

The predictive and external validity of the STarT Back Tool in Danish primary care

Lars Morsø · Peter Kent · Hanne B. Albert ·
Jonathan C. Hill · Alice Kongsted · Claus Manniche

Received: 12 September 2012/Revised: 17 December 2012/Accepted: 25 January 2013/Published online: 10 February 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Purpose The STarT Back Tool (SBT) was recently translated into Danish and its concurrent validity described. This study tested the predictive validity of the Danish SBT. **Methods** Danish primary care patients ($n = 344$) were compared to a UK cohort. SBT subgroup validity for predicting high activity limitation at 3 months' follow-up was assessed using descriptive proportions, relative risks, AUC and odds ratios.

Results The SBT had a statistically similar predictive ability in Danish primary care as in UK primary care. Unadjusted relative risks for poor clinical outcome on activity limitation in the Danish cohort were 2.4 (1.7–3.4) for the medium-risk subgroup and 2.8 (1.8–3.8) for the high-risk subgroup versus 3.1 (2.5–3.9) and 4.5 (3.6–5.6) for the UK cohort. Adjusting for confounders appeared to explain the lower predictive ability of the Danish high-risk group.

Conclusions The Danish SBT distinguished between low- and medium-risk subgroups with a similar predictive

ability of the UK SBT. That distinction is useful information for informing patients about their expected prognosis and may help guiding clinicians' choice of treatment. However, cross-cultural differences in the SBT psychosocial subscale may reduce the predictive ability of the high-risk subgroup in Danish primary care.

Keywords Classification · Predictive value of tests · Validation · Low back pain · STarT Back Tool

Introduction

The STarT Back Screening Tool (SBT) has shown promise in triaging non-specific low back pain (NSLBP) patients in primary care [1, 2]. In the (UK), the SBT has shown an ability in primary care to identify modifiable prognostic factors, classify people into prognostic subgroups [1] and improve patient outcomes through subgroup-matched treatment pathways [3].

The SBT was recently translated into Danish, using best-practice translation methods [4, 5]. The result was a linguistically accurate and culturally acceptable tool with adequate discriminative validity to be used in Danish primary care [6]. However, the predictive and external validity of SBT in Danish primary care has not been established [6] and therefore there is a need for those aspects of the SBT's measurement properties to be described [7].

There is an increasing focus on questionnaires that index prognostic risk in low back pain (LBP), such as the SBT [1] and the Orebro Musculoskeletal Pain Questionnaire [8]. Questionnaires that can be used in routine care settings require a trade-off between brevity, simplicity and precision [9] and adequate validation is important if clinicians are to be confident in their use. Recognising these

L. Morsø (✉) · P. Kent · H. B. Albert · C. Manniche
Research Department, Spine Centre of Southern Denmark,
Hospital Lillebaelt, Institute of Regional Health Services
Research, University of Southern Denmark, A Member
of the Clinical Locomotion Network, Oestre Hougvej 55,
Middelfart 5500, Denmark
e-mail: lars.morsoe@slb.regionyddanmark.dk

J. C. Hill
Arthritis Research UK Primary Care Centre, Keele University,
Keele, Staffordshire, UK

A. Kongsted
Nordic Institute of Chiropractic and Clinical Biomechanics,
A Member of the Clinical Locomotion Network, Odense,
Denmark

challenges, Justice et al. [10] proposed a multi-layered approach to external validation that includes a focus on reproducibility and transportability, and it has been recommended that validity testing occurs in multiple and setting-specific samples [11].

Validation of the predictive ability of the SBT in the cultural context of Danish primary care would determine whether prognostic stratification based on SBT subgroups is similar to that in the UK, and inform the implementation of this screening tool in Danish daily healthcare practice [7, 12]. Therefore, the aim of this study was to compare the predictive validity of the Danish version of the SBT in Danish primary care to the English version of the SBT in UK primary care.

Methods

Patient groups

Two Danish patient samples from general medical practice (GP) and physiotherapy primary care clinics were pooled into a cohort representing Danish primary care (Danish cohort). The analyses were performed on this cohort and compared to an existing NSLBP cohort from UK primary care (UK cohort) [13]. In both cohorts, almost all patients were initially triaged by GPs, some completing their study questionnaires during or after that consultation and others at physiotherapy practices following referral.

