

The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting

Lars Morsø · Peter Kent · Claus Manniche ·
Hanne B. Albert

Received: 18 February 2013/Revised: 16 May 2013/Accepted: 7 June 2013/Published online: 21 June 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Introduction The predictive ability of the STarT Back Tool (SBT) in secondary care settings has not been investigated. The aim of this study was to determine the SBT's predictive ability in a Danish secondary care setting and compare this to a Danish primary care setting.

Methods Poor clinical outcome at 6 months (>30 points on a 0–100 Roland Morris Disability Scale) was calculated in secondary care ($n = 960$) and primary care ($n = 172$) cohorts. The cohorts were stratified into SBT subgroups and estimates of additional risk for poor outcome were calculated [relative risk (RR), unadjusted and adjusted odds ratios]. The discriminative ability was determined using the area under the curve statistic.

Results In secondary care 69.0 % and in primary care 40.2 % had poor outcome on activity limitation. Although significant, the predictive ability of the SBT in secondary care (medium-risk RR 1.5, high-risk RR 1.7) was not as strong as in primary care (medium-risk RR 2.3, high-risk RR 3.5). Adjusting for episode duration and pain intensity only changed the predictive ability marginally in secondary care. The discriminative ability of the SBT was similar in both cohorts despite differences in the predictive ability.

Conclusion The SBT had less predictive ability in a Danish secondary care setting compared to a Danish primary care setting for persistent activity limitation at

6 months follow-up. SBT-targeted treatment implications in secondary care were not investigated in this study.

Keywords STarT Back Tool · Predictive ability · Secondary care · Targeted treatment

Introduction

The STarT Back Tool (SBT) is a screening tool for non-specific low back pain (LBP) that has been validated in primary care [1]. On the basis of potentially modifiable prognostic factors, the SBT classifies people into prognostic subgroups and identifies targeted treatment pathways for those subgroups [1]. The construct, concurrent and predictive validity of the SBT has been investigated [2–4]; it has been translated into several languages [5–7] and its cross-cultural validity and cross-cultural predictive ability have been described [7, 8]. In addition, a high-quality randomised controlled trial that matched treatment pathways to each SBT subgroup showed improved patient outcomes and cost-effectiveness [9]. Overall, this evidence suggests that the SBT can provide important prognostic information in primary care settings [8], changes in SBT overall scores may provide important clinical decision-making information for treatment monitoring [2], SBT-targeted treatment can be effective [10], and the SBT is well accepted by primary care clinicians. In addition, research is being undertaken into appropriate implementation strategies for the SBT in primary care [11].

The SBT was developed for primary care and consequently the majority of research regarding the SBT has been performed in primary care settings. However, evidence as to whether the SBT is appropriate for use in secondary care settings has not been established. In

L. Morsø (✉) · P. Kent · C. Manniche · H. B. Albert
Research Department, Spine Centre of Southern Denmark,
Oestre Hougvej 55, 5500 Middelfart, Denmark
e-mail: lars.morsoe@slb.regionyddanmark.dk

L. Morsø · P. Kent · C. Manniche · H. B. Albert
Institute of Regional Health Services Research, University
of Southern Denmark, A Member of the Clinical Locomotion
Network, Odense, Denmark

general, it is recommended that external validation of questionnaires across care settings should be undertaken due to differences in patients mix [12, 13], and this might be particularly true for LBP patients in secondary care. Definitions of secondary care vary across countries and in the context of the Danish health system, secondary care is defined as government-funded specialised care requiring a referral. LBP patients referred to Danish secondary care have longer duration, greater pain intensity and higher frequency of referred leg pain compared to those in primary care [14]. Predictably, only patients with more complex back problems and poorer prognosis are referred to secondary care.

The two central considerations about the usefulness of the SBT in secondary care are: (1) its predictive ability (prognostic accuracy) and (2) its ability to indicate appropriate subgroup-targeted treatment pathways. A number of studies have tested the predictive ability in primary care [1, 2, 8, 15], but no previous studies have investigated and reported the predictive ability of SBT in secondary care. Therefore, the aim of the current study was to extend previous investigations of the SBT by comparing the predictive ability of the SBT in a Danish secondary care setting and a Danish primary care setting.

