ and a modified
39
40 354 SBT have been tested to predict physical health outcome, using the SF-36⁵⁴ but this was the first time the
41
42 355 predictive validity of the SBT was studied for the outcomes of HRQoL and work ability for individuals with
43
44 356 both back and neck pain. This study showed that the SBT can identify acute/subacute back and neck pain
45
46 357 patients with a poor long-term HRQoL and work ability outcome. Therefore this study widens the usefulness
47
48 358 of the SBT compared to earlier studies^{17 55-58}. More research is needed to find appropriate treatments for
49
50 359 patients with nonspecific acute/subacute back and/or neck pain⁵⁹. The SBT is primarily designed as a
51
52 360 “stratified care tool” which involves targeting treatment to subgroups of patients based on their key
53
54 361 characteristics⁶⁰. Future studies are required to investigate whether the implementation of screening together

1
2 362 with matched treatment pathways improves HRQoL and work ability outcomes for these patients. The
3
4 363 results of this study suggest that the SBT may help clinicians in primary care to pay more attention to work
5
6 364 related factors at an early stage which is a priority in preventing chronicity⁶¹ and essential for a successful
7
8 365 rehabilitation process⁶².

12 366

16 367 **Authors Contributions**

19 368 All authors discussed the results and commented on the manuscript. MF, IP, KS and BG were responsible
20
21 369 for the study design, data analysis and interpretation. MF, BG and KS prepared and validated data. MF
22
23 370 collected data and drafted the manuscript. JH and CPS took part in study design, data analysis and
24
25 371 interpretation of data. All authors read and approved the final version of the manuscript.

29 372

33 373 **Acknowledgements**

37 374 Senior statistician Mikael Åström for statistical support, datamanager Chanchai Manuswin and the staff at
38
39 375 Data analysis and Register Centre, Region Skåne, Lund, Sweden, the research team at the Institute of
40
41 376 Primary Care and Health Sciences, Keele university, Stoke-on-Trent, UK and the staff at the Blekinge
42
43 377 Centre of Competence, Karlskrona, Sweden for help with support. All physiotherapists and patients involved
44
45 378 in this project, including the WorkUp project, REHSAM, in Region Skåne, Region Kronoberg and
46
47 379 Landstinget Blekinge, Sweden, for help with data collection.

51 380 **Data sharing statement**

55 381 The datasets analysed during the current study are available from the corresponding author on reasonable
56
57 382 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](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](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, Beudart 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.000000000000062 [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]

- 1
2 486 42. Morso L, Kent P, Albert HB, et al. The predictive and external validity of the STarT Back Tool in Danish primary
3 487 care. *European spine journal : official publication of the European Spine Society, the European Spinal*
4 488 *Deformity Society, and the European Section of the Cervical Spine Research Society* 2013;22(8):1859-67. doi:
5 489 10.1007/s00586-013-2690-z [published Online First: 2013/02/12]
- 6 490 43. George SZ, Fritz JM, Bialosky JE, et al. The effect of a fear-avoidance-based physical therapy intervention for
7 491 patients with acute low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976)*
8 492 2003;28(23):2551-60. doi: 10.1097/01.brs.0000096677.84605.a2 [published Online First: 2003/12/04]
- 9 493 44. Stigmar KG, Petersson IF, Joud A, et al. Promoting work ability in a structured national rehabilitation program in
10 494 patients with musculoskeletal disorders: outcomes and predictors in a prospective cohort study. *BMC*
11 495 *musculoskeletal disorders* 2013;14:57. doi: 10.1186/1471-2474-14-57 [published Online First: 2013/02/07]
- 12 496 45. Burstrom K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the EQ-
13 497 5D. *Qual Life Res* 2001;10(7):621-35. [published Online First: 2002/02/02]
- 14 498 46. Ahlstrom L, Grimby-Ekman A, Hagberg M, et al. The work ability index and single-item question: associations with
15 499 sick leave, symptoms, and health—a prospective study of women on long-term sick leave. *Scand J Work*
16 500 *Environ Health* 2010:404-12.
- 17 501 47. Lundin A, Leijon O, Vaez M, et al. Predictive validity of the Work Ability Index and its individual items in the
18 502 general population. *Scand J Public Health* 2017;45(4):350-56. doi: 10.1177/1403494817702759 [published
19 503 Online First: 2017/04/08]
- 20 504 48. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching
21 505 clinical outcomes. *BMJ (Clinical research ed)* 2013;346:e5595. doi: 10.1136/bmj.e5595 [published Online
22 506 First: 2013/02/07]
- 23 507 49. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research.
24 508 *BMJ (Clinical research ed)* 2013;346:e5793. doi: 10.1136/bmj.e5793 [published Online First: 2013/02/07]
- 25 509 50. Foster NE, Mullis R, Hill JC, et al. Effect of stratified care for low back pain in family practice (IMPACT Back): a
26 510 prospective population-based sequential comparison. *Ann Fam Med* 2014;12(2):102-11. doi:
27 511 10.1370/afm.1625 [published Online First: 2014/03/13]
- 28 512 51. Kuoppala J, Lamminpaa A. Rehabilitation and work ability: a systematic literature review. *J Rehabil Med*
29 513 2008;40(10):796-804. doi: 10.2340/16501977-0270 [published Online First: 2009/02/27]
- 30 514 52. Carroll C, Rick J, Pilgrim H, et al. Workplace involvement improves return to work rates among employees with
31 515 back pain on long-term sick leave: a systematic review of the effectiveness and cost-effectiveness of
32 516 interventions. *Disabil Rehabil* 2010;32(8):607-21. doi: 10.3109/09638280903186301 [published Online First:
33 517 2010/03/09]
- 34 518 53. Forsbrand M, Grahn B, Hill JC, et al. Comparison of the Swedish STarT Back Screening Tool and the Short Form of
35 519 the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck
36 520 pain. *BMC musculoskeletal disorders* 2017;18(1):89. doi: 10.1186/s12891-017-1449-9
- 37 521 54. Hill JC, Afolabi EK, Lewis M, et al. Does a modified STarT Back Tool predict outcome with a broader group of
38 522 musculoskeletal patients than back pain? A secondary analysis of cohort data. *BMJ Open*
39 523 2016;6(10):e012445. doi: 10.1136/bmjopen-2016-012445 [published Online First: 2016/10/16]
- 40 524 55. Hill JC, Dunn KM, Main CJ, et al. Subgrouping low back pain: a comparison of the STarT Back Tool with the Orebro
41 525 Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 2010;14(1):83-9. doi:
42 526 <http://dx.doi.org/10.1016/j.ejpain.2009.01.003>
- 43 527 56. Wideman TH, Hill JC, Main CJ, et al. Comparing the responsiveness of a brief, multidimensional risk screening tool
44 528 for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain*
45 529 2012;153(11):2182-91.
- 46 530 57. Beneciuk JM, Robinson ME, George SZ. Subgrouping for patients with low back pain: A multidimensional
47 531 approach incorporating cluster analysis and the start back screening tool. *Journal of Pain* 2015;16(1):19-30.
48 532 doi: <http://dx.doi.org/10.1016/j.jpain.2014.10.004>
- 49 533 58. Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool and
50 534 prognosis for people receiving physical therapy for low back pain. *Physical therapy* 2011;91(5):722-32.
- 51 535 59. Åsenlöf P, Michaelson P, Grahn B, et al. Acute neck and back pain: preventive interventions - Effects of physical
52 536 training, manual treatment and cognitive behavioural interventions. : Swedish Agency for Health technology
53 537 Assessment and Assessment of Social Services (SBU) 2016.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

538 60. Foster NE, Hill JC, O'Sullivan P, et al. Stratified models of care. *Best practice & research Clinical rheumatology*
539 2013;27(5):649-61. doi: 10.1016/j.berh.2013.10.005 [published Online First: 2013/12/10]
540 61. Rozenberg S, Foltz V, Fautrel B. Treatment strategy for chronic low back pain. *Joint Bone Spine* 2012;79(6):555-
541 59. doi: <http://dx.doi.org/10.1016/j.jbspin.2012.09.003>
542 62. Wahlin C, Ekberg K, Persson J, et al. Evaluation of self-reported work ability and usefulness of interventions
543 among sick-listed patients. *Journal of occupational rehabilitation* 2013;23(1):32-43. doi: 10.1007/s10926-
544 012-9376-y [published Online First: 2012/07/05]