The Danish GP sample was prospectively recruited as part of a GP audit conducted by the regional health authority from February to May 2011, while prospective data collection from 27 Danish physiotherapy clinics occurred from May to September 2011 until 200 baseline questionnaires were collected. There was a 72 % (GP) and 86 % (physiotherapy) follow-up at 3 months. As the physiotherapy sample was the smallest ($n = 172$), the total Danish cohort ($n = 344$) was formed by adding 172 patients randomly selected from the GP cohort to evenly weight the cohort across both professional disciplines. This balanced mix of GP and physiotherapy patients was to improve the generalisability of the results, as recommended by de Vet et al. [11]. A detailed flowchart for the Danish cohort is shown in Appendix 1.

The inclusion criteria were people 18–65 years of age with NSLBP identified either by: (1) specific diagnostic coding recorded in GP electronic patient records, or (2) by physiotherapists using the criteria contained in the European guidelines for NSLBP in primary care [14]. Participants completed questionnaires on basic demographic details, fear of movement (Tampa Scale of Kiniesophobia) [15], catastrophisation (Coping strategy questionnaire) [16], anxiety and depression (Hospital Anxiety and

Depression Scale) [17] and the SBT [1]. Data from both settings were independently double-entered into a database (Epidata 3.1, The EpiData Association, Odense, Denmark) by two research secretaries.

The UK data were from the BeBack Study [13] conducted by the Arthritis Research UK Primary Care Centre, Keele University, from which baseline and 3-month follow-up questionnaire scores were extracted. The study was a prospective cohort of consecutive patients, from a socioeconomically heterogeneous population, who consulted with low back pain in eight general practices in England. Six months' follow-up data from this cohort have previously been reported and the current study uses the full 3 months' follow-up data ($n = 856$) available from that cohort [13]. Details of the recruitment and data collected have also been previously published [13].

Three-month outcomes were chosen for our study as this has been shown to be the most important time point for the clinical course of LBP in primary care, marking the end of rapid improvement and heralding the onset of persistent pain [18].

Data analysis

Previous studies have tested the predictive validity of questionnaires using a variety of methods [19, 20]. In the current study, the predictive validity and external validity of the SBT met the criteria proposed by Justice et al. [10] for comparing cohorts across countries and at a different outcome time points than from that previously studied (in our case, 3 months in the current study and 6 months in the original SBT study in the UK).

Descriptive analysis of the baseline characteristics of the Danish and the UK cohort was performed (means and standard deviations, medians and inter-quartile ranges). Baseline differences between the two cohorts were examined using Mann–Whitney *U*, Chi-square or Kruskal–Wallis Tests, depending on the data type and distribution.

The three statistical methods that had been used to describe the predictive of the SBT in the original UK validation study [1] were mirrored in our study. The outcome measures used were all threshold scores on measures of activity limitation, pain and pain bothersomeness. These were the constructs chosen in the original study due to their being recommended in expert consensus statements [21]. We replicated these threshold scores so that results could potentially be compared across studies and time points.

Firstly, comparison was made between the proportions of patients with a poor clinical outcome on activity limitation at 3 months in both cohorts, stratified by SBT subgroup. Poor clinical outcome was defined in this context as a Roland Morris disability questionnaire (RMDQ) [22] sum score (0–100 scale) of 30 points or more [23] at 3 months'

follow-up. The cut point used in original SBT development study in the UK [1] was 7 on a 0–24 scale but as we used the proportional recalculation method to convert all RMDQ scores to a 0–100 scale, that threshold was recalculated to be 30 points or more. The proportional recalculation method has been shown to be more accurate in managing any missing RMDQ answers [23].

Secondly, for both cohorts, the same outcome was used to estimate the additional risk (relative risk) [24] for poor outcome resulting for people in the medium or high SBT risk subgroup compared to the low-risk subgroup.

Thirdly, the area under the curve (AUC) statistic from receiver operating characteristic (ROC) curves was used to describe the ability of the baseline SBT sum scores (0–9 scale) to discriminate (sensitivity/1–specificity) [24] between people with and people without the following outcomes at 3 months: (1) poor outcome on activity limitation as defined above, (2) LBP still being ‘severe’ (8–10 on a 0–10 point scale), and (3) LBP rated as ‘very’ or ‘extremely’ bothersome on a 5-point pain bothersomeness scale. All these criteria were used in the original UK validation study [1].