Materials and methods

This study was conducted at a specialised, multidisciplinary, secondary care setting—the Medical Department of the Spine Centre of Southern Denmark. Patients referred to the spine centre were almost all from the administrative region of Southern Denmark, which has a catchment population of 1.2 million people. Less than 5 % of referrals were from one of the other four regions in Denmark. Patients were referred to the spine centre for a comprehensive evaluation due to suboptimal improvement during assessment in primary care. Most referrals were received from GPs (>90 % of the referrals), with the remainder from chiropractors (approximately 5 %) or medical specialists (<5 %). While most patients are evaluated at the spine centre and referred back to primary care for further treatment, some have a very brief course of treatment at the spine centre and others are referred for surgical evaluation.

The secondary care data were self-reported by patients via electronic questionnaires using touch screen computers at the time of their first consultation at the spine centre. The electronic questionnaires were part of the SpineData database, which is a comprehensive registry of all patients attending the medical department. Prospective data were available for 960 consecutive low back pain patients with baseline and 6-months follow-up questionnaires from the period January 2012 to November 2012. The only inclusion

criteria for secondary care patients participating in the study were full electronic completion of the SBT at baseline. This inclusion criteria resulted in 20.8 % (250 patients) otherwise eligible patients being excluded. No diagnostic data were available in the database but using magnetic resonance imaging to document the presence of lumbar patho-anatomic findings, previous descriptive research on this clinical population showed that approximately 0.5 % or less have serious pathology (tumour, fracture, tuberculosis), 15 % have central stenosis, 29 % have nerve root compromise and the remainder have non-specific LBP [16].

The results from this secondary cohort were compared to those from an existing primary care physiotherapy cohort that had been collected for testing the predictive ability of SBT in Danish primary care [15]. That study combined data from GP and physiotherapy practices with outcome measured at 3 months, but in the current study, only the physiotherapy data were used, as 6-month follow-up data were only available for that sample. Details of the recruitment criteria and data collection used in that study have previously been reported [15]. Briefly, data were collected from May to September 2011 at 27 Danish physiotherapy clinics. Baseline data were available for 172 patients and for 83 % ($n = 144$) at 6 months follow-up. The variables used in the current study were extracted to match those available in the secondary care sample.

Data from the physiotherapy setting had been entered into a database (Epidata 3.1, The EpiData Association, Odense, Denmark) by a research secretary, while those from secondary care were entered directly into the SpineData database by the patients themselves. This study was approved by the Scientific Ethics Committee of the Region of Southern Denmark (S-20100036) and all patients gave written informed consent for research use of their data.

Data measurement

All data were collected in identical ways in both cohorts. Age (years) and gender (female, male) were extracted from each patient's unique social security (CPR) number. Duration of the current pain episode was calculated from the date of first consultation and the patient self-reported onset date of the current episode. Patients also self-reported: numbers of days off work during the last 3 months, number of previous LBP episodes, the SBT (9-item version), activity limitation [Danish 23-item version of the Roland Morris Disability Questionnaire (RMDQ)] [17], low back pain intensity and leg pain intensity (0–10 Numerical Rating Scale). The outcome measures collected at 6 months follow-up were: low back pain intensity, leg pain intensity and activity limitation.

Data analysis

Descriptive analysis (means and standard deviations, medians and inter-quartile ranges) of the baseline characteristics of both cohorts were tabulated at the level of the total samples and also stratified by SBT subgroup. Baseline differences between the three SBT subgroups were examined using Mann–Whitney *U*, Chi square or Kruskal–Wallis tests, depending on the data type and distribution.

Previous studies have tested the predictive validity of health questionnaires using a variety of methods [18, 19]. The current study mirrored the three statistical methods used in the original development study of the English language version of the SBT [1], which were also those used in the validation study of the SBT in Danish primary care [15].

The proportion of patients with a poor clinical outcome at 6 months was calculated, stratified by SBT subgroup. Poor clinical outcome was defined as persistent activity limitation measured by the Roland Morris Disability Questionnaire (RMDQ) [20] score at the 6-month follow-up. The cut point used in the original SBT development study in the UK was 7 points on a 0–24 scale [1], but as we used the proportional recalculation method to convert all RMDQ scores to a 0–100 scale, the threshold was recalculated to be 30 points or more. The proportional recalculation method has been shown to more accurately manage any missing RMDQ answers [21].