For peer review only

559 **Figure legends**

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

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

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

573

574

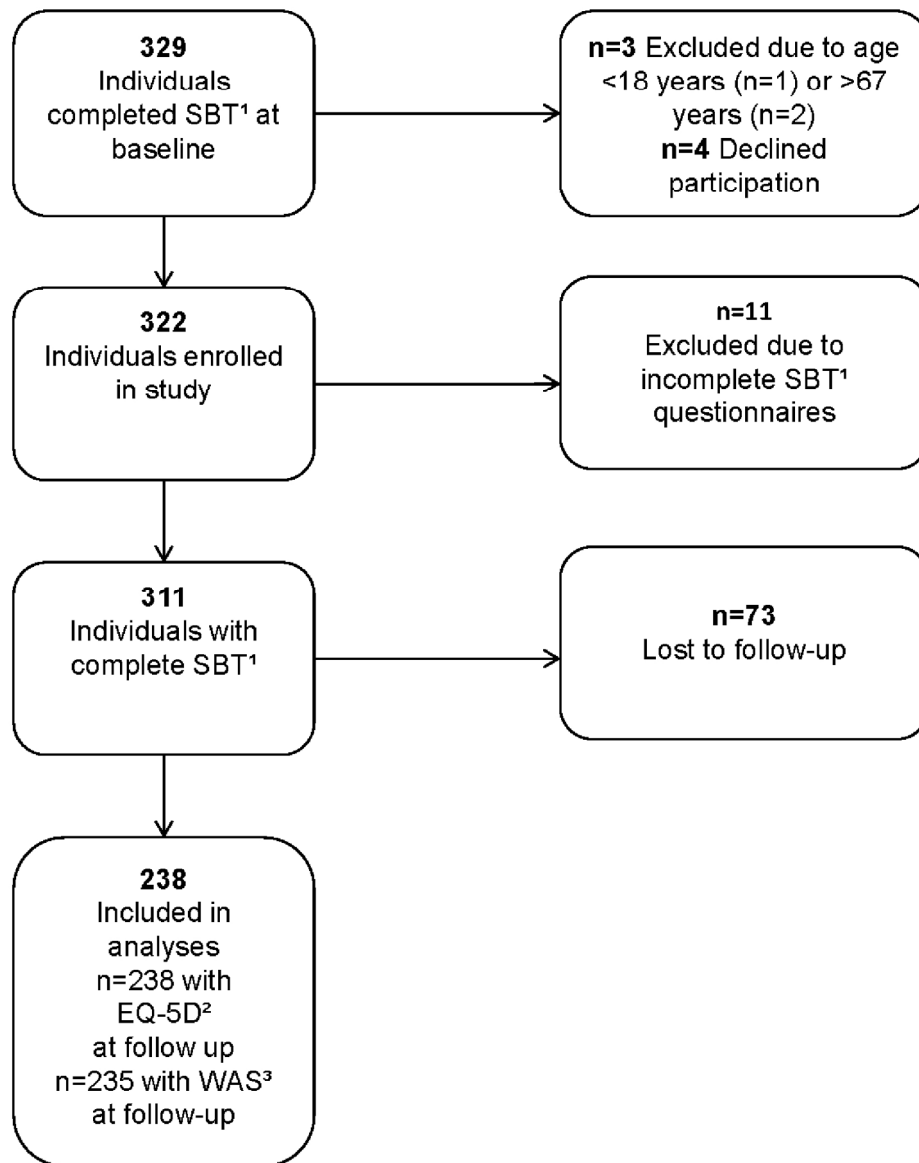
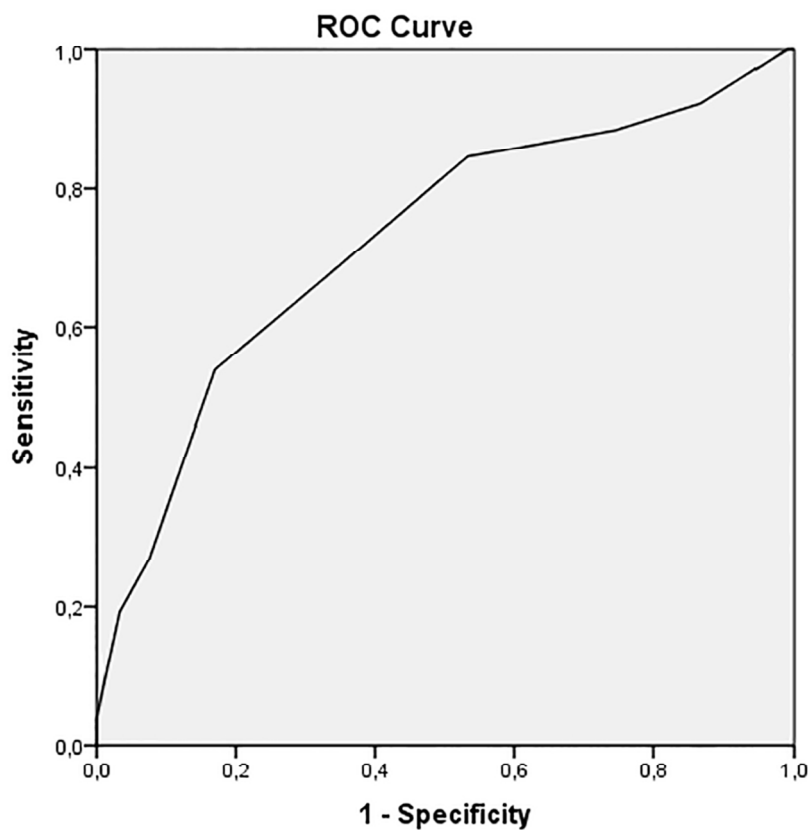


Fig.1. Flowchart of inclusion and exclusion of participants. ¹STarT Back Tool. ²EuroQol five-dimension. ³Work Ability Score.

151x181mm (300 x 300 DPI)



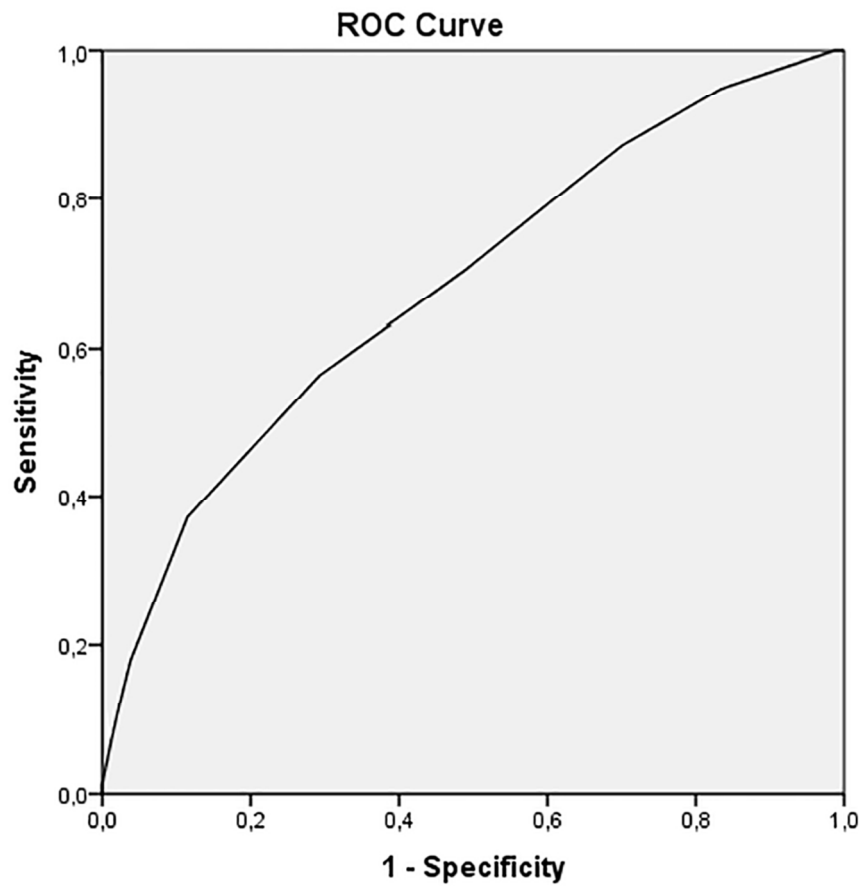
Diagonal segments are produced by ties.