In addition, as the proportions of patients with a poor clinical outcome on activity limitation and unadjusted relative risks suggested that the psychosocial subscale of the Danish SBT might not have the same predictive ability as the UK version, logistic regression was performed to explore for potential confounding. Due to suspected confounding by treatment exposure (approximately 60 % of the Danish cohort was referred for physiotherapy and approximately 18 % in the UK cohort) and by differential treatment effectiveness at modifying psychosocial risk factors (treatment being heterogeneous), these covariates were further explored. Adjusted odds ratios controlled for Danish care setting (GP and physiotherapy) and change in SBT psychosocial subscale risk factors (fear of movement, catastrophisation, anxiety, depression and pain bothersomeness), measured by their full reference standard questionnaires. An odds ratio greater than 1 in these regression models means that particular clinical characteristic increases the odds of having a poor outcome and an odds ratio less than 1 means that it is protective against a poor outcome. All covariates were initially entered into the model and reported, followed by a manual backwards stepwise reduction ($p < 0.05$ to remove) to the most parsimonious model.

Relative risk estimates and random number generation were performed using Microsoft Excel 2003 (Microsoft Corp, Redmond, WA, USA). Logistic regression was performed using STATA 12 (StataCorp, College Station, TX, USA). All other statistical analyses were conducted using PASW 13.0 (IBM Inc., Somers, NY, USA).

Results

Baseline differences between the Danish and the United Kingdom cohorts

The two cohorts were significantly different on a number of baseline characteristics (Table 1). On an overall cohort level, the Danish cohort reported higher pain intensity, higher activity limitation, slightly more prevalent leg pain and slightly higher catastrophisation. There were also significant differences between the two cohorts in the distribution of people across the three SBT groups, with a lower proportion of ‘low risk’ patients and a higher proportion of ‘high risk’ patients in the Danish cohort.

Comparison of the unadjusted risk of poor clinical outcome on activity limitation at 3 months

Overall 47 % in the Danish cohort and 36 % in the UK cohort had poor outcome at 3 months. Reassuringly, in both the Danish and UK cohort, the proportion of patients with a poor outcome was lowest in the low-risk subgroup and highest in the high-risk subgroup (Fig. 1). This is also reflected in the relative risks of the medium-risk and high-risk subgroups. Although the proportions and relative risks vary between the cohorts, the gradient in the trend line across risk subgroups was similar, indicating that predictive ability of the SBT broadly followed a comparable pattern in both countries.

At an SBT subgroup level, these unadjusted data suggest that the Danish high-risk subgroup does not have the same incremental step size in predictive validity compared with the median-risk subgroup, as in the UK cohort. While the proportion of patients with a poor outcome was quite similar in the low-risk subgroups (Danish 24 %; UK 17 %) and medium-risk subgroups (Danish 57 %; UK 54 %), it was considerably lower in the Danish high-risk subgroup (64 %) compared to that in the UK cohort (78 %). As a consequence, while the relative risks (RR) for the medium-risk subgroup are comparable across cohorts, there was only a marginal step up to the high-risk subgroup in the Danish cohort, whereas the step up was greater in the UK cohort (Fig. 1). As the distinction between the medium- and high-risk subgroups is that higher scores on the SBT psychosocial subscale are required to be classified as being high-risk, these unadjusted results could initially be interpreted as suggesting that the psychosocial subscale does not have the same predictive validity in the Danish cohort. Raising the threshold score on the subscale from 4 to 5 had almost no effect of the predictive strength of the Danish high-risk subgroup (from to RR 2.7 [1.8; 3.8] to RR 2.8 [1.8; 4.4]). Therefore, logistic regression was performed to control for potential confounding.