The same threshold was used to estimate the additional risk (relative risk) [22] for poor outcome for people classified into the medium or high SBT risk subgroup compared to the low-risk subgroup. Differences between the risk groups within each cohort were tested using Chi-square test for 2×2 tables.

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) [22] between people with and without the following outcomes at 6 months follow-up: (1) activity limitation as defined above and (2) LBP intensity still being ‘severe’ (8–10 on a 0–10 point scale). These criteria were those used in the original UK validation study [1].

In addition to those three statistical approaches, odds ratios for poor outcome on activity limitation for SBT subgroups were also calculated in unadjusted and adjusted form using logistic regression. This was performed to explore whether the predictive ability of the SBT in secondary care was confounded by baseline differences between the cohorts. All covariates were initially entered into the model, followed by a manual backwards stepwise reduction ($p < 0.05$ to remove) to the most parsimonious

model. An odds ratio greater than one in these regression models means that particular clinical characteristic increases the odds of having a poor outcome and an odds ratio less than one means that it is protective against a poor outcome.

As it eventuated that the relative risk estimates were lower in our secondary care data than in primary care and to have a predictive reference standard to compare those results to post hoc, we also calculated the relative risks of poor outcome on activity limitation using baseline pain intensity or baseline activity limitation as the predictor. The predictor of baseline pain intensity was formed by creating three categories that each contained 33 % of the participants based on the distribution of the cohort’s scores on the 0–10 pain intensity scale. The same distribution-based method was used for categorising the baseline activity limitation scores on the 0–100 RMDQ scale to create a three-category predictor variable.

The relative risk estimates were calculated using Microsoft Excel 2003 (Microsoft Corp, Redmond, WA, USA) and 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 secondary and primary care cohorts

The two cohorts were significantly different at baseline on duration of episode ($p < 0.001$), leg pain intensity ($p < 0.001$) and borderline significance on LBP intensity ($p = 0.059$). As seen in Table 1, within each cohort there were reassuringly significant differences between SBT subgroups, with increased LBP intensity, leg pain intensity and activity limitation across the low-risk to high-risk SBT subgroups. At baseline, the level of activity limitation was highest in the high-risk subgroup in both care settings with median RMDQ score of 78.3 (IQR 65–87) in secondary care and 77.8 (IQR 70–84) in primary care (0–100 scale).

Six-month outcome differences between the secondary and primary care cohorts

At 6 months follow-up, there were differences between cohorts in LBP intensity, leg pain intensity and activity limitation ($p < 0.001$) (Table 2). The higher values for LBP intensity, leg pain intensity and activity limitation in secondary care were also retained when stratified by SBT subgroup.

Table 1 Baseline characteristics for the Danish secondary and primary care cohorts