35
36
37
38
39
40
41
42
43

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.

44
45
46
47
48
49

111x113mm (300 x 300 DPI)



Diagonal segments are produced by ties.

35
36
37
38
39
40
41
42
43

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.

44
45
46
47
48
49

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:	<i>BMJ Open</i>
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, Grahm, Birgitta; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg Hill, Jonathan; Keele University, Research Institute of Primary Care and Health Sciences Petersson, Ingemar; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Skåne University Hospital Post Sennehed, Charlotte; Lunds Universitet, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund; Department of Research and Development, Region Kronoberg 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

SCHOLARONE™
Manuscripts

1
2
3 1 Can the STarT Back Tool predict health related quality of life and
4
5
6 2 work ability after an acute/subacute episode with back or neck pain? –
7
8
9 3 a psychometric validation study in primary care

10
11
12 4 M. Forsbrand^{1,2*}, B. Grahn^{1,3}, JC. Hill⁴, IF. Petersson^{1,6}, C. Post Sennehed^{1,3}, K. Stigmar^{5,6}

15 5 ¹Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Orthopedics, Lund,
16
17 6 Sweden, ²Blekinge Centre of Competence, Landstinget Blekinge, Karlskrona, Sweden, ³Department of
18
19 7 Research and Development, Region Kronoberg, Växjö, Sweden, ⁴Research Institute of Primary Care and
20
21 8 Health Sciences, Keele University, Stoke-on-Trent, United Kingdom, ⁵Department of Health Sciences,
22
23 9 Physiotherapy, Lund University, Lund, Sweden, ⁶Skåne University Hospital, Lund, Sweden.

24
25
26 10
27
28 11 ***Corresponding author:**

29
30
31 12 Malin Forsbrand, RPT, PhD student.

32
33 13 Address: Blekinge Centre of Competence, SE-371 81 Karlskrona, Sweden.

34
35 14 E-mail: malin.forsbrand@med.lu.se

36
37 15 Telephone: +46455735616

38
39 16 **E-mails authors (not corresponding):** birgitta.grahn@kronoberg.se, j.hill@keele.ac.uk,

40
41 17 ingemar.petersson@skane.se, charlotte.sennehed@kronoberg.se, kjerstin.stigmar@med.lu.se

42
43 18 **Word count:** 4108 words

44
45 19 **Category:** Research article

46
47 20 **Funding sources:** This work was supported by The Scientific Committee of Blekinge County Council,
48
49 21 Sweden (Dnr 2017-00686) and the County Councils Region Skåne, Kronoberg and Blekinge, REHSAM
50
51 22 research, Sweden (RS2011/005).

52
53 23 **No conflicts of interest declared**

24 Abstract

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

30 **Setting:** Prospective data from 35 primary care centers in south Sweden during 2013.

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

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

37 **Results:** Based on SBT risk group stratification, 103 (43%), 107 (45%) and 28 (12%) patients were
38 considered as low, medium and at high risk respectively. There were statistically significant differences
39 in HRQoL ($p<0.001$) and work ability ($p<0.001$) at follow-up between all three SBT risk groups. Patients
40 in the high risk group had a significantly increased risk of having poor HRQoL (OR 6.16, 95 % CI 1.50-
41 25.26) and poor work ability (OR 5.08, 95 % CI 1.75-14.71) vs the low risk group at follow-up. The
42 AUC was 0.73 (CI 0.61-0.84) for HRQoL and 0.68 (CI 0.61-0.76) for work ability.

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

46

1
2 47 **Keywords:** STarT Back Tool, health related quality of life, work ability, primary care, back pain, neck
3
4 48 pain.
5

6 49 Strengths and limitations of this study

- 9 50 • This is the first study to evaluate the predictive validity of SBT of the outcomes HRQoL and work
11 51 ability at long-term follow-up in a population with acute/subacute back and/or neck pain.
- 13 52 • In this prospective study we have recruited patients from 35 different primary care centers, where
15 53 many physiotherapists were engaged.
- 17 54 • The predictive validity of the SBT was examined in different ways.
- 19 55 • Limited baseline data was available for one part of the study population.
- 21 56 • Limitations of the study were the broad variation in time to follow-up.

67 Introduction

68 Musculoskeletal pain, especially back pain (BP) and neck pain (NP) are highly prevalent in the general
69 population^{1 2} causing disability for the individual and high costs for society³⁻⁵. Individuals with BP and
70 NP are mostly managed in primary care^{6 7} and patients presenting with these conditions are at risk of
71 sickness absence⁸ and poor health related quality of life (HRQoL)^{9 10}. To have concurrent BP and NP is
72 also common¹¹ and increases the risk of work disability further in the long-term¹². Whilst most
73 individuals with acute back pain improve quickly and return to work¹³, for some of them the pain is
74 more severe and lasts for a longer period^{14 15}. In a Swedish cohort of individuals with BP and NP about
75 half of the population reported pain and disability 5 years after onset¹⁶. There are recommendations for
76 the use of screening methods in health care to identify patients in early stages with the purpose to guide
77 them to the best treatment¹⁷⁻¹⁹, to support staying at work or for enhancing return to work^{20 21}. The UK
78 Nice guidance recommend using brief questionnaires to identify individuals of poor outcomes and
79 stratify care²² but there is a lack of such tools that can be used in primary care. The widely used STarT
80 Back Tool (SBT)²³, is a brief risk stratification tool that includes nine questions on predictors for long-
81 term disabling back pain, in order to match individuals to appropriate targeted treatments, according to
82 their prognostic profile. Using the SBT together with targeted treatment pathways has shown improved
83 efficiency regarding patients' clinical outcomes and reduced health care costs in the United Kingdom²⁴.
84 The SBT is cross-culturally adapted and validated in Swedish²⁵ and recently also for a population with
85 both back and neck pain in primary care²⁶. The SBT is developed and validated to predict future
86 disability due to low back pain of any duration^{23 27-30}, but it has not yet been studied for the outcomes of
87 HRQoL and work ability for a population with acute/subacute back and neck pain in primary care. The
88 aim of this study was therefore to evaluate the predictive validity of SBT of the outcomes HRQoL and
89 work ability at long-term follow-up in a population with acute/subacute back and/or neck pain.

90

91 **Methods**

92 **Design**

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

97 **Participants and procedure**

98 Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care
99 centers in the southern parts of Sweden, as part of an RCT³¹. Patients that all applied for physiotherapy
100 treatment on self-referral due to an episode of acute/subacute (<12 weeks) back and/or neck pain, who
101 were not currently on sick leave or had been on sick leave for less than 60 days and who had been
102 working ≥ 4 consecutive weeks last year were asked to participate. It could be either a first episode or a
103 recurrent episode of back and/or neck pain after a period of at least three months of no substantial pain.
104 Patients that were pregnant, had severe pathology (“red flags”)³² or were not able to understand the
105 Swedish language were not eligible to participate. At baseline, patients completed the “ÖMPSQ-short”³³
106 which was used for screening for inclusion to the RCT (≥ 40 points)³¹ and the SBT which was
107 administered only for the purpose of psychometric testing. Thereafter the SBT was not actively used by
108 the physiotherapists or any other professionals. In all, 329 patients completed the SBT questionnaire and
109 formed the population of this psychometric study. Patients that were older than 67 years or younger than
110 18 years (n=3), declined participation (n=4), had any missing item on the SBT (n=11) or those who were
111 lost to follow-up (n=73) were excluded. The final study population (n=238) consisted of patients
112 included in the RCT (RCT intervention, n=61 and RCT control, n=99) and patients not included in the
113 RCT (n=78). The analyses were restricted to those who had complete data for work ability (n=235) and