Table 1 Baseline characteristics of the Danish and UK primary care cohorts

	Danish cohort (<i>n</i> = 344)	UK cohort (<i>n</i> = 856)	Tests for differences between the Danish and UK cohorts*
Age in years			
Median, (IQR ^a)	50.0 (41–59)	46 (39–53)	<i>p</i> < 0.001
Female	199 (57.8 %)	305 (58.8 %)	<i>p</i> = 0.772
Duration (%)			
<4 weeks	149 (44.2 %)	327 (38.2 %)	<i>p</i> = 0.556
4–12 weeks	66 (19.6 %)	221 (25.8 %)	
>12 weeks	122 (36.2 %)	285 (33.3 %)	
STarT subgroup, proportions			
Low	121 (37.5 %)	460 (53.7 %)	<i>p</i> < 0.001
Medium	127 (39.3 %)	295 (34.5 %)	
High	75 (23.2 %)	90 (10.5 %)	
Pain intensity ^b (0–10 scale) ^c			
Overall median (IQR ^a)	7 (5–8)	5 (3–7)	<i>p</i> < 0.001
Mild (0–5)	130 (38.7 %)	527 (61.6 %)	
Moderate (6–7)	98 (29.2 %)	196 (22.9 %)	
Severe (8–10)	108 (32.1 %)	127 (14.8 %)	
Presence of referred leg pain	241 (72.2 %)	518 (60.5 %)	<i>p</i> < 0.001
Presence of comorbid neck or shoulder pain	155 (46.8 %)	463 (54.1 %)	<i>p</i> = 0.014
Activity limitation ^d (0–100 scale) ^e			
Median (IQR ^a)	60.9 (39–77)	33.3 (17–54)	<i>p</i> < 0.001
Pain bothersomeness			
Not at all	2 (0.6 %)	29 (3.4 %)	<i>p</i> < 0.001
Slightly	14 (4.2 %)	108 (12.6 %)	
Moderately	82 (24.7 %)	235 (27.5 %)	
Very much	178 (53.6 %)	307 (35.9 %)	
Extremely	56 (16.6 %)	166 (19.4 %)	
Fear of movement ^e (17–68 scale) ^c			
Median (IQR ^a)	36 (30–41)	40 (36–43)	<i>p</i> < 0.001
Catastrophisation ^g (0–36 scale) ^c			
Median (IQR ^c)	10 (5–15.8)	9 (4–14)	<i>p</i> = 0.009
Anxiety ^g (0–21 scale) ^c			
Median (IQR ^c)	5 (3–8)	8 (5–11)	<i>p</i> < 0.001
Depression ^g (0–21 scale) ^c			
Median (IQR ^c)	2 (1–5)	6 (3–9)	<i>p</i> < 0.001

* Tests for differences were the Man–Whitney *U*, Chi-square or Kruskal–Wallis procedures, depending on data type and distribution

^a Inter-quartile range

^b Numeric Rating Scale (0–10)

^c High scores are worse

^d Roland Morris disability questionnaire

^e Tampa Scale of Kinesophobia

^f Coping strategy questionnaire (catastrophisation subscale)

^g Hospital Anxiety and Depression Scale

Comparison of the adjusted risk of poor clinical outcome on activity limitation at 3 months

The unadjusted odds ratios (OR) in Table 2 mirror the distinction between the Danish and UK cohorts seen in the relative risk results. The predictive ability of the medium-risk subgroups was similar (Danish OR 4.2 [2.5; 7.3], UK OR 5.6 [4.0; 7.8]), whereas in the Danish cohort the high-risk subgroup added only a little predictive information (OR 5.6 [3.0; 10.5]) and was much more predictive in the UK cohort (OR 16.9 [9.7; 29.3]).

Adjustment for care setting and change scores in the psychosocial constructs resulted in the predictive ability of

the Danish ‘high risk’ group (adjusted OR 15.9 [5.2; 48.2]) approximating that of the UK cohort, when the regression model included all covariates. The parsimonious model only retained four covariates (care setting, change in anxiety, change in pain bothersomeness and the interaction between change in pain bothersomeness and care setting) with the predictive ability of the Danish ‘high risk’ group being almost identical to that observed in the UK cohort in this model (adjusted OR 15.7 [6.6; 37.5]). There was no significant (*p* < 0.05) non-linearity in the covariates included in the adjusted models (regression model not reported).