Secondary care cohort	Total cohort (n = 960)			SBT			Test for differences between SBT subgroups* (p value)
	Low risk (n = 252) (27.7 %)	Medium risk (n = 296) (32.5 %)	High risk (n = 363) (39.9 %)	Low risk (n = 54) (34.6 %)	Medium risk (n = 64) (41.0 %)	High risk (n = 38) (24.4 %)	
Age in years							
Mean (SD)	52.0 (14.1)	52.3 (14.6)	52.0 (13.8)	53.3 (14.6)	50.4 (14.6)	49.5 (16.7)	0.377
Female, proportion	521 (54.3 %)	116 (46 %)	179 (60.5 %)	30 (55.6 %)	41 (64.1 %)	17 (44.7 %)	0.163
Duration in months							
<1	47 (5.0 %)	5 (2.0 %)	18 (6.3 %)	18 (34.0 %)	29 (47.5 %)	13 (34.2 %)	0.246
1–3	139 (14.9 %)	36 (14.6 %)	56 (19.6 %)	17 (32.1 %)	13 (21.3 %)	7 (18.4 %)	
>3	746 (80.0 %)	205 (83.3 %)	212 (74.1 %)	18 (34.0 %)	19 (31.1 %)	18 (47.7 %)	
Pain intensity ^a (0–10) ^b							
Low back pain, median (IQR)	6 (4–7)	4.3 (2–6)	6.0 (5–7)	4.0 (3–6)	6.0 (4–8)	6.0 (5–7)	0.001
Leg pain, median (IQR)	5 (2–7)	3.0 (1–5)	5.0 (3–6)	1.5 (0–3)	3.0 (1–6)	4 (1–7)	0.001
Activity limitation ^c (0–100 scale) ^b							
Median (IQR)	60.9 (44–78)	34.8 (22–52)	60.9 (48–74)	34.8 (22–52)	60.9 (48–74)	78.3 (65–87)	<0.001
Proportion of patients >30	838 (88.6 %)	158 (64.2 %)	281 (95.9 %)	158 (64.2 %)	281 (95.9 %)	354 (98.9 %)	<0.001
Primary care cohort							
Total cohort (n = 172)							
Age in years							
Mean (SD)	52.0 (15.2)	53.3 (14.6)	50.4 (14.6)	53.3 (14.6)	50.4 (14.6)	49.5 (16.7)	0.377
Female, proportion	98 (57.0 %)	30 (55.6 %)	41 (64.1 %)	30 (55.6 %)	41 (64.1 %)	17 (44.7 %)	0.163
Duration in months, proportion							
<1	65 (38.9 %)	18 (34.0 %)	29 (47.5 %)	18 (34.0 %)	29 (47.5 %)	13 (34.2 %)	0.246
1–3	39 (23.4 %)	17 (32.1 %)	13 (21.3 %)	17 (32.1 %)	13 (21.3 %)	7 (18.4 %)	
>3	63 (37.7 %)	18 (34.0 %)	19 (31.1 %)	18 (34.0 %)	19 (31.1 %)	18 (47.7 %)	
Pain intensity ^a (0–10 scale) ^b							
Low back pain, median (IQR)	5.0 (4–7)	4.0 (3–6)	6.0 (4–8)	4.0 (3–6)	6.0 (4–8)	6.0 (5–7)	0.001
Leg pain, median (IQR)	3.0 (0–6)	1.5 (0–3)	3.0 (1–6)	1.5 (0–3)	3.0 (1–6)	4 (1–7)	0.001
Activity limitation ^c (0–100 scale) ^b							
Median (IQR)	60.9 (39–77)	34.8 (26–49)	67.4 (52–78)	34.8 (26–49)	67.4 (52–78)	77.8 (70–84)	<0.001
Proportion of patients >30	146 (86.3 %)	36 (66.7 %)	60 (93.8 %)	36 (66.7 %)	60 (93.8 %)	38 (100 %)	

SD standard deviation, IQR inter-quartile range

* Tests for differences between subgroups were ANOVA or Kruskal–Wallis procedures, depending on data type and distribution

^a Numeric Rating Scale (0–10)

^b High scores are worse

^c Roland Morris Disability Questionnaire

Table 2 Outcome at 6 months follow-up for the Danish secondary and primary care cohorts

Secondary care cohort	Total cohort (n = 960)			
	SBT	Low risk (n = 252) (27.7 %)	Medium risk (n = 296) (32.5 %)	High risk (n = 363) (39.9 %)
Pain intensity ^a (0–10 scale) ^b				
Low back pain, median (IQR) ³	4.7 (2–6)	3.0 (2–5)	4.3 (2–6)	5.7 (3–7)
Leg pain, median (IQR) ³	2.7 (0–5)	1.3 (0–4)	2.7 (0–5)	3.7 (1–6)
Activity limitation ^c (0–100 scale) ^b				
Median (IQR)	47.8 (28–70)	26.1 (13–48)	47.8 (22–70)	65.2 (35–83)
Proportion of patients >30 ^d	656 (69.0 %)	120 (47.8 %)	209 (71.3 %)	292 (81.3 %)
Primary care cohort	Total cohort (n = 144)			
	SBT	Low risk (n = 48) (36.9 %)	Medium risk (n = 52) (40.0 %)	High risk (n = 30) (23.1 %)
Pain intensity ^a (0–10 scale) ^b				
Low back pain, median (IQR)	3.0 (2–5)	3.0 (2–4)	2.0 (0–4)	4.5 (2–5)
Leg pain, median (IQR)	1.3 (0–4)	0.5 (0–3)	2.0 (1–5)	1.0 (0–5)
Activity limitation ^c (0–100 scale) ^b				
Median (IQR)	26.0 (9–57)	17.0 (4–26)	30.0 (9–55)	65.0 (25–75)
Proportion of patients >30 ^d	51 (40.2 %)	9 (20.0 %)	21 (46.7 %)	18 (69.2 %)

IQR inter-quartile range

^a Numeric Rating Scale

^b High scores are worse

^c Roland Morris Disability Questionnaire

^d Patients with a Roland Morris Disability Questionnaire score >30 were classified as having persistent activity limitation