1
2 114 HRQoL (n=238) outcomes at long-term follow-up. The reason we included both RCT and not RCT
3
4 115 patients was to ensure as broad a sample as possible for this SBT predictive validity study. RCT patients
5
6 116 received either structured physiotherapy treatment with a workplace intervention (RCT intervention) or
7
8 117 structured physiotherapy without a workplace intervention (RCT control)³¹ and were followed up at the
9
10 118 planned 12-months follow-up. Not RCT patients received usual primary care and were followed up by
11
12 119 postal questionnaires. Data from all questionnaires were manually entered into a SPSS 22.0 database and
13
14 120 were thoroughly checked and validated. All questionnaires were scored, and missing items handled,
15
16 121 according to the methods specified by the instrument developers.
17
18
19
20
21 122

23 123 **Baseline data**

24
25
26
27 124 Baseline questionnaire data included type of treatment received (RCT intervention, RCT control or usual
28
29 125 primary care) and self-reports of SBT, age and gender.
30
31
32
33 126

34 127 **STarT Back Tool**

35
36
37 128 The STarT Back Tool (SBT) is a 9-item questionnaire with questions relating to modifiable physical
38
39 129 (item 1–4) and psychosocial (item 5–9) risk factors for long-term disabling BP, designed to support
40
41 130 clinicians in directing individuals to different levels of care²³. The SBT has three risk subgroups which
42
43 131 classifies patients into low, medium or high risk for poor disability outcomes. The SBT overall score
44
45 132 ranges between 0 and 9. Item 1–4 is about referred leg pain, neck or shoulder pain, difficulties in walking
46
47 133 and difficulties in dressing. Item 5–9 form the psychosocial subscale which screen for fear of physical
48
49 134 activity, anxiety, pain catastrophizing, depressive mood and overall impact from their BP. Items 1–8 have
50
51 135 a dichotomous response option; “disagree” (0p) or “agree” (1p). Item 9 uses a 5-point Likert Scale from
52
53 136 “not at all” to “extremely”, where responses “very much” or “extremely” are counted as one point and
54
55 137 the other responses as zero. A total score of ≤ 3 points indicates low risk, a total score ≥ 4 points in
56
57
58
59
60

1
2 138 combination with <4 points on the psychosocial subscale (item 5–9) are medium risk and a psychosocial
3
4 139 subscale score of ≥ 4 points indicates high risk for poor disability outcomes²³.
5
6
7 140

141 **Long-term follow-up data**

143 *Health related quality of life*

144 Health-related quality of life (HRQoL) was measured by the EuroQol five-dimension (EQ-5D, 3L)
145 questionnaire³⁴ which is a generic, health-related quality of life instrument^{35 36}. The EQ-5D comprises
146 the EQ descriptive system which has 5 dimensions: mobility, self-care, usual activities, pain/discomfort
147 and anxiety/depression. The digits for the 5 dimensions are combined in a 5-digit number describing the
148 respondent's health state³⁷. The 5-digit number is given a value between -0.59 and 1.0 according to the
149 UK tariff³⁸, where 1 corresponds to full health and lower EQ-5D values reflect lower HRQoL. Health
150 Related Quality of Life was also dichotomized into "poor" HRQoL (EQ-5D <0.6) and "good" HRQoL
151 (EQ-5D ≥ 0.6), based on a proposed cut-off for having sufficient capacity to be able to work for a
152 population with back and neck pain³⁹.
153

154 *Work ability*

155 Work ability was measured by self-reports on the single item question ("current work ability compared
156 with the lifetime best") from the Work Ability Index (WAI)^{40 41}. This first item in the WAI is known as
157 the "Work Ability score" (WAS)⁴². It consists of a scale from 0 representing "cannot work at all right
158 now" to 10 representing "my work ability as at its best right now" and has been proposed to be used as a
159 simple indicator for assessing the status and progress of work ability^{43 44}. Work ability was also
160 dichotomized using a previously published cut-off score⁴³ into "poor" work ability (WAS <8 points) and
161 "good" work ability (WAS ≥ 8 points).
162

163 **Statistical analyses**

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

168

169 **Predictive performance of the SBT**

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

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

182 Third, we evaluated the ability of the SBT overall scores (0-9 points) to discriminate between individuals
183 with poor or good HRQoL/work ability in long-term follow-up. For that purpose, we used the area under
184 the curve (AUC) statistics from receiver operating characteristic (ROC) curves⁴⁵. The strength of
185 discrimination was set according to the following descriptors: 0.7-<0.8 acceptable discrimination, 0.8-
186 <0.9 excellent discrimination, and ≥ 0.9 outstanding discrimination⁴⁶.

1
2 187 In addition, the predictive validity of the SBT risk group cutoffs (low/medium and medium/high) was
3
4 188 assessed by calculating sensitivity, specificity, positive predictive values (PPV), negative predictive
5
6 189 values (NPV) and positive and negative likelihood ratios (LRs) against long-term HRQoL and work
7
8 190 ability outcomes. The SBT risk group cutoffs (low/medium and medium/high) were used in line with the
9
10 191 original study²³. The PPV is the probability that a poor outcome is present when the test is positive and
11
12 192 the NPV is the probability that a good outcome is present when the test is negative. Higher positive LRs
13
14 193 and lower negative LRs indicate better discrimination. Likelihood ratios above 5 or below 0.2 are
15
16 194 generally seen as supporting a strong test, whereas values close to 1 indicate poor test performance⁴⁷.
17
18
19
20
21 195

24 196 **Patient and Public Involvement**

26
27
28 197 Relevant patient organizations were involved in the development and design of the RCT, where this study was
29
30 198 embedded. For this psychometric study, no patients were involved. The results of this study will be disseminated to
31
32 199 study participants by the use of SBT in primary care.
33
34
35
36
37 200

40 201 **Ethics**

42
43
44 202 The study was approved by the Regional Ethical Review Board in Lund, Sweden (Dnr 2012/497,
45
46 203 2013/426, Dnr 2015/214). Prior to inclusion, all patients obtained written information about the purpose
47
48 204 of the study and each individual gave informed consent to participate in the study (opt-out). The
49
50 205 principles of the Declarations of Helsinki were followed.
51
52
53
54 206

207 **Results**

208 **Study population**

209 The inclusion and exclusion of participants in the study is presented in a flowchart (Figure 1).

210

211 **INSERT FIG 1 here**

212

213 The final sample consisted of 238/329 patients (72%) including 160 (67%) females and 78 (33%) males.

214 Baseline characteristics of the study population are summarized in Table 1. The patient sample included

215 103 (43%) patients at low risk, 107 (45%) patients at medium risk, and 28 (12%) patients at high risk.

216 The median time to long-term follow-up was 13 (range 11-27) months. For not RCT patients, the median

217 time to follow-up was 12 months (range 11-19) and for RCT patients, the median time was 22 months

218 (range 16-27).

219

220

221

222

223

224

225

226

227

228

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

Variable	Total population <i>n</i> =238	SBT risk group		
		Low <i>n</i> =103 (43%)	Medium <i>n</i> =107 (45%)	High <i>n</i> =28 (12%)
Age, median (range)	46 (19-67)	45 (22-64)	47 (21-67)	38 (19-63)
Sex, <i>n</i> (%) female	160 (67)	73 (71)	72 (67)	15 (54)
Area of pain ^a				
BP ^b , <i>n</i> (%)	91 (38)	42 (41)	41 (38)	8 (29)
NP + BP ^c , <i>n</i> (%)	147 (62)	61 (59)	66 (62)	20 (71)
Type of intervention, <i>n</i> (%)				
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

^aArea of pain Based on question number 2 (neck or shoulder pain) on SBT

^bBP Back pain

^cNP + 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).

244

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

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

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

248 ^aEQ-5D scores, range -0.59-1

249 ^b3 missing from the high risk group (total population: *n*=235 and *n*=25 for the high risk group)

250 ^cWhere 0 equates to "completely unable to work" and 10 equates to "work ability at its best"

251 ^dKruskal-Wallis test, ^eChi square test for trend

252

253 The regression analysis showed that the SBT high risk group could significantly predict poor HRQoL
 254 (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,
 255 *p*=0.003) at long-term follow-up also after adjusting for age, sex, treatment and time to follow-up (Table
 256 3). Our regression model was well adapted to the data material as a non-significant *p*-value >0.05 of
 257 Hosmer and Lemeshow's test indicates that the model is good⁴⁸ (Table 3).