As there was no significant interaction between the SBT subgroups and care setting (regression model not

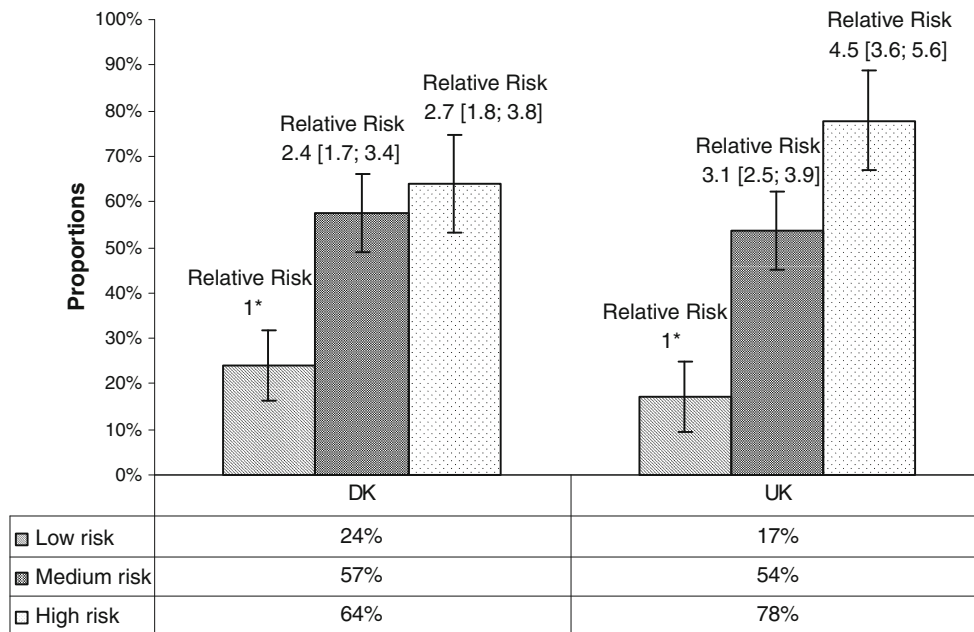


Fig. 1 Proportions of patients within each STarT Back Tool subgroup who had a poor clinical outcome on activity limitation (Poor clinical outcome defined as a Roland Morris disability

questionnaire score >30) at 3 months and their relative risk. Asterisk the low-risk STarT Back Tool subgroup is the reference category

reported), and care setting was not changed by study participation, we interpret the reduced odds (0.31 [0.13; 0.71] parsimonious model) of a poor outcome in the physiotherapy patients as evidence of confounding [25, 26]. As the psychosocial covariates may have changed as result of treatment and/or natural history, we believe they may be on the causal pathway and interpret their effect on altering risk as evidence of effect mediation [25, 26]. It was not possible to calculate adjusted ORs for the UK cohort as the same covariate data were not available.

Comparison of the ability of the total baseline SBT scores to identify people with outcomes above a clinical threshold at 3 months

The AUC statistics describing the ability of the baseline SBT scores (0–9 scale) to discriminate between people with and people without scores above threshold values on three different 3-month outcomes are shown in Table 3. For the outcomes of LBP ‘still being severe’ and LBP rated as ‘very or extremely bothersome’ the discriminative ability was similar across cohorts. However, for the outcome of a RMDQ score above 30 points (0–100 scale) the difference was more substantial (Danish AUC 0.71 [0.66; 0.77], UK AUC 0.81 [0.78; 0.84]).

Discussion

The aim of this study was to compare the predictive of the Danish version of the SBT in Danish primary care and the English version of the SBT in UK primary care. There were baseline differences between the cohorts, which reflect their being from different health care systems and cultures. However, this study did not aim to match the cohorts but to include samples that were likely to be representative of their clinical populations and to compare the SBT predictive ability in each cohort. Overall, the results of the current study indicate that the ability of SBT to predict increased risk of poor prognosis at 3 months in Danish primary care was similar to that seen in UK primary care for the low- and medium-risk SBT subgroups, whereas we initially observed almost no difference between the predictive strength of the medium- and high-risk subgroups in the Danish cohort. However, there was a very large difference between the cohorts in exposure to physiotherapy treatment and we found that after adjusting for this confounding [26] and also for significant effect mediation [26] due to change in two psychosocial characteristics, the predictive strength of the high-risk Danish subgroup was almost identical to the UK cohort (unadjusted estimate). Data were not available to perform adjusted analysis in the UK cohort. Whether this asymmetric analysis (Danish adjusted predictive estimates and UK unadjusted estimates)

Table 2 The odds of having poor clinical outcome on activity limitation (Roland Morris disability questionnaire score >30 (0–100 scale)) at 3 months by STarT Back Tool subgroup the Danish and UK cohorts, estimated using logistic regression