Unadjusted risk of poor clinical outcome on activity limitation at 6 months

At a cohort level, 69.0 % in secondary care and 40.2 % in primary care had a poor outcome on activity limitation (Table 2) 6 months after their index consultation. When stratified, the proportion of those increased from low-risk to high-risk SBT subgroup within each cohort, but with some distinct differences between the cohorts. Most notable was the large difference in patients with poor outcome on activity limitation in the low-risk subgroup (47.8 % in the secondary care cohort and 20.0 % in the primary care cohort) with almost half of the patients in the secondary care cohort still having an RMDQ score above 30 points. That pattern of a larger proportion in secondary care having a poor outcome was also retained across the other subgroups.

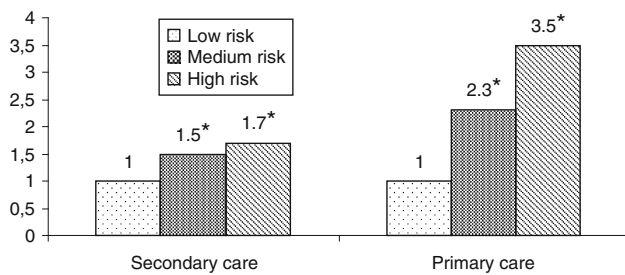


Fig. 1 Relative risk of poor clinical outcome (more than 30 Roland Morris Disability Questionnaire points on a recalculated 0–100 scale, * $p < 0.05$) on activity limitation at 6 months by SBT subgroup in the Danish secondary and primary care cohorts

Another important observation was that the gradient of relative risk across the three SBT subgroups was not nearly as steep in secondary care as in the primary care (Fig. 1). Though still significantly predictive of additional risk of poor outcome in the medium-risk [RR 1.5 (95 % CI 1.3, 1.7)] and the high-risk group [RR 1.7 (1.5, 2.0)], these unadjusted results indicate that the predictive ability for the SBT subgroups for 6 month outcome was not as strong in secondary care as it was in primary care [RR medium risk 2.3 (95 % CI 1.2, 4.5), high risk 3.5 (95 % CI 1.8, 6.6)].

It is likely that these two findings of (1) nearly half the low-risk subgroup in secondary care having a poor outcome and (2) the reduced predictive ability of the SBT subgroups in secondary care are inter-related, as the low-risk subgroup is the reference category for the predictive ability. As it was possible that this relationship was also confounded by the difference in baseline episode duration and pain intensity between the cohorts, an adjusted analysis was also performed.

Unadjusted and adjusted odds of poor clinical outcome on activity limitation at 6 months

The unadjusted odds ratios (OR) shown Table 3 reflect the difference between the cohorts already reported in the relative risk results. The predictive ability of the medium-risk subgroup across cohorts was not markedly different [secondary care OR 2.7 (1.9, 3.9), primary care 3.5 (1.4, 8.9)], whereas the difference was more distinct in the high-

Table 3 The odds of having a poor clinical outcome on activity limitation at 6 months follow-up in the Danish secondary care and primary care cohorts, estimated by STarT Back Tool subgroup using logistic regression

	Secondary care cohort (n = 903)		Primary care cohort (n = 116)	
	Odds ratio (95 % CI)	p value	Odds ratio (95 % CI)	p value
Unadjusted model				
STarT Back Tool low-risk subgroup ^a	1.00		1.00	
STarT Back Tool medium-risk subgroup	2.72 (1.91, 3.87)	<0.001	3.50 (1.37, 8.93)	0.009
STarT Back Tool high risk	4.76 (3.31, 6.84)	<0.001	9.00 (2.97, 27.25)	<0.001
Constant	0.92 (0.72, 1.17)	<0.488	0.25 (0.12, 0.52)	<0.001
Model-adjusted episode duration and baseline low back pain				
STarT Back Tool low-risk subgroup ^b	1.00		1.00	
STarT Back Tool medium-risk subgroup	2.36 (1.61, 3.47)	<0.001	5.22 (1.69, 16.13)	0.004
STarT Back Tool high risk	3.31 (2.18, 5.03)	<0.001	6.88 (1.65, 28.63)	0.008
Episode duration (months)				
0–1	1.00		1.00	
1–3	1.08 (0.52, 2.23)	0.835	3.60 (0.86, 15.09)	0.080
>3	2.10 (1.10, 4.02)	0.025	4.49 (1.31, 15.41)	0.017
Low back pain intensity baseline				
1.17 (1.09, 1.26)	<0.001	1.56 (1.24, 1.97)	<0.001	
Constant	0.252 (0.12, 0.53)	<0.001	0.018 (0.004, 0.088)	<0.001

^a Roland Morris Disability Questionnaire score >30 (0–100 scale)

^b Reference category

risk groups [secondary care OR 4.8 (3.3, 6.8), primary cohort 9.0 (3.0, 27.6)].