258

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.

Coefficient	HRQoL			Work ability		
	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	χ^2-test	P-value	df	χ^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

^aPoor HRQoL measured by EuroQol five-dimension questionnaire (EQ-5D) <0.6

^bPoor work ability measured by Work ability score (WAS) <8

HRQoL: Cox-Snell R²=0.12. Nagelkerke R²=0.21, n=238.

Work ability: Cox-Snell R²=0.11. Nagelkerke R²=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 (≥ 0.7) 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.

280

281 Discussion and Conclusions

282 This is the first study to evaluate the predictive validity of SBT for HRQoL and work ability outcomes at
283 long-term follow-up in a population with acute/subacute back and/or neck pain. The findings of this
284 study support the ability of the SBT risk groups to predict future poor HRQoL or poor work ability, for
285 patients presenting with an episode of acute/subacute back and/or neck pain in primary care. Individuals
286 classified as SBT high risk had a significantly increased risk of having poor HRQoL (OR 6.2) and poor
287 work ability (OR 5.1) in the long-term compared to individuals classified as SBT low risk. The
288 population studied was relatively homogenous including only patients with acute or subacute pain, not
289 individuals with chronic pain. This study population differs from the original UK development
290 population for SBT by excluding chronic back pain and including neck pain. As might be expected, the
291 distribution between the SBT risk groups at baseline differed compared to the UK development
292 population²³. In our study population, the percentage of individuals at high risk were lower (12%)
293 compared to the original UK sample (15%)²³ which may be due to our sample including patients with
294 acute/subacute pain. However, there is still a clear and statistically significant difference in HRQoL and
295 work ability outcomes between the three risk groups in the expected direction in our Swedish sample.

296

297 Strengths of this study include the prospective design of a well characterized group of individuals from
298 35 different primary care centers. The SBT was used and administered by many different
299 physiotherapists which makes this setting real and clinically relevant. Another strength is that we
300 analyzed the predictive validity in different ways, for example we studied both the established SBT risk
301 groups and the SBT overall score to predict the outcomes of HRQoL and work ability. We also analyzed
302 the outcomes HRQoL and work ability both on the continuous scale (Kruskal-Wallis) and as
303 dichotomized (logistic regression).

1
2 304
3
4
5 305 A weakness of this study is that we had limited access to baseline data from patients not included in the
6
7 306 RCT (n=78/238) compared to RCT patients (n=160/238). For not RCT patients, we did not have access
8
9 307 to baseline data from HRQoL and work ability questionnaires. For that reason, we were not able to do
10
11 308 comparative analyzes on baseline and follow-up data. When recommending tools for use in primary care
12
13 309 settings, preferably they should have been validated in large trials within this specific setting. However,
14
15 310 as is the case with this study of the SBT, information from smaller studies is still of scientific value. We
16
17 311 accept that our study population (n=329) is unlikely to be representative of all individuals consulting
18
19 312 primary care for acute/subacute BP and/or NP. However, even if they are a selected group of participants,
20
21 313 we don't think that this will have substantially affected the psychometric validation questions examined
22
23 314 in this study.

24
25
26
27 315
28
29
30 316 The time to follow-up varied between patients in our study which may have influenced the results. The
31
32 317 optimal time point for identifying patients at risk of developing persistent back pain may vary and is a
33
34 318 forum for discussion⁴⁹. In our study, two third of the study population (n=160) were in the RCT and
35
36 319 were followed-up at a planned physiotherapy visit at 12 months. For not RCT patients (n=78) the
37
38 320 ambition was also to follow-up at 12 months but these patients were followed-up with postal
39
40 321 questionnaires and due to practical reasons there were a wider variation on the time for follow-up. This is
41
42 322 of course a limitation, but did not have impact on the results in the regression analyses. However, we had
43
44 323 access to information about tentative confounding factors and we investigated several of these factors
45
46 324 (age, sex, treatment and time to follow-up) that may have potentially influenced the prognostic ability of
47
48 325 the SBT. In this study we included both patients with neck pain and back pain. Since this group of
49
50 326 patients often have concurrent pain from the back or neck¹¹, we decided to not include this in the
51
52 327 regression analysis. In another SBT non-stratified primary care setting where they studied different
53
54 328 influences (care setting, episode duration and time to follow-up) on the prognostic ability of the SBT for

1
2 329 disability outcomes⁵⁰ they found that the only factor that modified the prognostic ability of the SBT risk
3
4 330 groups was episode duration with SBT being less predictive in very acute patients (<2 weeks duration).

5
6
7 331
8
9
10 332 The ability of the SBT overall score to discriminate between patients with poor or good HRQoL and
11
12 333 work ability differed slightly between the two outcomes with a slightly better discrimination for HRQoL
13
14 334 (0.73) than for work ability (0.68). The AUC values are not very high, but still around 0.7, which is
15
16 335 considered as acceptable⁴⁶. In a recent systematic review, Karran et al⁵¹ investigated how well
17
18 336 prognostic screening instruments for BP, including the SBT, discriminate between patients who develop
19
20 337 a poor outcome and those who do not⁵¹. Prognostic screening tools tend to perform poorly at assigning
21
22 338 higher risk scores to individuals who develop chronic pain compared to those who do not and they also
23
24 339 tend to predict disability outcomes better than most other outcomes⁵¹. The discriminative performance of
25
26 340 SBT for work ability outcomes in this study (AUC 0.68) was higher than for other prognostic tool's
27
28 341 reported abilities to discriminate pain outcomes (pooled AUC= 0.59)⁵¹ and the SBT discriminative
29
30 342 performance for HRQoL outcomes in this study (AUC 0.73) was in line with the pooled disability
31
32 343 predictive performance (pooled AUC=0.74). In comparison to the original UK sample and a Danish
33
34 344 sample in primary care, where participants had variable duration of back pain and the primary outcomes
35
36 345 were disability at 3 months follow-up^{23 28}, the predictive ability of the SBT in our study was not as
37
38 346 strong as in the UK population (AUC 0.81) but similar to the Danish population (AUC 0.71). In our
39
40 347 study, as in the Danish study, the physiotherapy treatment was not targeted to SBT risk groups and
41
42 348 treatment was therefore likely to be heterogeneous. A variation of values are expected as the AUC
43
44 349 (derived from the ROC curve: sensitivity/1-specificity), depends on the characteristics of the population
45
46 350 and possible explanations might be cultural and differences in treatment. Another possible explanation in
47
48 351 variation of AUC values may be that a ROC curve analysis requires dichotomization of outcomes and the
49
50 352 definitions of poor outcome may also have affected the results. The discriminative ability of the SBT risk
51
52 353 groups to predict poor HRQoL and work ability outcome was affected of how the three risk groups were
53
54
55
56
57
58
59
60

1
2 354 merged and dichotomized (low vs medium/high or low/medium vs high). Similar differences in
3
4 355 discrimination were also found in the original study for disability outcomes²³. But regardless of which
5
6 356 cutoff that was used, the results of the LRs indicate a slightly better discrimination of the SBT for poor
7
8 357 HRQoL than for poor work ability and that the NPVs were consistently high for both outcomes which
9
10 358 indicate a high probability that a good outcome is present when patients are classified as low risk. The
11
12 359 proportion of patients with poor HRQoL and poor work ability was significantly higher in higher SBT
13
14 360 risk groups at long-term follow-up, but not all patients were correctly classified. When patients are
15
16 361 misclassified as low risk they may be undertreated and when patients are misclassified as high risk they
17
18 362 may be overtreated. It is important for clinicians to be aware of the potential of misclassification as costs
19
20 363 for misclassification and overtreatment of patients with a good prognosis can be high²⁴ and also
21
22 364 detrimental in patients with acute back pain⁵².