	Danish cohort (<i>n</i> = 322)		UK cohort (<i>n</i> = 845)	
	Odds ratio [CI 95 %]	<i>p</i> value	Odds ratio [CI 95 %]	<i>p</i> value
Unadjusted model				
STarT Back Tool low-risk subgroup ^a	1.00		1.00	
STarT Back Tool medium-risk subgroup	4.24 [2.45; 7.32]	<0.001	5.56 [3.99; 7.76]	<0.001
STarT Back Tool high-risk	5.57 [2.97; 10.47]	<0.001	16.88 [9.71; 29.34]	<0.001
Constant	0.32 [0.21; 0.48]	<0.001	0.22 [0.16; 0.26]	<0.001
Full model adjusted for care setting and change on STarT Back Tool psychosocial constructs (<i>n</i> = 213)				
STarT Back Tool low-risk subgroup ^a	1.00			
STarT Back Tool medium-risk subgroup	8.16 [3.44; 19.30]	<0.001		
STarT Back Tool high-risk	15.85 [5.22; 48.17]	<0.001		
Included covariates				
Care setting ^b	0.36 [0.13; 1.01]	0.052		
Change in fear of movement ^c	0.99 [0.90; 1.09]	0.879		
Change in catastrophisation ^d	0.87 [0.78; 0.97]	0.013		
Change in anxiety ^e	0.81 [0.65; 1.02]	0.078		
Change in depression ^e	1.03 [0.81; 1.34]	0.077		
Change in pain bothersomeness ^f	0.36 [0.19; 0.65]	<0.001		
Interaction between care setting and change in fear of movement	0.92 [0.81; 1.04]	0.195		
Interaction between care setting and change in catastrophisation	1.14 [1.00; 1.13]	0.055		
Interaction between care setting and change in anxiety	1.17 [0.88; 1.55]	0.284		
Interaction between care setting and change in depression	0.77 [0.55; 1.09]	0.142		
Interaction between care setting and change in pain bothersomeness	1.89 [0.90; 3.97]	0.092		
Constant	1.04 [0.42; 2.58]	0.933		
Parsimonious model adjusted for care setting and change on STarT Back Tool psychosocial constructs (<i>n</i> = 296), using manual backwards stepwise procedure				
STarT Back Tool low-risk subgroup ^a	1.00			
STarT Back Tool medium-risk subgroup	7.89 [3.87; 16.11]	<0.001		
STarT Back Tool high-risk	15.73 [6.60; 37.47]	<0.001		
Care setting ^b	0.31 [0.13; 0.71]	0.006		
Change in anxiety ^c	0.81 [0.73; 0.89]	<0.001		
Change in pain bothersomeness ^f	0.27 [0.17; 0.43]	<0.001		
Interaction between care setting and change in pain bothersomeness	2.48 [1.41; 4.34]	0.002		
Constant	1.02 [0.49; 2.15]	0.951		

Bold values indicate significant level of $p < 0.05$

^a Reference value

^b Patient recruitment in the Danish physiotherapy setting compared with the GP setting (reference value)

^c Tampa Scale of Kiniesophobia

^d Coping strategy questionnaire (catastrophisation subscale)

^e Hospital Anxiety and Depression Scale

^f Bothersome question

appropriately explains the differences between cohorts in SBT performance requires further discussion.

One potential reason for this difference in risk prediction could have been that the psychosocial subscale that classifies people as high-risk, rather than medium-risk, was not as precise in the Danish population. This could have been

an influence, as the concurrent validation study of the Danish translation showed that, compared to the original UK cohort, the discriminative ability of three of the psychosocial subscale questions was less strong [6]. On average, the association between a 'yes' response on a psychosocial subscale question and scores on their

Table 3 Discriminative ability of total baseline scores on the STarT Back Tool to correctly classify people with high scores on three different dichotomised outcomes at 3 months follow-up

	Danish primary care cohort AUC [95 % CI]	UK primary care cohort AUC [95 % CI]
People with a Roland Morris disability questionnaire score >30 at 3 months	0.71 [0.66; 0.77]	0.81 [0.78; 0.84]
People with severe back pain at 3 months (8–10 on a 0–10 scale)	0.79 [0.68; 0.89]	0.81 [0.78; 0.84]
People with ‘very’ or ‘extremely’ bothersome pain at 3 months	0.70 [0.64; 0.76]	0.72 [0.68; 0.76]

AUC area under the curve statistic from receiver operating characteristic (ROC) curves

respective full reference standard questionnaires had an AUC 0.115 less than in the UK cohort. These differences were believed to be due to disparity in severity between the cohorts (the Danish cohort in that study being more severe and chronic) and the Danish reference standard questionnaires not having been validated [6]. However, imprecision in the psychosocial subscale, and/or cultural differences in the influence of psychosocial factors on outcome, cannot be ruled out as explanatory influences on the predictive ability (unadjusted analysis) of the Danish high-risk SBT subgroup.