However, adjustment for baseline duration of episode and pain intensity resulted in only marginally reduced ORs in the medium-risk and high-risk subgroups. Episode duration made statistically significant contributions to the models in both cohorts, and baseline LBP intensity to the model in secondary care. In some cases those changes increased the ORs by 10 %, but as they occurred in both cohorts, they did not account for the reduced predictive ability of the SBT subgroups in secondary care. There were no statistically significant interactions between SBT subgroups and episode duration and this was also reflected in the correlation between SBT total scores and episode duration being very weak (-0.005 in secondary care and 0.037 in primary care). There was also no significant interaction between pain intensity and SBT subgroups. Therefore, for predicting persistent activity limitation at 6 months, both baseline episode duration and baseline low back pain intensity were predictive in both cohorts and had an influence that was independent of the predictive ability of the SBT subgroups.

To gain a sense of what the predictive ability of other reference standard predictors would be in secondary care, post hoc analyses were performed using the three-category distribution-based predictors of baseline pain intensity and activity limitation. The RRs for baseline pain intensity were medium risk 1.5 (1.3, 1.7) and high risk 1.6 (1.4, 1.8); for activity limitation were medium risk 1.6 (1.4, 1.8) and high risk 1.8 (1.6, 2.1). These were nearly identical to those obtained when using the SBT subgroups as predictors.

The ability of the baseline SBT total scores to identify people with outcomes above a clinical threshold at 6 months follow-up

The AUC statistics describing the ability of baseline SBT total scores (0–9 scale) to discriminate between people with and without scores above threshold values on two different 6-month outcomes are shown in Table 4. For both outcomes, activity limitation ‘still being present’ and, LBP ‘still being severe’, the discriminative ability was similar across cohorts.

Discussion

The aim of this study was to compare the predictive ability of the SBT in a Danish secondary care setting and a Danish primary care setting for the outcome of persistent activity limitation at 6 months follow-up. The results indicate that the SBT subgroups were not as strongly predictive of poor outcome in the Danish secondary care setting, but were as predictive as similarly categorised baseline pain intensity or activity limitation scores.

The results also show very similar proportions of patients across the cohorts having poor activity limitation at baseline, both at an overall cohort level and also when stratified into SBT subgroups. However, at 6 months follow-up these proportions were quite different, reflecting that the recovery trajectories were less favourable in secondary care, a finding which is in concordance with earlier findings [23]. While the large proportion of secondary care patients with poor outcome in the high-risk group was similar to the primary care cohort and to that found in other primary care studies [1], 47.8 % in the secondary care low-risk subgroup and 71.3 % in the medium-risk subgroup who had a poor outcome were clearly different from that in primary care [15].

The proportion of patients classified into the SBT low-risk subgroup who nonetheless experienced persistent activity limitation was much larger in secondary care (47.8 % of ‘low-risk’ patients in secondary care, 20.0 % in primary care). While this higher proportion in secondary care might be expected, it has the consequence of attenuating the relative risk estimates that were possible, because this subgroup is the reference category (the denominator in the relative risk formula). This is seen in the results showing that similarly categorised baseline pain and activity limitation were no stronger at prediction in this cohort, despite it being well recognised that these are strong predictors [24] and that baseline values are the best predictors for the same outcome [2]. It therefore seems that prediction in this setting is challenging, perhaps due to a combination of more frequent poor outcome and a wider variability of outcome relative to baseline presentation.