23
24
25
26 365
27
28
29 366 The EQ-5D was applied to measure HRQoL because it has been found to have good prediction of return
30
31 367 to work and the cut-off ≥ 0.6 on EQ-5D has been proposed to be a limit for having sufficient capacity to
32
33 368 work for patients with back and neck pain³⁹. Another cut-off has been used in a study of patients with
34
35 369 musculoskeletal pain taking part in a national rehabilitation program in Sweden where ≥ 0.5 on EQ-5D at
36
37 370 start showed reduced sick leave days after the rehabilitation⁵³. Our population had a median EQ-5D
38
39 371 score of 0.80 which is just below the mean scores for a Swedish normal population (0.84)⁵⁴. The fact that
40
41 372 our sample included patients at an early stage of their pain (acute/subacute) with no or short time of sick
42
43 373 leave may have influenced the high level of HRQoL in our study sample. To measure work ability, we
44
45 374 used the WAS which is the first item in the WAI, a widely used questionnaire for measuring the health
46
47 375 and functional capacity dimension of work ability⁴¹. The cut-off (WAS $< 8/\geq 8$) chosen in this study
48
49 376 represents poor or moderate (poor) and good/excellent (good) work ability based on the same
50
51 377 categorization as for the whole WAI⁴². The WAS has shown to be a good alternative to the whole WAI
52
53 378⁵⁵ even though the whole WAI is superior compared to its individual items⁵⁶.

1
2 379
3
4 380 SBTs concurrent validity has earlier been studied for patients with back and/or neck pain²⁶ and a
5
6 381 modified SBT have been tested to predict physical health outcome, using the SF-36⁵⁷ but this was the
7
8 382 first time the predictive validity of the SBT was studied for the outcomes of HRQoL and work ability for
9
10 383 individuals with both back and neck pain. Therefore this study widens the usefulness of the SBT
11
12 384 compared to earlier studies^{23 58-61}. There is also need for short questionnaires that are easy-to-use in
13
14 385 clinical to distribute and interpret, especially in primary care. The SBT is primarily designed as a
15
16 386 “stratified care tool” which involves targeting treatment to subgroups of patients based on their key
17
18 387 characteristics⁶² but in this study, we wanted to study if the SBT could predict the important outcomes
19
20 388 HRQoL and work ability when applied in an RCT of neck and back pain. In this study, the
21
22 389 physiotherapists did not target treatment based on SBT. However, we accept that some of the constructs
23
24 390 within the SBT may have been addressed by the intervention provided which may have affected SBTs
25
26 391 ability to predict the above mentioned outcomes. The results of this study suggest that the SBT can be
27
28 392 used as a prognostic tool in primary care for subgroup identification of acute/subacute back and/or neck
29
30 393 pain patients at risk of poor long-term HRQoL and/or work ability outcome. This information about
31
32 394 important risk factors may help clinicians in primary care to develop personalized treatment strategies
33
34 395 which are a priority in research⁶³. Future studies are required to investigate whether the implementation
35
36 396 of screening together with matched treatment pathways have an effect on HRQoL and work ability
37
38 397 outcomes for these patients.
39
40
41
42
43
44 398

47 399 **Authors Contributions**

48
49
50
51 400 All authors discussed the results and commented on the manuscript. MF, IP, KS and BG were
52
53 401 responsible for the study design, data analysis and interpretation. MF, BG and KS prepared and validated
54
55
56
57
58
59

1
2 402 data. MF collected data and drafted the manuscript. JH and CPS took part in study design, data analysis
3
4 403 and interpretation of data. All authors read and approved the final version of the manuscript.
5
6
7
8 404

11 405 **Acknowledgements**

15 406 Senior statistician Mikael Åström for statistical support, datamanager Chanchai Manuswin and the staff
16
17 407 at Data analysis and Register Centre, Region Skåne, Lund, Sweden, the research team at the Institute of
18
19 408 Primary Care and Health Sciences, Keele university, Stoke-on-Trent, UK and the staff at the Blekinge
20
21 409 Centre of Competence, Karlskrona, Sweden for help with support. All physiotherapists and patients
22
23 410 involved in this project, including the WorkUp project, REHSAM, in Region Skåne, Region Kronoberg
24
25
26 411 and Landstinget Blekinge, Sweden, for help with data collection.
27
28
29
30 412

33 413 **Data sharing statement**

37 414 The datasets analysed during the current study are available from the corresponding author on reasonable
38
39 415 request.
40
41
42
43 416
44
45
46
47 417
48
49
50
51 418
52
53
54
55 419
56
57
58
59
60

420

421 **References**

- 422 1. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and
423 injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010.[Erratum appears
424 in *Lancet*. 2013 Feb 23;381(9867):628 Note: AlMazroa, Mohammad A [added]; Memish, Ziad A [added]].
425 *Lancet (London, England)* 2012;380(9859):2163-96. doi: [http://dx.doi.org/10.1016/S0140-
426 6736\(12\)61729-2](http://dx.doi.org/10.1016/S0140-6736(12)61729-2)
- 427 2. Bergman S, Herrstrom P, Hogstrom K, et al. Chronic musculoskeletal pain, prevalence rates, and
428 sociodemographic associations in a Swedish population study. *J Rheumatol* 2001;28(6):1369-77.
429 [published Online First: 2001/06/21]
- 430 3. Gerdle B, Bjork J, Coster L, et al. Prevalence of widespread pain and associations with work status: a population
431 study. *BMC musculoskeletal disorders* 2008;9:102. doi: 10.1186/1471-2474-9-102 [published Online First:
432 2008/07/17]
- 433 4. Gustavsson A, Bjorkman J, Ljungcrantz C, et al. Socio-economic burden of patients with a diagnosis related to
434 chronic pain--register data of 840,000 Swedish patients. *Eur J Pain* 2012;16(2):289-99. doi:
435 10.1016/j.ejpain.2011.07.006 [published Online First: 2012/02/11]
- 436 5. Alexanderson K, Norlund A. . Aim, background, key concepts, regulations, and current statistics. *Scandinavian
437 journal of public health* 2004;32(63 suppl):12-30.
- 438 6. Kinge JM, Knudsen AK, Skirbekk V, et al. Musculoskeletal disorders in Norway: prevalence of chronicity and use
439 of primary and specialist health care services. *BMC musculoskeletal disorders* 2015;16:75. doi:
440 10.1186/s12891-015-0536-z [published Online First: 2015/04/19]
- 441 7. Jordan KP, Joud A, Bergknut C, et al. International comparisons of the consultation prevalence of
442 musculoskeletal conditions using population-based healthcare data from England and Sweden. *Ann
443 Rheum Dis* 2014;73(1):212-8. doi: 10.1136/annrheumdis-2012-202634 [published Online First:
444 2013/01/25]
- 445 8. Holtermann A, Hansen JV, Burr H, et al. Prognostic factors for long-term sickness absence among employees
446 with neck-shoulder and low-back pain. *Scand J Work Environ Health* 2010;36(1):34-41.
- 447 9. Nolet PS, Cote P, Kristman VL, et al. Is neck pain associated with worse health-related quality of life 6 months
448 later? A population-based cohort study. *Spine Journal: Official Journal of the North American Spine
449 Society* 2015;15(4):675-84. doi: <https://dx.doi.org/10.1016/j.spinee.2014.12.009>
- 450 10. Nolet PS, Kristman VL, Cote P, et al. Is low back pain associated with worse health-related quality of life 6
451 months later? *European Spine Journal* 2015;24(3):458-66. doi: [https://dx.doi.org/10.1007/s00586-014-
452 3649-4](https://dx.doi.org/10.1007/s00586-014-3649-4)
- 453 11. Leijon O, Wahlstrom J, Mulder M. Prevalence of self-reported neck-shoulder-arm pain and concurrent low
454 back pain or psychological distress: time-trends in a general population, 1990-2006. *Spine*
455 2009;34(17):1863-8. doi: <http://dx.doi.org/10.1097/BRS.0b013e3181ab3397>
- 456 12. Nyman T, Grooten WJ, Wiktorin C, et al. Sickness absence and concurrent low back and neck-shoulder pain:
457 results from the MUSIC-Norrtaalje study.[Erratum appears in *Eur Spine J*. 2007 May;16(5):639-40].
458 *European Spine Journal* 2007;16(5):631-8.
- 459 13. Pengel LH, Herbert RD, Maher CG, et al. Acute low back pain: systematic review of its prognosis. *BMJ (Clinical
460 research ed)* 2003;327(7410):323.
- 461 14. Dunn KM, Hestbaek L, Cassidy JD. Low back pain across the life course. *Best practice & research Clinical
462 rheumatology* 2013;27(5):591-600. doi: 10.1016/j.berh.2013.09.007 [published Online First: 2013/12/10]
- 463 15. Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of
464 general patient populations. *European spine journal : official publication of the European Spine Society,
465 the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*
466 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](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]