Another reason for the difference in (unadjusted) risk prediction observed in the current study for the high-risk subgroup could have been the difference between the cohorts in exposure to physiotherapy treatment (approximately 60 % of the Danish cohort and approximately 18 % of the UK cohort). Exposure to physiotherapy was a confounder, as patients in the physiotherapy group had a substantially lower risk of poor outcome than in the GP group (OR 0.31 [0.13; 0.71]) and that effect was the same across SBT subgroups because there was no significant interaction between care setting and SBT groups.

A further reason for the difference in (unadjusted) risk prediction could have been due to differential treatment effectiveness at modifying psychosocial risk factors, in either the GP or physiotherapy care settings. For example, the physiotherapy treatment was not targeted to SBT subgroup and was likely to be heterogeneous. There was evidence to support this effect mediation, as change in two psychosocial constructs (anxiety and pain bothersomeness) were significant in the adjusted models, and there was a significant interaction between change in pain bothersomeness and care setting. These data do not allow the distinction between treatment effects and change due to natural history, but the interaction between change in pain bothersomeness and care setting is suggestive of a differential treatment effect.

There is some evidence that such a differential effect is explanatory of a weakening of the SBT predictive ability in unadjusted analysis. Secondary analysis of data from the UK randomised controlled trial [3] comparing the

predictive ability of the SBT groups in the targeted treatment group to that in the usual care group showed that the predictive ability was reduced by effective treatment (unpublished data). Put simply, when treatment is effective, the predictive ability of the SBT is reduced due to the natural history of the condition being modified.

Direct comparison of these results with those of other translations is not possible, as this degree of validation has not been published for other versions. However, our results are comparable to those found in a USA validation study using the English language version [27].

Strengths of this study are the rigour of the validation method, the ability to compare results across both cultures, and the Danish cohort consisting of most professions commonly consulted for back pain. A weakness of this study was the inability to perform adjusted analyses in the UK cohort. An additional consideration is that as the cohorts were different at baseline, some of the differences in SBT predictive validity might be due to factors other than change in the psychosocial characteristics. We did not explore these, as the only substantive differences between the cohorts were in the predictive ability of the high-risk (psychosocial subscale) subgroups and those differences were explained by adjusting for psychosocial change. However, differences between countries/settings are to be expected and are an appropriate reason for testing the SBT in different cohorts to determine whether the predictive ability of the SBT is robust to these differences.

Conclusion

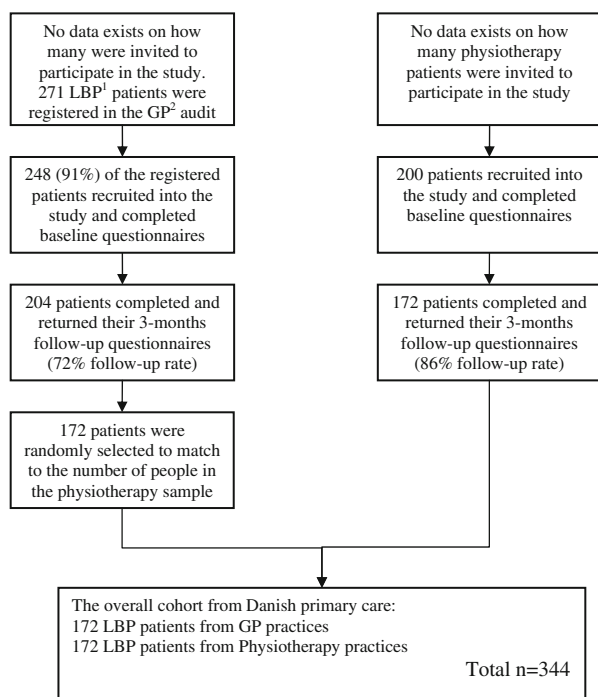
In conclusion, based on previous results from the concurrent validation study [6] and these current results on predictive ability, the SBT is suitable as a triage tool of LBP patients in Danish primary care. The Danish SBT distinguished between low- and medium-risk subgroups with a similar predictive ability to the UK SBT. That distinction is useful information for informing patients about their expected prognosis and may help guiding clinicians' choice of treatment. However, cross-cultural differences in