The results also indicated that the predictive ability of the psychosocial subscale component, which is the

Table 4 Discriminative ability of the STarT Back Tool to correctly classify people with high scores on two different dichotomised outcomes at 6 months follow-up

	Danish secondary care cohort AUC (95 % CI)	Danish primary care cohort AUC (95 % CI)
People with a Roland Morris Disability Questionnaire score >30 (0–100 scale) at 6 months	0.69 (0.66, 0.73)	0.73 (0.64, 0.82)
People with severe back pain at 6 months (8–10 on a 0–10 scale)	0.72 (0.68, 0.77)	0.66 (0.46, 0.85)

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

distinction between medium and high-risk subgroups, was lower in secondary care. This was previously noted in an earlier primary care study that compared the predictive ability of the SBT in the UK and Denmark [15]. In that instance, those differences were explained by changes in the psychosocial factors during the treatment period, probably due to differences in treatment exposure. In the current study, confounding may also have occurred due a difference in the management of psychosocial factors, but the available data did not allow statistical adjustment for change in these factors.

Another explanation for the different predictive ability of the SBT in these primary and secondary care cohorts could be differences in case mix. Although diagnostic codes were not available in the data from either setting, SBT was originally validated in people diagnosed by GPs as having non-specific LBP. In our secondary care setting, approximately 45 % have MRI evidence of central stenosis or nerve root compromise [16] and this may have affected their recovery trajectories and, thereby, the predictive ability of the SBT. Another potential factor affecting the predictive ability could be a social class bias that we believe results in an over-representation of lower socio-demographics in the secondary care cohort. In pregnancy-related pelvic pain, it has been shown that socio-demographics are influential on outcome [25]. The SBT does not measure these characteristics and it may be that for it to have better predictive ability in secondary care, these factors would need to be included.

In the regression models that adjusted for baseline differences, only episode duration and baseline pain intensity were retained as an independent predictive factor alongside the SBT subgroups in secondary care. Previous studies have shown that both influence outcome and return to work [26, 27], but our findings indicate that, in this context, neither exerted an influence that could explain the differences between care settings in the SBT predictive ability.

Paradoxically, the results in both cohorts of our AUC analysis show similar discriminative ability of the SBT 9-item sum scores to correctly classify patients on two dichotomised outcome measures (persistent activity limitation and severe LBP) at 6 months follow-up, despite differences in the predictive ability of the SBT subgroup classification. This might be interpreted to indicate that the predictive ability potentially would improve by changing the SBT cut points, but such post hoc analysis revealed that neither changing these cut points nor using median baseline activity limitation in secondary care as the outcome criterion, or both, more than marginally altered the predictive ability (results unreported).

Previous studies of primary care in the UK and Denmark indicate that 17–24 % of people classified into the low-risk group nonetheless had a poor outcome [15]. Therefore, it is

to be expected that some ‘failed’ low-risk patients who do not improve are referred to secondary care. However, given that almost half of the ‘low-risk’ patients in secondary care had a poor outcome, perhaps we need to reframe the language in this setting so that this subgroup is referred to as ‘low complexity’ compared to ‘medium and high complexity’ subgroups.

A strength of this study was the use of a pre-existing validation model to test the predictive ability of the SBT classification categories, as this allowed the comparison of results across care settings and two previous studies [1, 15]. Two other primary care studies used different methodological approaches to assess the predictive ability of SBT [2, 8]. In one study, SBT sum scores were used as a continuous scale in longitudinal modelling of a non-uniform outcome period [8]. In the other study the SBT sum scores and the outcome measures were used as continuous scales in multiple linear regression modelling to monitor of change during treatment and avoid the borderline misclassification of cases [2]. Our study was not designed to monitor change, but to investigate the predictive ability of the baseline SBT classification categories (low-, medium- and high-risk subgroup) and therefore we mirrored the method used in the original validation studies [1, 15].

A limitation of this study is that it was not designed to investigate the treatment implications of the SBT. Although the SBT predictive ability was not as strong as in primary care, it was investigated by us in secondary care where care pathways were uninfluenced by the SBT subgroup. Therefore, it is possible that an ‘SBT type’ of classification might have clinically useful treatment implications in secondary care, although such risk-based classification may require including different constructs. Theoretically, this might be achieved by extending the original SBT with additional questions on constructs relevant to prognosis and stratified care in secondary care. For example, social constructs or different psychological constructs may be more relevant in secondary care. However, the construction and validation of a ‘secondary care SBT’ would be a substantial project and was beyond the brief of the current project. In addition, as secondary care settings and the characteristics of their patients vary greatly within and between countries, caution should be exercised in generalising these results.