- 1
2 519 33. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Orebro Musculoskeletal Pain
3 520 Screening Questionnaire. *Spine* 2011;36(22):1891-5. doi:
4 521 <http://dx.doi.org/10.1097/BRS.0b013e3181f8f775>
5 522 34. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33(5):337-
6 523 43.
7 524 35. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-
8 525 208. [published Online First: 1990/11/05]
9 526 36. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37(1):53-72. [published Online First:
10 527 1996/06/06]
11 528 37. Greiner W, Weijnen T, Nieuwenhuizen M, et al. A single European currency for EQ-5D health states. *The*
12 529 *European Journal of Health Economics, formerly: HEPAC* 2003;4(3):222-31.
13 530 38. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108. [published Online
14 531 First: 1997/11/21]
15 532 39. Hansson E, Hansson T, Jonsson R. Predictors for work ability and disability in men and women with low-back
16 533 or neck problems. *European Spine Journal* 2006;15(6):780-93.
17 534 40. Tuomi K, Ilmarinen J, Jahkola A, et al. Work ability index, Finnish Institute of Occupational Health. 2: Helsinki,
18 535 1998.
19 536 41. Ilmarinen J, Tuomi K, Seitsamo J. New dimensions of work ability. *International Congress Series* 2005;1280:3-
20 537 7. doi: <http://dx.doi.org/10.1016/j.ics.2005.02.060>
21 538 42. Gould R, Ilmarinen J, Järvisalo J, et al. Dimensions of work ability. 2008
22 539 43. Ahlstrom L, Grimby-Ekman A, Hagberg M, et al. The work ability index and single-item question: associations
23 540 with sick leave, symptoms, and health--a prospective study of women on long-term sick leave. *Scand J*
24 541 *Work Environ Health* 2010;36(5):404-12. [published Online First: 2010/04/08]
25 542 44. El Fassi M, Bocquet V, Majery N, et al. Work ability assessment in a worker population: comparison and
26 543 determinants of Work Ability Index and Work Ability score. *BMC Public Health* 2013;13:305. doi:
27 544 10.1186/1471-2458-13-305 [published Online First: 2013/04/10]
28 545 45. Kirkwood B, Sterne J. Medical Statistics. 2nd: Oxford, England.: Blackwell Publishing, 2003.
29 546 46. Hosmer Jr D, Lemeshow S, Sturdivant RX. Introduction to the Logistic Regression Model. *Applied Logistic*
30 547 *Regression* 2000
31 548 47. Mehling WE, Avins AL, Acree MC, et al. Can a back pain screening tool help classify patients with acute pain
32 549 into risk levels for chronic pain? *Eur J Pain* 2015;19(3):439-46. doi: 10.1002/ejp.615 [published Online
33 550 First: 2014/11/11]
34 551 48. Cole T. Applied logistic regression. DW Hosmer and S. Lemeshow, Wiley, New York, 1989. : Wiley Online
35 552 Library 1991.
36 553 49. Melloh M, Elfering A, Kaser A, et al. What is the best time point to identify patients at risk of developing
37 554 persistent low back pain? *Journal of back and musculoskeletal rehabilitation* 2015;28(2):267-76. doi:
38 555 10.3233/bmr-140514 [published Online First: 2014/08/07]
39 556 50. Morso L, Kongsted A, Hestbaek L, et al. The prognostic ability of the STarT Back Tool was affected by episode
40 557 duration.[Erratum appears in *Eur Spine J*. 2016 Aug;25(8):2677-8; PMID: 27160824]. *European Spine*
41 558 *Journal* 2016;25(3):936-44. doi: <https://dx.doi.org/10.1007/s00586-015-3915-0>
42 559 51. Karran EL, McAuley JH, Traeger AC, et al. Can screening instruments accurately determine poor outcome risk
43 560 in adults with recent onset low back pain? A systematic review and meta-analysis. *BMC Med*
44 561 2017;15(1):13. doi: 10.1186/s12916-016-0774-4 [published Online First: 2017/01/20]
45 562 52. George SZ, Fritz JM, Bialosky JE, et al. The effect of a fear-avoidance-based physical therapy intervention for
46 563 patients with acute low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976)*
47 564 2003;28(23):2551-60. doi: 10.1097/01.brs.0000096677.84605.a2 [published Online First: 2003/12/04]
48 565 53. Stigmar KG, Petersson IF, Joud A, et al. Promoting work ability in a structured national rehabilitation program
49 566 in patients with musculoskeletal disorders: outcomes and predictors in a prospective cohort study. *BMC*
50 567 *musculoskeletal disorders* 2013;14:57. doi: 10.1186/1471-2474-14-57 [published Online First:
51 568 2013/02/07]
52 569 54. Burstrom K, Johannesson M, Diderichsen F. Swedish population health-related quality of life results using the
53 570 EQ-5D. *Qual Life Res* 2001;10(7):621-35. [published Online First: 2002/02/02]

- 1
2 571 55. Ahlstrom L, Grimby-Ekman A, Hagberg M, et al. The work ability index and single-item question: associations
3 572 with sick leave, symptoms, and health—a prospective study of women on long-term sick leave. *Scand J*
4 573 *Work Environ Health* 2010;404-12.
- 5 574 56. Lundin A, Leijon O, Vaez M, et al. Predictive validity of the Work Ability Index and its individual items in the
6 575 general population. *Scand J Public Health* 2017;45(4):350-56. doi: 10.1177/1403494817702759
7 576 [published Online First: 2017/04/08]
- 8 577 57. Hill JC, Afolabi EK, Lewis M, et al. Does a modified STarT Back Tool predict outcome with a broader group of
9 578 musculoskeletal patients than back pain? A secondary analysis of cohort data. *BMJ Open*
10 579 2016;6(10):e012445. doi: 10.1136/bmjopen-2016-012445 [published Online First: 2016/10/16]
- 11 580 58. Hill JC, Dunn KM, Main CJ, et al. Subgrouping low back pain: a comparison of the STarT Back Tool with the
12 581 Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 2010;14(1):83-9. doi:
13 582 <http://dx.doi.org/10.1016/j.ejpain.2009.01.003>
- 14 583 59. Wideman TH, Hill JC, Main CJ, et al. Comparing the responsiveness of a brief, multidimensional risk screening
15 584 tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its
16 585 parts. *Pain* 2012;153(11):2182-91.
- 17 586 60. Beneciuk JM, Robinson ME, George SZ. Subgrouping for patients with low back pain: A multidimensional
18 587 approach incorporating cluster analysis and the start back screening tool. *Journal of Pain* 2015;16(1):19-
19 588 30. doi: <http://dx.doi.org/10.1016/j.jpain.2014.10.004>
- 20 589 61. Fritz JM, Beneciuk JM, George SZ. Relationship between categorization with the STarT Back Screening Tool
21 590 and prognosis for people receiving physical therapy for low back pain. *Physical therapy* 2011;91(5):722-
22 591 32.
- 23 592 62. Foster NE, Hill JC, O'Sullivan P, et al. Stratified models of care. *Best practice & research Clinical rheumatology*
24 593 2013;27(5):649-61. doi: 10.1016/j.berh.2013.10.005 [published Online First: 2013/12/10]
- 25 594 63. Rozenberg S, Foltz V, Fautrel B. Treatment strategy for chronic low back pain. *Joint Bone Spine*
26 595 2012;79(6):555-59. doi: <http://dx.doi.org/10.1016/j.jbspin.2012.09.003>