the SBT psychosocial subscale may reduce the predictive ability of the high-risk subgroup in Danish primary care. Whether SBT subgroup-matched treatment pathways are as effective in the Danish population as in the UK requires subsequent research.

Acknowledgments The authors thank the Danish Quality Unit of General Practice, the GPs and physiotherapists for collecting data, and the research secretaries at the Spine Centre of Southern Denmark and NIKKB for assistance in handling that data. The authors are also grateful for funding received by the Region of Southern Denmark and the University of Southern Denmark.

Conflict of interest None.

Appendix 1. Formation of the Danish primary care cohort



¹LBP = low back pain.

²General practitioner.

References

- Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE et al (2008) A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthr Rheum* 59(5): 632–641
- Hill JC, Foster NE, Hay EM (2010) Cognitive behavioural therapy shown to be an effective and low cost treatment for subacute and chronic low-back pain, improving pain and disability scores in a pragmatic RCT. *Evid Based Med* 15(4):118–119
- Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE et al (2011) Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet* 378(9802):1560–1571
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 25(24):3186–3191
- Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S et al (1998) Translating health status questionnaires and evaluating their quality: the IQOLA project approach. *International quality of life assessment. J Clin Epidemiol* 51(11):913–923
- Morso L, Albert H, Kent P, Manniche C, Hill J (2011) Translation and discriminative validation of the STarT Back Screening Tool into Danish. *Eur Spine J* 20(12):2166–2173
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
- Linton SJ, Boersma K (2003) Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain* 19(2):80–86
- Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 159(9):882–890
- Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. *Ann Intern Med* 130(6): 515–524
- de Vet HCW, Terwee CB, Mokkink LB, Knol DL (2011) *Measurement in medicine*, 1st edn. Cambridge University Press, New York, pp 150–201
- Altman DG, Vergouwe Y, Royston P, Moons KG (2009) Prognosis and prognostic research: validating a prognostic model. *BMJ* 28(338):b605
- Foster NE, Bishop A, Thomas E, Main C, Horne R, Weinman J et al (2008) Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome? *Pain* 136(1–2):177–187
- van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A et al (2006) Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 15(Suppl 2):S169–S191
- Swinkels-Meewisse EJ, Swinkels RA, Verbeek AL, Vlaeyen JW, Oostendorp RA (2003) Psychometric properties of the Tampa Scale for kinesiophobia and the fear-avoidance beliefs questionnaire in acute low back pain. *Man Ther* 8(1):29–36
- Rosenstiel AK, Keefe FJ (1983) The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain* 17(1):33–44
- Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. *Acta Psychiatr Scand* 67(6):361–370
- Pengel LH, Herbert RD, Maher CG, Refshauge KM (2003) Acute low back pain: systematic review of its prognosis. *BMJ* 327(7410):323
- Cleland JA, Fritz JM, Brennan GP (2008) Predictive validity of initial fear avoidance beliefs in patients with low back pain receiving physical therapy: is the FABQ a useful screening tool for identifying patients at risk for a poor recovery? *Eur Spine J* 17(1):70–79
- Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majoowski GR et al (2004) A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med* 141(12): 920–928

21. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B et al (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976)* 23(18): 2003–2013
22. Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 8(2):141–144
23. Kent P, Lauridsen HH (2011) Managing missing scores on the Roland Morris disability questionnaire. *Spine (Phila Pa 1976)* 36(22):1878–1884
24. Kirkwood BRSJAC (1988) Measurement error: assessment and implications, *Essential Medical Statistics*, 2nd edn. Blackwell Science Ltd., Oxford, pp 429–446
25. Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D (2001) How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry* 158(6):848–856
26. Kraemer HC, Wilson GT, Fairburn CG, Agras WS (2002) Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 59(10):877–883
27. Fritz JM, Beneciuk JM, George SZ (2011) Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Phys Ther* 91(5):722–732