Conclusion

In our multidisciplinary Danish secondary care setting, the SBT classification subgroups were less able to predict persistent activity limitation at 6 months follow-up than in a Danish physiotherapy primary care setting. This finding remained even after adjusting for baseline differences in

episode duration and LBP intensity. In both settings, both episode duration and baseline low back pain intensity were predictive factors that were independent of SBT subgroup classification. The usefulness of SBT subgroup-targeted treatment in secondary care was not investigated in this study.

Conflict of interest None.

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. *Arthritis Rheum* 59(5): 632–641
- Beneciuk JM, Bishop MD, Fritz JM, Robinson ME, Asal NR, Nisenzon AN et al (2012) The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Phys Ther* 93:321–333
- Hill JC, Dunn KM, Main CJ, Hay EM (2010) Sub grouping low back pain: a comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 14(1): 83–89
- Wideman TH, Hill JC, Main CJ, Lewis M, Sullivan MJ, Hay EM (2012) Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. *Pain* 153(11):2182–2191
- Bruyere O, Demoulin M, Brereton C, Humblet F, Flynn D, Hill JC et al (2012) Translation validation of a new back pain screening questionnaire (the STarT Back Screening Tool) in French. *Arch Public Health* 70(1):12
- Gusi N, Del Pozo-Cruz B, Olivares PR, Hernandez-Mocholi M, Hill JC (2011) The Spanish version of the “STarT Back Screening Tool” (SBST) in different subgroups. *Aten Primaria* 43:356–361
- 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
- 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
- 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
- Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P et al (2012) Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). *Fam Pract* 29(1):50–62
- Foster NE, Mullis R, Young J, Doyle C, Lewis M, Whitehurst D et al (2010) IMPaCT Back study protocol. Implementation of subgrouping for targeted treatment systems for low back pain patients in primary care: a prospective population-based sequential comparison. *BMC Musculoskelet Disord* 11:186
- Moons KG, Altman DG, Vergouwe Y, Royston P (2009) Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 338:b606
- de VetHenrica CW, Terwee CB, Mokkink LB, Knol DL (2011) *Validity. Measurement in medicine*, 1st edn. Cambridge University Press, New York, pp 150–201
- Morso L, Kent P, Albert HB, Manniche C (2013) Is the psychosocial profile of people with low back pain seeking care in Danish primary care different from those in secondary care? *Man Ther* 18(1):54–59
- Morso L, Kent P, Albert HB, Hill JC, Kongsted A, Manniche C (2013) The predictive and external validity of the STarT Back Tool in Danish primary care. *Eur Spine J*
- Albert HB, Briggs AM, Kent P, Byrhagen A, Hansen C, Kjaergaard K (2011) The prevalence of MRI defined spinal path anatomies and their association with modic changes in individuals seeking care for low back pain. *Eur Spine J* 20(8):1355–1362
- Albert HB, Jensen AM, Dahl D, Rasmussen MN (2003) Criteria validation of the Roland Morris Questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica. *Ugeskr Laeger* 165(18):1875–1880
- Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski 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
- 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
- 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
- Kent P, Lauridsen HH (2011) Managing missing scores on the Roland Morris Disability Questionnaire. *Spine (Phila Pa 1976)* 36(22):1878–1884
- Kirkwood BRSJAC (1988) *Measurement error: assessment and implications. Essential Medical Statistics*, 2nd edn. Blackwell Science Ltd., Oxford, pp 429–446
- Grotle M, Vollestad NK, Brox JI (2006) Clinical course and impact of fear-avoidance beliefs in low back pain: prospective cohort study of acute and chronic low back pain: II. *Spine (Phila Pa 1976)* 31(9):1038–1046
- Chou R, Shekelle P (2010) Will this patient develop persistent disabling low back pain? *JAMA* 303(13):1295–1302
- Albert HB, Godskesen M, Korsholm L, Westergaard JG (2006) Risk factors in developing pregnancy-related pelvic girdle pain. *Acta Obstet Gynecol Scand* 85(5):539–544
- Grotle M, Foster NE, Dunn KM, Croft P (2010) Are prognostic indicators for poor outcome different for acute and chronic low back pain consulters in primary care? *Pain* 151(3):790–797
- Pengel LH, Herbert RD, Maher CG, Refshauge KM (2003) Acute low back pain: systematic review of its prognosis. *BMJ* 327 (7410):323