596

597

598

599

600

601

602

603

604

605 **Figure legends**

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

608

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

614

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

619

620

621

622

623

624

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

625

For peer review only

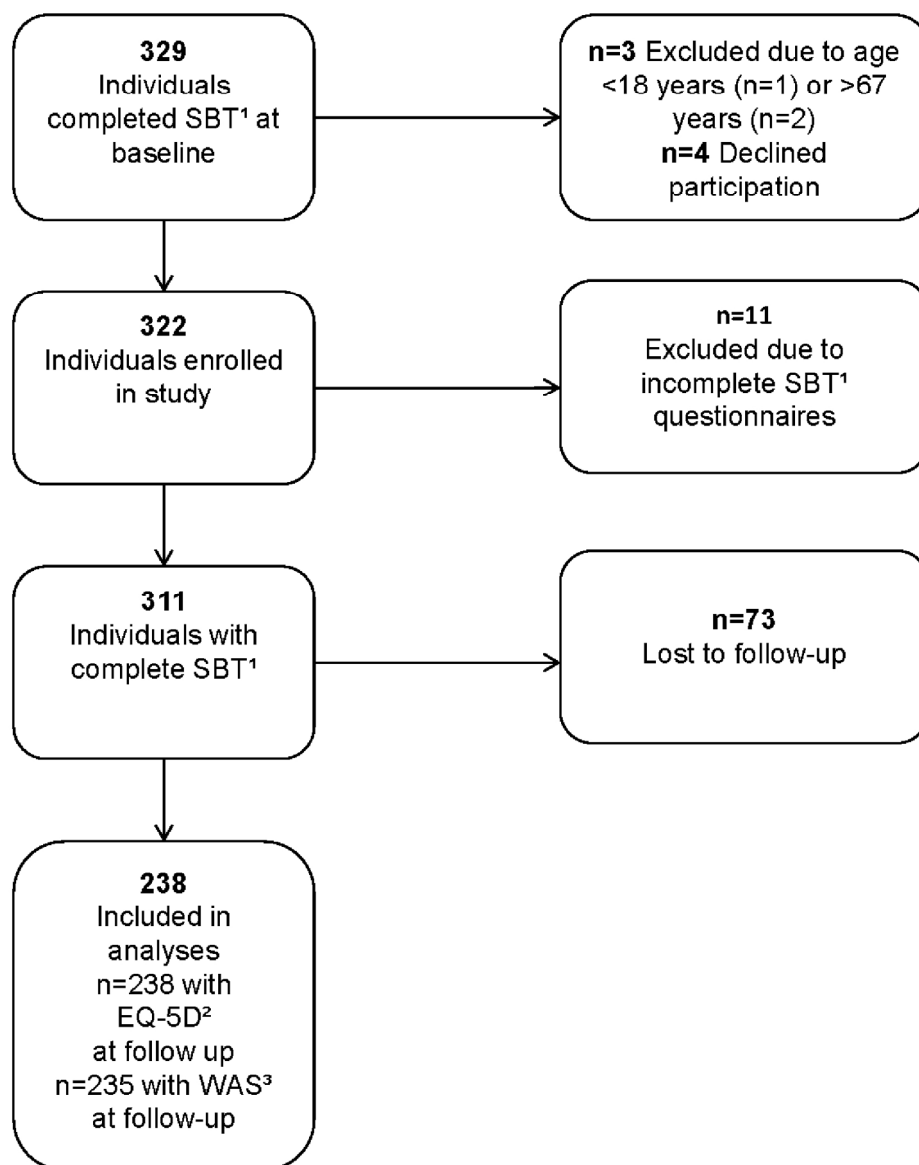
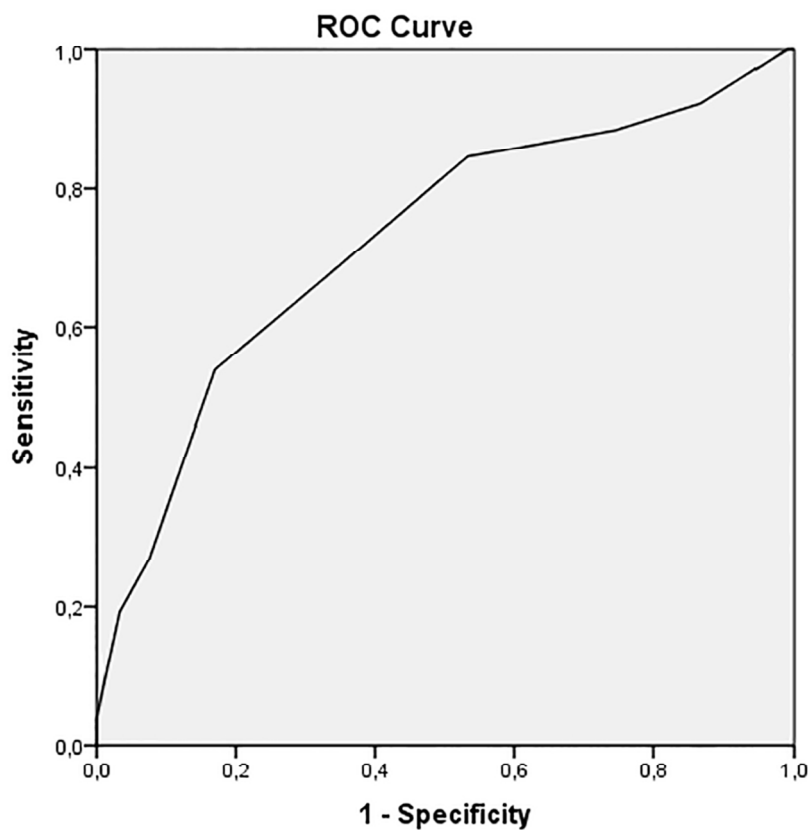


Fig.1. Flowchart of inclusion and exclusion of participants. ¹STarT Back Tool. ²EuroQol five-dimension. ³Work Ability Score.

151x181mm (300 x 300 DPI)



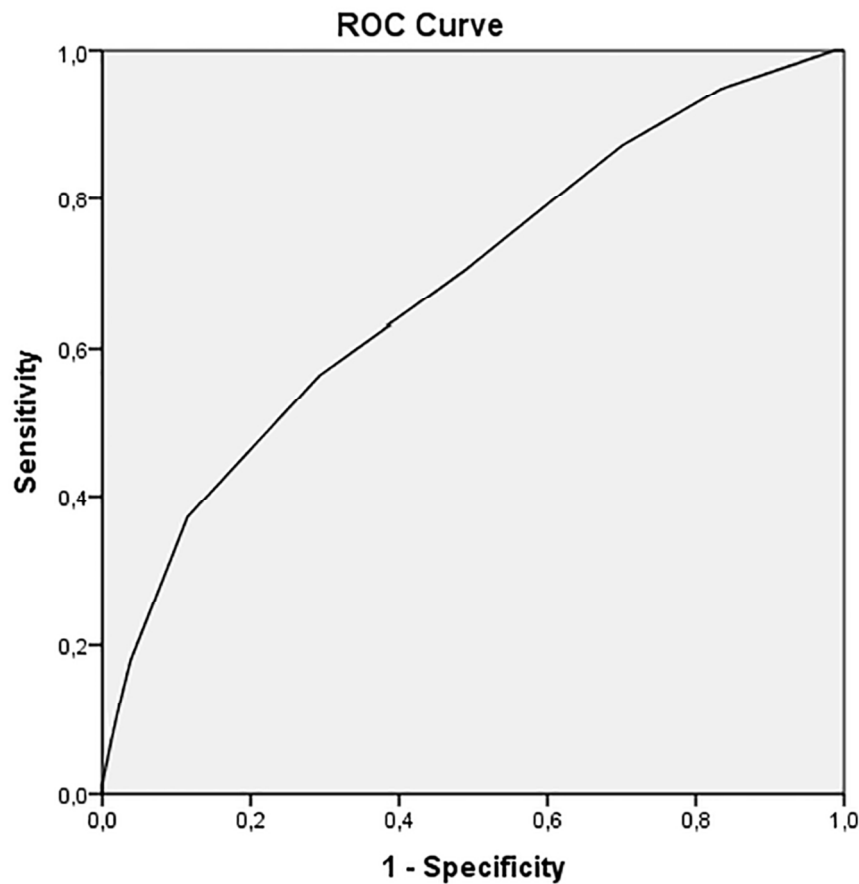
Diagonal segments are produced by ties.

35
36
37
38
39
40
41
42
43

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.

44
45
46
47
48
49

111x113mm (300 x 300 DPI)



Diagonal segments are produced by ties.

35
36
37
38
39
40
41
42
43

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.

44
45
46
47
48
49

98x99mm (300 x 300 DPI)